CMUG Deliverable

Number:	**D3.1_A**
Due date:	**March 2012**
Submission date:	**30 August 2012**
Version:	**1.2**

Climate Modelling User Group

Deliverable 3.1_A

Technical note on CMUG ECV Quality Assessment Report

Centres providing input: MOHC, MPI-M, ECMWF, MétéoFrance

Version nr.	Date	Status
0.1	25 Nov 2011	Initial Template RS
0.2	17 Feb 2012	Updates from ESA comments RS
0.3	5 Apr 2012	Complete draft for comment by CMUG
0.4	19 Apr 2012	Added comments from CMUG prior to release RS
1.0	20 Apr 2012	Final version for ESA. Not to be distributed wider.
1.1	13 Jun 2012	Added MPI additions + response to ESA comments
1.2	30 Aug 2012	Added MPI additional text and revised ozone section

CMUG Deliverable
Number: D3.1_A
Due date: March 2012
Submission date: 30 August 2012
Version: 1.2

Deliverable 3.1

Technical note

CMUG ECV Quality Assessment Report

Table of Contents

CMUG ECV Quality Assessment Report

1. Purpose and scope of the Technical note

The purpose of this activity is to assess the quality of the Climate Data Records (CDRs) delivered by the ESA CCI and use them with coupled Earth System Models. To provide added value for climate modelling activities such as initialisation, assimilation, model evaluation and development, trend analysis and monitoring, the CDRs must have the 'climate quality' and meet the requirements which have been given in the URDs.

The assessment of the CDRs will not be a repeat of the validation performed by the CCI teams which primarily will concern data products from Level 2 to Level 4.

The first version of this document (V1) will only be a proof of methodology by assessing “Precursor Products” taken from existing datasets. Later versions will report on the actual CCI CDRs produced in phase 1 of the CCI.

2. Terminology used

To aid the reader and avoid confusion the definition of the main terms used in this report are given here.

Essential Climate Variable (ECV) defines a specific variable defining the atmospheric, ocean or land surface state. One ECV can include several different climate data records (e.g. ozone total column and ozone profile). They have been defined by GCOS (2011) for ECVs measured by satellites.

Climate Data Record (CDR) is a level 2 or 3 dataset for an ECV which has been processed to a standard sufficient for climate monitoring purposes. Level 1 datasets (e.g. top of atmosphere radiances) are referred to as FCDRs (Fundamental Climate Data Record).

Pre-cursor refers to a CDR which has similar characteristics to the planned CCI CDRs. It may not be “climate quality”. The ESA GlobXXX series datasets are examples of precursors. The main requirement for this purpose is that it can be assessed in a similar way to the CCI CDRs to demonstrate the methodology.

Assessment here is a generic term which refers to a variety of different ways to determine the fidelity of a CDR. The various methods for assessment are given in section 3.

Assimilate here refers to a CDR being used within an atmospheric, ocean or land surface model to adjust the state variables to better fit the observations taking into account the uncertainties of the observations and model first guess.

Hindcast is a where a NWP model is run in the past to verify the accuracy of its forecasts with observations.

Nudging in data assimilation means to add a term to the state vector that is proportional to the difference of the calculated meteorological variable and the observed value. This term "keeps" the calculated state vector closer to the observations.

Uncertainty refers to a combination of random and systematic (bias) errors for each variable in a CDR. It normally refers to an individual observation but can refer to area and time averaged quantities.

Consistency refers to the consistency of related ECVs (e.g. fire and aerosols) in space and time. This is important for relationships between different ECVs and also between different CDRs for the same ECV.

Climate model is a numerical representation of the climate system based on the physical, chemical and biological properties of its components, their interactions and feedback processes, and accounting for some of its known properties.

Ensemble A group of parallel model simulations used for climate projections or predictions. Variation of the results across the ensemble members gives an estimate of uncertainty. Ensembles made with the same model but different initial conditions only characterise the uncertainty associated with internal climate variability, whereas multi-model ensembles including simulations by several models also include the impact of model structural differences.

CMIP-5 is an exercise to compare the current state of art climate models and has provided an ensemble of different predictions.

Reanalyses are estimates of historical atmospheric and oceanic temperature, wind, current, and other meteorological and oceanographic quantities, created by processing past meteorological and oceanographic data using fixed state-of-the-art weather forecasting models (atmospheric reanalysis), ocean monitoring and forecasting models (ocean reanalysis) and data assimilation techniques.

3. Methodology applied to assess climate data records

For climate modelling the four key applications of the CCI datasets are to *Enable Model-Observation Confrontation, Provide Boundary Conditions, Provide Initial Conditions, and Provide Observations capable of assimilation*. Model-Observation Confrontation is the natural first step for a new dataset to be used with climate models and this will be the primary activity performed by CMUG in a number of ways as listed below. Model-Observation Confrontation plays a significant role in the decision process that determines whether a dataset is deemed suitable (from the user's perspective) for the other 3 key applications.

The CMUG assessment will encompass the following aspects for a selection of the CCI climate data records:

Confront

- *consistency* of Global Satellite Data Products in *time* (e.g. stability, uncertainty of bias)

- *consistency with independent observations* (e.g. limb view, in-situ, ground-based remote sensing)
- *consistency with precursor datasets* to understand the differences and assess if the CCI datasets are better representations of the atmospheric/surface state
- *consistency compared to reanalysis fields*
- *consistency across ECVs*
- *ability to capture climate variability and small climate change signals* (e.g. observed trends) for their use in Climate Monitoring and Attribution.

Assimilate and boundary condition

- *impact* in Model and Data Assimilation Systems (for a few ECVs where appropriate).

There is not a single methodology that can be used universally but several approaches from different science teams and tailored for each ECV so only general comments are given here with the details in section 4 for each ECV. In many cases an observation operator is required to compare the measured quantities with the actual model variables although often this operator is fairly trivial. A simple operator would be interpolation from model grid to observation point in space and time. A more complex operator would be a radiative transfer model to compare measured top-of-atmosphere radiances (level 1 data) with model equivalents. If higher level 2 or 3 products are used the operator is usually simpler as the variables are closer to the model output although the error characteristics of the products can be more complex.

When the products are used in model analyses, the correction of their systematic and random errors may be required. When they are used for direct comparison, the way they are used could be refined and this will be a topic of research in the assessment. In particular the assessment of the uncertainties provided with the data will need to be assessed in an objective manner.

Data used for assessment of CDR	Advantages	Drawbacks
Climate Model (single, ensemble)	Spatially and temporally complete	Model has uncertainties Not all variables available
Re-analyses	Spatially and temporally complete	Analysis has uncertainties Not all variables available
Precursors	Comparing like with like	Some precursors may have large uncertainties
Independent satellite or in situ measurements	Different ‘view’ of atmosphere/surface	May have much larger uncertainty than CDR, need to include representativity errors
Related observations (surface and TOA fluxes, temperature, water vapour)	Assures consistency with other model variables	May not be spatially or temporally complete

Table 1. The various options for assessing the CDR and their advantages and drawbacks

For many atmospheric ECVs a comparison with the ERA-Interim reanalysis is appropriate to assess the overall fidelity of the CDRs. A comparison with other independent measurements, in situ, ground based remote sensing and other satellite products is also important. Ideally they should exhibit some differences in time sampling or are measurements using a different technique (e.g. limb viewing in infrared or microwave, aircraft sampling). There are also some related products which can be linked to some CCI ECVs (e.g. CO for biomass burning and aerosols, humidity or precipitation for clouds) which should be used in the assessments. Table 1 summarises the advantages and disadvantages of the various assessment datasets.

The consistency across ECVs is something that has been specifically identified as being important to the climate modelling community (the CCI project's targeted user community) and the CMUG will look at this aspect of the CCI datasets, drawing attention where necessary to inconsistencies between related ECVs. Increasingly, the climate modelling community approaches consistency from an integrated perspective which includes consistency across ECV product levels, e.g. from Level-1 radiances to Level-2 swath-based geophysical products to Level-3 gridded products, and also extends to ancillary data products such as bias corrections and homogenization terms. It is therefore important that the CCI continues its commitment to open access and traceability, which will entail preserving and making available all such products generated during the project.

An important requirement of an observational dataset for reanalysis is that when assimilated it improves (or at least does not degrade) the short range forecasts of relevant meteorological variables. Assimilation of the CCI products is a longer term goal in the context of reanalysis projects (e.g. ERA-CLIM) and represents a critical test for some CDRs, but given that such tests are expensive to perform they must first be preceded by extensive quality assurance on the observational datasets in order to maximize the prospects for demonstrating beneficial impact.

The following sections describe initial assessments of 'precursor' datasets to demonstrate the methodology of assessing the CCI CDRs when they become available. A variety of methods are employed for the different ECVs. Not all CCI ECVs could be covered here.

4. Assessment of climate data records for CCI ECVs

4.1 Sea Surface Temperature

Sea surface temperature was assessed with two different precursors, the ARC ATSR dataset and the *p*-HadISST2 historical analysis of SST. Two different approaches were also used as described below.

4.1.1 ARC SST assessment

The obvious pre-cursor for the SST ECV data is the ATSR Re-processing for Climate (ARC) datasets. These have recently become available from the project team. The ARC project has developed an accurate SST data set using the (A)ATSR instruments which is aimed for use in climate change analyses. Further information about the ARC project can be found in Merchant, et. al. (2008). The ARC data is from August 1991 to December 2009 using ATSR-1 on ERS-1, ATSR-2 on ERS-2 and AATSR on EVISAT. The dataset includes a two and

three channel retrieved skin SST and the corrections needed to produce the sub-skin (a depth of about 1mm) temperature and temperature at depths of 0.2m, 1m and 1.5m. The 1m measurement is taken here to typically represent the depth at which the buoys measure. Several other fields for each grid point are included such as the total column water vapour, solar flux and wind speed from the ERA-Interim reanalysis. Finally the uncertainty of the SST retrieval which is a combination of the theoretical performance of the SST retrieval, number of pixels present and the variability in each cell is also included and it was felt important to try and validate this.

In order to verify the random and systematic errors of the ARC data, a comparison was made using drifting buoys. Two sources of buoy data were used for the collocation process and only drifting buoys were selected. From 1991-1996, the International Comprehensive Ocean-Atmosphere Data Set (ICOADS) version 2.5 was used (Woodruff et al. 2011). The ICOADS contains data acquired over a time scale of centuries and includes drifting and moored buoys, ship and platform observations. The locations of buoy observations in the ICOADS files used here were only recorded to an accuracy of 0.1°. From 1997 drifting buoy data from the Global Telecommunications System (GTS) were used. More observations can be found in the GTS files in comparison to the ICOADS files and locations of measurements are given to three decimal places.

SSTs from the buoys and ARC were compared using a collocation technique where a buoy observation is matched to a satellite observation if certain criteria are met. During the analysis, biases were investigated and their variation as a function of variables such as wind speed and insolation. In addition a three way error analysis was carried out which, with the inclusion of a third SST data source from a microwave radiometer, AMSR-E, allows the estimation of the standard deviation of the error in the ARC, buoy and AMSR-E measurements. AMSR-E version 5 data were obtained in daily averaged files and fields at a resolution of 0.25° were used. The AMSR-E SSTs correspond to an ocean depth of a few millimetres.

The quality of the ARC data was investigated in a number of ways. Global and regional statistics and zonal trend and seasonal variations concerning the ARC – buoy bias were studied. Histograms were constructed to gain a better understanding of the spread of the biases and comparisons between match-ups from the different satellites were made. In addition, investigations were carried out to look at the biases as a function of wind-speed, insolation, total column water vapour, the time difference between the collocated measurements and various other fields.

In order to account for the spatial variation when considering the global statistics, biases were collected into 1° x 1° cells and measurements within each box were averaged. This helped to ensure that cells towards the high latitudes containing fewer match-ups were not overwhelmed by the higher densities in lower latitudes. Using a simple cosine, area weights were calculated for each cell which were related to the area of the box on the surface of the Earth. This gives each cell an equal contribution to the overall mean so that grid boxes with larger areas on the globes do not skew the mean. Before collecting the data onto the 1° grid, a mean and standard deviation were calculated from the differences. Those match-ups exceeding ±3 times the

standard deviation from the mean were rejected and successful collocations were then averaged in the 1° grid boxes. Error bars on the plots represent the 95% confidence limits calculated from the standard error. Values were taken from Student's t-distribution (Lyons *et al.* 2005) to multiply the standard error to the required confidence level.

The method of the three-way error analysis enables the calculation of the standard deviation of random error on each observation type. Work carried out by O'Carroll *et. al* (2008), using AATSR three channel night-time retrievals from 2003 (and similarly comparing with AMSR-E and drifting and moored buoy SSTs) found that the error variance can be calculated by the following equation:

$$\sigma_x^2 = 0.5 * (V_{xy} + V_{zx} - V_{yz}) \qquad (1)$$

where σ_x^2 is the variance of the random error in observation type x and V_{xy} is the variance between two observation types, x and y. The derivation and discussion of this result can be found in O'Carroll *et. al* (2008). The standard deviation of the differences between each combination of two data sources was calculated and substituted in equation (1). Note the method relies on using scales for which the covariances of the errors of representativeness (errors concerning the "difference between the value of the variable on the space/time scale on which it is actually measured and its value on the space/time scale on which we wish to analyse it") are negligible compared to the standard error covariances. This assumption allows simplification to the equation above.

Global Statistics

The number of drifting buoy observations per day increases quite dramatically throughout the ARC period leading to a corresponding big rise in the number of collocations. As a result statistics derived from data in the ATSR-1 period (before 1996) should be treated with more caution as the confidence intervals for the statistics are comparatively large.

Figure 1 shows the mean bias for each monthly 3° latitude band of data. The two channel night retrieval is used, to show more collocations from the ATSR-1 period, however it is quite representative of the values and patterns observed in the other retrieval types. The full Bayesian cloud mask has been used for the ATSR-2 and AATSR data. The transition between the ATSR-2 and AATSR periods during 2002 appears to be seamless. There is a noticeable difference in the stability to more extreme values in the earlier, ATSR-1 years.

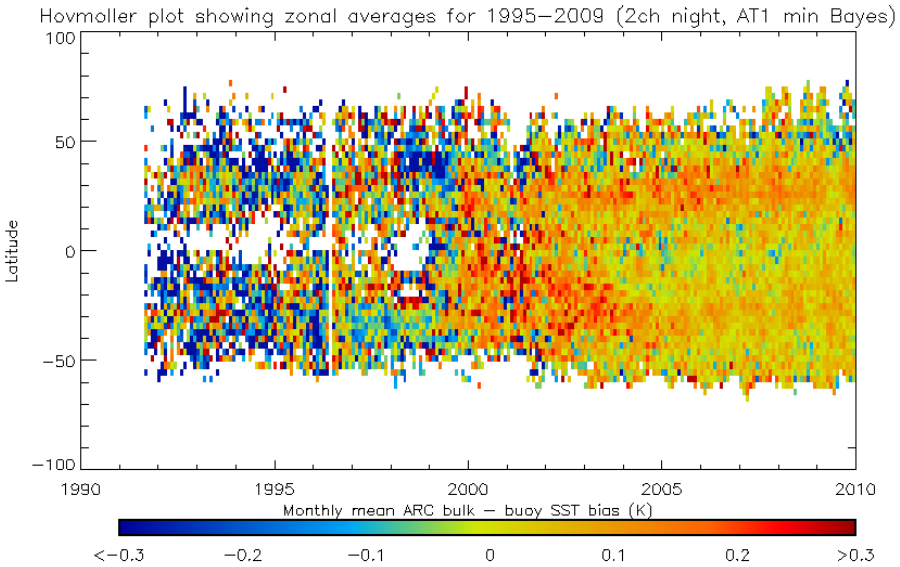


Figure 1. Hovmöller diagram showing the corresponding average bias. Two channel night-time retrievals were used and the full Bayesian cloud mask was used for all ATSR-2 and AATSR data.

A summary of the mean global biases and standard deviations for the different instrument periods is presented in Table 2. All the overall mean biases for the different instruments are found to have a magnitude of less than 0.1K although there is an element, particularly in the earlier years, of having a balance between warm and cold biases rather than collocations with a consistently small bias. The two channel night-time retrievals produce the lowest overall bias compared to the drifting buoys followed by the three channel retrieval. However, the three channel retrieval consistently gives the lowest standard deviation over the complete time series.

The reduction in standard deviation of the bias seen in the later years is partly correlated with the dramatic increase in the number of match-ups which also leads to narrower 95% confidence limits. The use of the full Bayesian cloud mask appears to produce slightly lower standard deviations. This supports the concept that the full cloud mask should be more effective in cloud detection and therefore reduces the number of cloud contaminated observations.

Channel selection	ATSR-1 bm = Bayes min mask bf = Bayes full mask		ATSR-2		AATSR		Whole time series 2ch: 1991-2009 3ch: 1995-2009	
	Mean (K)	Std. dev (K)	Mean (K)	Std. dev (K)	Mean (K)	Std. dev (K)	Mean (K)	Std. dev (K)
3 ch night only	-	-	0.043	0.266	0.059	0.143	0.054	0.151
2 ch night only	-0.085 (bm)	0.478 (bm)	0.040	0.296	0.053	0.159	0.044	0.182
2 ch day only	0.008 (bm) 0.008 (bf)	0.470 (bm) 0.471 (bf)	0.066	0.321	0.072	0.158	0.064	0.184

Table 2. Summary of mean global biases and standard deviations using 1° grid boxes for the different instrument periods and the complete time series

It is worth noting that throughout the analysis, although the bulk SST was used, statistics calculated using the ARC SSTs from 0.2m and 1.5m were not significantly different – the night-time retrievals using two or three channels showed very little change while the statistics for the daytime retrieval at 0.2m displayed the largest difference as would be expected due to diurnal thermoclines.

Regional statistics

There are varying trends in the biases in the different ocean regions. The statistics of the different regions were investigated and the annual means are presented in Figure 2. This plot shows the results from the two channel night-time retrieval in order to include results from ATSR-1 but the three channel retrieval behaves in an almost identical way in the later years.

Generally, the spread in the mean biases and the standard deviations (not shown) from the different regions decrease in more recent years. The transition from the ATSR-1 instrument is quite noticeable with a much larger divergence of the means from the different regions before 1996. Throughout most of the time period, the North Atlantic and Southern Ocean regions tend to have lower mean biases – further evidence that SSTs around the higher latitudes often show cooler biases.

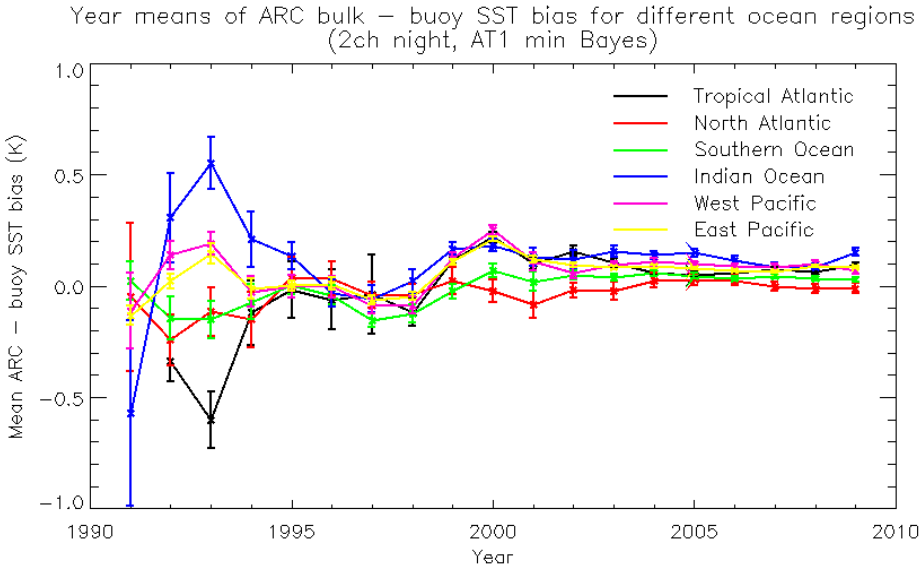


Figure 2. Plot comparing the year mean biases of different ocean regions. Two channel night-time ARC SSTs were used.

Correlations between SST bias and the estimated uncertainty and other meteorological parameters

The biases seen in the ARC SST product were investigated to see what, if any, relationships there were with other meteorological parameters. The results presented here only include ATSR-2 and AATSR data but in each case the trend in the ATSR-1 data was the same or the confidence intervals were too large to discern any dependence. In addition the estimated uncertainty available with each measurement was also assessed.

For the bias as a function of wind speed, there is a slight trend for lower biases at higher wind speeds which is more defined for the daytime data. Thermoclines should have been removed in the pre-processing but this result could be associated with their occurrence at low wind speeds. The wind data used were taken from the ERA-Interim reanalysis rather than measurements taken by the drifting buoys. The *in situ* observations provided few measurements and an uneven spread of wind speeds making it difficult to make any conclusions.

Virtually no trend is observed for the dependence of the two channel daytime retrieval on insolation (provided by the ERA-Interim reanalysis) although a small decrease in bias can be seen for lower solar fluxes. The collocations in these low flux regions are mostly found in the higher latitudes. A stricter threshold to screen potential thermoclines of 0.05K gave no discernable difference in the results.

A small increase in bias in the two channel daytime retrieval is observed for increasing total column water vapour (TCWV). Dependences are not present in the night-time two or three channel retrievals although for both day and night, at very high TCWV the value of the average bias is very unstable. The higher values of TCWV occur mainly in the tropics where the increased cloud cover can make retrievals more challenging. The larger amount of water vapour absorption may cause difficulty in calculating an accurate estimate of radiation from the sea surface. However, no relationship was found in two or three channels retrievals between the bias and the number of cloudy pixels in the field of view within each ARC grid cell. The dependence of the bias on the time difference between the buoy and satellite measurement was also investigated. There is no overall trend in the bias as the time difference was increased.

The relationship between the bias and the estimated uncertainty (random error) of the ARC SST retrieval was analysed. The uncertainty was estimated from a combination of the theoretical performance of the physical retrieval, the number of pixels present and the variability in each cell. The two channel night-time (Figure 3a) and daytime biases remain quite stable up to an uncertainty value of around 0.6K and then start to fluctuate for larger values. This suggests the bias in global SST will be unaffected at least up to estimated uncertainties of 0.7K. However, in the case of the three channel retrieval (Figure 3b) the average bias only remains stable up to uncertainties of around 0.35K before decreasing for values up to around 0.6K and fluctuating beyond that. This suggests more work is needed in estimating the uncertainties of 3-channel retrievals. The locations of the match-ups with high uncertainty values are evenly distributed over the globe and in virtually all cases of uncertainties greater than around 0.32K the cells were adjacent to cloud edges. Removing the ARC SSTs with large uncertainty in the retrieval may improve the accuracy. The large majority of observations have retrieval uncertainties less than 0.4K so there is little impact on the overall statistics. This analysis of the uncertainties provided is a good model for the CCI datasets where it is a clear requirement that uncertainties are provided with all CDRs generated in the CCI.

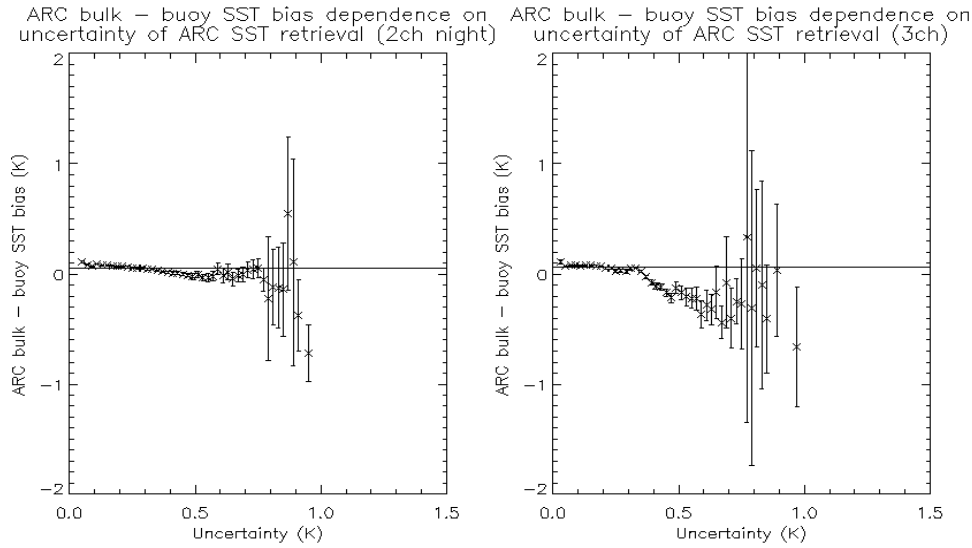


Figure 3(a) Relationship between uncertainty of ARC SST retrieval and the two channel night-time retrieval, (b) Relationship between uncertainty in ARC SST retrieval and the three channel retrieval. Crosses mark the mean for that bin and the error bars show the 95% confidence interval for the mean of that bin. The overall mean of the biases used in the bins is also marked as a horizontal line. No ATSR-1 data were included in these plots.

Three way error analysis

An estimation of the measurement errors using a 3-way comparison as described earlier is a powerful way to estimate the overall mean errors for a particular observation type. The years 2003 to 2009 were chosen for this analysis of the observation errors of the ARC SSTs in order to assess whether there are any trends in the standard deviation of error for any of the instruments. The 2003 results allow comparison with those obtained in O'Carroll *et. al* (2008).

Table 3 shows the standard deviation of errors calculated for the different instruments. The same error is found for the ARC data in 2003 as was found for the AATSR data in the study of O'Carroll *et. al* (2008). Similar errors are also found for the AMSR-E SST in both studies while the error for the buoy SST in this report is slightly lower in comparison. The three-way error analysis was also carried out for all AATSR years except 2002 using the same criteria. The standard deviations of errors are shown in Table 3. The observation errors for the ARC SSTs are consistent over the seven years and do not have any obvious trend. However, for the AMSR-E SSTs there seems to be a slight rise in error as the years progress which could be due to the instrument degrading with age. The buoy SST errors fall slightly in the early years then become more stable. This is an example where access to the level 1 AMSR-E data is important to understand the changes in instrument behaviour. The Data Buoy Cooperation Panel (DBCP) had a campaign to put out over 1250 drifting buoys to improve the network and this work was completed in 2005. The gradual introduction of these new buoys may be the cause of the decreasing error.

Instrument	Standard deviation of error for each year (K)						
	2003	2004	2005	2006	2007	2008	2009
ARC bulk SST	0.137	0.129	0.139	0.137	0.138	0.136	0.134
AMSR-E SST	0.468	0.462	0.462	0.466	0.482	0.489	0.500
Buoy SST	0.189	0.174	0.155	0.152	0.149	0.149	0.153

Table 3. Standard deviation of error for 2003 – 2009 for the ARC bulk, AMSR-E and buoy SSTs determined using the 3-way matchups.

4.1.2 Assessment of pre-final HadISST2

What is HadISST2 and why is it interesting for the international reanalysis community?

The HadISST2 dataset analysed here is a pre-final version that is not in the public domain but was shared with some climate centres for testing and evaluation. To distinguish the draft dataset evaluated here from the final published version a "*p*-" is appended to the name denoting 'pre-final'. The dataset is currently (April 2012) under preparation by the Met Office Hadley Centre and when finished and released will be an update of HadISST1 (Rayner et al., 2003). The *p*-HadISST2 dataset is globally complete at 0.25 degree spatial resolution. The geophysical parameters (essential climate variables) reported in *p*-HadISST2 are Sea Surface Temperature and Sea Ice Concentration for the period 1899 to 2010, i.e. more than 100 years, at daily temporal resolution.

The SST data sources include both in situ observations (version 2.5 ICOADS data) and satellite retrievals (from NASA's AVHRR Pathfinder version 5 data and from ESA's ATSR-2 and AATSR products but not the ARC data described in sec 4.1.1). The bias adjustments for the in situ data account for changes in the measurement method and those for the AVHRR data account for aerosol contamination and diurnal drift by comparison with coincident ATSR and in situ observations and measurements of aerosol optical depth. The HadISST2 algorithm involves large-scale interpolation of SST data using Empirical Orthogonal Functions, blending of smaller–scale spatial variability, time-interpolation from monthly to daily fields, and adjustments to account for sea-ice.

When released, the final HadISST2 product will be available at ECMWF within the frame of the ERA-CLIM project, and will consist of an ensemble of 10 realizations delivered in NetCDF format. The ensemble members represent "equally likely" realizations of the sea surface temperature and sea ice. There is no "control" member that is recognized by the data providers as being the "best estimate", and this is a feature that is in contrast to other approaches to ensemble datasets.

The attempt to represent uncertainty via an ensemble of equally plausible realizations of SST evolution is a novel feature of the HadISST2 dataset. The CMUG work reported here explores the usefulness of this approach, and complements other R&D in this area.

Description	Period	Resolution	Comments
AMIP ensemble	1900-2011	T159L91 (~128 km grid) 10 members	Reference for data assimilation activities (reanalysis production). L91: The topmost layer of the 91 level model is around 0.01hPa.
Reanalysis based on Ensemble of Data Assimilations	As above	T159L91 10 members	Using surface observations only.
Early Decades reanalysis	2 early decades (pre-satellite era) of the 20th century	T511L91 (~40 km grid)	Using all available observations, surface and upper air. One realisation of SST/Sea Ice evolution required as lower boundary condition.
ERA-Interim replacement reanalysis	1979-present (i.e. satellite era)	T511L91	Upgraded assimilation system. Using reprocessed observations where suitable. One realisation of SST/Sea Ice evolution required as lower boundary condition.

Table 4. ERA-CLIM activities which seek to make use of HadISST2.

The preliminary version of the ensemble currently available at ECMWF is being assessed for suitability for use in ERA-CLIM reanalysis (modelling and data assimilation) activities. The primary application of HadISST2 within these activities is as a lower boundary condition for the atmospheric model. The ERA-CLIM reanalysis activities use the atmospheric model in a number of different contexts (see Table 4 for a subset). Note the prominence of century-long ensembles (the AMIP ensemble and the Ensemble of Data Assimilations using surface-only observations) and the attention given to both the pre-satellite and satellite eras (the Early Decades reanalysis and the ERA-Interim replacement).

The international reanalysis community has a tradition of sharing experience and knowledge about datasets. The ECMWF experience with HadISST2 and future CCI datasets will inform/influence the decisions on whether to accept such datasets in other reanalyses (including those to be undertaken in the frame of projects such as MACC-II).

How is ECMWF assessing the *p*-HadISST2 and future CCI products? Why develop a database environment?

As explained above, there is a need to assess HadISST2 for suitability for use as a lower boundary condition for the ECMWF atmospheric model in ERA-CLIM reanalysis (modelling and data assimilation) activities, and assessing the *p*-HadISST2 is a valuable step towards that goal. Part of this assessment comprises a set of Quality Assurance procedures addressing specific questions summarized in Table 5 and described below. It should be noted that Quality Assurance is necessarily influenced by the nature of the product and the intended application. We have formulated the questions in a way that should make them relevant to many ECV products arising from the CCI in future, and described the methods of assessment in a way

that should be generically applicable, but the details may vary from one ECV dataset to another. Furthermore, the assessment procedures invoked here on the *p*-HadISST2 are not intended to be exhaustive but rather a valuable contribution to a wider Quality Assessment that will include complementary evaluation criteria developed in parallel.

Quality Assurance question	Assessment method	Assessment findings	Outcome/recommendations
Does the dataset contain unphysical values?	Examine range of diagnostics to detect unphysical values	Realization 19 has unphysical values, SST exceeding 1500K. Detectable in diagnostics such as Maximum Value or Maximum of Ensemble Mean/Spread (Figure 10).	Correction required, to be discussed with data providers.
Is there consistency with existing reanalysis products?	Compare dataset with a range of diagnostics from existing reanalyses, from decadal to monthly timescales, to identify potential anomalies or inconsistencies	Relative to existing reanalyses, *p*-HadISST2 (realization 103) shows a sharp drop in Tropical SST around 2nd half of 1991 (Figure 6).	Further investigation of the root cause required: could be a physical feature not previously captured in reanalyses, or an artefact in *p*-HadISST2 due to sudden presence of stratospheric aerosol from Mt Pinatubo eruption.
Are the provided uncertainty estimates (if any) consistent with current understanding?	Examine uncertainty diagnostics (ensemble spread for HadISST2, which must first be computed/derived from the provided product)	Sub-monthly variations in ensemble spread are apparent, and challenge current understanding (Figure 8/9).	Further investigation of the root cause required: in particular how the ensemble-spread representation of uncertainty is affected by time-interpolating monthly SST fields to daily resolution.

Table 5. Assessment of p-HadISST2 dataset

To lay the foundations for assessing CCI datasets in future, ECMWF has started to develop a database environment and give an initial demonstration of its usefulness by assessing *p*-HadISST2. Table 6 gives the main objectives of the assessment environment and explains why a database system is the natural choice for its implementation.

The database schema ultimately envisaged for this work is shown in Figure 4. A step-wise incremental implementation is underway, starting with a reduced prototype version for the assessment of *p*-HadISST2. The different datasets held in the prototype database are identified by the Product stream table. A separate geophysical parameter table enables one dataset/product to be associated with multiple ECVs. For the time dimension, a date-time table has been implemented in such a way that it is possible to store time series at various temporal resolutions ranging from hourly through to decadal. It also permits the holding of climatologies on different timescales (annual cycles, seasonal means, multi-year averages etc.). Each time series is associated with a geographical region defined in the region table. Subject to technical constraints of the overall database size, the region could in principle be as small as a single model grid point or satellite pixel; in the current implementation the ingested dataset time series are defined on standard regions ranging from continental-scale land masses

to ocean basins to zonally-averaged climate zones (Tropics, Mid-latitudes, etc.). One essential element upstream of (or as part of) the input side of the database system is the set of tools for computing regional diagnostics from (gridded) ECV products.

Objective	Relevance of database system	Comments
Provide a unifying and easily extensible environment for assessment of multiple datasets including future CCI datasets	1) A well-designed database enables holding diagnostics from multiple datasets and is easily extensible. 2) Assessment methods previously developed in an ad-hoc manner can be unified within a common environment.	1) Effectiveness demonstrated by the prototype database designed/implemented for assessing *p*-HadISST2: the database holds diagnostics from multiple datasets and is easily extensible. 2) Implementation makes use of common tools for comparing many different quantities,
Facilitate comparison against reanalysis diagnostics on climate-relevant timescales (e.g. ranging from daily/monthly to decadal)	Two essential components: 1) a database that holds diagnostics from reanalyses as well as from observational datasets, ranging from daily/monthly to decadal timescales, 2) tools to enable the comparison (starting with visualization and subsequently quantitative metrics)	1) Prototype database already developed during this work. 2) Visualization tools have proved effective for assessment purposes, paving the way for quantitative metrics (Figs 5-9).
Provide user-interactivity through a flexible interface	An interactive database interface gives the user flexibility to decide what to retrieve/plot/evaluate.	1) Web-based interface implemented for prototype database. For examples of the pull-down menus, see Figure 5. Menu selections have been cropped from Figures 6-9. 2) Capability to perform batch processing is also desirable - has been implemented via Python scripts.

Table 6. Advantages of a climate data record monitoring system.

Details of the Quality Assessment of *p*-HadISST2

In Table 5 above we stated the specific questions to be addressed by Quality Assurance procedures, and summarized the main findings. Here we provide the details that support those findings. The assessment addresses both the decadal timescale and the monthly/annual timescale. In the plots where a single ensemble member is shown together with equivalent reanalysis quantities, the member shown is realisation 103 which has been chosen (at random) for control simulations in ERA-Clim.

Figure 5 shows the *p*-HadISST2 time series (member 103) over its full temporal extent, 1899 to 2009/10, with the daily values averaged to monthly means and averaged over the Tropical Oceans (20S to 20N excluding land). Over plotted are 3 reanalyses (ERA-40 and ERA-Interim from ECMWF, JRA-25 from JMA). Overall the agreement is good. The period from 1960 to 2000 shows that *p*-HadISST2 is repeatedly colder at its annual minimum, by a few tenths of a Kelvin. Also *p*-HadISST2 exhibits a warming trend over the past century by

around 0.5K, but confidence in any derived trend estimate needs to take into account a wider appreciation of the uncertainties in the dataset. This is beyond the scope of the current work on precursors.

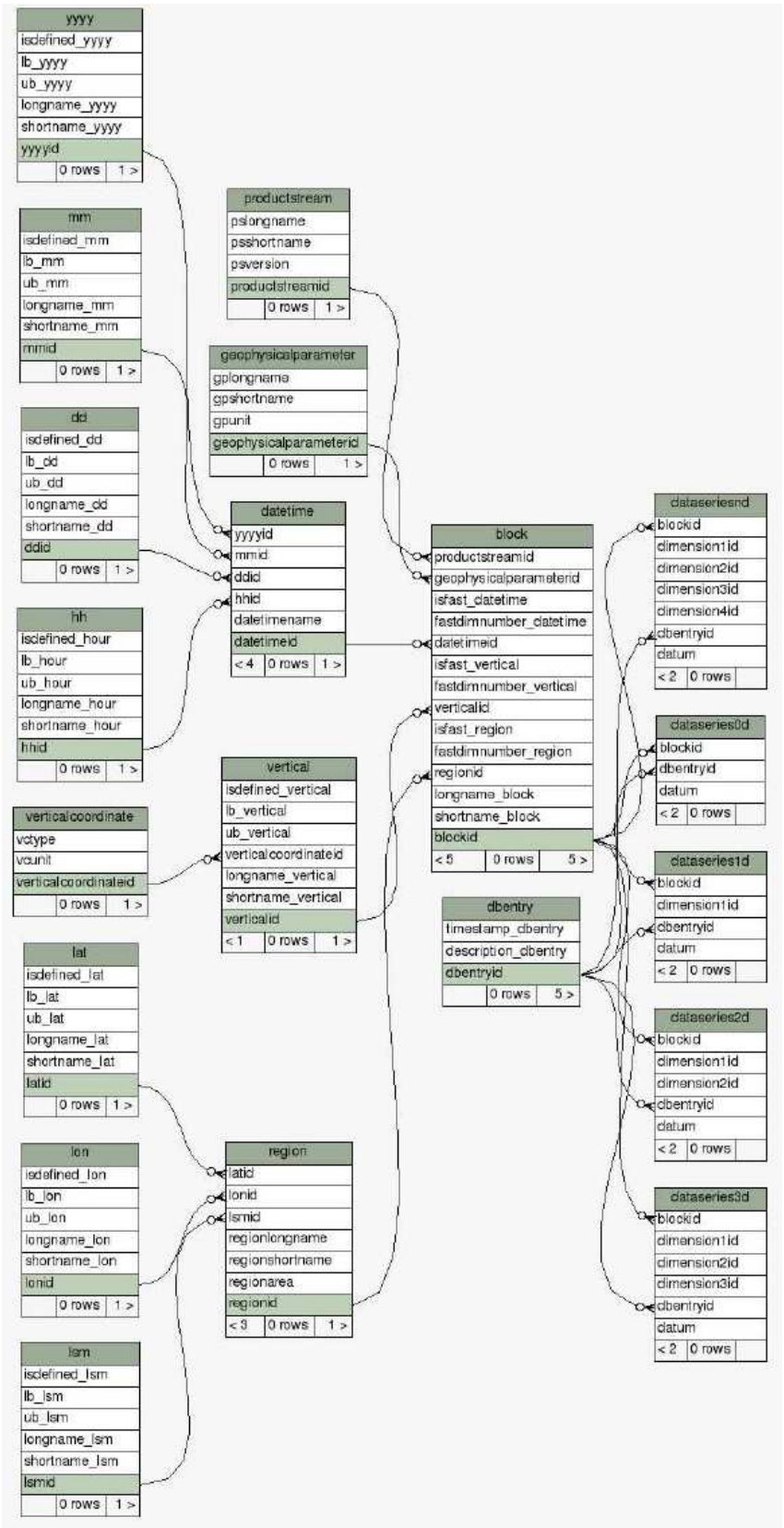


Figure 4. Database schema for climate data record monitoring

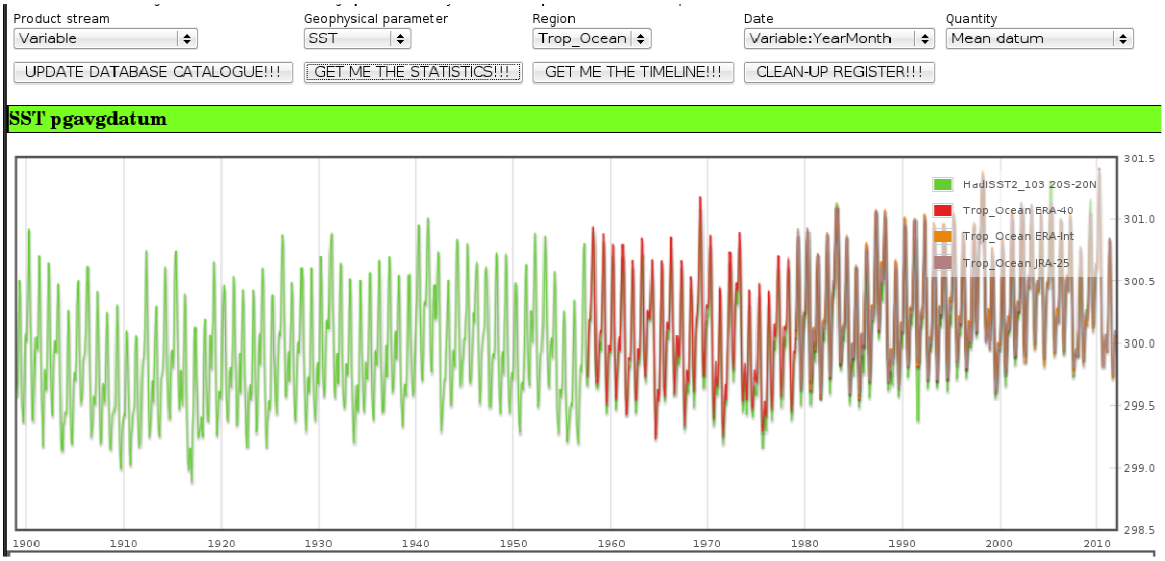


Figure 5. p-HadISST2 time series for tropical ocean compared with recent reanalyses. The web-based interface to the climate monitoring database implements menus to select Geophysical parameter and region etc., which are shown at the top of this Figure but omitted from subsequent one.

Figure 6 again shows *p*-HadISST2 (member 103) and the 3 reanalyses, but now showing the anomalies of each with respect to its own annual climatology (defined for the period 1979 to 2010), again computed for the monthly mean time series over the Tropical Oceans. Relative to the 3 reanalyses, *p*-HadISST2 has a stronger El Nino in 1998, and some cooler periods in 1986, 1988 and 1991. Overall the anomaly differences are smaller than the interannual variations due to El Nino/La Nina. The *p*-HadISST2 cooler period in 1991 coincides with an increase in stratospheric aerosol from the eruption of Mount Pinatubo, and ECMWF's own experience with reanalysis during this period suggests that it warrants further investigation. In principle, increased aerosol could lead to a cooling of the oceans, however the sharpness of the cooling in *p*-HadISST2 and its subsequent recovery are relatively rapid and could indicate a deficiency in the aerosol corrections of the SST satellite retrievals in this period. At the very least, this period provides an opportunity for CCI SST to demonstrate cross-ECV consistency with aerosol products and sub-surface ocean data.

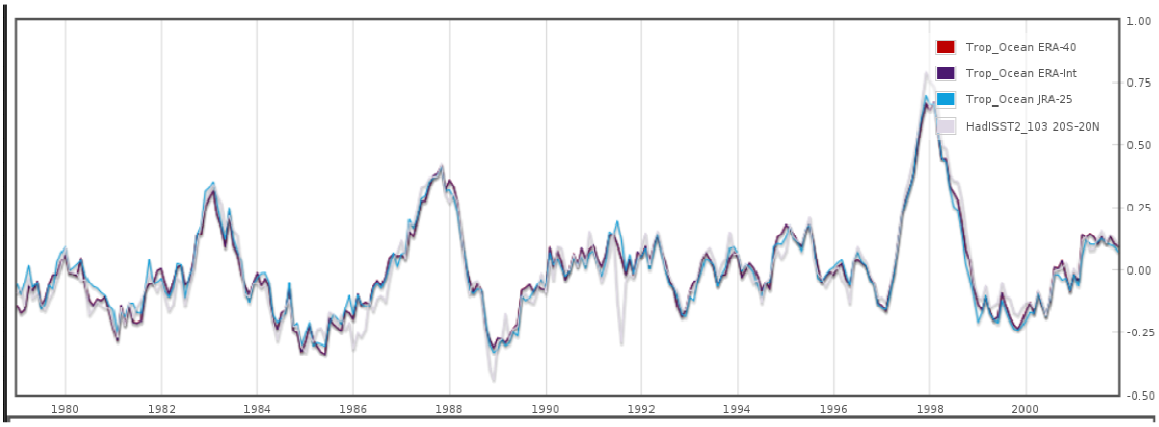


Figure 6. Time series of SST anomalies for p-HadISST2 and several reanalyses for latitudes from 20S to 20N.

Figure 7 shows the annual cycle of *p*-HadISST2 over the Tropical Oceans, averaged over 3 distinct periods: 1899 to 1920 (orange, "early-century"), 1939 to 1960 (green, "mid-century"), and 1979 to 2010 (purple, "later-century"). The horizontal axis is the month of the year (1=January, etc). A systematic increase in SST is apparent in moving from one period to the next (corroborating the description of Figure 5 above) but the amplitude of the annual cycle is broadly similar. For all 3 periods, the maximum of the annual cycle occurs in April. The minimum of the annual cycle occurs in September for the early period but in August for the later period. This could reflect some climatic change, or alternatively some uncertainty in the dataset (e.g. the generally greater uncertainty in earlier periods).

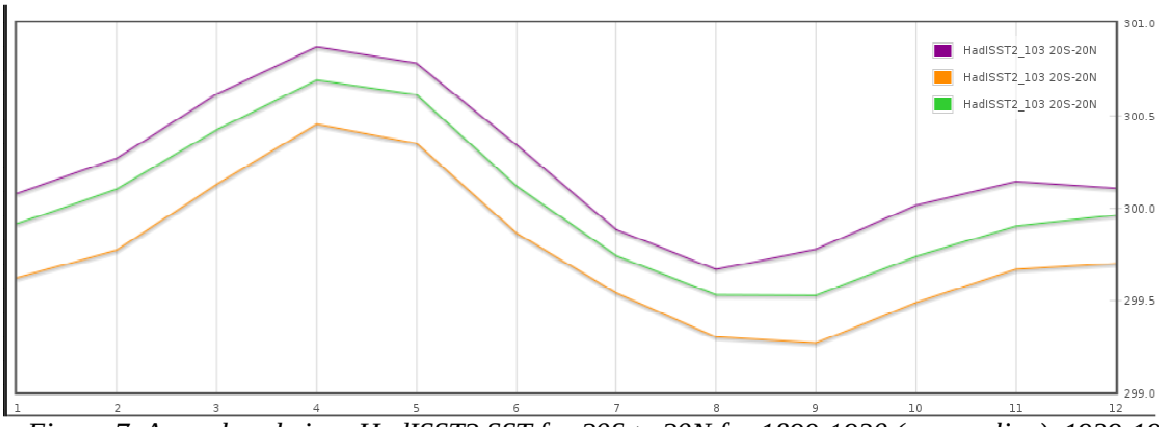


Figure 7. Annual cycle in p-HadISST2 SST for 20S to 20N for 1899-1920 (orange line), 1939-1960 (green line) and 1979-2010 (purple line)

Figure 8 shows the spread (standard deviation) of the 10 member *p*-HadISST2 ensemble for the period 2000-2009. The spread (measure of uncertainty) is computed at each grid point, and then spatially averaged over the Tropical Oceans. Note the elevated values for April 2001 and February 2005, which are comparable to the spread values before 1985 (not shown). Thus, they almost certainly correspond to a reduction in the satellite data used as input to *p*-HadISST2. The purple dots show the spread on the 16th of each month, which is close to the nominal date of the monthly *p*-HadISST2 analysis. The green dots show the spread on the 1st of each month, and are systematically lower than for the nominal analysis date. This feature seems to be the consequence of the procedure for producing daily *p*-HadISST2 fields from the

monthly analyses: one step of this procedure is a linear time interpolation which can only reduce the spread. A different result can be expected if the analysis procedure were shifted by half a month, i.e. monthly analysis output for the first of each month and linear interpolation to the 16th. This raises some questions of consistency between the use of spread as a representation of uncertainty and its compatibility with linear time interpolation.

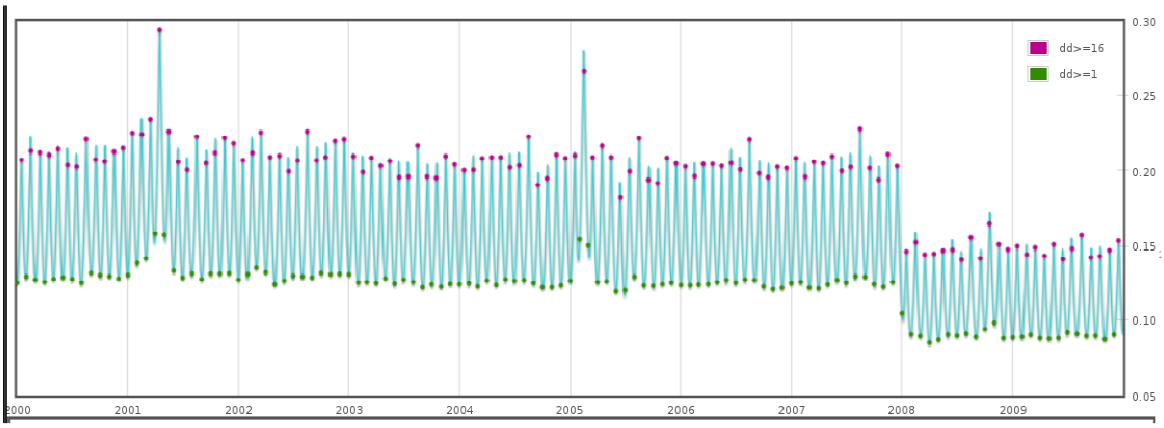


Figure 8. Ensemble spread for p-HadISST2 SST from 20S-20N from 2000 to 2009. Purple points are spread on 16th of month and green dots are the spread on the first of the month.

Figure 9 is similar to Figure 8 in that it shows the spread (standard deviation) of the 10 member *p*-HadISST2 ensemble, but now for the single year 2007 and averaged over the Baltic Sea. The reduction in spread on the 1st of each month is again apparent, especially in the ice-free months (July to November). A further feature is evident in other months, namely sharp jumps in spread on the 1st. This does not occur in ice-free regions such as the Tropical Oceans (figures not shown), and can thus be attributed to the way the *p*-HadISST2 SST product takes account of sea-ice changes on a monthly rather than daily basis. It should be noted that the 10 ensemble members currently have identical sea-ice fields so that uncertainty in the sea-ice distribution is not transferred to spread in the SST ensemble. This may change in future versions of *p*-HadISST2.

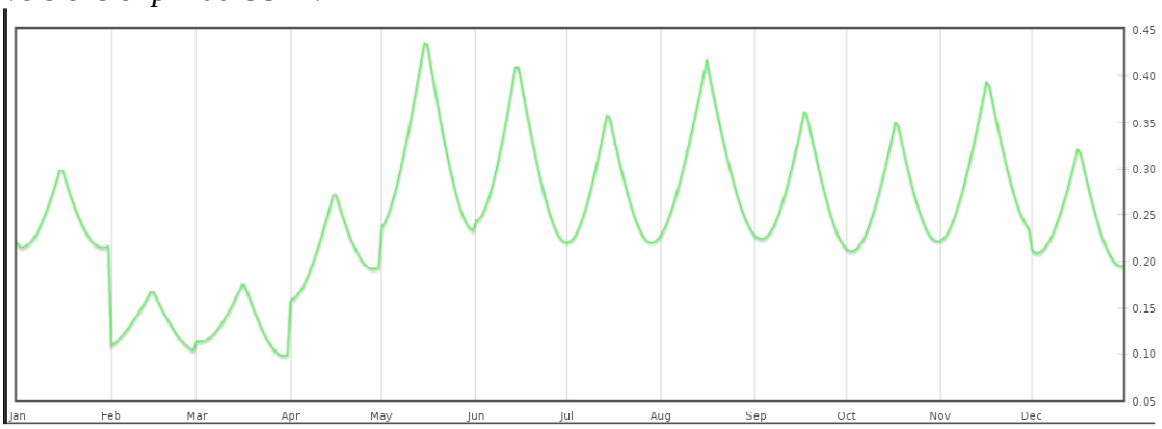


Figure 9. Ensemble spread of p-HadISST2 SST for the Baltic Sea in 2007.

Figure 10 is a map of *p*-HadISST2 ensemble spread (standard deviation) for 15th October 2010. Careful inspection reveals that a single grid point in the eastern Atlantic is anomalous –

the maximum spread exceeds 382K. The anomalous grid point is also visible in maps of the ensemble mean which has a maximum value exceeding 426K (figure not shown). These anomalies are traceable to an anomaly in Ensemble member 19, where the maximum value for SST exceeds 1573K and is clearly unphysical. Note that single-grid point anomalies such as these do not show up in the regional averages currently contained in the prototype Climate Monitoring Database and were in fact detected by complementary assessment work conducted by Hans Hersbach (ECMWF) in the frame of ERA-CLIM. Nonetheless, this highlights the need for further Quality Assurance checks on the data products, in this case on the minimum and maximum values, and for automatic alerts when reasonable thresholds are exceeded. Arguably, many such checks can and should be part of the Quality Assurance implemented by data providers before release of the products. In practice it is difficult for data providers to guarantee that their checks will be sufficiently comprehensive, which means that the Quality Assurance must also be implemented by product users.

Complementary assessment results of *p*-HadISST2

As was mentioned above, the database environment used to assess *p*-HadISST2 is intended to be complemented by other assessment activities. Some of these are conducted in the frame of the ERA-CLIM project itself, resulting in additional findings:

One encouraging development is that a number of lakes and smaller seas are now resolved: specifically, the US Great Lakes; the Baltic; the Caspian Sea; and the Black Sea, including its northern extension, the Sea of Azov.

By visual inspection of maps, an anomaly was identified in an initial test product, consisting of an evolving patch of ocean close to Antarctica where SST was incorrectly set to missing values during the period February to June 1956. Following early feedback to the data providers, the anomaly was fixed in subsequent deliveries.

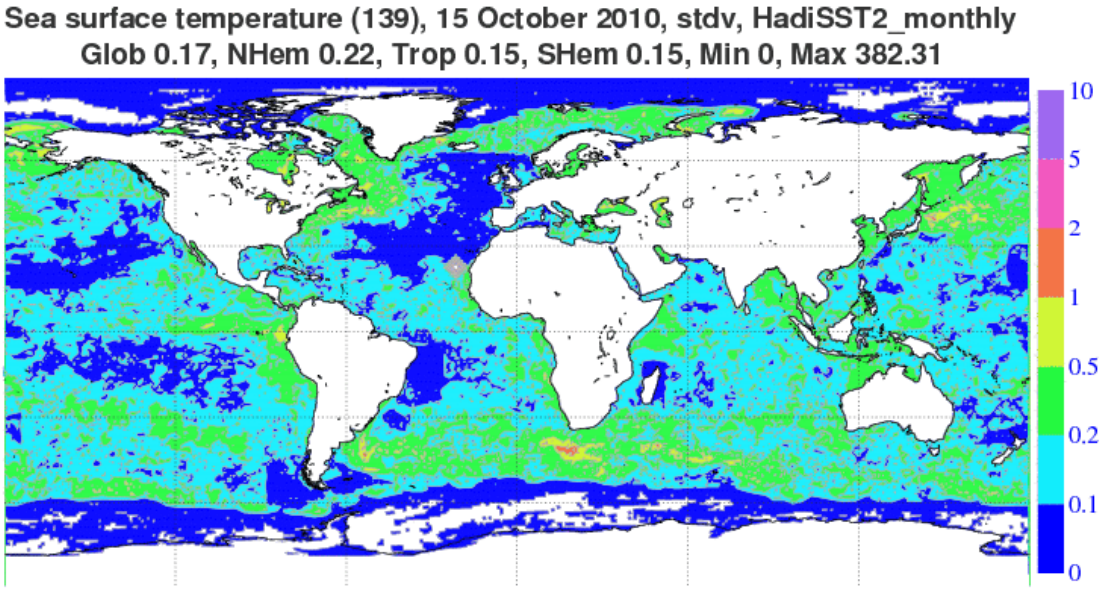


Figure 10. Map of ensemble spread in p-HadISST2 SST for 15th Oct 2010.

Such findings remind us that Quality Assurance benefits from a multiplicity of approaches, including visual inspection of maps.

Outlook for future CCI products for SST and other ECVs

This work with the *p*-HadISST2 "precursor" has demonstrated different types of anomalies that can/do arise in ECV datasets. It provides an opportunity for the CCI_SST team (and the other ECV teams) to benefit from this experience, by implementing their own Quality Assurance procedures to minimize the amount/severity of the anomalies that are present in the products they deliver to users.

One anticipated application for CCI_SST products is that they will be accepted as input for future versions of HadISST. In this context, there is a significant opportunity for CCI_SST products to demonstrate that they reduce uncertainty in future HadISST products, as represented by ensemble spread (see above) or otherwise.

The assessment above noted the uncertainty regarding the root cause of the sharp drop in *p*-HadISST2 SST in the period following the 1991 Mt Pinatubo volcanic eruption. This is another opportunity for sufficiently good CCI_SST products to reduce uncertainty (in its qualitative sense) and thus benefit the climate modelling community. It is also an opportunity to demonstrate cross-ECV consistency with aerosol products and sub-surface ocean data.

Lessons learned for the quality assurance and evaluation process, future development of the database (schema and tools).

The database environment has proven to be a useful approach/methodology for conducting the Quality Assurance procedures. It provides a unifying and easily extensible environment for a range of assessment methods that, in the past, were typically implemented on a dataset-by-dataset and project-by-project basis.

The precursor assessment confirms our view that the database environment is an essential component of ECMWF's assessment of future CCI datasets. The Quality Assurance that it provides, and its ability to facilitate confrontation/intercomparison with other reprocessed datasets, will be an important filter before accepting datasets for use in reanalysis and climate modelling applications. To realize its full potential CMUG recommends further development of the database system ranging from:

- evolution of database schema,
- ingestion of additional datasets and/or a user-upload capability,
- development of further database tools for visualization and time series analysis, including homogeneity-testing

It is of course desirable to implement Quality Assurance as close as possible to the source of the data, and ideally as part of the production process. CMUG recommends that further discussions take place to ensure that this is adequately addressed by all ECV teams.

4.2 Ocean Colour

We have assessed the impact of assimilating the ESA GlobColour chlorophyll data, derived from ocean colour, into FOAM-HadOCC, a physical-biogeochemical coupled model run pre-operationally at the UK Met Office.

The FOAM (Forecasting Ocean Assimilation Model) system is based on the NEMO physical model (for full details the reader is referred to Storkey *et al*., 2010). FOAM routinely assimilates remotely sensed (surface only) and *in situ* (surface and vertical profiles) observations of temperature, salinity, sea surface height and sea ice concentration. The data assimilation scheme is of optimal interpolation (OI)-type, and is described in detail in Martin *et al.* (2007).

The biogeochemical component of the coupled model is the Hadley Centre Ocean Carbon Cycle Model (HadOCC; Palmer and Totterdell, 2001). It is a relatively simple nutrient, phytoplankton, zooplankton and detritus (NPZD) model, which also includes dissolved inorganic carbon (DIC) and alkalinity. Recent upgrades to this model include the use of a variable carbon to chlorophyll ratio based on Geider *et al.* (1997).

The coupled model assimilates chlorophyll derived from the level three merged ocean colour data provided by GlobColour. The chlorophyll observations used are global, daily averaged fields (with associated error estimates and confidence flags) and they are assimilated using the nitrogen balancing scheme described in Hemmings *et al.* (2008), which directly updates all (observed and unobserved) biogeochemical model state variables. The observation operator performs a comparison between observations and model values at the observation time by using the FGAT (First-Guess at the Appropriate Time) technique, and this information is very useful for verifying the biological model, in addition to being used in the assimilation. For the merged level three GlobColour products, where no time information is supplied, the chlorophyll observations are taken to be valid at 12:00 UTC.

In order to assess the impact of the biological assimilation, we have performed two short hindcasts (from January to December 2008), after spinning the model up for one year (from January to December 2007). The first hindcast is the control run (hereafter referred to as “Control”) and did not assimilate any chlorophyll data. The second hindcast assimilated the derived chlorophyll data from GlobColour (hereafter referred to as “Assim”).

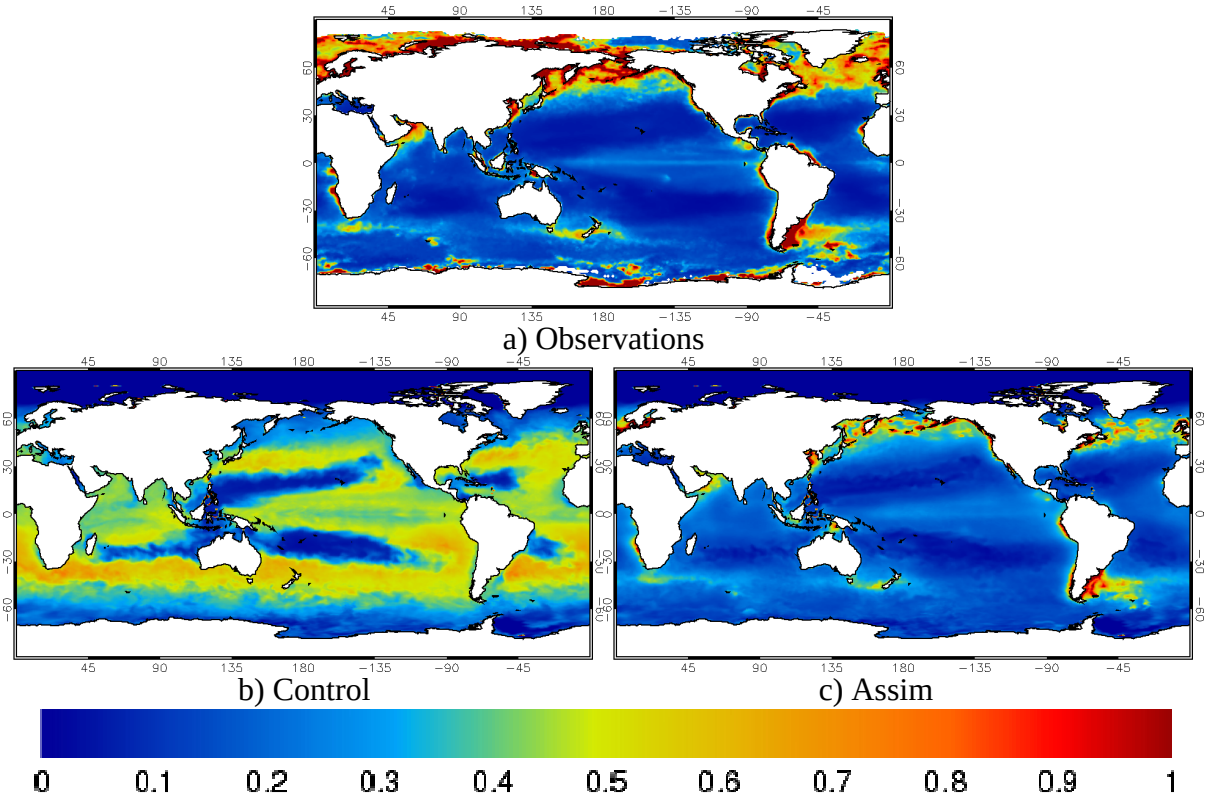


Figure 11. Mean surface chlorophyll (mg m^{-3}) for 2008.

Maps of annual average surface chlorophyll for Control, Assim and the GlobColour observations are shown in Fig. 11. In this simple visual comparison it can be clearly seen that Control is very different from the observations, whereas Assim matches them much more closely, in terms of both spatial pattern and magnitudes. In this sense, the assimilation can be considered a success. Control has too much chlorophyll across most of the ocean, but too little chlorophyll in the Brazil-Malvinas confluence off the Patagonian coast, as well as north of about 50°N. This is accentuated in Fig. 11 because most of the observations at high northern latitudes are taken during the northern hemisphere summer. An in-depth discussion of the reasons for these biases is outside the scope here, but the overestimation of chlorophyll in most regions is linked to excess nutrient concentrations at the surface. The chlorophyll assimilation is able to counteract this bias somewhat, propagating the increments such that model chlorophyll concentrations are either increased or decreased in a realistic manner. Assim is not a perfect match for the observations, however chlorophyll patterns in regions such as the Brazil-Malvinas confluence, which are not reproduced by Control, are captured well by Assim.

Figure 12 shows time series of daily mean global bias and Root Mean Squared Error (RMSE) for Control and Assim. Included for comparison are the equivalent mean absolute and root mean squared errors of the observations themselves, calculated from the values given in the GlobColour products. It is clear that both model runs have too much chlorophyll compared to the observations. However, both the bias and RMSE are much lower for Assim than for Control, which indicates that the assimilation is having a positive impact on the modelled

chlorophyll concentrations, as intended. The mean global bias for 2008 is 0.398 $\log_{10}$(mg m^{-3}) for Control and 0.119 $\log_{10}$(mg m^{-3}) for Assim. The mean global RMSE is 0.586 $\log_{10}$(mg m^{-3}) for Control and 0.314 $\log_{10}$(mg m^{-3}) for Assim. The correlation is also improved, from 0.261 for Control to 0.619 for Assim. This improvement is immediate, with the error considerably reduced after only a single day of assimilation. The error for Assim remains lower, and fairly constant, throughout the year, suggesting that Assim is performing well at capturing the seasonal cycle.

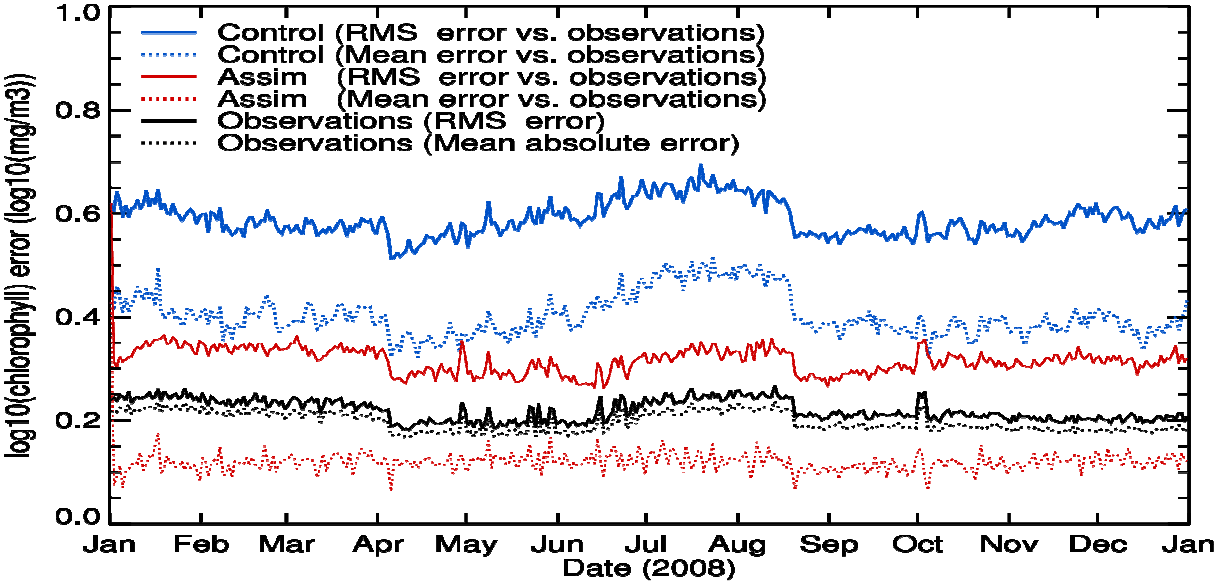


Figure 12. Time series of global model and observation error for 2008. The solid lines represent RMSE, the dotted lines represent bias. The blue and red lines represent the error in Control and Assim respectively when compared to the GlobColour observations on the observation operator step prior to assimilation each day. The black lines are the observation errors specified in the GlobColour files. Mean absolute error is given for the observations because the signs of the errors are not known.

A Taylor plot for $\log_{10}$(chlorophyll), using the same model-data comparisons as Fig. 12, is shown in Fig. 13. This type of plot provides a way to show the improvement of the fit to the observations before and after assimilation. As well as a global average, Fig. 13 provides a comparison for different regions, to see how the assimilation affects the model in each ocean basin. Across all regions, both unbiased RMSE and correlation are improved in Assim, with similar values obtained in each basin, indicating that Assim has comparable skill across the entire model domain, which is less clearly the case for Control. However, whilst the unbiased RMSE and correlation are universally improved, the normalised standard deviation generally remains similar, and is even made worse in some regions, including for the global average. In all cases the standard deviation is too low for Assim, suggesting that the assimilation may be smoothing out too much of the variability in the model.

The assimilation has clearly improved the model's simulation of surface chlorophyll compared to the assimilated observations, throughout the year and across all ocean basins. It has also considerably improved the bias, RMSE and correlation with the GlobColour observations compared to a control run. This improvement was immediate, and sustained over

the entire year and in every ocean basin. Errors against independent in situ observations were also reduced, and there was evidence of improvement in nutrient concentrations, zooplankton biomass and sea surface pCO_2 (not shown here).

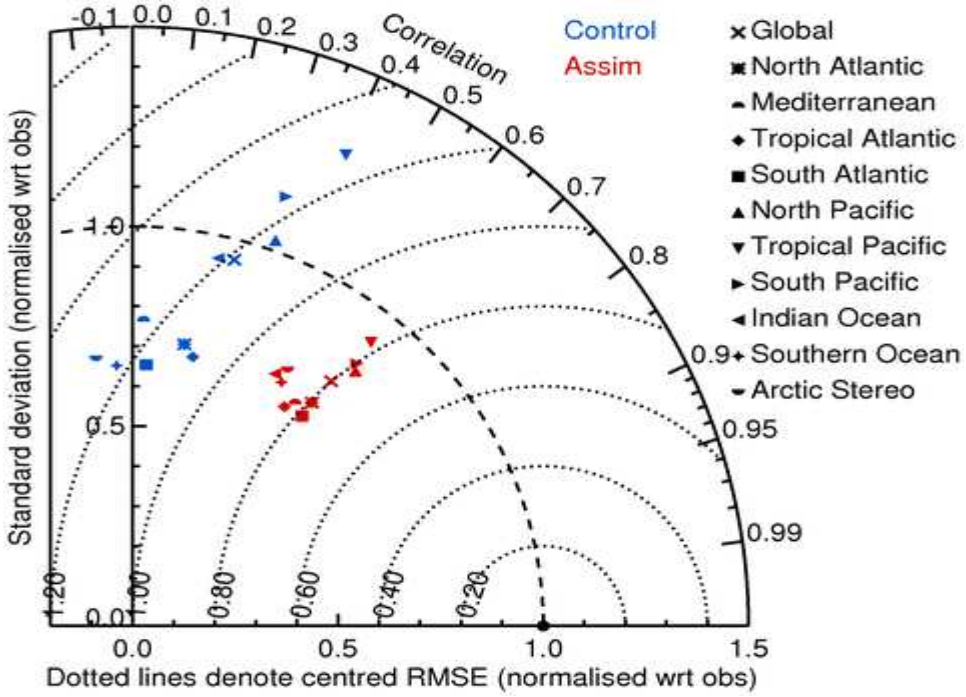


Figure 13. Taylor plot showing skill in each ocean basin for Control and Assim when compared to the GlobColour observations on the observation operator step prior to assimilation each day.

4.3 Sea Level

The added value of a satellite-derived ECV within the context of climate model evaluation depends on the ability of the climate model to reproduce this variable. The model performances must be accurate enough in terms of temporal and spatial mean and/or variability in order to appreciate the increment between a new satellite product and its precursors. As a first step of the development of a methodology to assess the SSH CCI record, we have started an assessment of the performances of a coupled regional climate covering the Mediterranean Sea with a precursor SSH dataset.

The choice of a regional climate model is guided by the fact that one potential improvement of the SSH CCI record comes from the resolution of the gridded products that should reach 25km. This resolution is more compatible with the resolution of a regional climate model than with those of current atmosphere-ocean general circulation models. The choice of the region is guided by the fact that this region will be the focus of the MedCORDEX sub-project of the CORDEX (A COordinated Regional climate Downscaling Experiment) international simulation exercise. Within this context simulations of the last decades will be performed with regional climate models coupled to Mediterranean Sea models. This region will also be the focus of the so-called HyMEX experiment (2011-2015) aiming at studying the hydrological cycle of the Mediterranean area. Within this context, the observational network, including over the Mediterranean Sea, will be reinforced giving access to in-situ independent data to compare with the models outputs and the ECV datasets.

The coupled regional model used for the assessment is the coupled ALADIN-Climat/Nemomed8 model from CNRM (Centre National de Recherches Météorologiques, Météo-France). At this stage, only the Mediterranean Sea component Nemomed8, that includes a calculation of the free sea level surface, with a horizontal resolution of 1/8°, has been used in the confrontation exercise. This model is a regional adaptation of the NEMO global ocean model mainly developed at IPSL (Sevault et al., 2009).

The precursor is the SSALTO/DUACS SSH (Dibarboure et al, 2009), combining altimetric data from several satellites (Topex/Poseidon, ERS-1/2, Jason-1, Envisat and OSTM/Jason-2). This dataset was assimilated in the MERCATOR ocean assimilation system to produce the GLORYS1v1 ocean reanalysis spanning the period 2002-2009 at a resolution of ¼°. The comparison of the simulated sea level to the precursor is here only made through the GLORYS1v1 SSH. A comparison with an independent SSH product inferred from the COMBINE oceanic re-analysis (Balmaseda et al, 2010) is also made. This last re-analysis was performed with the ECMWF ocean assimilation system for the period 1958-2008 at a resolution of 1°. There is no assimilation of satellite-derived SSH in it.

The mean sea level over the Mediterranean Sea inferred from the two re-analyses, at the monthly time scale, is reproduced in figure 14. Also shown on this figure is the SSH inferred from a recently produced new version of the GLORYS ocean re-analysis (GLORYS2v2) that includes some improvements and covers the whole period of satellite altimetry missions (since 1992). However as this last re-analysis is only available for a few months, it couldn't be used to constrain the Mediterranean Sea model at its boundary (see below). It is shown in order to give some idea of the expected impact on the SSH of an update of the assimilation system, including an improved altimetric dataset.

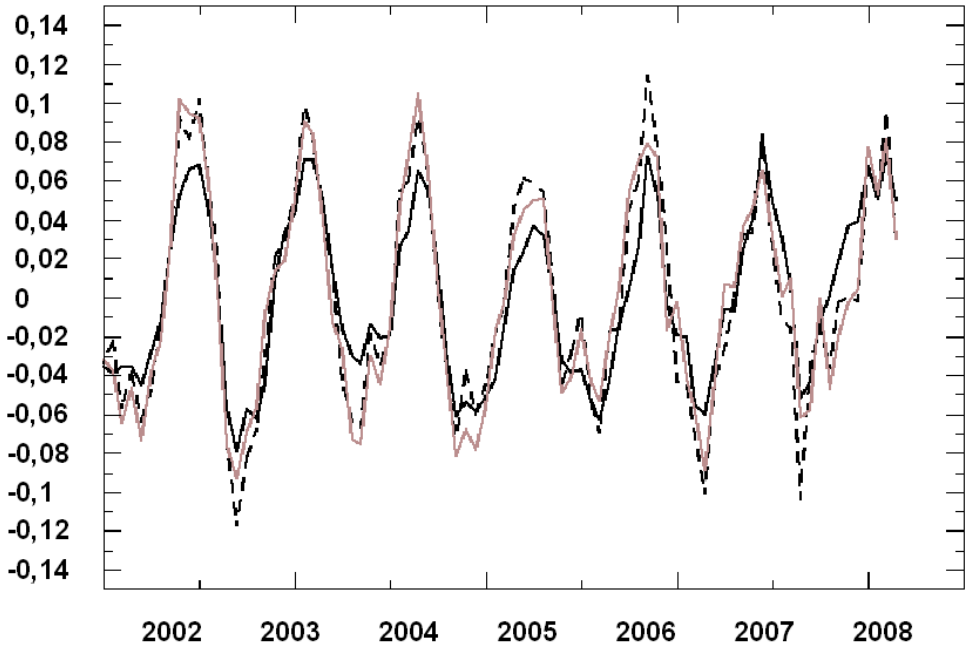


Figure 14: Averaged SSH over the Mediterranean basin from COMBINE (black solid line), GLORYS1v1 (black dashed line) and GLORYS2v1 (brown solid line) ocean reanalyses.

The plot shows a very close agreement between the three re-analyses at the interannual time scale. The correlation coefficients between the two GLORYS reanalyses is equal to 0.96 and it is equal to 0.92 when the COMBINE re-analysis is compared to either one of the two GLORYS reanalyses. Some more significant differences can be noted when considering the mean seasonal cycle that is close between the two GLORYS reanalyses (respectively 14.27m and 13.82m for version 1 and 2) but is reduced with the COMBINE reanalysis (11.07m). The impact of the satellite product is thus clearly apparent even on this aggregated indicator.

The Nemomed8 model was integrated for the 2002-2008 period using the ERA-interim atmospheric reanalysis to specify the radiative and non-radiative air-sea fluxes. Two simulations of this kind were performed, one for which the simulated SSH is nudged towards the GLORYSv1 SSH in a buffer zone located in the near Atlantic, and one for which this nudging is made towards the COMBINE SSH. The two mean SSH over the Mediterranean Sea inferred from these two simulations are compared to the mean SSH of the two corresponding re-analyses in Figure 15 (also at the monthly time scale).

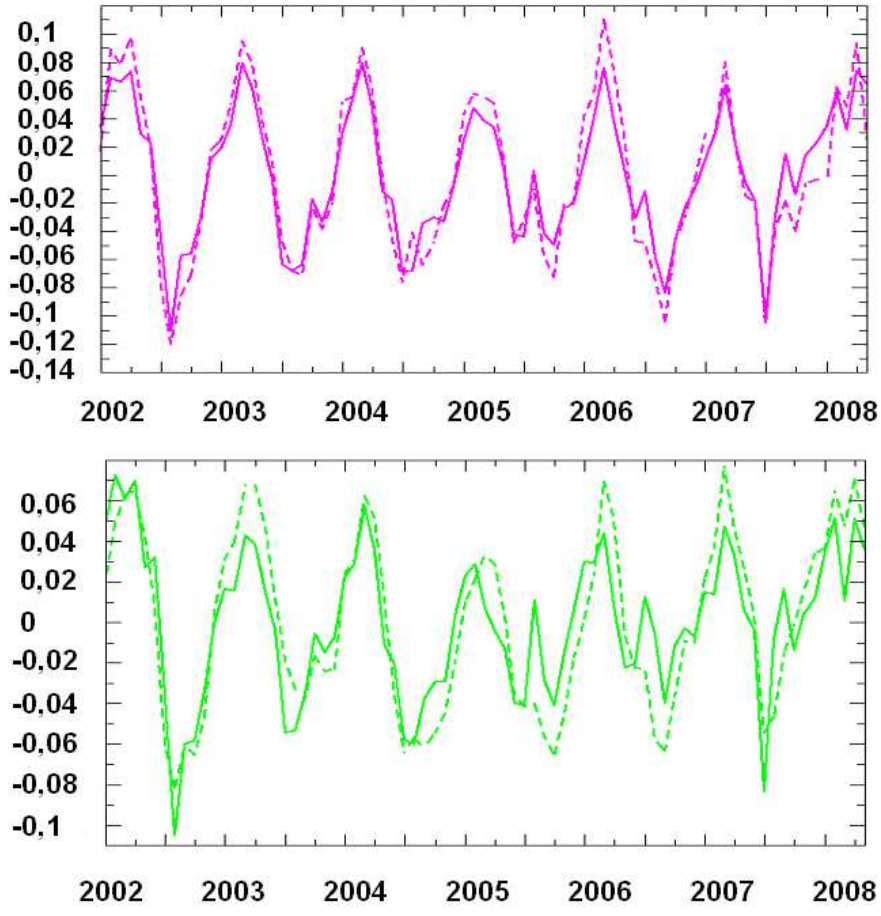


Figure 15: Averaged SSH over the Mediterranean basin from NEMOMED8 simulations (solid lines) constrained over the near Atlantic by GLORYS1v1 (pink) and COMBINE (green) reanalyses and from the corresponding reanalyses (dashed lines).

The agreement between the simulated mean SSH, constrained towards the GLORYSv1 outside the Mediterranean Sea, and the GLORYSv1 mean SSH is very good (upper panel).

The correlation coefficient between the two is equal to 0.95. This gives some confidence in the model results even if this only a preliminary comparison that needs to be pursued with more complete statistics accounting for the spatial variability. The agreement between the model and the reanalysis is significantly lower when the model is constrained with the COMBINE SSH at its boundary (lower panel). The correlation coefficient in this case reduces to 0.86. This result suggests that a lower physical content in the reference variable (here the COMBINE SSH doesn't include the information coming from the altimetry) may result in a degradation of the scores of the comparison between the simulated variable and this reference. If this is confirmed, this gives more confidence in the use of the model as a tool to assess different observational products.

This is of course a very preliminary analysis that needs to be completed with other diagnostics. The next step will consist in confronting the model outputs directly with the satellite product (not through the MERCATOR assimilation system). This next step comprises a comparison with coupled simulations performed with ALADIN-Climat/Nemomed8 in the context of MedCORDEX. This will make possible in this case to assess the consistency of CCI SST and CCI SSH by jointly analysing the agreement of the simulated variables with the two ECVs.

4.4 Clouds

Here we consider the use of precursor data sets for evaluating climate model simulations of cloud properties. The focus is on cloud parameters of specific relevance to the Earth Radiation Budget (ERB) and on the evaluation of global climate models. It should be noted that the cloud CCI data set will include a range of cloud properties, e.g. areal coverage, cloud top altitude, ice/water path and droplet size. As a result it is necessary to consider more than one precursor data set, the choice depending on the parameter being considered (see below).

In relation to the CCI clouds the ISCCP data set has a similar scope and aims and has several advantages as a precursor:

- It provides a long record, from 1983 to the present, allowing both climatologies and interannual variability (e.g. ENSO) to be examined.
- Its use for climate model evaluation is mature and its strengths/weaknesses are understood.
- Directly comparable diagnostics are available in the new generation of CMIP5 climate models.
- It is independent of the inputs to the proposed CCI clouds products.

In what follows we consider CCI clouds products as described in the current version of the PSD (dated 6th April 2011). It should be noted that there is no directly comparable precursor of the proposed merged products, i.e. the ATSR-MODIS-AVHRR Level 3c product. We therefore consider single-sensor products only. The aim here is to illustrate how these single sensor products could eventually be used for model evaluation purposes.

The most basic cloud parameter one can consider is the total cloud coverage, although by itself its utility as a diagnostic of model performance is, of course, somewhat limited. Figure 16 compares the observed annual mean ISCCP total cloud amount with the five models

currently available in the CMIP5 archive. These climate models have all used the COSP simulator (Bodas-Salcedo et al., 2011) to derive a direct equivalent of the ISCCP-retrieved cloud amount. This is important because a satellite-retrieved cloud amount will not correspond to any individual model's standard cloud amount parameter, making the comparison difficult to interpret. It also allows the observational data to be used for making consistent comparisons between different models.

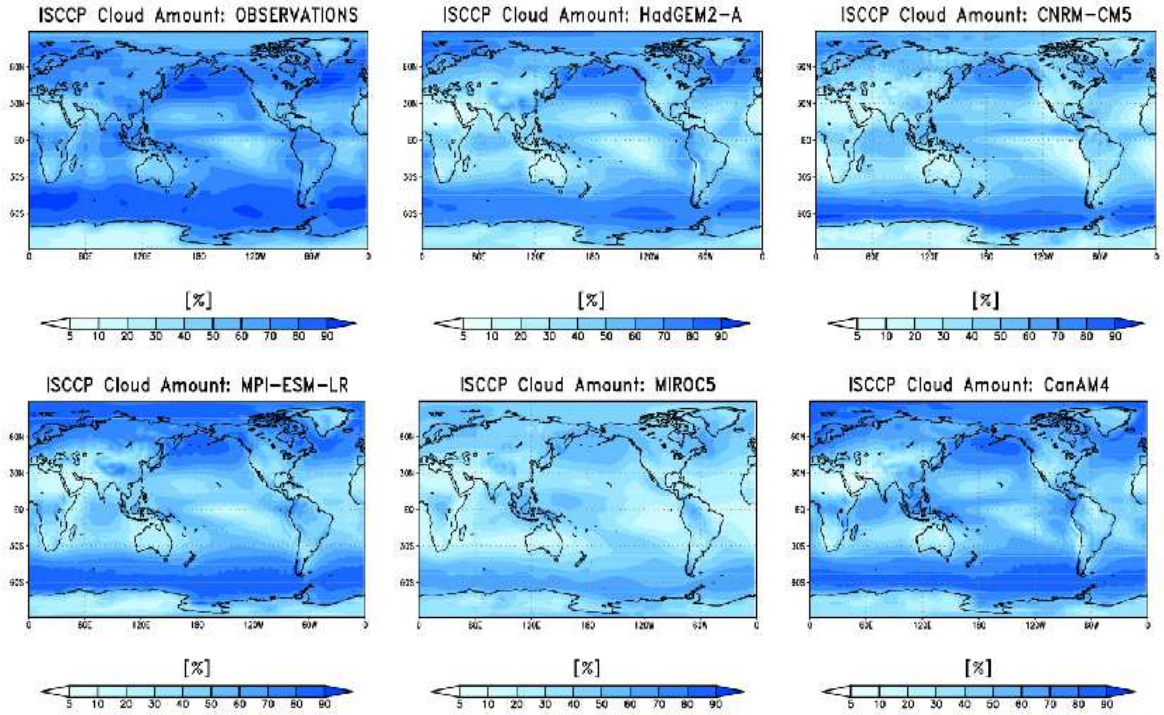


Figure 16: Comparison of the observed annual mean ISCCP total cloud coverage (%) with the five climate models currently available in the CMIP5 archive. The model simulations are atmosphere-only models forced with the observed SSTs and sea-ice distributions for 1979-2008. The COSP simulator has been employed in these models to enable a direct comparison with the ISSCP cloud retrieval.

Differences between the observations and the models are most apparent over the mid-latitude oceans in both hemispheres, the ITCZ and SPCZ in the tropics, the sub-tropical marine stratocumulus regions off the west coast of continents (e.g. South America and California) and over land regions generally. In all cases the models underestimate cloud coverage relative to ISCCP. This lack of cloud is compensated by an overestimate of the cloud albedo in the models (Fig. 17). Such compensations of errors enable models to reproduce the observed cloud radiative effects much better than might be expected given these deficiencies (e.g. Martin et al., 2006). As an example, Fig. 18 shows the comparison of the models' shortwave cloud radiative effect (SW CRE, which depends on both cloud amount and cloud albedo) with data from the NASA CERES instrument. The differences with the observations are clearly much smaller than the errors in the cloud albedo would suggest. In a similar manner the relationship between the simulations of cloud top height – and of cloud amounts at the three proposed vertical layers – and the longwave CRE can also be examined (not shown).

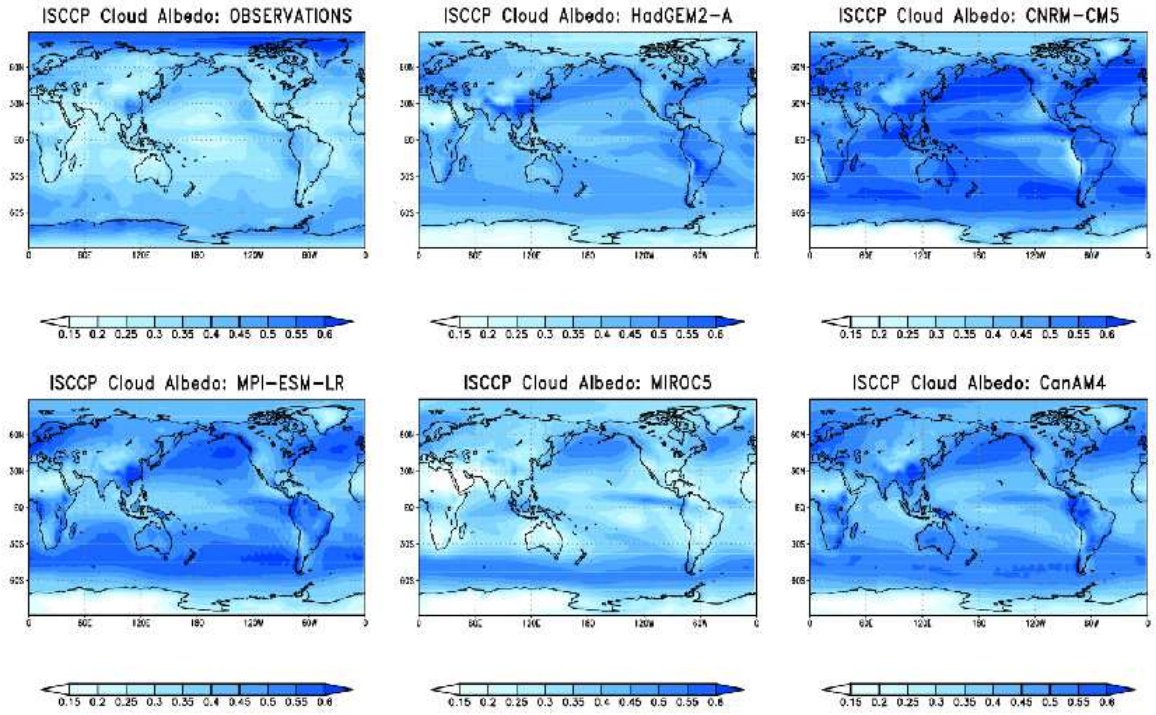


Figure17: As Fig.16 but for ISCCP cloud albedo (unitless). Again, the COSP simulator has been used to calculate a direct model equivalent of the ISCCP-retrieved cloud albedo.

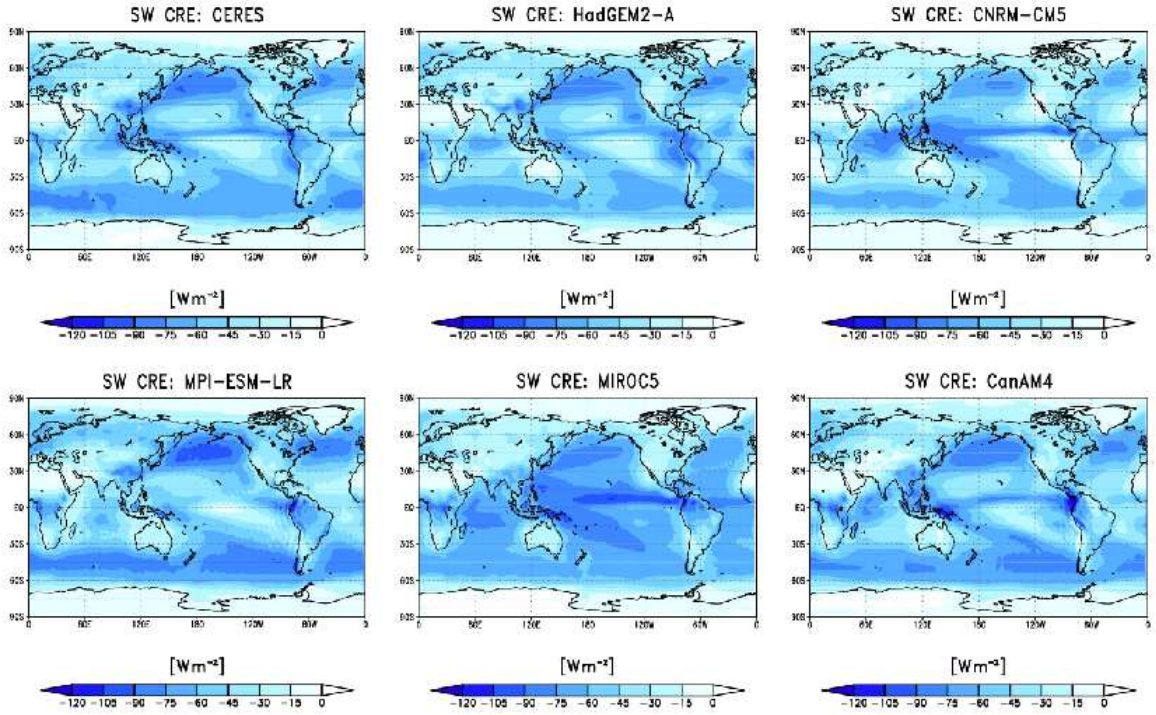


Figure18: Comparison of the observed annual mean shortwave cloud radiative effect from CERES (Wm^{-2}) with the CMIP5 models.

These relatively simple comparisons nonetheless highlight two important points:

- From the model evaluation perspective they show how the cloud data need to be combined with radiation budget data in order to reliably determine the quality of the simulation.
- From the observational perspective they also enable the relationship between the cloud properties and the top-of-atmosphere radiative fluxes to be determined and checked for consistency.

The clouds CCI will also derive cloud liquid water and ice content products. In common with other retrievals using visible and infra-red measurements, e.g. ISCCP, MODIS, ATSR, these will not be directly retrieved but calculated from the cloud optical depth and the droplet effective radius. This is clearly quite different from retrievals using passive microwave observations (SSM/I) or space-borne radar (CloudSat). Figure 19 compares the ISCCP cloud liquid water path (LWP) with retrievals from MODIS, ATSR, CloudSat and SSM/I. Differences between ISCCP and the other two VIS/IR retrievals are positive everywhere and the geographical patterns of these differences are very similar. At least part of the reason for this is related to the fact that ISCCP uses a fixed effective droplet size. Comparisons of ISCCP with CloudSat and SSMI/I are, however, somewhat different, particularly in the tropics and sub-tropics but also over the mid-latitude oceans in the Northern Hemisphere. Such differences between these observationally-derived estimates of LWP have important consequences for their use for model evaluation. This is illustrated in Fig. 20, which compares the Hadley Centre model (HadGEM2) with the different satellite products. All four data sets indicate an overestimate of LWP at mid-latitudes (consistent with the overestimate in cloud albedo seen above), although they differ in the magnitude of the discrepancy. In the tropics, however, the assessment of the model's performance clearly depends on the observed product being used for the comparison. We can also note that (i) the CCI clouds product will use both ATSR and MODIS as inputs and (ii) it will be important to assess the CCI cloud liquid water path products against those from instruments such as SSM/I and CloudSat as these are already being used by the modelling community.

In a similar manner, Fig. 21 shows comparisons of the HadGEM2 model's simulation of liquid water droplet effective radius (r_{eff}) with retrievals from MODIS and ATSR. As noted above, ISCCP is not a suitable precursor as its standard products assume a fixed value for this quantity. Again we note the quite different interpretation of the model's simulation according to the data set to which it is being compared and the need for the CCI product to be assessed against currently-available satellite estimates of r_{eff}. Similar comments also apply to the CCI ice particle size products. It should again be noted that both MODIS and ATSR will be inputs to the cloud CCI products.

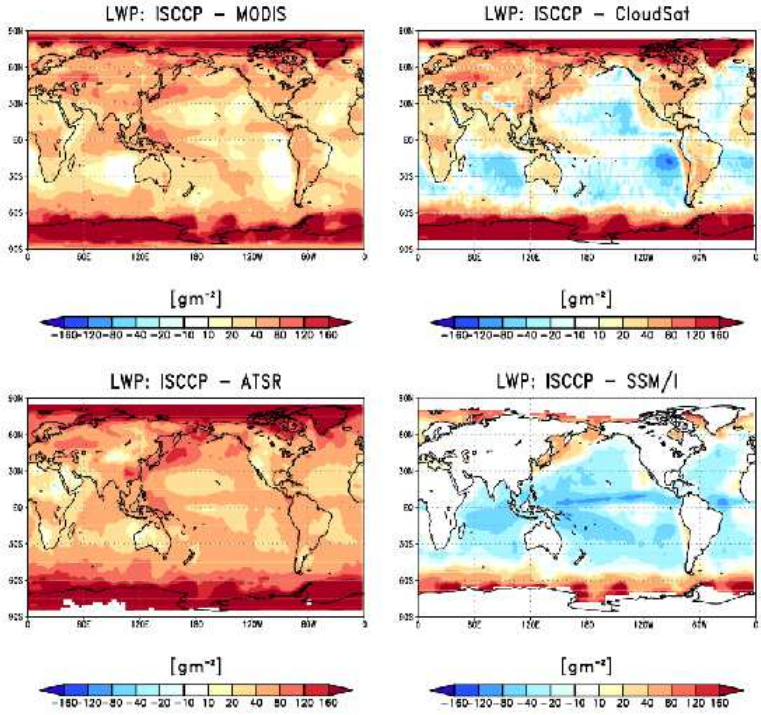


Figure 19. Comparison of annual mean ISCCP cloud liquid water path (gm^{-2}) with retrievals from MODIS, ATSR-GRAPE, CloudSat and SSM/I.

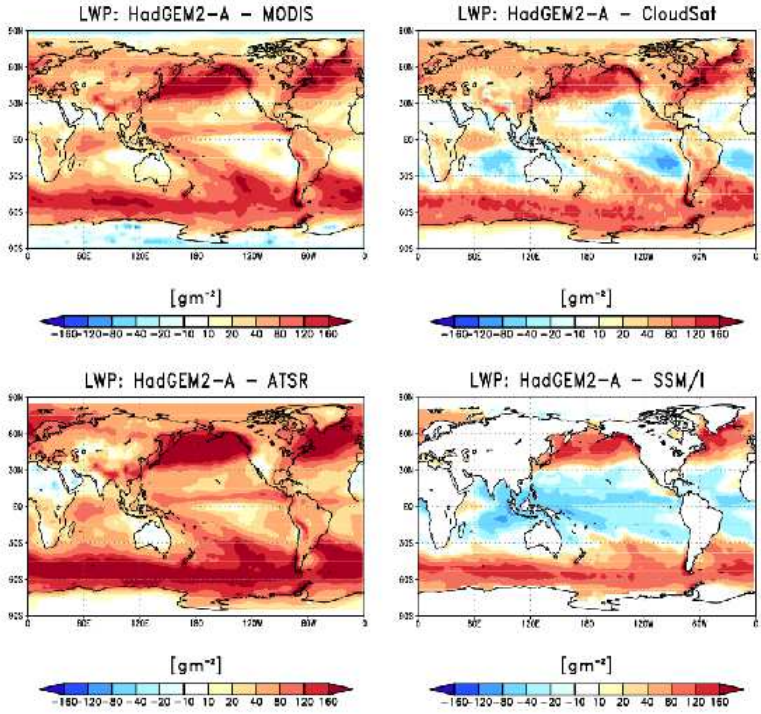


Figure 20: Comparison of HadGEM2 cloud liquid water path (gm^{-2}) with retrievals from MODIS, ATSR-GRAPE, CloudSat and SSM/I.

A key objective of the clouds CCI project is to produce a long record of cloud properties suitable for examining long-term variability (and possibly trends). The ISCCP record extends from 1983 to the present and thus allows interannual variations to be considered and compared to model simulations (it is not considered suitable for trend detection). As an example, Fig. 22 compares cloud amount anomalies from ISCCP for the 1997-98 El Niño event with those simulated by the CMIP5 models. As the models are all driven by the observed SSTs this provides an evaluation of their atmospheric responses to the warmer temperatures at this time. It is clear that the ability of the models to reproduce the observed cloud anomalies varies considerably, even across this small ensemble. Similar anomalies in the other cloud properties (e.g. cloud height, optical depth, etc) and for other anomalously warm or cold periods can also be constructed. Here we can note that:

- This is a valuable method of model evaluation which provides additional information to simply comparing mean climatologies or the seasonal cycle.
- As mentioned above, in combination with, for example, radiation budget and other cloud data, this provides a further consistency check on the cloud CCI products.
- Similar analysis of other ECVs can also provide a useful consistency check across the range of CCI products.

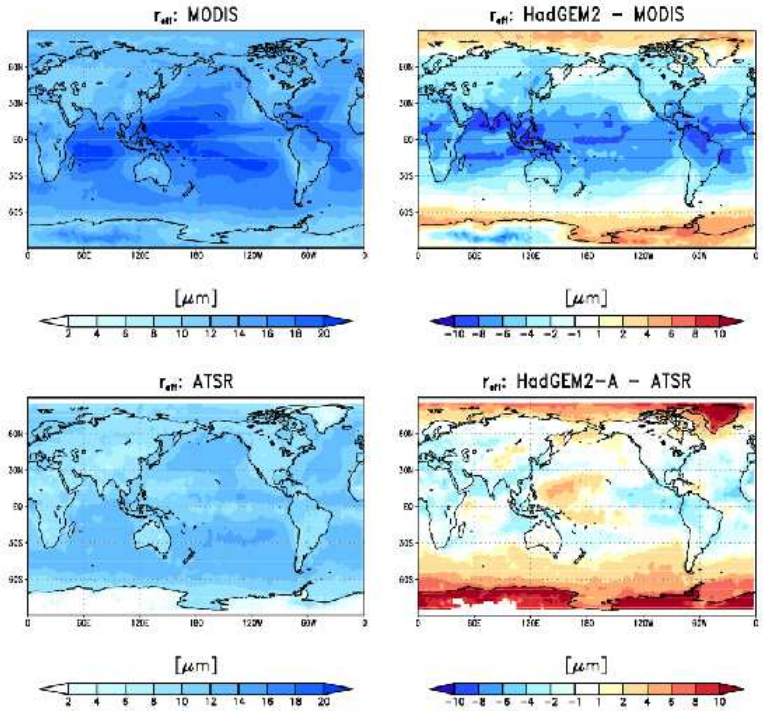


Figure 21: Cloud liquid droplet size retrievals (μm) from MODIS and ATSR-GRAPE and comparison with HadGEM2.Note that the model-simulated parameter is r_{eff} "as seen from space" in order to enable a reliable comparison with the observationally-derived product.

To summarise we can make the following comments:

- It will be important for the clouds CCI to demonstrate the added value of the proposed products compared to what is currently available and being used by the climate modelling community. This includes both independent data sets (CloudSat, SSMI/I, etc) and those which already exist and are based on the same instruments as the CCI

clouds products (e.g. MODIS, ATSR-GRAPE). In the latter case, it will be especially important to assess the value of the cloud CCI's proposed single-sensor products, i.e. the Level 3a/b data sets as described in the PSD. This will presumably be done as part of the cloud CCI's PEP – the point here is simply that it will be relevant to the applications for climate modelling.

- A useful addition to the product list would be ISCCP-like histograms. These are based on a classification of clouds according to cloud top pressure and optical depth and are already available for other sensors, including MODIS. Because these cloud types are radiatively defined they greatly assist the use of data in combination with radiation budget information, e.g. from CERES.
- The proposed provision of retrieval and sampling errors should aid the interpretation and application of the CCI clouds data sets.
- The eventual production of a long (i.e. 30 years plus) cloud record will be useful for evaluating model performance, e.g. interannual variability.
- For long (i.e. 30 years plus) cloud records based on multiple instruments, the identification and elimination of spurious temporal inconsistency remains an important challenge. Improving consistency in the form of better temporal homogeneity would be a valuable CCI activity. The methodologies for assessing progress on temporal homogeneity may differ from those for examining radiative impacts, to the extent that some homogeneity assessments can be conducted without a COSP simulator.
- It will be very difficult to make a definitive assessment of the cloud CCI products in the modelling context based on the three years of data currently proposed to be the output from Phase 1.

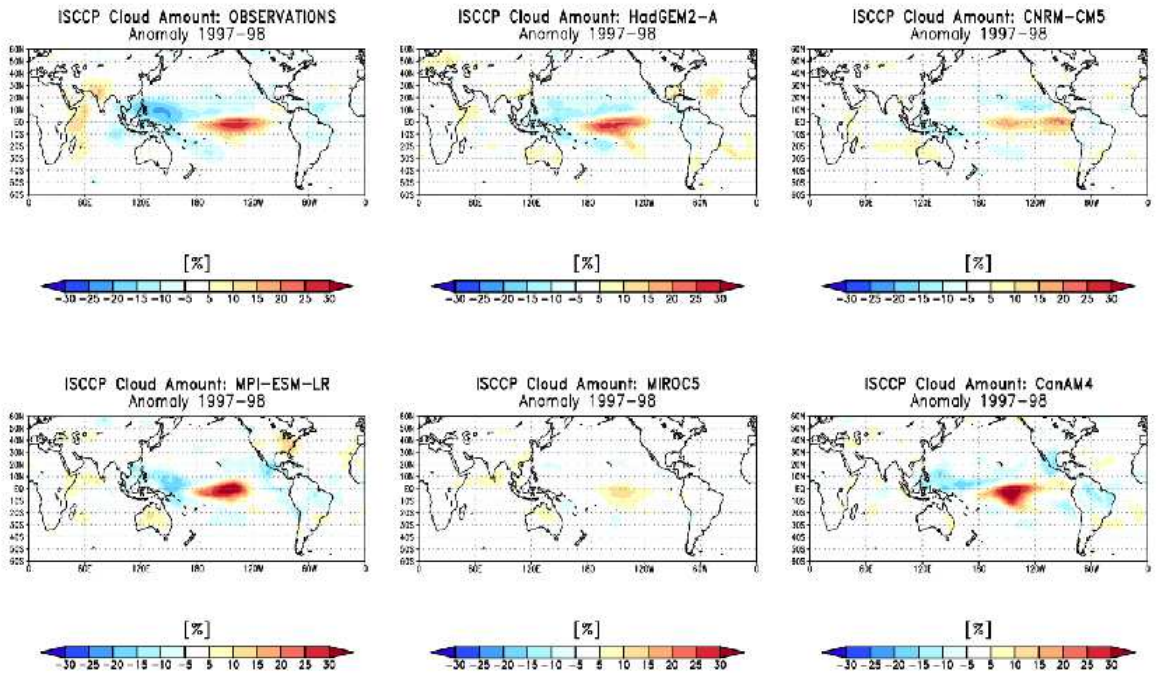


Figure 22: Comparison of anomalies in ISCCP total cloud amount (%) during the 1997-98 El Niño event with the CMIP5 models. Note that the models are all forced with the same, observed SST distribution.

4.5 Ozone

Ozone is an important reactive gas in the atmosphere and so monitoring its spatial and temporal variability over a multi-decadal period provides an important climate relevant indicator. There are a number of different ways the ozone CDRs can be assessed which include:

- Monitoring trends, variability, anomalies and attribution
- Model validation
- Assimilation in models

and this report gives examples of each of these. The CCI products to be assessed will be the ozone total column amount, the nadir ozone profiles and the limb view ozone profiles. As the ozone CCI project will deliver L2 and L3 products, all have to be assessed. The CMUG is primarily interested in validating the products with respect to their use in models.

At CNRM, this assessment is initially performed for the following three precursors:

- NIWA products from satellite UV spectrometers (TOMS to GOME-2) for total column ozone
- EUMETSAT IASI ozone retrievals for nadir profiles
- MIPAS and MLS retrievals for limb profiles

These provide all three types of ozone derived products which will be produced by the CCI. The subsections below give examples of assessments for the various applications. Table 7 gives a summary of the experiments which were run to assess the satellite ozone products.

Application	Model/Reanalysis	Exp ID	IASI	MLS	OMI	NIWA Dobson
Global trends	MOCAGE T42, T170 T42 T170 T42, T170	 Ml Mh All Alh Ahl Ahh	 None None X (310 km) X (78 km) X (310 km) X (78 km)	 None None X X X X		
Regional trends	MERRA				X	
Regional trends	MOCAGE T42, T170		X	X	X	
Model validation	CNRM-CCM CNRM-ACM					X
Assimilation	MOCAGE T42, T170	As for global trends	X (310 km) X (78 km	X (310 km) X (78 km)		

Table 7. Summary of assimilation experiments with ozone products. Ml and Mh are the control runs with no satellite data assimilated. Axx are the experiments with IASI data assimilated at different spatial resolutions of the model and the data. The MLS assimilation experiments were not assigned a separate id.

4.5.1 Monitoring trends, anomalies and attribution.

The IASI total column ozone and profiles in the stratosphere, in the troposphere or in the UTLS are used for this assessment. This product is available from EUMETSAT since the launch of Metop in 2006. The goal of this assessment is a comparison of level 3 or level 4 products with an ensemble of model predictions, over a long time period, to study interannual variations. A study of one typical year to understand seasonal variations and transport in relation to ENSO etc has value in assessing future cross-ECV consistency.

i) Global scale

The global distribution of ozone and when/where the mean maxima and minima are occurring can be monitored by satellites. An example is shown with IASI retrievals (Eumetsat total column ozone L2 product with cloud mask in the CNRM-CM model grid) in Figure 23.

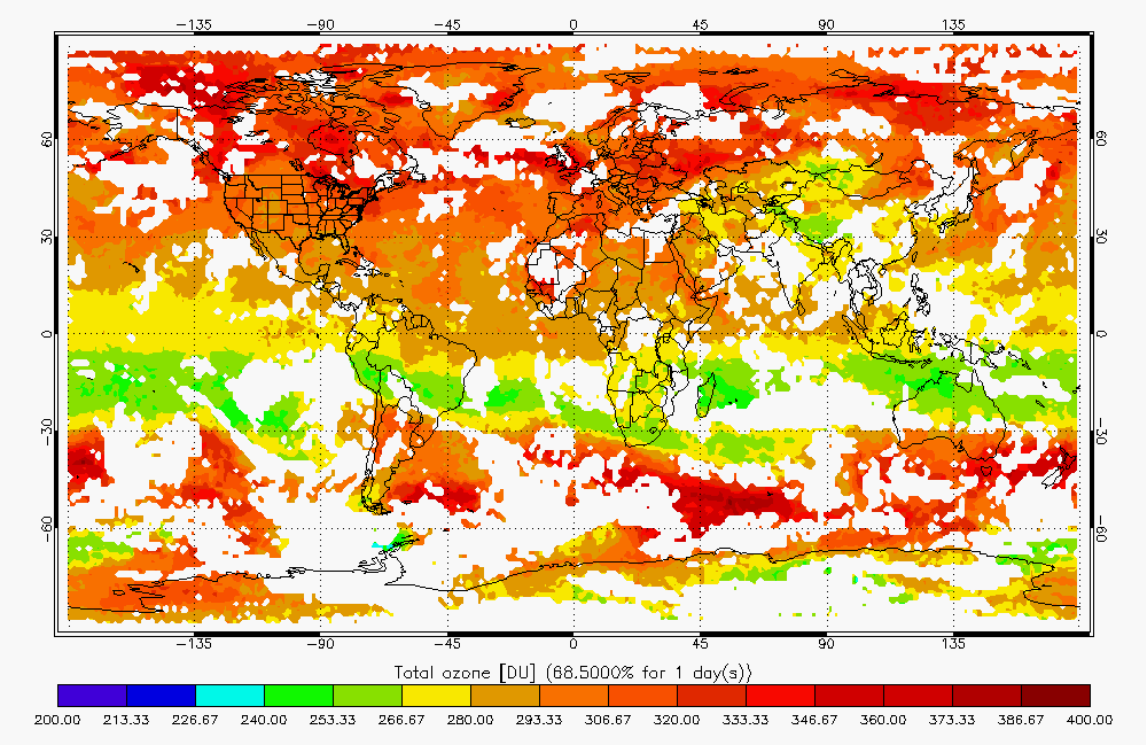


Figure 23. Global distribution of ozone from IASI retrievals

An innovative way to assess the time series is to examine the energy spectra for retrieved total column and ozone profiles (sub-columns or concentration at given heights) derived from a FFT of the global time series. This can also be compared with a model reanalysis. The typical wavenumbers of the ozone distribution are analysed and compared (through the spectral differences) as shown in Figure 24 (left panel). The MOCAGE model without assimilation of ozone data is first used to produce ozone fields at horizontal resolutions of T42 (Ml) and T170 (Mh). Then the two versions of the model (at low and high resolution) assimilate the IASI ozone observations and the MLS observations respectively averaged in low or high resolution grid meshes giving *All* and *Alh* fields and at high model resolution giving *Ahh* fields. In Figure 24a, the energy spectra for ozone fields Ml and Mh are *eh* and *el*. The plots show up to wavenumber 42 of the monthly average of daily differences (*eh*-*el*) normalized by the mean (*eh*+*el*)/2. Large differences are observed at all wavenumbers (dotted line), with values

increasing with the wavenumber. This shows that the Mh has changed the large scale structure of ozone in the model as well as the small scale compared to Ml. The amplitude of the waves is higher on Mh. The comparison of the analyzed fields *Alh* and *All* demonstrates that at the same model resolution, the change coming from the higher resolution satellite observations impact only smaller scale structures (i.e. high wavenumbers). In figure 24b, the zonal difference of total column ozone is shown from Ml and Mh normalized by ERA Interim. The figure shows a significant bias of 2% in the tropics. At mid and high latitudes the variability is higher and the bias becomes significant only above 70°S. Examining the zonal means is instructive when comparing various products as in Figure 24 (right panel).

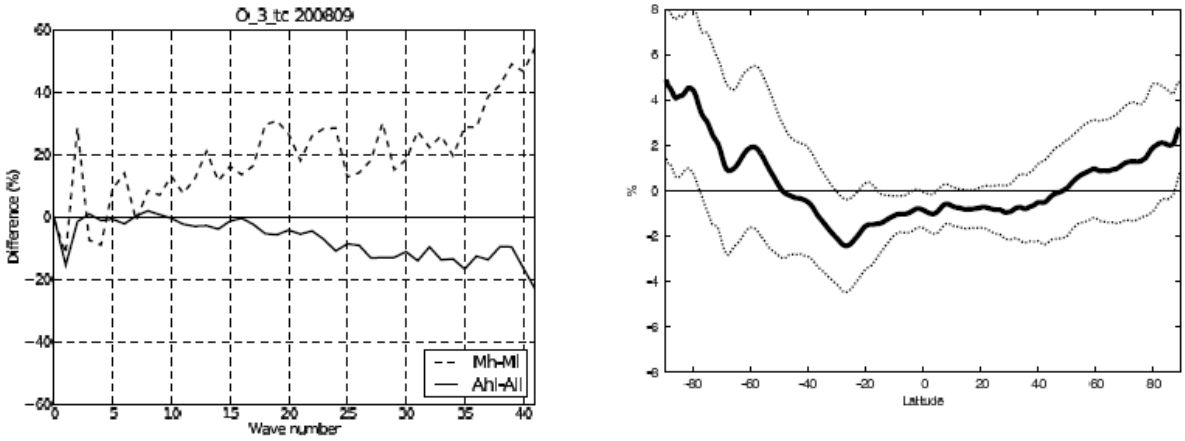


Figure 24. Left panel shows the monthly average in % of daily differences between the energy spectra (dashed lines) normalised in different ways. Right panel shows the zonal difference between total ozone columns in % computed from the experiments. The dashed lines represent the standard deviation of the differences (after Pajot et. al.; 2011).

Other indicators can also be used in the assessment of the data. An example is the frequency of stratospheric intrusion events, their location, duration and the mass of ozone which is transported into the troposphere.

ii) Regional scale

Some regions are of particular interest for ozone studies such as the ozone hole over Antarctic and some specific indicators can be proposed here:

a) Time indicators : Date of start, and recovery, time span, date of minimum
b) Surface indicators : size of the vortex (in km^2) and volume of ozone destroyed (mass deficit)
c) Minimum value
d) Potential vorticity
e) Ozone loss

Some of these variables are shown in Figure 25 for the OMI UV pre-cursor ozone data and MERRA reanalysis fields. Plots of the regional distribution are very useful to show the discrepancies between the various fields. Note that at a regional scale for ozone, the timescale is often seasonal and satellite data are often assimilated into models to update the analyses as shown in Figure 26. This work has also been performed using data from the Concordiasi

experiment by L. ElAmraoui at CNRM who assimilated IASI, MLS and OMI ozone products in the MOCAGE model (at T42 and T170 horizontal resolution). Figure 26 shows some similarities with a high bias for IASI but a more detailed description of the ozone field. Where the improvements with OMI are small, use of high resolution IASI data in a high resolution model appears to make small scale structures which are useful to help understand the dynamics of the ozone hole.

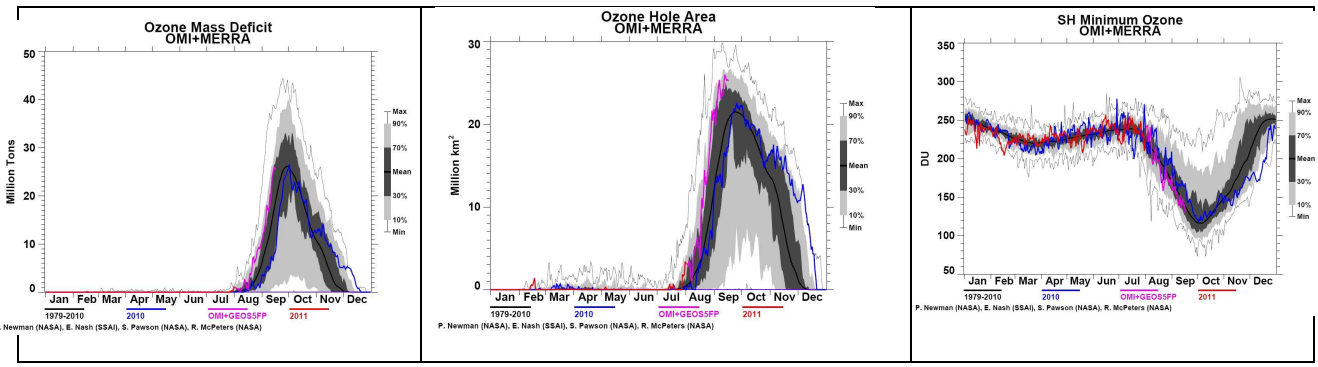


Figure 25. Plot of various ozone parameters for different periods for MERRA and OMI measurements.

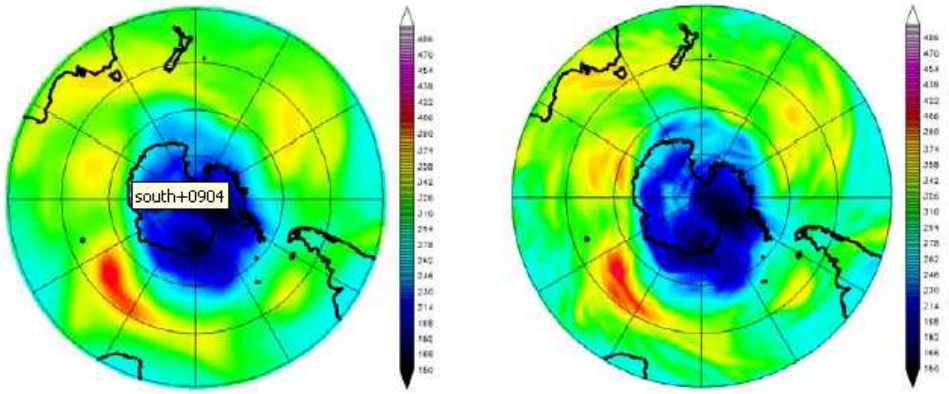


Figure 3 : colonnes totales d'ozone (DU) issues d'une analyse des données IASI le 4 septembre 2008 à 00 UTC ; à gauche pour une résolution T42 (~2.8°) et à droite pour une résolution T170 (~0.7°)

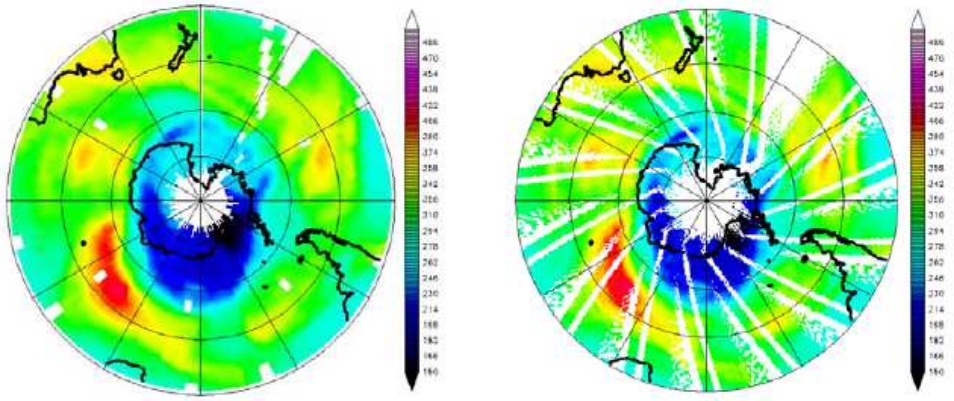


Figure 4 : colonnes totales d'ozone (DU) issues de l'instrument OMI le 4 septembre 2008 ; à gauche moyennées sur une grille de résolution 2° et à droite sur une grille de résolution 0.5°

Figure 26. Total column ozone from different ozone analyses (low resolution T42 left; high resolution T170 right) in the top panels using IASI data and measurements from OMI in the lower panels.

4.5.2 Use in model assessments

Climate and NWP ozone model fields are assessed through comparison with ERA Interim, with ozone sondes or with other ozone precursors.

At CNRM, this has been demonstrated by Michou *et al*, (2011). In this work the new version of the CNRM CCM climate model, with a more accurate radiation scheme and an in-line interactive chemistry scheme was compared with the previous of the CNRM atmospheric chemistry model (ACM) where chemistry was provided by a non coupled chemical transport model. These data have been provided to the CCMVal-2 exercise. Monthly zonal-means of NIWA total column observations have been compared to the CNRM-ACM and CNRM-CCM simulations valid for the 1980–1990 and 1990–2000 periods. There are general patterns in the observations that models should reproduce well. In this case the CNRM-CCM reproduces patterns better than the CNRM-ACM, e.g. spring mid-latitude gradients, temporal evolution throughout the year of the tropical total column, and confinement of the Antarctic polar vortex.
The negative trend in the column ozone at middle and high latitudes (>45º) is evident in both observational and model outputs. The two CNRM models overestimate Antarctic ozone depletion, in all periods shown, to a lesser extent however for CNRM-CCM than for CNRM-ACM. These results are confirmed by looking at climatological annual cycles of the total column ozone or profiles over high, mid and tropical latitude bands as in Tian *et al.* (2010) (see Figure 27).

As in the trend analysis similar metrics are used to assess the models:
- energy spectra of total ozone (monthly average of daily differences normalized by mean),
- plots of zonal differences (on monthly means)
- standard deviation in zonal mean as a function of latitude and altitude (profiles).

Maps (movie loops) at specific pressure levels of ozone volume mixing ratio over the south pole vortex to show the ozone gains or losses of different model versions (see Figure 28.)

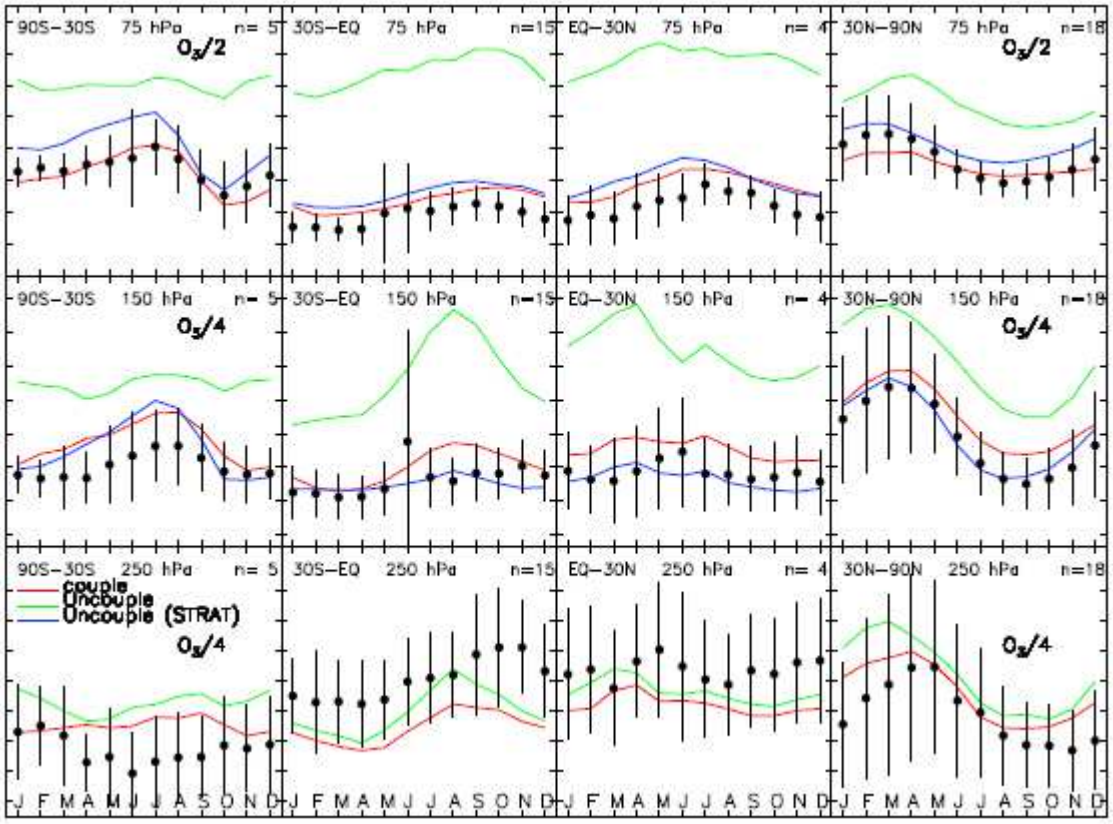


Figure 27. Comparison of the annual cycle of monthly mean ozone observations (black dots) and model runs TROP (green), STRAT (blue) and COUPLED (red) sampled for different latitude bands (90S-30S; 30S-0S; 0N-30N; 30N-90N) and different pressure levels (75,150 and 250hPa). Each panel is the mean of many years from several sites (number of sites is in top right corner). The model and observations are sampled in the same way. The vertical bars are the average of the interannual st. dev. at each station.

4.5.3 Assimilation of ozone

CERFACS are working on assimilation of IASI ozone products into models. The aim is to improve the representation of the forcing and ozone distribution in the models. The MOCAGE model is used with a T170 grid and a T42 grid. The most suitable method to assimilate ozone is to use the traditional 4D-Var methodology employed in many NWP centres. The value of high resolution IASI ozone products in a model with a fine mesh is demonstrated but the limited vertical distribution must be described by the IASI averaging kernels. Note that MLS horizontal resolution is not adequate for capturing the detail. This work will be extended to other seasons, and to tropospheric ozone. Figure 28 illustrates the ozone distribution from MOCAGE stand alone at T42 (Ml), T170 (Mh), from ERA-Interim, and from three assimilation experiments with IASI at coarse or high resolution. The details brought by high resolution IASI observations assimilated in a T170 model are obvious.

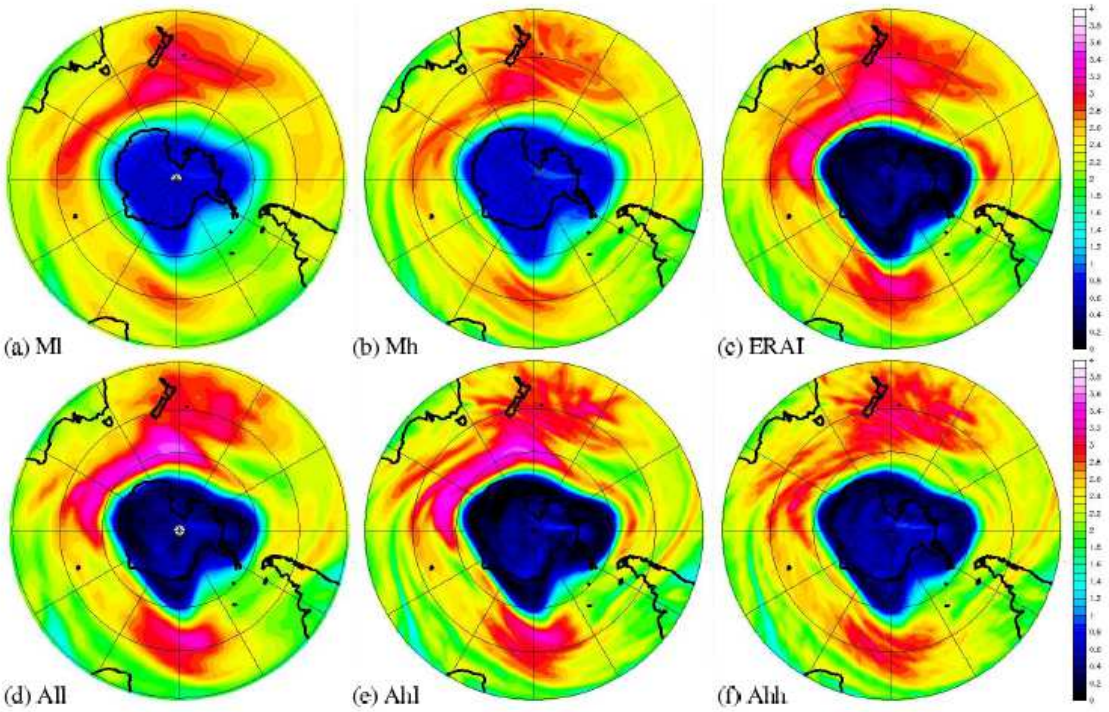


Figure 28. Ozone volume mixing ratio (in ppmv) centred over the S. hemisphere at 55hPa on Sept 30th 2010 at 12UTC. Plots from several assimilation experiments and ERA-Interim.(a) and (b) are MOCAGE low and high resolution runs without ozone assimilation, (c) is the ERA-Interim field, (d) is the MOCAGE low resolution model with low resolution IASI data assimilated, (e) is the same using the high resolution MOCAGE model and (f) is the high resolution model with high resolution data.

4.6 Greenhouse Gases

Not assessed by CMUG to date.

4.7 Aerosols

Not assessed by CMUG to date.

4.8 Land Cover

Land cover information is essential to provide information about the spatial patterns of the biosphere and related land surface parameters for a global dynamic vegetation model. Its distribution will directly affect the energy, water and carbon fluxes as simulated by a complex Earth System Model, like the Max-Planck-Institute for Meteorology Earth System Model (MPI-ESM). Land cover information is required either as *initial* or *boundary* conditions in a global dynamic vegetation model (DGVM).

The CCI land cover project will provide new, high resolution land cover information for at least one decade. An assessment of the potential impacts of the novel dataset in a coupled Earth-System model has been done by MPI-M using a pre-cursor data set. The natural pre-cursor for the CCI landcover product is the GlobCover (v2.2) product

(http://ionia1.esrin.esa.int/), generated from ENVISAT MERIS data at global scale with 300m resolution.

4.8.1 Integration of land cover information into MPI-ESM

The GlobCover data was integrated into the land surface component (JSBACH) of MPI-ESM. It is used to replace the current distribution of plant functional types (PFT) in JSBACH which were originally obtained from USGS EROS Data Center land cover classification (USGS, 2001; Olson 1994a; Olson 1994b).

In order to use the new information in the JSBACH model, the satellite classification schema of GlobCover had to be translated into model compliant PFT classes. PFTs reduce the diversity of species to some key plant types with similar physical and biogeochemical properties. PFT specific parameters, like e.g. canopy albedo, LAI, carbon content of leaves or roughness length are used to parameterize the specific properties of a PFT in the model.

In general, the concept of PFTs used by Earth-System Models is not consistent with the approach for classification of land cover from remote sensing measurements. While PFTs aim at describing characteristic functional behaviour of vegetation in terms of climatological or ecological parameters and might consist of aggregation of multiple species into a single category, remote sensing based classifications stratify the land surface in terms of spectral and temporal differences (Bonan *et. al*., 2002 and Poulter *et. al.* 2011). A summary of the land cover classes provided by GlobCover are given in Table 8. These are very similar to the classes that will be provided by the new ESA CCI landcover product.

To preserve the maximum information content of the original data product and allow for an integration of the new remote sensing information into MPI-ESM, a sequence of pre-processing steps are required which will be briefly summarized in the following. The general workflow is shown in Figure 29.

1. Spatial aggregation: The aggregation of the initial satellite data (300m) to a resolution of 0.5° is performed to compress the data volume while preserving the information on fractional coverage of the different land cover classes.
2. Transformation to plant physiology classes: The GlobCover classes are then converted into plant-physiology classes, which are provide the first steps towards the generation of PFTs from the landcover classes. The scheme developed by Poulter *et. al*. (2011) is used for that purpose. As the (spectrally defined) land cover classes are typically mixed classes, a re-classification step is needed. The different land cover classes are assigned to different biomes by splitting them into different biomes using expert knowledge. As this mapping is rather subjective it introduces uncertainties in the resulting PFT distributions. Table 9 shows the mapping matrix developed at MPI-M for the mapping of GlobCover classes into biomes.

11 - Post-floodingor irrigated croplands
14 - Rainfedcroplands
20 - MosaicCropland (50-70%) / Vegetation (grassland, shrubland, forest) (20-50%)
30 - MosaicVegetation (grassland, shrubland, forest) (50-70%) / Cropland (20-50%)
40 - Closed toopen (>15%) broadleaved evergreen and/or semi-deciduous forest (>5m)
50 - Closed(>40%) broadleaved deciduous forest (>5m)
60 - Open(15-40%) broadleaved deciduous forest (>5m)
70 - Closed (>40%) needleleaved evergreen forest (>5m)
90 - Open (15-40%) needleleaved deciduous or evergreen forest (>5m)
100 -Closed to open (>15%) mixed broadleaved and needleleaved forest (>5m)
110 - Mosaic Forest/Shrubland (50-70%) / Grassland (20-50%)
120 - Mosaic Grassland (50-70%) / Forest/Shrubland (20-50%)
130 - Closed to open (>15%) shrubland (<5m)
140 - Closed to open (>15%) grassland
150 - Sparse (>15%) vegetation (woody vegetation, shrubs, grassland)
160 - Closed(>40%) broadleaved forest regularly flooded - Fresh water
170 -Closed (>40%) broadleaved semi-deciduous and/or evergreen forest regularly flooded- Saline water
180 - Closed to open (>15%) vegetation (grassland, shrubland, woody vegetation) on regularly flooded or waterlogged soil - Fresh, brackish or saline water
190 – Artificial surfaces and associated areas (urban areas >50%)
200 - Bare areas
210 - Water bodies
220 - Permanent snow and ice

Table 8. Landcover classes of GlobCover data product

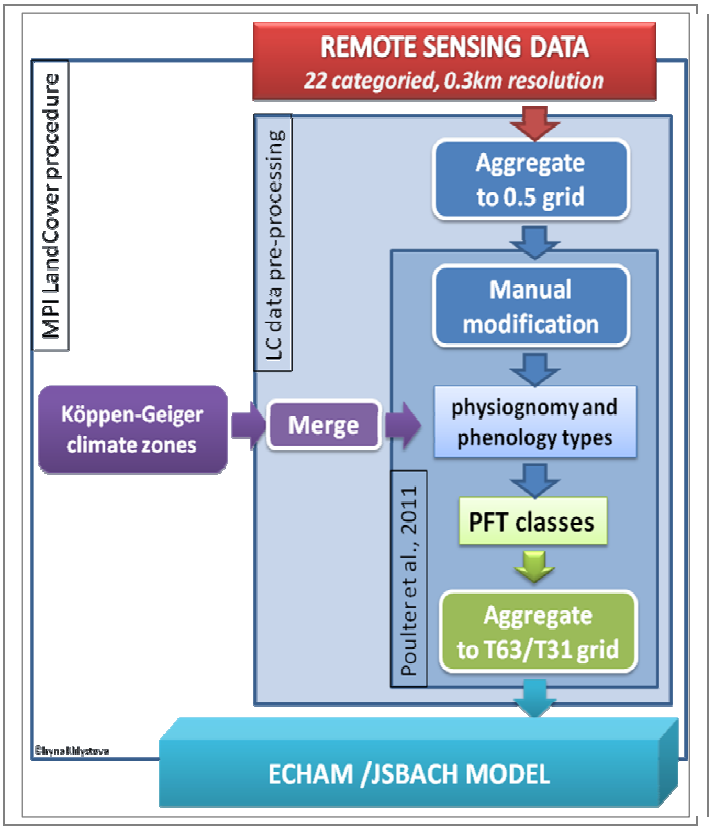


Figure 29: Schematic workflow of Plant Function Types mapping

3. PFT generation: Additional information on climate conditions is required for an appropriate mapping into PFT classes. This allows for instance to disentangle C3 and C4 grasses, as needed for the model simulations. As the information on tropical and extratropical vegetation is not provided by GlobCover, ancillary climate information is needed. The GlobCover v2.2 vegetation classes were finally remapped using Köppen-Geiger classification maps.
4. Inland water bodies: A large portion of the model grid cells contain a considerable amount of inland water bodies. Figure 33 shows the distribution of inland water bodies as derived from the ESA GlobCover product. The fractional coverage of inland water bodies typically ranges between 1% and 10%, but can exceed 20% of the area of a grid cell for instance in the Northern latitudes. As the present version of the MPI-ESM land surface scheme is not capable in simulating the dynamics of open water bodies explicitly, the PFT fraction in each grid cell needs to be normalized to ensure that all PFT fractions sum up to unity.

	BrEv	BrDe	NeEv	NeDe	NatGr	ManGrPs	ManGrCrop	ShrubAll	bare	water
glbcov_type11	0	0	0	0	0	0	1	0	0	0
glbcov_type14	0	0	0	0	0	0.5	0.5	0	0	0
glbcov_type20	0	0	0.1	0.1	0.1	0.25	0.45	0	0	0
glbcov_type30	0	0	0	0	0.25	0.25	0.25	0.25	0	0
glbcov_type40	0.5	0.5	0	0	0	0	0	0	0	0
glbcov_type50	0	1	0	0	0	0	0	0	0	0
glbcov_type60	0	1	0	0	0	0	0	0	0	0
glbcov_type70	0	0	1	0	0	0	0	0	0	0
glbcov_type90	0	0	0.5	0.5	0	0	0	0	0	0
glbcov_type100	0.25	0.25	0.25	0.25	0	0	0	0	0	0
glbcov_type110	0.2	0.2	0.2	0.2	0	0	0	0.2	0	0
glbcov_type120	0.1	0.1	0.1	0.1	0.2	0.2	0	0.2	0	0
glbcov_type130	0	0	0	0	0	0	0	1	0	0
glbcov_type140	0	0	0	0	0.5	0	0	0.5	0	0
glbcov_type150	0.1	0.1	0.1	0.1	0.4	0	0	0.2	0	0
glbcov_type160	0.5	0.5	0	0	0	0	0	0	0	0
glbcov_type170	0.5	0.5	0	0	0	0	0	0	0	0
glbcov_type180	0.1	0.1	0.1	0.1	0.2	0.2	0	0.2	0	0
glbcov_type190	0	0	0	0	0	0	0	0	1	0
glbcov_type200	0	0	0	0	0	0	0	0	1	0
glbcov_type210	0	0	0	0	0	0	0	0	0	1
glbcov_type220	0	0	0	0	0	0	0	0	1	0

Table 9. Weighting factor matrix to map GlobCover land cover classes to biomes

4.8.2 Novel PFT distribution

Figure 30 shows the distribution of major vegetation classes (forested, herbaceous, crops) as in the original data set currently used by JSBACH and the new generated maps obtained from GlobCover.

Overall, the two datasets show similar spatial patterns for the different major classes. The total forested area in JSBACH amount to 4140 Mha while it is 4186 Mha for GlobCover (+1.1%). Herbaceous and crop areas amount to 4633.8 Mha (4772.8 = +3%) and 1849.5 Mha (1780 Mha = - 4%).

While the general global statistics and patterns match well between the PFT distributions of the two sources, distinct differences can be observed on the basis of individual PFT classes (Figure 32). Figure 32 shows differences between Olson (1994) and GlobCover as well as area weighted zonal means of 3 major vegetation types. The GlobCover data set shows a larger fraction of forest in the boreal zone than Olson (1994). A larger extent of cropland area can is observed from GlobCover in China as well as Argentina, where strong cropland expansion occurred during the last decades. Especially the larger extend in forested areas is expected to result in substantial difference in the surface radiation budget due to changes in the surface albedo. This might pronounce geophysical feedbacks in MPI-ESM like the snow-albedo feedback.

The differences between the two different PFT distributions may have the following major reasons:

- *Change in land cover*: The classification of Olson (1994) is based on 1km AVHRR data from the years 1992 and 1993, which were classified into 94 ecosystem classes using an unsupervised classification approach (Loveland et al., 2000) while the ESA GlobCover product used was derived for the year 2005. Within the 13 years, land cover might have changed considerably in parts of the globe.

- *Differences in land cover product*: since the different products are based on different spectral information, classification schemes and classification legends, the resulting products differ in their information content. They can also differ due to viewing geometry (sun elevation, scan angle) and to spatial resolution of sensor.
- *PFT mapping*: The Olson (1994) classification contains a total of 94 ecosystem types which are mapped to 11-14 PFTs for the use in MPI-ESM (Hagemann and Loew, 2012). As has been discussed, the mapping of landcover/ecosystem types into PFT classes is a subjective process, including expert knowledge. Different procedures to translate spectral classes into PFT classes might therefore result in different PFT distributions.

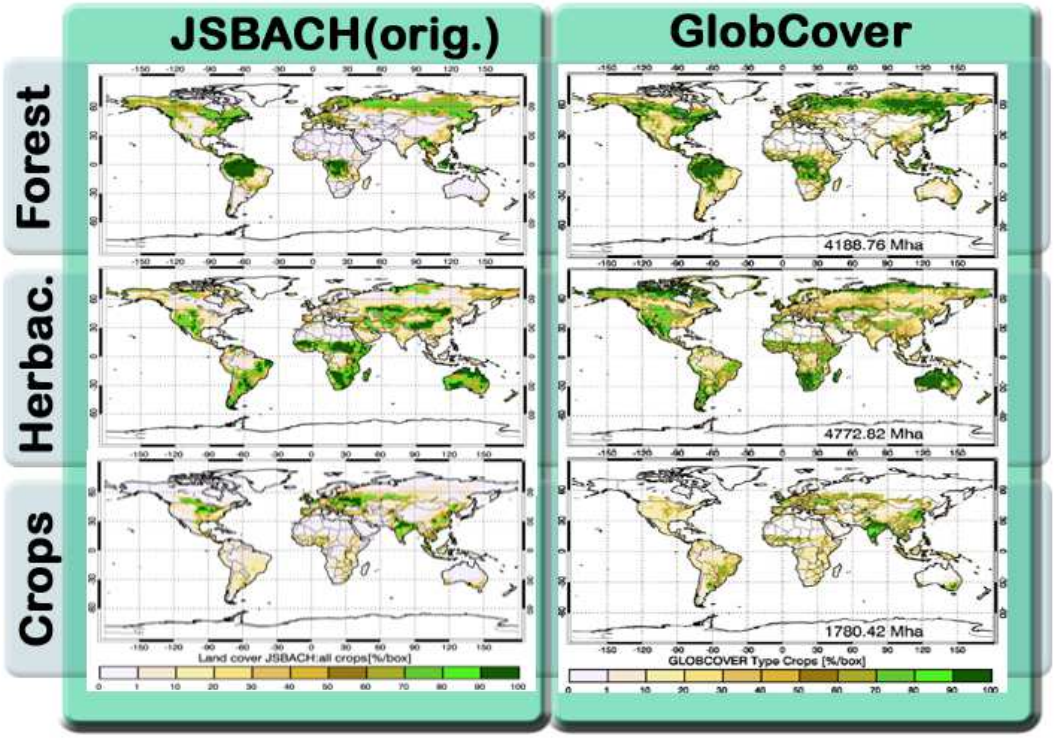


Figure 30: Distribution of principal land cover classes (forest, herbaceous, crops) for original JSBACH and GlobCover datasets.

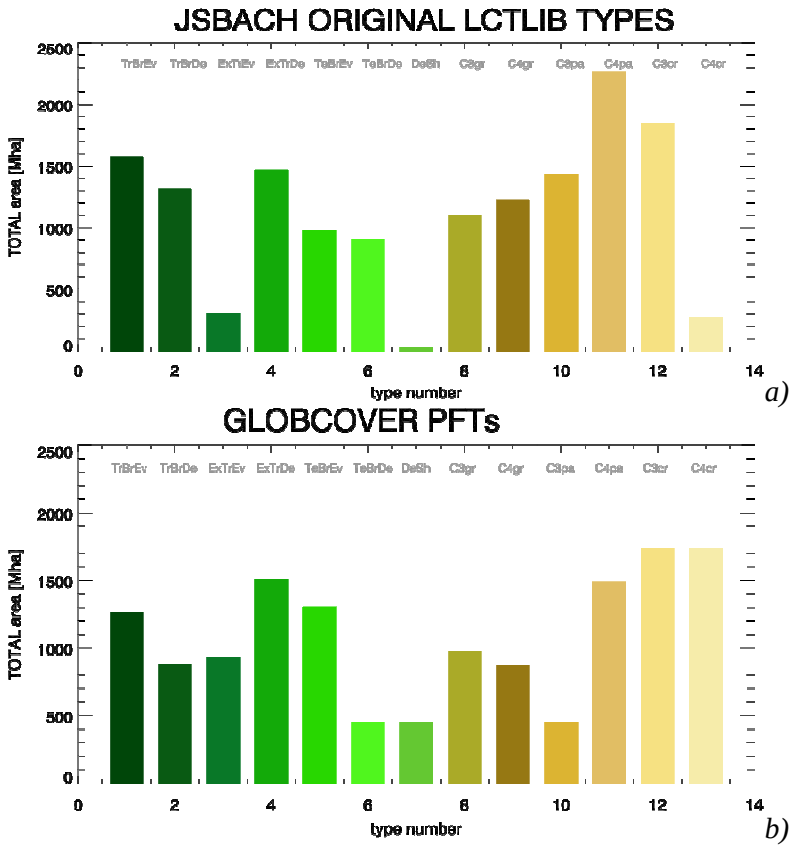


Figure 31 : Total area of different PFTs as in current PFT distribution of MPI-ESM and GlobCover derived PFT maps

4.8.3 Model performance metric and framework

Changes in PFT distribution will have a direct effect on biogeophysical variables like e.g. surface albedo or LAI as simulated by MPI-ESM. To assess the impact of using GlobCover and CCI landcover as new landcover information in the MPI-M instead of the currently used PFT distributions, a series of dedicated model experiments will be conducted. The simulation results will be compared against a) reference simulations from CMIP5 experiments, b) offline model simulations and c) independent datasets. To compare results from different model setups and data sets, a model benchmarking framework was developed (Figure 34).

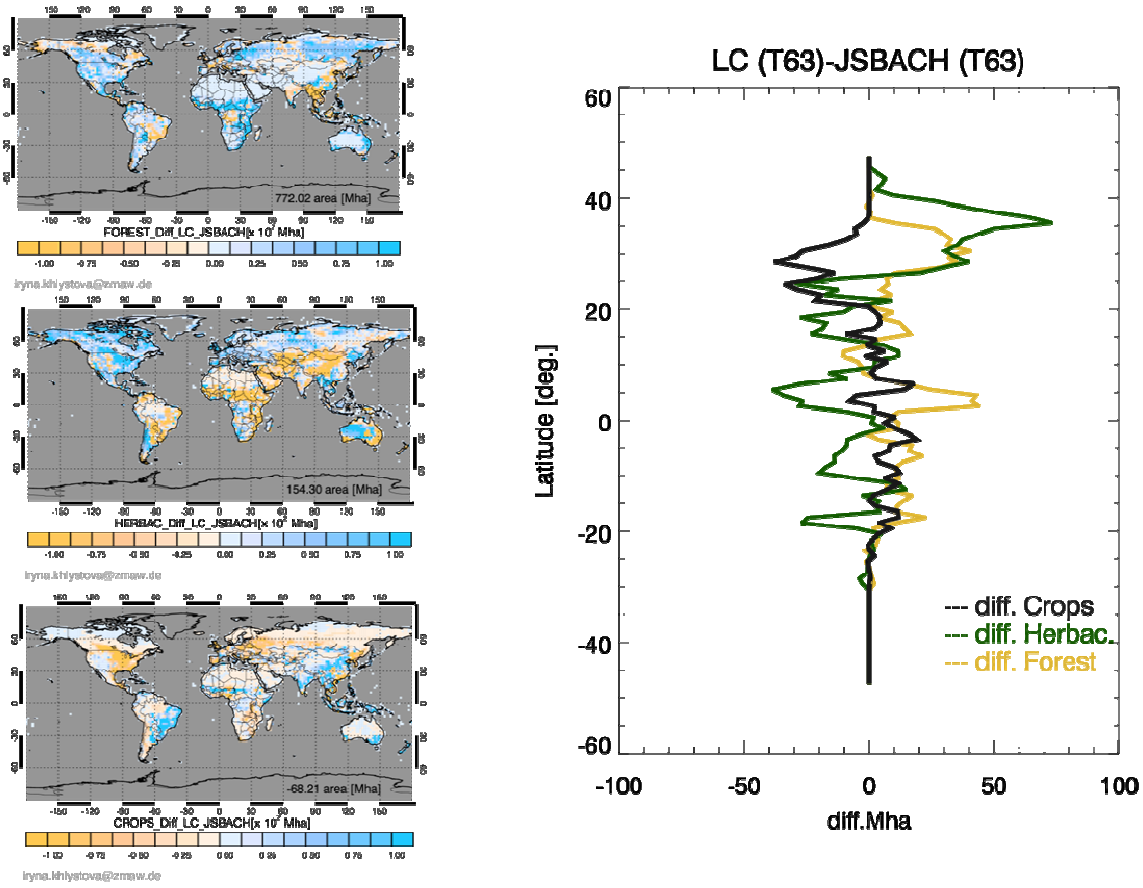


Figure 32: Difference between Globcover major biome types (foresee, herbaceous, crops) and currently used JSBACH input (left) and zonal area weighted differences (right)

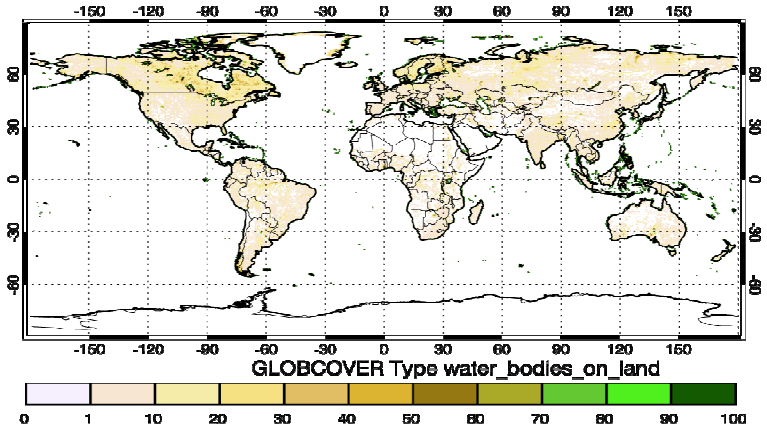


Figure 33: Distribution of fractional inland water bodies, based on the GlobCover classification

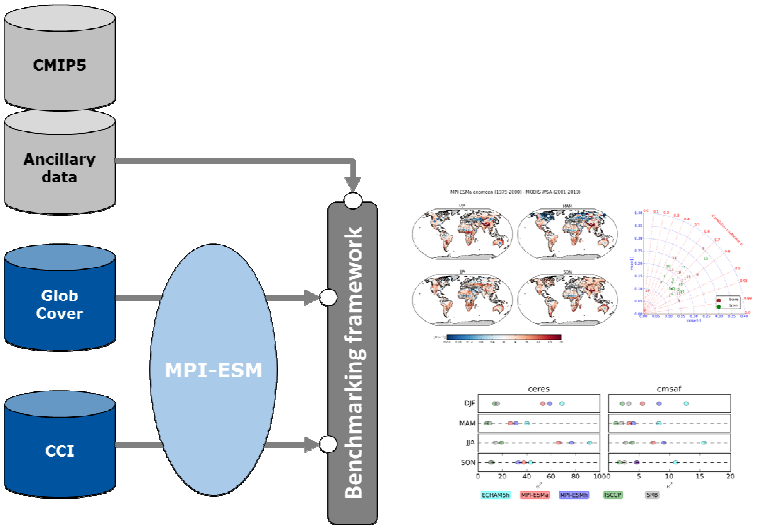


Figure 34. MPI-ESM land surface benchmarking framework to assess impact of CCI land cover dataset

This framework allows to easily compare a multitude of datasets and model experiments and assess their (relative) performance using different metrics like those developed by Reichler & Kim (2008) and Glecker et al. (2012). The accuracy of different land surface variables, e.g. surface albedo, tree cover, vegetation phenology and basic climate variables like temperature and precipitation can be analysed in a flexible manner. The framework will allow to quantify the impact of a) GlobCover and b) CCI landcover on offline and coupled climate model simulations in MPI-ESM for selected parameters. It has been already successfully applied for the evaluation of MPI-ESM CMIP5 simulation results (Hagemann & Loew, 2012) where it could be shown that the current version of MPI-ESM clearly outperforms its predecessor ECHAM5 in simulating the surface albedo dynamics. Figure 35 shows examples of analysis plots of surface albedo as simulated by the model, compared to a MODIS based climatology of land surface albedo. Hagemann & Loew (2012). It could be shown, that the model simulations with pre-scribed SST data shows a relative improvement of 39% compared to a multi-model mean. The full coupled model (ocean, atmosphere, land) of MPI-ESM was still 15% better than on average, while the older version of the MPI model (ECHAM5) which had a constant surface albedo scheme, performed worse (54%). While the implemented performance metrics only provide relative measures of model accuracy, they will allow to answer the question if MPI-M model performance is improved by a) integration of GlobCover data and b) quantify the potential impact of an improved CCI landcover product on the model performance.

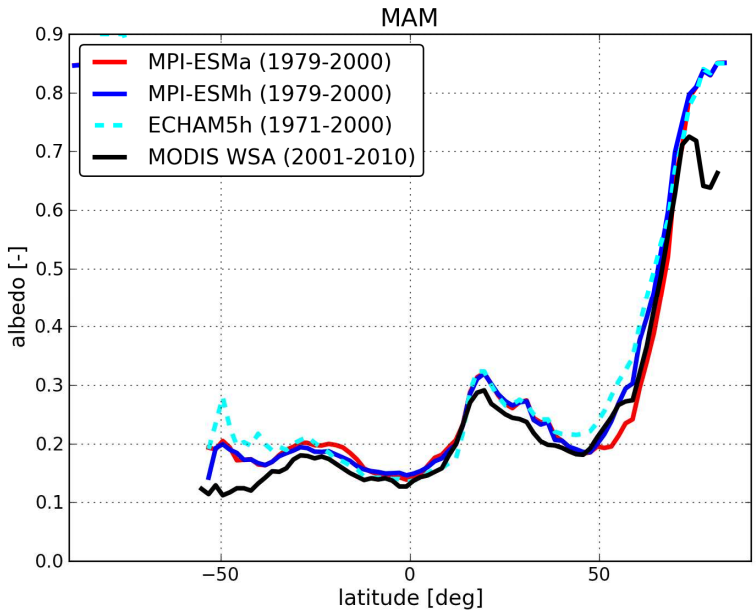
A series of model experiments with GlobCover and CCI landcover datasets will be conducted. First, the effect of the new PFT distribution on biogeophysical variables, simulated by JSBACH will be analysed in an offline simulations (CRU/NCEP forcing). Results will be analysed and compared first against a control simulation obtained with the current JSBACH model setup. Next, a coupled model-experiment (atmosphere-land) similar to the CMIP5 AMIP experiments will be conducted to investigate the impact of the new PFT distributions on land-atmosphere interactions. These simulations will reveal details on the effect of the PFT

distribution on e.g. global temperature. These simulations will be performed for a) the current JSBACH model setup (ctrl), b)GlobCover (exp_glob) and c) CCI landcover (exp_cci). Results will be compared against independent climate observations where possible to quantify a potential reduction in model error when using the CCI landcover dataset and validate its impact on important ECVs like air temperature (see Table 9).

Variable	*JSBACH offline*	*MPI-ESM (AMIP like simulations)*	*Observations*
Surface			
LAI	X	X	Phenology: start/end of season
Surface albedo	X	X	MODIS surface albedo, Meteosat albedo
Latent heat flux	X	X	
Sensible heat flux	X	X	
faPAR	X	X	
Snow cover	X	X	(GlobSnow)
Soil Moisture	X	X	
Surface temperature	X	X	(CERES EBAF)
Atmosphere			
Temperature 2m		X	CRU
Precipitation		X	CRU, GPCP v2.2

Table 9. Overview of variables analysed in offline and coupled model simulations for CCI landcover assessment

It is planned to perform the actual assessment of the ESA CCI landcover dataset in close collaboration with the ESA CCI land cover climate research users to assess the quality of the new landcover product and to disentangle the influence of different data products versus uncertainties in the pre-processing of the land cover data, as discussed in section 4.8.1, as well as the impact of different models..



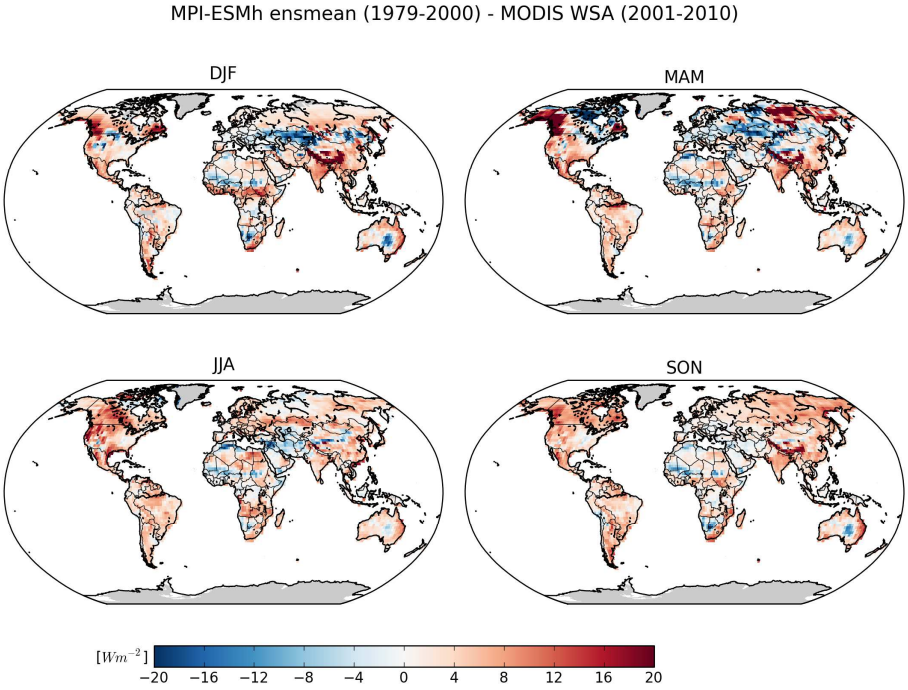


Figure35: Surface albedo benchmarking results for coupled MPI-ESM CMIP5 simulations. Zonal plots of 3 model versions compared to MODIS (top) and seasonal difference maps of surface upward solar flux between coupled MPI-ESM result and MODIS climatology (lower)

4.9 Fire

4.9.1 Fire in the Earth System

Fire is an important Earth System process, which impacts climate via multiple pathways, including atmospheric chemistry, aerosols, global vegetation patterns, land surface albedo and the carbon and nutrient cycles. At the same time fires are controlled by climate and the frequency of fires is expected to increase with future climate change. As such, fires form a complex feedback cycle in the Earth system which potentially forms an important contribution to the climate sensitivity of the Earth System. The net effect of fires on the climate system remains unclear as depending on the process fires can cool or warm the Earth System (Bowman et al, 2009).

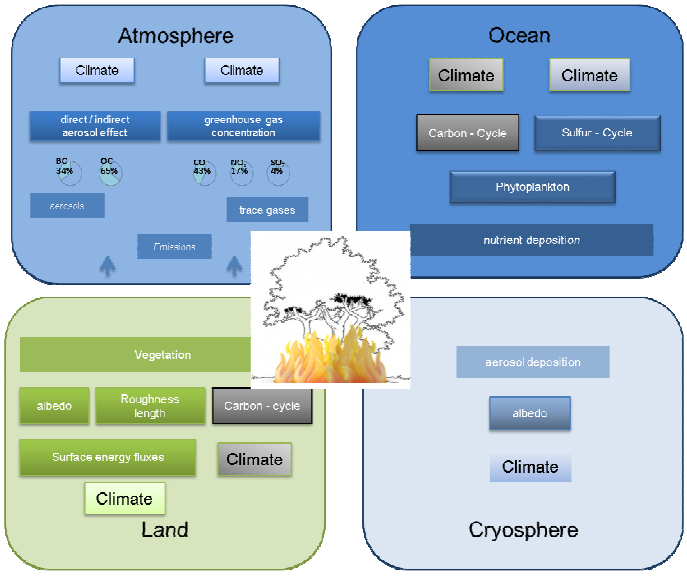


Figure 36: Fires influence the different compartments of an earth system model through changes in carbon and nutrient pools and land surface properties, e.g. land cover, albedo.

4.9.2 Fire in MPI-ESM

The burned area fire CCI dataset is assessed by CMUG for prescribing the *boundary conditions* in a global dynamic vegetation model. An additional application of the burned area dataset is the evaluation of fire models used in global dynamic vegetation models. For this purpose the JSBACH land surface model is used, which is part of the MPI-M Earth System Model.

Within JSBACH, fire is an important perturbation that impacts the vegetation distribution and carbon cycle. The fire algorithm of intermediate complexity in JSBACH simulates the fire occurrence probability dynamically as a function of soil moisture, biomass available for burning and the probability of an ignition source. The fire probability scales the potential burned area, a function of moisture and wind speed, to an actual burned area. Combined with measured estimates of the completeness of combustion CO_2 emissions are computed. (Figure 37). The CO_2 emissions are transferred to the atmospheric pools and the land carbon pools are accordingly reduced.

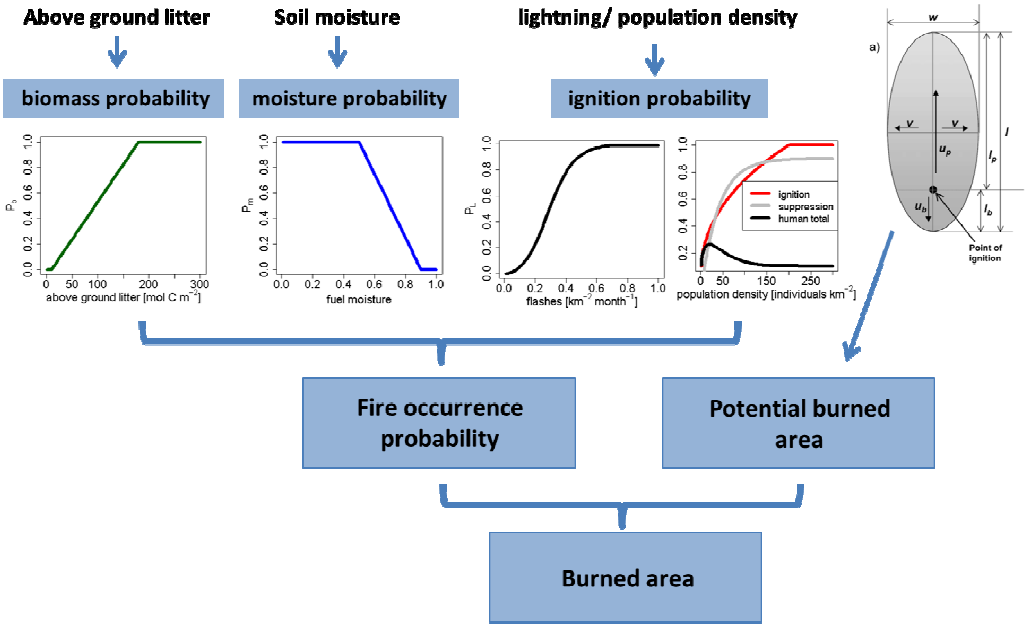


Figure 37. General scheme for simulating burned area in JSBACH

To estimate the actual amount of carbon emissions from fire, the burned area needs to be translated into the amount of carbon emitted into the atmosphere. Together with the fuel load of the vegetation, emissions of various trace gases and aerosol species can be estimated using emission factors (Figure 38). These emission factors are land cover dependent and are derived from observational data reported in the literature. Figure 39 shows a characteristic pattern of simulated burned area by MPI-ESM for present day climate conditions and their related carbon emissions into the atmosphere. These simulations show prominent fire occurrence in the semiarid areas of Africa and Eurasia as well in North America and Australia. As the model accounts for man made influence only as a function of population density and neglects socioeconomic or cultural differences, strong regional biases occur for the simulated burned area when compared to observations.

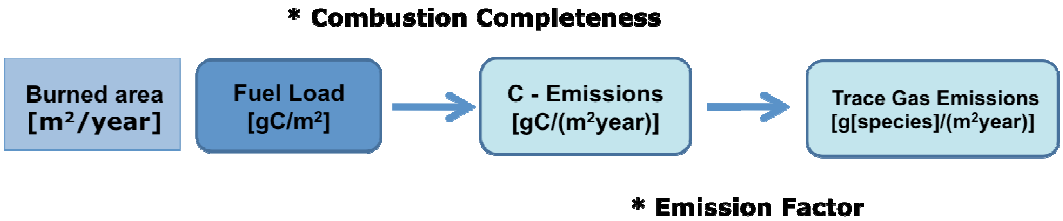


Figure 38: The conversion of burned area reported in the CCI dataset to the emissions of trace gases in the global vegetation model JSBACH.

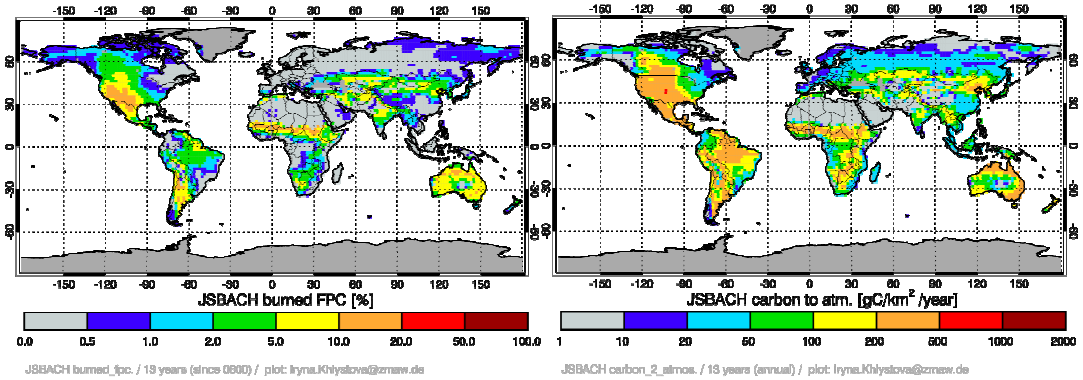


Figure 39. Characteristic pattern of burned area in MPI-ESM (left) and related simulated carbon emissions (right)

4.9.3 Prescribing satellite fire products in MPI-ESM

To assess the impact of the fire CCI dataset on MPI-ESM, an interface was developed which allows prescribing burned area from an external data source in the model as a boundary condition, instead of simulating it interactively. The interface is combined with a pre-processing procedure that converts the input data into a format suitable for JSBACH. As the definition of land and ocean boundaries typically differ between models and satellite data products, the interface includes a consistent treatment of the land-sea mask in the input data as well as in the JSBACH model to preserve the full information content of the observational data.

As a precursor dataset for the fire CCI the Global Fire Emission Database (GFED, version 3) was chosen (van der Werf *et. al.,* 2010, Giglio *et. al*., 2010). GFEDv3 reports burned area for the time period 1997 – 2010 on a monthly basis with a spatial resolution of 0.5 x 0.5 deg. For the application in JSBACH the burned area was mapped to the current model standard grid resolution of T63 (~1.8 x 1.8 deg).

The satellite based burned area product is then prescribed during the model simulations as an external boundary condition, similar to prescribing SST or CO2 concentration. As a first order approach, it is assumed that area burned for a particular PFT in a model grid cell is proportional to its fractional coverage of the particular grid cell.

As the carbon cycle is characterized by long timescales, such as the carbon uptake in the soil, vegetation model simulations typically require long spin-up periods before an equilibrium state is reached. A spin-up simulation using periodic climate forcing (representative for a pre-industrial climate) and periodic burned area boundary conditions for the years 1997-2010 is therefore performed until the carbon pools in the model reach equilibrium. The length of the spin-up period is in the order of 600-years, which is computationally very demanding and is therefore performed using an offline version of JSBACH, forced by periodic input from results of a GCM. Figure 40 shows simulated fire carbon emissions throughout the spin-up period when pre-scribing GFEDv3 burned area. It is observed that the fire emissions reach equilibrium after approximately 200 years. However, the different carbon pools considered in JSBACH require different spin up times. The soil carbon pool has the lowest turnover rate and

needs the longest to reach an equilibrium state. This gets important when burning of soil organic carbon will be considered in the model in the future.

The resulting fire carbon emissions are shown in Figure 40.

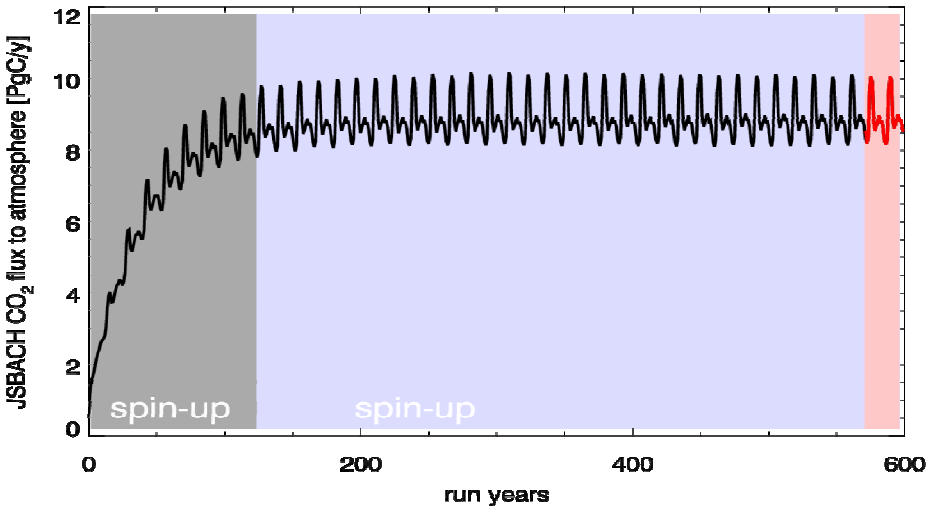


Figure 40: Fire Carbon emissions simulated within JSBACH using GFEDv3 reported burned area used for periodic boundary conditions.

Figure 41 shows the global distribution of burned area reported in GFEDv3 together with the fire carbon emissions as simulated within JSBACH averaged for the years 1997 to 2010 (one full forcing period of GFEDv3). Clearly evident are the hotspot regions for fires, such as Africa and South America.

Globally, the emissions amount to 8.9 PgC/year, which is higher than other reported estimates (~2-4 PgC/year). GFEDv3, for example, reports fire carbon emissions of 2.1 PgC/year. GFEDv3 uses a similar approach to the one we used within JSBACH. The burned area is prescribed in a vegetation model. In the case of GFEDv3 the CASA model is used. One reason for this substantial difference between the fire carbon emissions might be related with the partitioning of the burned area among different PFTs in a grid cell. In JSBACH we distributed the burned area equally over the prevailing PFTs. GFEDv3, however, combines burned area information with land cover information at high spatial scales for the partitioning. This information is available in the GFEDv3 pre-cursor data set and will be implemented in JSBACH in the future as well. Ongoing work includes a more detailed analysis of the differences arising from the different vegetation models applied. This will allow an assessment of the fire CCI data, in the context of uncertainties arising from the vegetation models used to convert burned area into fire carbon emissions.

It should also be emphasised that related satellite atmospheric datasets such as global maps of CO from IASI, CO_2 from GOSAT and O_3 from Sciamachy can be used to help verify the location and intensity of remote fires.

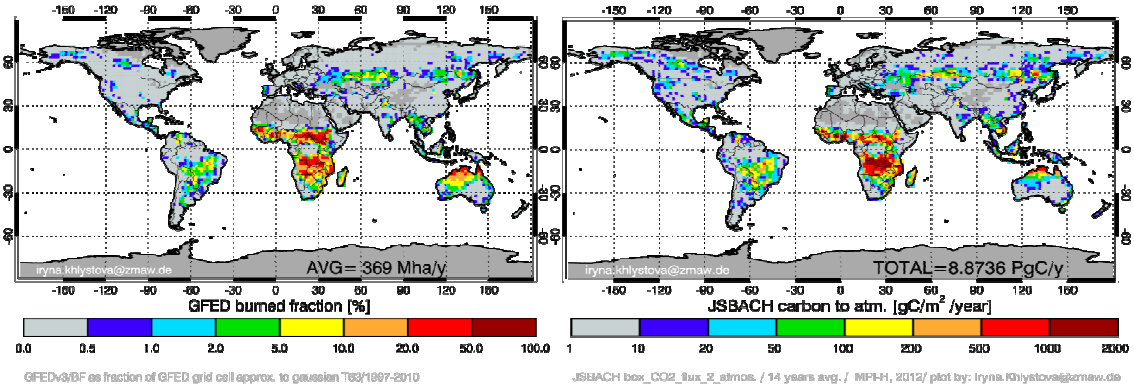


Figure 41. GFEDv3 burned area (left) and simulated fire carbon emissions (right) in [gC/m^2/year] averaged for one full forcing cycle using GFEDv3 burned area for the years 1997 to 2010 .

4.10 Glaciers

Not assessed by CMUG to date.

5. Summary of key points from a climate modelling perspective

This report documents some assessments of precursor products of the ESA CCI climate data records many of which the CCI datasets will be compared to when they become available. Table 10 lists the precursors used in the report and Table 11 summarises the assessment of the ECVs selected in this report. A few of the key points in this assessment are summarised here.

An assessment of SST by comparison with in-situ observations (drifting buoys) showed the importance of the coverage of the in-situ observing network in assessing the satellite product. The statistics clearly improved when the coverage of the in-situ network improved. This suggests caution should be exercised when assessing a satellite dataset with a sparse in-situ observing network.

A three way matchup technique is a powerful tool to independently assess observation errors for satellite and in-situ observations and this has been demonstrated for the SST dataset. This technique should be extended to other datasets where three independent measures of the same quantity are available. Also shown for SST was a validation of the uncertainties provided with the SST dataset by comparing with the bias of the dataset and it is clear this will be an important assessment for all the CCI datasets where a measure of uncertainty will be provided.

Level 4 SST analysis fields (i.e. HadISST2) were assessed by comparison with the ERA-Interim model fields and it was demonstrated how a number of important, but not immediately obvious, anomalies can be identified in the dataset which were not removed by the dataset provider. This kind of assessment is important for modellers to decide on the validity of a particular dataset for use in climate modelling and reanalysis applications. The SST dataset analysed was an ensemble of fields which were intended to cover the range of uncertainty in the measurements. This allows assessment of not only the primary

Pre-cursor Dataset	Variable	Area	Time span	Uncertainty included?
ARC SST	SST	Global	1991-2010	Yes
p-HadISST2	SST	Global	1899-2010	Yes
GlobCOLOUR	Chlorophyll concentration	Global	1995-2010	Yes
SSALTO/DUACS	SSH	Global	2002-2009	Yes
ISCCP	Clouds	Global	1982-2010	No
EUM IASI L2 ESA MIPAS	Ozone column Ozone profile	Global Global	2008-2009 2008-2009	No No
GlobCOVER	Land cover classification	Global	2005, 2009	No
GFEDv3	Fire burned area	Global	1997-2010	No

Table 10. A summary of pre-cursor datasets used in this report

measurement, in this case SST, but also the spread of the ensemble of measurements. Ensembles of retrieved climate data records of the same variable to encompass the uncertainty in the measurements will be more prevalent in the future and so it is important to strengthen collaboration between data providers and data users on developing methods to assess the spread of these ensembles as a means of validating the uncertainty. The SSH dataset assessment with regional reanalyses is another example shown for the Mediterranean Sea where it is planned to have an intense observing campaign in the next few years to allow more detailed validation of models and satellite datasets.

Methodology used for assessment of ECVs	Assessment of ECVs in this report
Climate Model (single, ensemble)	Clouds, Ozone, Land Cover, Fire
Re-analyses	SST, SSH, Ozone
Precursor datasets	N/A
Independent satellite or in situ measurements	SST, Clouds, Ozone
Related observations (surface and TOA fluxes, temperature, water vapour)	Clouds
Assimilation	Ocean colour, SSH, Ozone

Table 11. A summary of the type of assessments performed to assess ECVs in this report.

The assessment of a dataset through assimilation is illustrated with the GlobColour dataset where model analysed fields are clearly much closer to the observed fields after the

assimilation. The challenge is to show the improvement in the predictions after the assimilation of the dataset and in the case of ocean colour this is more challenging as the model itself clearly needs development. Assimilation studies using ozone datasets are also shown where the impact on the annual cycle and regional variations such as the polar vortex are studied.

The need for observation operators to represent the satellite measurements from climate model fields is shown for the clouds ECV where ISCCP data are used as a precursor dataset and the model fractional cloud cover from the COSP simulator are compared. Here the consistency with the associated radiation budget fields, cloud liquid water path and cloud liquid drop size from satellites and the model are used to demonstrate the importance of not only considering one variable in isolation when using satellite datasets for validation of model processes. The availability of the CMIP5 archive now allows datasets to be compared with an ensemble of different climate models which can help to better assess observational datasets as the model biases are more evident using this archive. Finally an assessment of datasets during an anomalous period (e.g. 1997-98 El Niño) is important as these are often used by modellers to asses how their models react to anomalous situations.

For the surface datasets (e.g. land cover and fire) the assessments are direct comparison between the model and new satellite derived fields, which provide model boundary conditions, and involve exploring the reasons for the differences and impacts on model simulations. The performance of the climate model with the new boundary conditions for example to show the changes in the carbon emissions can be used as a way to assess new surface datasets. It is critical here that all the surface variables are consistent with each other between datasets as the model will struggle to provide consistent surface analyses if not.

Finally it should be clear that what was not attempted in this study was to look at cross-ECV consistency which is also an important property for climate modelling applications. It is hoped this aspect can be explored in future studies by the CMUG using the CCI datasets where particular attention will be paid to this aspect. It should be noted that the individual ECV teams through their climate research groups will also be making complementary studies of their datasets but in general focussing less on climate modelling applications. A review of these CCI validation plans is planned as a separate CMUG document D2.3.

6. References

Balmaseda, A.B., K. Mogensen, F. Molteni, and A. Weaver, 2010: The NEMOVAR-COMBINE ocean re-analysis. *COMBINE Technical Report N°1, ECMWF, Reading.*

Bodas-Salcedo, A. et al., 2011: COSP: satellite simulation software for model assessment. *Bull. Amer. Meteor. Soc.*, **92***(8), 1023-1043. DOI: 10.1175/2011BAMS2856.1.*

Bonan, G.B., K. W. Oleson, M. Vertenstein, S. Levis, X. Zeng, Y. Dai, R.E. Dickinson, Z. Yang, 2002: The Land Surface Climatology of the Community Land Model Coupled to the NCAR Community Climate Model*. *J. Climate*, **15**, 3123–3149.

Bowman, D. M. J. S., Balch, J. K., Artaxo, P., Bond, W. J., Carlson, J. M., Cochrane, M. A., D'Antonio, C. M., DeFries, R. S., Doyle, J. C., Harrison, S. P., Johnston, F. H., Keeley, J. E., Krawchuk, M. A., Kull, C. A.,

Marston, J. B., Mortiz, M. A., Prentice, I. C., Roos, C. I., Scott, A. C., Swetnam, T. W., van der Werf, G. R., and Pyne, S. J.: Fire in the Earth System, *Science*, **324,** *481, doi:10.1126/science.1163886, 2009.*

Dibarboure G, Lauret O, Mertz F, Rosmorduc V, Maheu C (2009) SSALTO/DUACS user handbook, (M)SLA and (M)ADT near-real time and delayed time products. *CLS-DOS-NT-06.034. Ramonville-Saint-Agne, France: Aviso Altimetry*.

Geider, R. J., MacIntyre, H. L., and Kana, T. M. 1997 Dynamic model of phytoplankton growth and acclimation: responses of the balanced growth rate and the chlorophyll a: carbon ratio to light, nutrient-limitation and temperature, *Mar. Ecol.-Prog. Ser.*, **148**, *187-200.*

Giglio, L., Randerson, J. T., van der Werf, G. R., Kasibhatla, P. S., Collatz, G. J., Morton, D. C., and DeFries, R. S. 2010. Assessing variability and long-term trends in burned area by merging multiple satellite fire products, *Biogeosciences*, **7**, *1171-1186, doi:10.5194/bg-7-1171-2010*

Glecker, P.J., Taylor, K.E., Doutriaux, C. 2012: Performance metrics for climate models. *Journal of Geophysical Research, vol. 113, D06104, doi:10.1029/2007JD008972*

GCOS, 2011: Systematic observation requirements for satellite-based products for climate, 2011 update. *GCOS Report 154. WMO.*

Hagemann, S. and Loew, A. (2012): Combined evaluation of MPI-ESM land surface water and energy fluxes. *(In preparation.)*

Hemmings, J. C. P., Barciela, R. M., and Bell, M. J. 2008 Ocean color data assimilation with material conservation for improving model estimates of air-sea CO_2 flux. *J. Mar. Res.*, **66**, *87-126.*

Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, J, Yang, L., and Merchant, J.W., 2000, Development of a Global Land Cover Characteristics Database and IGBP DISCover from 1-km AVHRR Data, *Int. J. Remote Sens.* **21**, *1303-1330*

Martin, G. M. et al., 2006: The Physical Properties of the Atmosphere in the New Hadley Centre Global Environmental Model (HadGEM1). Part I: Model Description and Global Climatology. *J. Climate*, **19**, *1274–1301. doi: http://dx.doi.org/10.1175/JCLI3636.1*

Martin, M. J., Hines, A., and Bell, M. J. 2007 Data assimilation in the FOAM operational short-range forecasting system: a description of the scheme and its impact. *Q. J. Roy. Meteor. Soc.*, **133**, *981-995*.

Merchant C J, D Llewellyn-Jones, R W Saunders, N A Rayner, E C Kent, C P Old, D Berry, A R Birks, T Blackmore, G K Corlett, O Embury, V L Jay, J Kennedy, C T Mutlow, T J Nightingale, A G OCarroll, M J Pritchard, J J Remedios and S Tett 2008, Deriving a sea surface temperature record suitable for climate change research from the along-track scanning radiometers, *Adv. Sp. Res*, **41** *(1), 1-11. doi:10.1016/j.asr.2007.07.041*

M. Michou et al., 2011: A new version of the CNRM Chemistry-Climate Model, CNRM-CCM: description and improvements from the CCMVal-2 simulations, *Geosci. Model Dev. Discuss.*, **4**, *1129–1183, 2011*

O'Carroll, A.G., Eyre, J.R. & Saunders, R.W. 2008. Three-way error analysis between AATSR, AMSR-E and *in situ* sea surface temperature observations. *J. Atmos. and Oceanic Tech.*, **25**(7), 1197-1207.

Olson, J.S., 1994a Global ecosystem framework-definitions *USGS EROS Data Center Internal Report, Sioux Falls, SD*

Olson, J.S., 1994b Global ecosystem framework-translation strategy *USGS EROS Data Center Internal Report, Sioux Falls, SD*

B. Pajot et al 2012 High resolution assimilation of IASI Ozone data with a global CTM, *Submitted to Atmos. Chem. Phys., Oct 2011*

Palmer, J. R., and Totterdell, I. J. 2001 Production and export in a global ocean ecosystem model, *Deep-Sea Res. Pt. I*, **48**, *1169-1198.*

Poulter, B., Ciais, P., Hodson, E., Lischke, H., Maignan, F., Plummer, S., and Zimmermann, N. E.: Plant functional type mapping for earth system models, *Geosci. Model Dev.*, **4**, *993-1010, doi:10.5194/gmd-4-993-2011, 2011*

Poulter, B., Ciais, P., Hodson, E., Lischke, H., Maignan, F., Plummer, S., and Zimmermann, N. E.: Plant functional type mapping for earth system models, *Geosci. Model Dev.*, **4**, *993-1010, doi:10.5194/gmd-4-993-2011, 2011*

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan. 2003 Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *J. Geophys. Res.* **108***(D14), 4407, doi:10.1029/2002JD002670, 2003.*

Reichler, T. and Kim, J. (2008): How well die coupled models simulate today's climate? *Bull. Amer. Meteorol. Soc*., March 2008. 303-311. doi:10.1175/BAMS-89-3-303

Sevault, F., Somot, S., Beuvier, J., 2009. A regional version of the NEMO ocean engine on the Mediterranean Sea: NEMOMED8 user's guide. *Technical note 107, CNRM, Météo-France, Toulouse.*

Storkey, D., Blockley, E. W., Furner, R., Guiavarc'h, C., Lea, D., Martin, M. J., Barciela, R. M., Hines, A., Hyder, P., and Siddorn, J. R. 2010 Forecasting the ocean state using NEMO: The new FOAM system. *Journal of Operational Oceanography*, **3**, *3-15.*

Tian et al., 2009 : Effects of stratosphere/troposphere chemistry coupling on tropospheric ozone, *J.Geophys. Res.*, **115**, *D00M04, doi:10.1029/2009JD013515.*

U.S. Geological Survey, 2001 Global land cover characteristics data base version 2.0 *http://edcdaac.usgs.gov/glcc/globdoc2_0.htm*

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y., and van Leeuwen, T. T., 2010, Global fire emissions and the contribution of deforestation, savannah, forest, agricultural, and peat fires (1997–2009), *Atmos. Chem. Phys*., 10, *11707-11735, doi:10.5194/acp-10-11707-2010.*

Woodruff, Scott D.; Worley, Steven J.; Lubker, Sandra J.; Ji, Zaihua; Freeman, J. Eric; Berry, David I.; Brohan, Philip; Kent, Elizabeth C.; Reynolds, Richard W.; Smith, Shawn R.; and Wilkinson, Clive 2011 ICOADS Release 2.5: extensions and enhancements to the surface marine meteorological archive (2011). *Publications, Agencies and Staff of the U.S. Department of Commerce. Paper 332. http://digitalcommons.unl.edu/usdeptcommercepub/332*