Benchmarking CMIP5 models with a subset of ESA CCI Phase 2 data using the ESMValTool

Article

Accepted Version

Lauer, A., Eyring, V., Righi, M., Buchwitz, M., Defourny, P., Evaldsson, M., Friedlingstein, P., de Jeu, R., de Leeuw, G., Loew, A., Merchant, C. J. ORCID: https://orcid.org/0000-0003-4687-9850, Müller, B., Popp, T., Reuter, M., Sandven, S., Senftleben, D., Stengel, M., van Roozendael, M., Wenzel, S. and Willen, U. (2017) Benchmarking CMIP5 models with a subset of ESA CCI Phase 2 data using the ESMValTool. Remote Sensing of Environment, 203. pp. 9-39. ISSN 0034-4257 doi: 10.1016/j.rse.2017.01.007 Available at https://centaur.reading.ac.uk/68615/

To link to this article DOI: http://dx.doi.org/10.1016/j.rse.2017.01.007

Publisher: Elsevier

Benchmarking CMIP5 models with a subset of ESA CCI Phase 2 data using the ESMValTool

Axel Lauer[a,*], Veronika Eyring[a], Mattia Righi[a], Michael Buchwitz[b], Pierre Defourny[c], Martin Evaldsson[d], Pierre Friedlingstein[e], Richard de Jeu[f], Gerrit de Leeuw[g,h], Alexander Loew[i], Christopher J. Merchant[j], Benjamin Müller[i], Thomas Popp[k], Maximilian Reuter[b], Stein Sandven[l], Daniel Senftleben[a], Martin Stengel[m], Michel Van Roozendael[n], Sabrina Wenzel[a], and Ulrika Willén[d]

[a] Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[b] Institute of Environmental Physics (IUP), University of Bremen FB1, Bremen, Germany
[c] Université catholique de Louvain, Earth and Life Institute, Belgium
[d] Swedish Meteorological and Hydrological Institute (SMHI), 60176 Norrköping, Sweden
[e] College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, EX4 4QE, UK
[f] Transmissivity/VanderSat, Space Technology Center, Noordwijk, Netherlands
[g] Finnish Meteorological Institute (FMI), Climate Research Unit, 00101 Helsinki, Finland
[h] Department of Physics, University of Helsinki, 00014 Helsinki, Finland
[i] Department of Geography, Ludwig Maximilians Universität, Munich, Germany
[j] University of Reading, Reading, UK
[k] Deutsches Zentrum für Luft- und Raumfahrt (DLR), Deutsches Fernerkundungsdatenzentrum, Oberpfaffenhofen, Germany
[l] Nansen Environmental and Remote Sensing Center, Bergen, Norway
[m] Deutscher Wetterdienst (DWD), Offenbach, Germany
[n] Belgian Institute for Space Aeronomy, Brussels, Belgium

[*] Corresponding author email: axel.lauer@dlr.de

Abstract

The Coupled Model Intercomparison Project (CMIP) is now moving into its sixth phase and aims at a more routine evaluation of the models as soon as the model output is published to the Earth System Grid Federation (ESGF). To meet this goal the Earth System Model Evaluation Tool (ESMValTool), a community diagnostics and performance metrics tool for the systematic evaluation of Earth system models (ESMs) in CMIP, has been developed and a first version (1.0) released as open source software in 2015. Here, an enhanced version of the ESMValTool is presented that exploits a subset of Essential Climate Variables (ECVs) from the European Space Agency's Climate Change Initiative (ESA CCI) Phase 2 and this version is used to demonstrate the value of the data for model evaluation. This subset includes consistent, long-term time series of ECVs obtained from harmonized, reprocessed products from different satellite instruments for sea surface temperature, sea ice, cloud, soil moisture, land cover, aerosol, ozone, and greenhouse gases. The ESA CCI data allow extending the calculation of performance metrics as summary statistics for some variables and add an important alternative data set in other cases where observations are already available. The provision of uncertainty estimates on a per grid basis for the ESA CCI data sets is used in a new extended version of the Taylor diagram and provides important additional information for a more objective evaluation of the models. In our analysis we place a specific focus on the comparability of model and satellite data both in time and space. The ESA CCI data are well suited for an evaluation of results from global climate models across ESM compartments as well as an analysis of long-term trends, variability and change in the context of a changing climate. The enhanced version of the ESMValTool is released as open source software and ready to support routine model evaluation in CMIP6 and at individual modeling centers.

1 Introduction

Earth system models (ESMs) are essential tools for improving our understanding of the climate system as well as for assessing the response of the climate system to different natural or anthropogenic perturbations. Understanding of the capabilities and limitations of ESMs is a cornerstone for the interpretation of model results as well as for improving the models and is obtained through a comprehensive evaluation of the models with observations. Both improved models and an improved process understanding of the climate are important steps towards reducing the uncertainties in projections of future climate change and providing more trustworthy information for policy guidance. The number of models participating in the Coupled Model Intercomparison Project (CMIP) that is internationally coordinating ESM simulations is growing and the models participating are increasing in complexity and resolution. Traceable evaluation of the CMIP model ensemble with observations is therefore a challenging task.

The experimental design of the sixth phase of the Coupled Model Intercomparison Project (CMIP6) is now finalized. A central goal of CMIP6 is an improved and more routine evaluation of the participating climate models with observations (Eyring et al., 2016a). The CMIP Diagnostic, Evaluation and Characterization of Klima (DECK) experiments and CMIP historical simulations will provide the basis for the documentation of the model simulation characteristics. The aim is in particular to diagnose and improve the understanding of the origins and consequences of systematic model errors and inter-model spread.

To support this goal, the Earth System Model Evaluation Tool (ESMValTool, Eyring et al., 2016b) has been developed. The ESMValTool is a community diagnostics and performance metrics tool for systematic evaluation of Earth system models in CMIP, which has been developed by multiple institutions in several international projects. A first version of the

ESMValTool has been released as open source software in 2015 and is rapidly developing to include additional evaluation diagnostics and technical improvements. The ESMValTool will be - together with other software packages such as the Program for Climate Model Diagnostics and Intercomparison (PCMDI) metrics package (PMP, Gleckler et al., 2016) and the NCAR Climate Variability Diagnostic Package (CVDP, Phillips et al., 2014) that is included in the ESMValTool as a separate namelist - applied to CMIP6 results to provide a broad and comprehensive characterization of the CMIP6 models as soon as the output is published to the Earth System Grid Federation (ESGF). The foundation that will enable this is an efficient infrastructure (Eyring et al., 2016c) and the community-based experimental protocols and conventions of CMIP, including their extension to obs4MIPs (Teixeira et al., 2014; Ferraro et al., 2015) and ana4MIPs (https://www.earthsystemcog.org/projects/ana4mips/).

The Climate Change Initiative of the European Space Agency (ESA CCI) is a large international effort that provides global, long-term satellite data sets to the climate community that can be used to evaluate and improve the models (Hollmann et al., 2013). The ESA CCI is exploiting a large number of satellite observations to create robust long-term global records of selected essential climate variables (ECVs; GCOS, 2010; Bojinski et al., 2014) from numerous satellites and instruments. In this study, a subset of the ESA CCI Phase 2 ECVs has been implemented into the ESMValTool. This enhanced version of the ESMValTool is then used to evaluate CMIP5 models. ESA CCI data sets implemented so far include sea surface temperature, sea ice, cloud, soil moisture, land cover, aerosol, ozone, and greenhouse gases.

This paper is organized as follows: section 2 provides a brief description of the ESA CCI data used in this study to evaluate CMIP5 models. Section 3 summarizes the models and model simulations that are evaluated with the ESA CCI and other data, section 4 demonstrates the usage of the implemented ESA CCI data in summary statistics applied to CMIP5 models by calculating relative space-time root-mean-square deviations (RMSDs) from climatological mean seasonal cycles of selected ECVs. A specific focus is placed on the consideration of uncertainty information provided with the ESA CCI data, which is displayed in extended Taylor diagrams (Taylor, 2001) that are widely used to assess the performance of large model ensembles in reproducing observed quantities. Further insights into the evaluation of CMIP5 models with ESA CCI data and comparisons of ESA CCI data with alternative observational data sets are presented in section 5. A summary and a discussion of the main results and conclusions are given in section 6.

2 Brief description of the ESA CCI data

The datasets from the ESA CCI Phase 2 implemented into the enhanced version of the ESMValTool presented in this study are briefly described in the following. We would like to note that these datasets are only a subset of all CCIs available. It is planned to implement additional CCIs such as ocean color, sea level, ice sheets and fire as well as additional ECVs from the CCIs included here into future releases of the ESMValTool.

2.1 Sea surface temperature

The ESA CCI sea surface temperature (SST) data set (Merchant et al., 2014a,b) provides multi-decadal products of SST derived from infrared brightness temperatures measured from satellites. SST products (Rayner et al., 2015) are generated at full sensor resolution (1 to > 4 km) and are averaged on a regular latitude-longitude grid (0.05°). A gap-filled (Level 4 SST analysis) product covering the time 1992-2010 is currently used with the ESMValTool diagnostics. The Level 4 (L4) SST analysis is a daily 3-dimensional variational analysis of satellite data with a grid resolution of 0.05°. The analysis system is the Operational Sea surface Temperature and sea-Ice

Analysis (OSTIA) with improved covariance parameterization (Roberts-Jones et al., 2016). The L4 SST analysis has relatively good feature resolution, which is nonetheless lower than the grid resolution, and varies with the density of satellite coverage (Reynolds et al., 2013). Unlike the operational OSTIA products (Donlon et al., 2012) and the older OSTIA-based observational re-processing (Roberts-Jones et al., 2012), no in situ data are used in this CCI product. The product represents the daily value of SST at a nominal depth of 20 cm, representative of the SST measured by drifting buoys and bucket observations. This is possible because the lower-level SST CCI products contain both the skin (radiometric) temperature of the ocean surface at the time of satellite observation estimated based on radiative transfer physics (e.g., Embury et al., 2012a), and a turbulence-model-based adjustment to the 20 cm depth SST at a standardized time of day. The adjusted SST estimate is used as input to the L4 SST analysis. This means that the L4 SST analysis can be treated as independent of in situ data, and useful as a comparison point for the many SST products that are tuned to and/or incorporate in situ data. The standardization of the adjustment with respect to time of day is intended to reduce aliasing of the diurnal cycle into false long-term trends, as satellite overpass times vary (Embury et al., 2012b). All SSTs are provided with estimates of total uncertainty, and the L4 SST analysis product includes an operationally produced estimate of sea ice concentration (Good and Rayner, 2014).

Merchant et al. (2014a,b) provide an assessment of the accuracy of this product by comparison with more than 2.4 million buoys from different observational networks. A global median difference against drifting buoys of +0.05 K is observed, with a standard deviation (including the ~0.2 K uncertainty in the drifting buoy measurements) of 0.28 K. The comparison with Argo measurements at ~5 m depth (only from the latter part of the record) gives +0.04 K and 0.26 K respectively. Systematic regional errors on spatial scales of ~1000 km range from -0.5 K to +0.5

K, with positive bias of +0.09 K across equatorial regions overall (relative to measurements of the global tropical moored buoy array). Regions persistently affected by mineral atmospheric aerosol, particularly Saharan dust, appear negatively biased.

2.2 Sea ice

The ESA CCI sea ice data set provides observational data for sea ice concentration (sic) and sea ice thickness (sit) that are based on satellite retrievals for both Arctic and Antarctic sea ice. The sic data set is based on passive microwave data from Special Sensor Microwave Imager (SSM/I) covering the time period 1992 to 2008 and the Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E) covering the time period 2003-2010 (Lavergne and Rinne, 2014). The data sets are provided as daily gridded sic fields for both northern hemisphere and southern hemisphere on an equal area grid with 25 km grid spacing. Separate data sets for SSM/I and AMSR-E are available, where the SSM/I product is more mature, while the AMSR-E data can provide higher-resolution products. In addition, daily maps of total standard error and quality control flags are provided. The ESA CCI sea ice data set is built upon the algorithms and processing software originally developed at the EUMETSAT OSI SAF for their sic data set (RD-11). The algorithm used to produce the sic data sets is based on an extensive algorithm intercomparison study (Ivanova et al., 2015), aiming at identifying the optimal algorithm for producing sic data sets. In their study, a systematic comparison of 30 algorithms was done for different ice conditions, seasons and regions. The result was an implementation of a new algorithm for sic retrieval. It is based on a combination of previous algorithms and use of dynamic tie points and atmospheric correction of input brightness temperatures. Error sources of the sic products are particularly related to the marginal ice zone, areas of thin ice, melt-ponds in the summer season (Kern et al., 2016) and land contamination in coastal regions.

So far, only sea ice concentration and its standard error from the ESA CCI sea ice data set have been implemented into the ESMValTool. Sea ice thickness data sets from radar altimeter are also developed in the CCI project, but a final data set is not yet available. The ice thickness retrieval is based on sea ice freeboard measurements from altimeter that are converted to thickness using the hydrostatic equilibrium assumption and a priori knowledge about snow thickness, snow and ice density and penetration depth of the radar signal (Kern et al., 2015). The first ice thickness data set from ENVISAT for the period 2002 to 2012 has been presented (Lavergne and Rinne, 2014) as monthly mean thickness for the winter months in the Arctic. There are significant uncertainties in these results so far, which requires further studies to obtain a reliable product. Results from CryoSat thickness retrievals from 2010 to present, however, show promising results (e.g., Ricker et al., 2014, Kwok and Cunningham, 2015).

2.3 Cloud

The ESA CCI cloud data sets contain cloud property data retrieved from the passive satellite imager sensors AVHRR, MODIS, ATSR-2, AATSR and MERIS (Stengel et al., 2016a). Depending on the particular data set, time periods of 9 to 33 years between 1982 and 2014 are covered. In this study we used the Cloud_cci AVHRR-PM v2.0 data set (Stengel et al., 2016b), which is composed of data from AVHRR on-board NOAA-7, -9, -11, -14, -16, -18 and -19 and represents a nearly seamless time series from 1982 through 2014. The ESA CCI cloud data sets include cloud fraction (or cloud mask), thermodynamic phase, cloud top pressure (also converted to temperature and height), cloud optical thickness, cloud effective radius, cloud albedo and cloud liquid/ice water path. Various processing levels are available from Level 2 (pixel-based data) to daily sampled data (Level 3U) and monthly averages and histograms (Level 3C). All cloud properties are accompanied by pixel-based uncertainty estimates. While for most variables these estimates are based on optimal estimation theory, cloud mask uncertainty is based on hit rate scores against measurements from the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP). All pixel level uncertainties are propagated in a mathematically consistent way into the Level 3C products.

In this study monthly mean cloud fraction data (inferred from Level 3C data product with 0.5° resolution on a latitude-longitude grid) are used for comparison with CMIP5 model results. Cloud fraction represents the monthly summary of the results of Community Cloud retrieval for CLimate (CC4CL) cloud detection scheme (Sus et al., 2016; McGarragh et al., 2016). The monthly mean cloud detection uncertainty is also inferred from Level 3C products.

CC4CL cloud detection results have been validated against CALIOP space-based lidar measurements, with a global Kuipers score of 0.66 and a global hit rate of 81% (Karl-Göran Karlsson, personal communication) demonstrating the high quality of the cloud detection in the AVHRR-PM v2.0 data set.

It needs to be noted, that the Cloud_cci AVHRR-PM data set has a few limitations of which particularly the underrepresentation of optically very thin clouds (with optical thicknesses of below 0.15) and the sparse temporal sampling (twice a day for non-polar regions) is of relevance when using this data set for model evaluation. Particularly difficult conditions for cloud detection are polar night periods, for which the detection scores decrease significantly in the current version of the data set. Furthermore, the monthly cloud fraction data and the corresponding uncertainties of the Cloud_cci AVHRR-PM data set used in this study have not undergone any further processing such as satellite drift correction.

2.4 Soil moisture

The ESA CCI soil moisture product is the first ever multi-decadal satellite-based soil moisture product and is currently available for the time period 1978-2015 on a daily basis and at a spatial resolution of 0.25°x0.25°. The ESA CCI product represents soil moisture of the first centimeters of the soil and has been generated by merging active and passive microwave-based soil moisture products from multiple satellite missions (Liu et al., 2011, 2012; Wagner et al., 2012).

Dorigo et al. (2014) provide a comprehensive validation of the ESA CCI soil moisture using 932 in situ observation sites from 29 different observing networks (Dorigo et al., 2011, 2013). Despite the large difficulties in validating coarse resolution satellite soil moisture products with in situ point like observations (Crow et al., 2012), they conclude that the ESA CCI soil moisture product has an average unbiased root-mean-square error (RMSE) of 0.05 m^3 m^{-3}. It was shown that trends in the CCI observations largely agree with those obtained from various reanalysis products as well as precipitation, and vegetation vigor observations (Albergel et al., 2012; Dorigo et al., 2012). In addition, over the last seven years the ESA CCI soil moisture data set has been used for the yearly State of the Climate Reports issued by the National Oceanic and Atmospheric Administration (NOAA; e.g., Dorigo et al., 2016). Within these studies strong similarities were found between the spatial annual anomalies of ESA CCI soil moisture, and the terrestrial water storage from the Gravity Recovery and Climate Experiment (GRACE; e.g., Willet et al., 2014).

The ESA CCI soil moisture data set provides a multitude of quality flags and only soil moisture estimates considered reliable are used to create the data product. Snow covered areas and frozen ground are typically masked as well as dense or heterogeneously vegetated areas with high optical depth that are not expected to provide reliable soil moisture estimates (Loew, 2008; Parinussa et al., 2011).

2.5 Land cover

The ESA CCI land cover time series is the first consistent 300 m global land cover data set providing a characterization of the land surface from 1998 to 2012. The ESA CCI land cover product (v1.6.1) corresponds to high resolution global land cover information representative of three 5-year periods, referred to as epochs, for 2000 (1998-2002), 2005 (2003-2007) and 2010 (2008-2012). The three global land cover maps describe all the terrestrial areas by 22 land cover classes explicitly defined by a set of classifiers according to the United Nations Land Cover Classification System, each classifier referring to vegetation life form, leaf type and leaf longevity, flooding regime, non-vegetated cover types and artificiality (Di Gregorio, 2005).

The whole archive of full (300 m) and reduced resolution (1000 m) MERIS data acquired from 2003 to 2012 was first pre-processed and successfully fused as surface reflectance thanks to a set of improved algorithms for radiometric calibration, geometric and atmospheric corrections, and advanced cloud screening. A per pixel classification process, combining machine learning and unsupervised algorithms, was then applied to the full time series to serve as a baseline to derive land cover maps corresponding to each epoch. As temporal consistency was found as the most important requirement for the climate modeling community, a multi-year integration strategy was chosen for its better performance in reducing variability and improving stability (Bontemps et al., 2012). Detected from the full-resolution Satellite Pour l'Observation de la Terre (SPOT) vegetation time series (1998-2012), the land cover change corresponding to each epoch was applied through back- and up-dating methods but only concerning the main macroscopic changes observed for the forest classes (Li et al., 2016).

Inland open-water bodies and coastlines were mapped using wide-swath mode, image mode at medium-resolution (150 m) acquired by the Advanced Synthetic Aperture Radar sensor aboard ENVISAT satellite for a single period (2005-2010) (Santoro and Wegmüller, 2014) and then largely complemented with ancillary data.

The accuracy of the 2010 land cover product was estimated to 74.1% using the 2308 samples globally distributed and interpreted by regional experts. Further information on the accuracy of the ESA CCI land cover product in comparison to other existing global land cover data sets is provided by Tsendbazar et al. (2015).

In order to transform the ESA CCI land cover in Plant Functional Types (PFTs) distribution useable in ESMs, a CCI land cover user tool available from the visualization interface (http://maps.elie.ucl.ac.be/CCI/viewer/) can be used to apply a default or user-defined cross-walking table converting each land cover class into the corresponding proportions of PFT at the pixel level. This conversion also includes an aggregation of the different PFT distribution to coarser resolution grid cell in various projection systems.

In addition, the ESA CCI land cover products include information on the land surface seasonality at 1 km resolution which comprise climatological information of the vegetation greenness from Normalized Differenced Vegetation Index (NDVI) data as well as probabilities of snow and fire occurrences on a weekly basis at the pixel level. These were derived from SPOT vegetation daily observations from 1998 to 2012 as well from the corresponding MODIS time series. The inter-annual variability of these land surface seasonality variables was also computed from these 15-year time series on a weekly basis that can be used for comparison with models.

2.6 Aerosol

The ESA aerosol CCI team produces several long-term aerosol data sets (Popp et al., 2016) in response to Global Climate Observing System (GCOS) requirements, including variables such as aerosol optical depth (AOD) (from two Along-Track Scanning Radiometers (ATSR), the MEdium Resolution Imaging Spectrometer (MERIS) and the POLarization and Directionality of the Earth's Reflectances (POLDER) instrument), and stratospheric vertical extinction profiles (using stellar occultation by the Global Ozone Monitoring by Occultation of Stars (GOMOS) instrument). In response to the AEROCOM (http://aerocom.zmaw.de/) modeling community needs, also information on aerosol composition such as fine-mode AOD (from radiometers) or dust AOD (from the Infrared Atmospheric Sounding Interferometer IASI) and absorption AOD (from ATSR and POLDER) are derived from the retrieved mixing ratio of various aerosol components. All products include uncertainty estimates and are validated versus ground-based reference data (AERONET, Holben et al., 1998) by independent experts. For the retrieval of aerosol parameters from ATSR and IASI observations several algorithms are used, each of which applies different physical principles and mathematical methods and thus different solutions to the inversion problem. In the case of the ATSR radiometers, three algorithms (ADV from FMI, ORAC from Oxford University and RAL and SU from Swansea University) do perform very similarly, but with regional differences in both coverage and quantitative results, with none of them performing better than the others everywhere (de Leeuw et al., 2015).

The ESA CCI aerosol product used in this paper is the 17-year climate data record including total AOD and fine mode AOD, both at 550 nm, produced by SU (version 4.21) using data from two similar sensors: the ATSR-2 on the European Remote Sensing Satellite 2 (ERS-2-ATSR-2), covering the time period 1995-2003, and the Advanced ATSR (AATSR) on ESA's

Environmental Satellite (ENVISAT-AATSR, from 2002 to April 2012). Level 3 (L3) monthly mean data are used and only years with a full 12 months of data coverage are considered. Incomplete years from either platform (1995, 1996 and 2003) are not taken into account, restricting our analysis to the time period 1997-2011. The agreement of the data from the two ATSR instruments during the overlapping period 2002-2003 was found to be very good making it easy to combine the two data sets into a single time series. Here, we focus on total aerosol optical depth (od550aer), fine mode AOD (od550lt1aer), absorption optical depth at 550 nm (abs550aer) and AOD at 870 nm (od870aer). As an alternative observational data set, we use the L3 collection 6 data from the Moderate Resolution Imaging Spectroradiometer (MODIS; Levy et al., 2013) onboard Terra covering the time period 2003-2014.

2.7 Ozone

The ESA ozone CCI team produces a large number of L2 and L3 ozone data sets derived from various satellite sensors operating in nadir, limb and solar/stellar occultation geometries (see e.g. Miles et al., 2015; Lerot et al., 2014; Sofieva et al., 2013). In this work we use the total column ozone (toz) data sets which consist of combined and harmonized L3 data covering the time period between 1997 and 2010 (Coldewey-Egbers et al., 2015). Data from three platforms/instruments, the Global Ozone Monitoring Experiment (GOME) onboard the European Research Satellite 2 (ERS-2/GOME, 1996-2003), ENVISAT/SCIAMACHY (2003-2007), and GOME-2 onboard the Metop satellites (METEOP/GOME-2, 2007-2011) are provided as a merged gridded data set.

In additions to the total ozone data sets, we also include the ESA CCI limb gridded profile data, which consist of merged L3 monthly and zonally averaged data covering the time period 2007-2008 based on six different sensors, the MIPAS, SCIAMACHY, and GOMOS instruments onboard the ESA ENVISAT platform, the Optical Spectrograph and InfraRed Imaging System (OSIRIS) and the Sub-Millimetre Radiometer (SMR) onboard of Odin, and the Atmospheric Chemistry Experiment (ACE) instruments on Canadian SciSat platform.

The ozone CCI data sets used in this work have been extensively validated against ground-based networks of Dobson and Brewer total ozone spectrophotometers (Koukouli et al., 2015), as well as reference profile data sets from ozone sonde and lidar instruments (Hubert et al., 2016; Keppens et al., 2015). These studies have demonstrated that CCI total column ozone data sets closely match the accuracy and stability requirements defined by GCOS. Ozone profile data also comply with GCOS requirements but only in a limited range of altitudes, covering the mid- to upper stratosphere. In the upper-troposphere and lower stratosphere, the accuracy, precision and stability of current data sets are still to be improved. Validation studies concentrating on L3 products have shown that the main source of uncertainty in gridded or merged data sets is related to the limited sampling of satellite instruments. This source of uncertainty is especially significant in polar spring conditions when the ozone field is characterized by a large variability in space and time.

As an alternative reference data set for total ozone columns, we use data from the combined NIWA data set (Bodeker et al., 2005) covering the time period 1980-2010.

2.8 Greenhouse gases (GHG): XCO_2

The ESA CCI GHG product XCO_2 is retrieved from measurements of the two satellite instruments SCIAMACHY/ENVISAT (Bovensmann et al., 1999; Burrows et al., 1995) and TANSO-FTS/GOSAT (Kuze et al., 2009). XCO_2 is a dimensionless quantity (unit: ppm) defined as the vertical column of CO_2 divided by the vertical column of dry air (see Buchwitz et al.

(2005) for details). The XCO_2 distribution, the number of observations, the reported XCO_2 uncertainty and the XCO_2 standard deviation are available for 2003-2008 (land only) and 2009-2014 (land and ocean).

XCO_2 is retrieved from radiance spectra in the near-infrared/short-wave infrared (NIR/SWIR) spectral range using (mostly) optimal estimation (Rodgers, 2000) retrieval algorithms. Each retrieval algorithm used to generate the corresponding Level 2 (L2) product has an underlying radiative transfer model and a number of fit parameters (the so-called state vector elements), which are iteratively adjusted until the simulated radiance spectrum gives an optimal fit to the observed radiance spectrum (considering, e.g., instrument noise and a priori knowledge of relevant atmospheric parameters). For details we refer to the Algorithm Theoretical Basis Documents (ATBDs) available from the GHG-CCI website (http://www.esa-ghg-cci.org/sites/default/files/documents/public/documents/GHG-CCI_DATA.html) for each individual L2 data product. For the generation of the gridded L3 obs4MIPs product at monthly time resolution a spatial resolution of 5°x5° has been selected (instead of, e.g., 1°x1°) to ensure better noise suppression (note that the underlying individual satellite retrievals as contained in the L2 products are sparse due to very strict quality filtering).

The gridded L3 obs4MIPs products have been generated from the individual sensor/algorithm L2 XCO_2 input data. In order to correct for the use of different CO_2 a priori assumptions in the independently retrieved products, all products have been brought to a common a priori using the Simple Empirical CO_2 Model (SECM) described by Reuter et al. (2012). After this, a gridded L3 product is generated from each L2 product by averaging all soundings onto a 5°x5° monthly grid. Only those grid cells are further considered having a standard error of less than 2 ppm. The grid cell uncertainty is computed from the reported L2 uncertainties and a term accounting for potential regional and temporal biases. To avoid potential discontinuities in the obs4MIPs time series, each L3 product has been offset corrected to have the same mean value of all overlapping grid boxes. The obs4MIPs XCO_2 value in a given grid cell is computed as the mean of the individual L3 values. Finally a filtering procedure has been applied to remove “unreliable” grid cells considering the overall noise error originating e.g. from instrumental noise (1.6 ppm) and total uncertainty (1.8 ppm) of each cell.

The obs4MIPs XCO_2 product has been validated by comparison with ground-based XCO_2 retrievals from the Total Carbon Column Observation Network (TCCON, Wunch et al., 2011) using version GGG2014 as a reference (Wunch et al., 2015). In short, the following has been found: for XCO_2 the mean difference (satellite minus TCCON) is 0.3 ppm and the standard deviation of the difference to TCCON is 1.2 ppm. The total uncertainty of the obs4MIPs product is therefore about 1.5 ppm (1-sigma, per monthly 5°x5° grid cell, obtained via linear adding instead of root-sum-square to be on the safe side). Details are given in Buchwitz and Reuter (2016).

Due to the gridding / averaging process applied to generate obs4MIPs products detailed time/location information is not available in the obs4MIPs data product and also averaging kernels are not (yet) part of these products. Typically, however, the satellite XCO_2 averaging kernel is close to unity. This is especially the case in the lower troposphere, where the CO_2 variability is typically largest. Therefore applying the averaging kernels typically changes the XCO_2 values by less than 1 ppm (Dils et al., 2014) and other error sources are likely more relevant for using the obs4MIPs product such as the representativity error. A representativity error originates from the fact that the GHG field from the obs4MIPs data set are derived by averaging spatially and temporally sparse satellite observations, i.e., are not representative for the "true" monthly mean value of a given grid cell. Note that the validation results reported in the previous paragraph have also been obtained without considering the averaging kernels. The differences given above include to some extent the representativity error as well as other error sources such as the uncertainty of the TCCON reference observations, which is 0.4 ppm (1-sigma). It is recommended to use the reported overall uncertainty range of 0.3 ± 1.2 ppm (1-sigma) and/or the reported uncertainties for each grid cell as given in the obs4MIPs product file.

3 Models and simulations

In this study we use output from almost 50 global climate models (Table 1) that participated in CMIP5 (Taylor et al. 2012). The model data were obtained from the World Climate Research Programme's (WCRP) CMIP5 data archive made available through the Earth System Grid Federation.

Table 1. CMIP5 coupled models used in this study (historical simulations extended beyond 2005 with RCP4.5 results). The models marked with asterisks (*) also provided model experiments with interactive ozone chemistry, models marked with daggers (†) also provided emission driven experiments with an interactive carbon cycle (historical emission driven simulations extended beyond 2005 with RCP8.5 results).

Model(s)	**Host Institute**	**Resolution (atmosphere)**	**References**
ACCESS1.0, ACCESS1.3	CSIRO (Commonwealth Scientific and Industrial Research Organisation, Australia) and BOM (Bureau of Meteorology, Australia)	1.9°x1.3°, L38	Bi et al. (2013)
BCC-CSM1.1, BCC-CSM1.1-M	Beijing Climate Center, China Meteorological Administration, China	2.8°x2.8°, L26; 1.1°x1.1°, L26	Wu et al. (2010), Wu (2012)
BNU-ESM†	College of Global Change and Earth System Science (GCESS), BNU, Beijing, China	2.8°x2.8°, L26	Ji et al. (2014)
CanCM4, CanESM2†	Canadian Center for Atmospheric Research, Canada	2.8°x2.8°, L35	Arora et al. (2011)
CCSM4	National Center for Atmospheric Research (NCAR), United States	1.3°x0.9°, L26	Gent et al. (2011)
CESM1-BGC†, CESM1-CAM5, CESM1-CAM5-1-FV2, CESM1-FASTCHEM,	NSF/DOE NCAR (National Center for Atmospheric Research) Boulder, CO, United States	1.3°x0.9°, L26; 1.3°x0.9°, L26; 1.3°x0.9°, L26; 1.3°x0.9°, L26;	Long et al. (2013) Hurrel et al. (2013)

CESM1-WACCM*		2.5°x1.9°, L66	Marsh et al. (2013)
CMCC-CM, CMCC-CMS	Centro Euro-Mediterraneo per i Cambiamenti Climatici (CMCC), Bologna, Italy	0.8°x0.8°, L31; 1.9°x1.9°, L95	http://www.cmcc.it/models; Scoccimarro et al. (2011)
CNRM-CM5*	Centre National de Recherches Météorologiques (CNRM), Météo-France and Centre Européen de Recherches et de Formation Avancée en Calcul Scientifique (CERFACS), France	1.4°x1.4°, L31	Voldoire et al. (2013)
CSIRO-Mk3.6.0	Australian Commonwealth Scientific and Industrial Research Organization (CSIRO) Marine and Atmospheric Research, Queensland Climate Change Centre of Excellence (QCCCE), Australia	1.9°x1.9°, L18	Rotstayn et al. (2010)
EC-EARTH	EC-Earth (European Earth System Model)	1.1°x1.1°, L62	Hazeleger et al. (2010)
FGOALS-g2, FGOALS-s2	Institute of Atmospheric Physics, Chinese Academy of Sciences (LASG)	2.8°x3°, L26; 2.8°x1.7°, L26	Li et al. (2010), http://www.lasg.ac.cn/FGOALS/CMIP5
FIO-ESM†	The First Institution of Oceanography, SOA, Qingdao, China	2.8°x2.8°, L26	Qiao et al. (2013)
GFDL-CM3*, GFDL-ESM2G†, GFDL-ESM2M† GFDL-CM2p1	Geophysical Fluid Dynamics Laboratory (NOAA GFDL), United States	2.5°x2.5°, L48; 2.5°x2°, L24; 2.5°x2°, L24; 2.5°x2°, L24	Donner et al. (2011); http://nomads.gfdl.noaa.gov/
GISS-E2-H*, GISS-E2-H-CC, GISS-E2-R*, GISS-E2-R-CC	Goddard Institute for Space Studies (NASA/GISS), United States	2.5°x2°, L40	Schmidt et al. (2006)
HadCM3, HadGEM2-AO HadGEM2-CC, HadGEM2-ES	Met Office Hadley Centre, United Kingdom	3.8°x2.5°, L19; 1.9°x1.3°, L60; 1.8°x1.3°, L38	Collins et al. (2001); Collins et al. (2008); Collins et al. (2011)
INM-CM4	Institute for Numerical Mathematics (INM), Russia	2°x1.5°, L21	Volodin et al. (2010)
IPSL-CM5A-LR, IPSL-CM5A-MR, IPSL-CM5B-LR	Institut Pierre Simon Laplace (IPSL), France	3.8°x1.9°, L39; 2.5°x0.6°, L39; 3.8°x1.9°, L39	Dufresne et al. (2013), Hourdin et al. (2013)
MIROC-ESM†, MIROC-ESM-CHEM*, MIROC4h, MIROC5	Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Atmosphere and Ocean Research Institute (AORI), University of Tokyo and National Institute for Environmental Studies (NIES), Japan	2.8°x2.8°, L80; 2.8°x2.8°, L80; 0.6°x0.6°, L56; 1.4°x1.4°, L40	Watanabe et al. (2011); Sakamoto et al. (2012); Watanabe et al. (2010)
MPI-ESM-LR†, MPI-ESM-MR, MPI-ESM-P	Max Planck Institute for Meteorology, Germany	1.9°x1.9°, L47	Roeckner et al. (2006); Stevens et al. (2013)
MRI-CGCM3, MRI-ESM1†	Meteorological Research Institute (MRI), Japan	1.1°x1.1°, L48	Yukimoto et al. (2011)
NorESM1-M, NorESM1-ME†	Norwegian Climate Centre, Norway	2.5°x1.9°, L26	Bentsen et al. (2013)

For all variables except for column average CO_2 (XCO_2), we analyze the concentration driven CMIP5 historical simulations - twentieth-century simulations for 1850-2005 conducted with the best record of natural and anthropogenic climate forcing. In order to extend the model runs beyond the year 2005, we use results from simulations forced under the Representative Concentration Pathways 4.5 for the years 2006-2014. RCP4.5 is a scenario applied within

CMIP5 prescribing future greenhouse gas concentrations and resulting in a radiative forcing of 4.5 W m^{-2} in the year 2100 relative to pre-industrial values (Clarke et al., 2007; Smith and Wigley, 2006; Wise et al., 2009). The differences in the forcings between 2006 and 2014 for the different emission scenarios (RCP2.6, RCP4.5, RCP8.5) are rather small and negligible compared with the variability of the ensemble members of an individual model. We chose the RCP for which the most data for the analyzed ECVs were available, which is RCP4.5.

For aerosol and ozone, the evaluation is only performed for the subset of CMIP5 models that has interactive aerosols and chemistry, respectively.

Since CO_2 is prescribed in the concentration driven historical simulations, we analyze the emission driven historical simulations (esmHistorical) for XCO_2. In this case, the simulations were extended beyond 2005 with the corresponding RCP8.5 (esmrcp85) simulations because emission driven simulations for RCP4.5 were not part of the CMIP5 experiment design.

If there are multiple ensemble members available for any given model, we only consider the ensemble member "r1i1p1" in our analysis. The only exceptions to this are the EC-EARTH model, for which complete data sets were only available for "r6i1p1" and the GISS-E2-H and GISS-E2-R models for which we used ensemble members with interactive ozone chemistry ("r1i1p2"; see section 5.7).

4 CMIP5 summary statistics

An assessment of the agreement of simulated climatological mean state and seasonal cycle for key variables such as ECVs with observations is commonly seen as a reasonable starting point for the evaluation of ESMs (e.g., Gleckler et al., 2008; Flato et al., 2013; Hagemann et al., 2013; Eyring et al., 2016b). Following Gleckler et al. (2008) and similar to Fig. 9.7 of Flato et al.

(2013), we start the evaluation of the models by calculating the normalized relative space-time RMSD of the climatological seasonal cycle from CMIP5 simulations compared with observations for selected variables (section 4.1) and extended Taylor diagrams summarizing the multi-year annual mean performance (section 4.2). For land use variables, no summary statistics are calculated because the observations are rather static, i.e. do not provide a seasonal cycle.

All variables except for sea ice concentration are averaged over the whole globe. Sea ice concentration is averaged over the latitude band 60°N to 90°N (Arctic, "NHpolar") and 60°S to 90°S (Antarctic, "SHpolar"). The model results are compared to a reference data set (marked with asterisks in Table 2) and - where other data are available - to an alternative observationally based data set. Table 3 gives an overview of the variables and the corresponding CMOR names used while the observationally based data sets used for the evaluation are summarized in Table 2. For the models, results are averaged over the years with observational data available given in Table 2. Note that if alternative observationally based data are available, only years covered by both, the reference and the alternative observations, are used.

Table 2. Observationally based data sets used for the model evaluation. The data sets marked with asterisks (*) are used as reference data sets in Figure 1 (lower right triangles), the other data sets are used as alternative data sets (upper left triangles in Figure 1, red stars in Figure 2). The variable names are defined in Table 3. The years specify the periods analyzed in Figure 1 and Figure 2.

Data set	**Type**	**Variable(s)**	**Resolution**	**Years (Figure 1, Figure 2)**	**Estimate of systematic errors**	**Reference(s)**
AIRS_L3_RetStd-v5	satellite	hus	1°x1°	2003-2010	~25%	Tian et al. (2013), Susskind et al. (2006)
BDBP	ozonesondes	tro3	-	2006-2007		Hassler et al. (2008, 2009)
CERES-EBAF*	satellite	rlut, rsut, sw_cre, lw_cre	1°x1°	2001-2012	~5 $W m^{-2}$	Loeb et al. (2009, 2012)
CLARA-A2	satellite	clt	0.5°x0.5°	1982-2014		Karlsson et al. (2013),

						http://dx.doi.org/10.5676/EUM_SAF_CM/CLARA_AVHRR/V002
ERA-Interim*	reanalysis	ta, tas, ua, va, zg; hus; clt	0.75°x0.75°	1980-2005; 2003-2010; 1982-2014		Dee et al. (2011)
ESA CCI Aerosol*	satellite	od550aer; od870aer, od550lt1aer, abs550aer	1°x1°	2003-2011; 1997-2011		Popp et al. (2016)
ESA CCI Cloud*	satellite	clt	0.5°x0.5°	1982-2014		Stengel et al. (2016b)
ESA CCI Greenhouse Gases	satellite	xco2	5°x5°	2009-2014	~1.5 ppm	Buchwitz and Reuter (2016)
ESA CCI Ozone*	satellite	toz; tro3	1°x1°; 360°x10°	1997-2010; 2007-2008		Van Roozendael et al. (2015)
ESA CCI Land Cover	satellite	lccs_class	300 m	2000, 2005, 2010		Defourny et al. (2015)
ESA CCI Sea Ice*	satellite	sic	25 km x 25 km	1992-2008		Sandven et al. (2015)
ESA CCI Sea Surface Temperature*	satellite-based/analysis	ts	0.05°x0.05°	1992-2010	~0.05 K (global median)	Merchant et al. (2014a,b)
ESA CCI Soil Moisture*	satellite	sm	0.25°x0.25°	1988-2005	~0.05 m^3 m^{-3}	Liu et al. (2011, 2012), Wagner et al. (2012)
GPCP_L3_v2.2*	satellite + gauge	pr	2.5°x2.5°	1980-2005	0-2 mm day^{-1}	Adler et al. (2003), Huffman and Bolvin (2013)
HadISST	satellite-based/analysis	ts	1°x1°	1992-2010		Rayner et al. (2003)
IGAG/SPARC	satellite + ozonesondes + model+ analysis	tro3	5°x5°	1960-2008		Cionni et al. (2011)
MODIS	satellite	clt; od550aer	1°x1°	2003-2014; 2003-2011		Platnick et al. (2003); Remer et al. (2005); doi: 10.5067/MODIS/MYD08_M3.006;
NCEP	reanalysis	ta, tas, ua, va, zg	2.5°x2.5°	1980-2005		Kalnay et al. (1996)
NIWA	satellite analysis	toz	1.25°x1°	1997-2010		Bodeker et al. (2005)
NSIDC-NT, NSIDC-BT	satellite	sic	25 km x 25 km	1992-2008		Cho et al. (1996), Comiso and Nishio (2008), Comiso (1995)
PATMOS-x	satellite	clt	1°x1°	1982-2014		Heidinger et al. (2014)

Table 3. Variables used.

Variable	Name	Unit	Comment
abs550aer	Ambient aerosol absorption optical thickness at 550 nm	1	

clt	Total cloud fraction	%	For the whole atmospheric column, as seen from the surface or the top of the atmosphere; includes both large-scale and convective clouds
hus	Specific humidity	1	
lccs_class	Land cover class	-	
lw_cre	Longwave cloud radiative effect	W m^{-2}	At the top of the atmosphere
mrsos	Soil moisture in upper portion of soil column	kg m^{-2}	Mass of water in all phases in a thin surface soil layer
od550aer	Ambient aerosol optical thickness at 550 nm	1	AOD from the ambient aerosols (i.e., includes aerosol water); does not include AOD from stratospheric aerosols if these are prescribed but includes other possible background aerosol types
od550lt1aer	Ambient fine aerosol optical thickness at 550 nm	1	od550 due to particles with wet diameter less than 1 µm ("ambient" means "wetted"); when models do not include explicit size information, it can be assumed that all anthropogenic aerosols and natural secondary aerosols have diameter less than 1 µm
od870aer	Ambient aerosol optical thickness at 870 nm	1	AOD from the ambient aerosols (i.e., includes aerosol water); does not include AOD from stratospheric aerosols if these are prescribed but includes other possible background aerosol types
pr	Precipitation	kg m^{-2} s^{-1}	At surface; includes both liquid and solid phases from all types of clouds (both large-scale and convective)
rlut	TOA outgoing longwave radiation	W m^{-2}	At the top of the atmosphere
rsut	TOA outgoing shortwave radiation	W m^{-2}	At the top of the atmosphere
sic	Sea ice area fraction	%	Fraction of grid cell covered by sea ice
sm	Volumetric soil moisture in upper portion of soil column	m^3 m^{-3}	Volume of water in all phases in a thin surface soil layer
sw_cre	Shortwave cloud radiative effect	W m^{-2}	At the top of the atmosphere
ta	Air temperature	K	
tas	Near-surface air temperature	K	
toz	Total ozone column	DU	Equivalent thickness at standard temperature and pressure (stp) of atmosphere ozone content
tro3	Ozone volume mixing ratio	ppbv	
ts	Surface temperature	K	"skin" temperature (i.e., SST for open ocean)
ua	Eastward wind	m s^{-1}	
va	Northward wind	m s^{-1}	
xco2	column average CO_2 concentration	ppm	
zg	Geopotential height	m	Geopotential height

4.1 Portrait diagram

Figure 1 provides a synoptic overview of the relative quality of the CMIP5 models' representation of simulated climatological mean state and the seasonal cycle for ECVs compared with the multi-model median. The figure shows the relative space-time root-mean-square deviation (RMSD) from the climatological mean seasonal cycle assessing whether a specific model performs better or worse than the other models. The model data have been regridded to the grid of the reference data using a higher order patch recovery interpolation (Khoei and Gharehbaghi, 2007; Hung et al., 2004) and normalized with the centered median (i.e., subtracting the median and then dividing by the median). For the calculation of the RMSD, only grid cells with observational data available for at least 95% of the time period are taken into account.

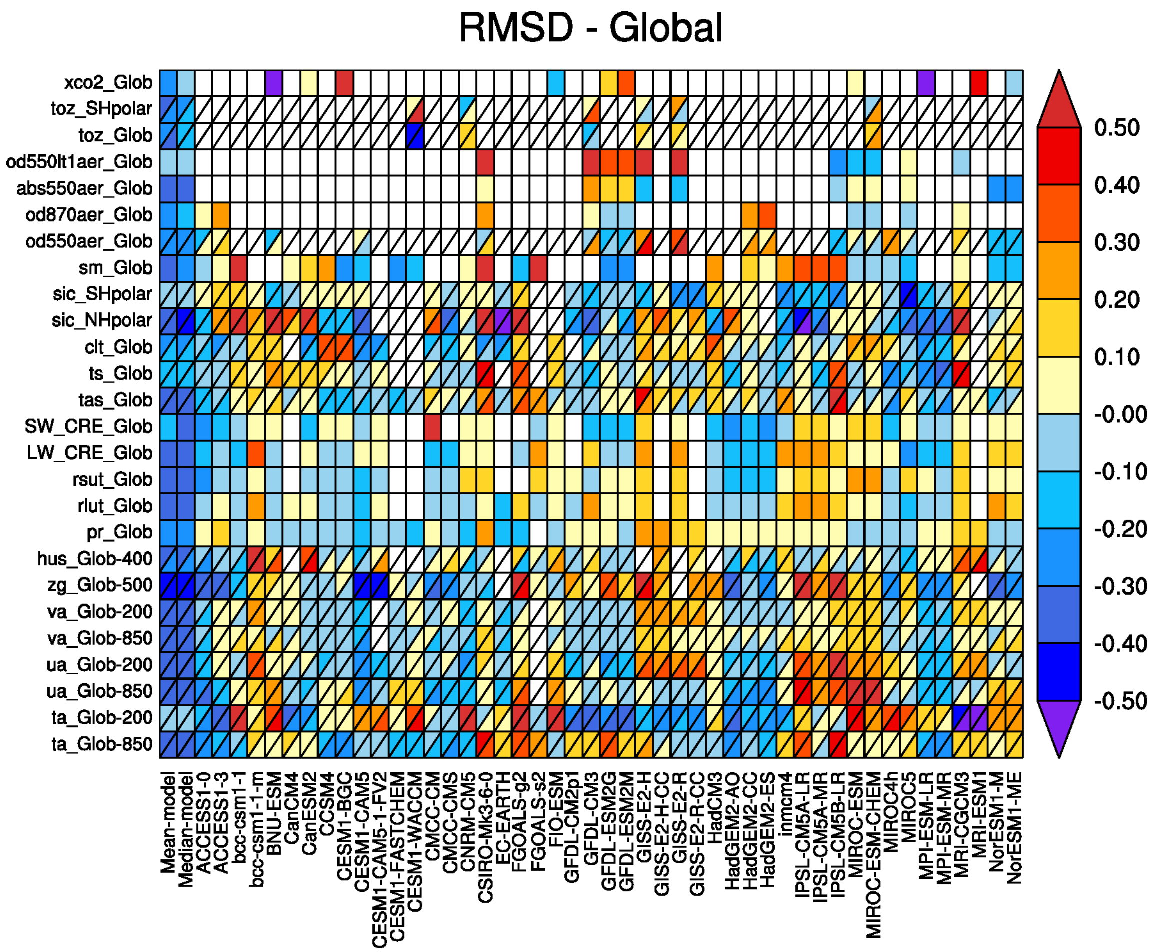


Figure 1. Relative space-time root-mean-square deviation (RMSD) calculated from the climatological seasonal cycle of the CMIP5 simulations. The years averaged depend on the years with observational data available and are summarized in Table 2. A relative performance is displayed, with blue shading indicating better and red shading indicating worse performance than the median of all model results. A diagonal split of a grid square shows the relative error with respect to the reference data set (lower right triangle, data sets marked with asterisks in Table 2) and the alternative data set (upper left triangle). White

boxes are used when data are not available for a given model and variable. The variable names are defined in Table 3.

As such it can be seen as a starting point of model evaluation while the reasons for differences between model and observations need to be further investigated in additional analyses. The figure includes all variables that are shown in figure 9.7 of Flato et al. (2013) and adds variables with ESA CCI data now available. We would like to note that some differences compared to the portrait diagram of Flato et al. (2013) are introduced by using a different set of models, time range and observationally base reference data sets.

As found in previous studies, the performance varies across the models and variables, with some models comparing better with observations for one variable and another model performing better for a different variable. Typically, the multi-model mean outperforms any individual model, which also holds for many of the newly added ECVs. Exceptions to this are, for example, global average temperatures at 200 hPa (ta_Glob-200), sea ice (sic_NHpolar, sic_SHpolar), aerosol optical depth of fine particles at 550 nm (od550lt1aer_Glob), and column average CO_2 (xco2_Glob). In the following we discuss the results only for the variables that are compared to the ESA CCI data sets and refer to Flato et al. (2013) for results on the other variables.

SST: typical biases in the geographical distribution of the simulated SST include a warm bias in the subtropical stratocumulus regions as well as a cold bias in the equatorial Pacific. Individual models performing worse than the multi-model mean (Figure 1) include, for instance, the CSIRO, the FGOALS, and the MRI models. The reasons for this are rather different, for example the CSIRO model shows a cold bias in the subtropical North Pacific whereas the FGOALS model shows a warm bias in the subtropical Southeast Pacific.

Sea ice: for sea ice concentration (sic), the ESA CCI SI SSM/I and the National Snow and Ice Data Center NSIDC-NT (Walsh et al., 2015) observations are used for comparison with the CMIP5 models. Figure 1 shows that the choice of the reference data set does not impact the results for the model performance in reproducing the observed sea ice concentration significantly. This is expected as the two sea ice data sets are in rather good agreement.

Cloud: for total cloud cover (clt), the choice of the reference data set can make some difference for the calculated performance of the individual models. A number of models such as, for instance, the GFDL-CM3 and some of the HadGEM2 models have a larger RMSD when compared against the ESA CCI data set than against the data from Pathfinder Atmospheres Extended (PATMOS-x). The ESA CCI cloud data show slightly higher values (10-15%) for total cloud cover in the subtropical stratocumulus regions off the west coasts of North and South America as well as off the coast of Australia. In contrast, cloud amounts in the ESA CCI data are smaller over the tropical Pacific with frequent deep convection (-10 to -20%). These are also regions in which the models typically struggle to reproduce the observations. The average model bias is therefore larger when the models are compared with the ESA CCI data rather than the PATMOS-x data. An exact quantitative assessment, however, requires application of a satellite simulator in the models to take into account satellite overpass times and lower cut-off thresholds (Bodas-Salcedo et al., 2011), which is beyond the scope of this study. The comparison of total cloud cover done here should therefore only be seen as a starting point for further evaluation of the ESMs.

Soil moisture: the inter-model spread for soil moisture (sm) is large and most models tend to systematically over- or underestimate soil moisture throughout the globe compared with the ESA CCI data. It should be noted, however, that a quantitative comparison is difficult as the layer thickness considered is not consistent among the models and the satellite observations (see discussion in section 5.4). Qualitatively, many models such as the FGOALS, GFDL, HadGEM, and MIROC models overestimate the soil moisture particularly in higher latitudes in Asia, as well as Alaska and the northern part of Canada.

Aerosol: performance metrics for the four aerosol variables od550aer, od870aer, abs550aer, and od550lt1aer are calculated with respect to the ESA CCI data set (see section 2.1) as reference data set (lower triangles) and an alternative data set from MODIS. Shown are only CMIP5 models with interactive aerosols (ACCESS1-0, ACCESS1-3, BNU-ESM, CESM1-CAM5, CSIRO-Mk3-6-0, GFDL-CM3, GFDL-ESM2G, GFDL-ESM2M, GISS-E2-H, GISS-E2-R, HadGEM2-CC, HadGEM2-ES, IPSL-CM5B-LR, MIROC4h, MIROC5, MIROC-ESM, MIROC-ESM-CHEM, MRI-CGCM3, NorESM1-M, NorESM1-ME), models using pre-scribed aerosol climatologies have not been taken into account. Except for od550lt1aer, the multi-model mean outperforms the individual models. Because of differences in the two satellite data sets for AOD, which are largest over the continents (see section 5.1), the choice of the reference data set can make a difference in the resulting model grading with most models performing slightly better against MODIS than the ESA CCI data set. Additional analysis with the ESMValTool (not shown) reveals that even though most models agree on the basic properties of the AOD distribution (od550aer), the relative spread among the models for absorption AOD (abs550aer) and AOD of fine particles ($d < 1$ μm, od550lt1aer) is large. It should be noted that the observational uncertainties for these quantities are also larger than for AOD at 550 nm. For CMIP5, only the latter was evaluated whereas od550lt1aer, abs550aer, and od870aer are shown for the first time here.

Ozone: the performance metric of total column ozone with respect to the ESA CCI (lower triangle) and NIWA (upper triangle) data is shown only for models with interactive chemistry (CESM1-WACCM, CNRM-CM5, GFDL-CM3, GISS-E2-H, GISS-E2-R, MIROC-ESM-CHEM). The performance of the individual CMIP5 models for global total column ozone is quite similar for the two observational data sets. This is not surprising as both reference data sets are based on the same satellite observations from GOME-2 and SCIAMACHY (Bodeker et al., 2005; Loyola et al., 2009). However, in the polar regions (toz_SHpolar) there are significant differences that likely occur because the ESA dataset has gaps in polar winter whereas these are filled in the NIWA data set. Typical biases in CMIP5 models with interactive chemistry include, for instance, an overestimation of total ozone in high northern latitudes (> 60°N) throughout the year and an underestimation of ozone in Antarctica during summer (November to January) (Eyring et al., 2013).

CO_2: only results from emission driven simulations are included in the performance metric shown for XCO_2 in Figure 1. The BNU-ESM and the MPI-ESM-LR models outperform the multi-model mean, which is biased high compared with the ESA CCI data as most models systematically overestimate the column average CO_2 concentrations. This overestimation could be possibly caused by slightly too weak CO_2 sinks in some models (Friedlingstein et al., 2014).

For most variables, the choice of reference data set does not make a big difference when using global averages for comparison with the CMIP5 models. This is, however, not necessarily the case when looking into more details such as individual regions or seasons. More on the comparison of the ESA CCI data with alternative observationally based data sets are given in the individual subsection of section 5.

4.2 Taylor diagrams

Another widely used way to summarize comparisons of results from a number of different models with observations are Taylor diagrams (Taylor, 2001). The Taylor diagrams shown in Figure 2 give the standard deviation and linear pattern correlation with observations of the total spatial variability calculated from multi-year annual means, so in contrast to the space-time RMSD evaluated in section 4.1, here only the geographical pattern is evaluated. For the calculation of the Taylor diagrams, all data have been regridded to a regular 1°x1° latitude-longitude grid using a patch recovery interpolation method. For each variable, a common masking to exclude missing values has been applied to all data sets.

The standard deviations are normalized by the observed standard deviations, so the observed climatology is represented in each panel by the filled black dots on the x-axis at x = 1. The pattern correlation is given in this polar projection by the angular coordinate. The linear distance between the observations and each model is proportional to the root-mean-square error (RMSE) and can be estimated in multiples of the observed standard deviation with the gray circles centered on the observational dots. The multi-model mean values have been calculated over all models with data available (black star). Where available, an alternative reference data set (see Table 2) is also shown in Figure 2 (red star).

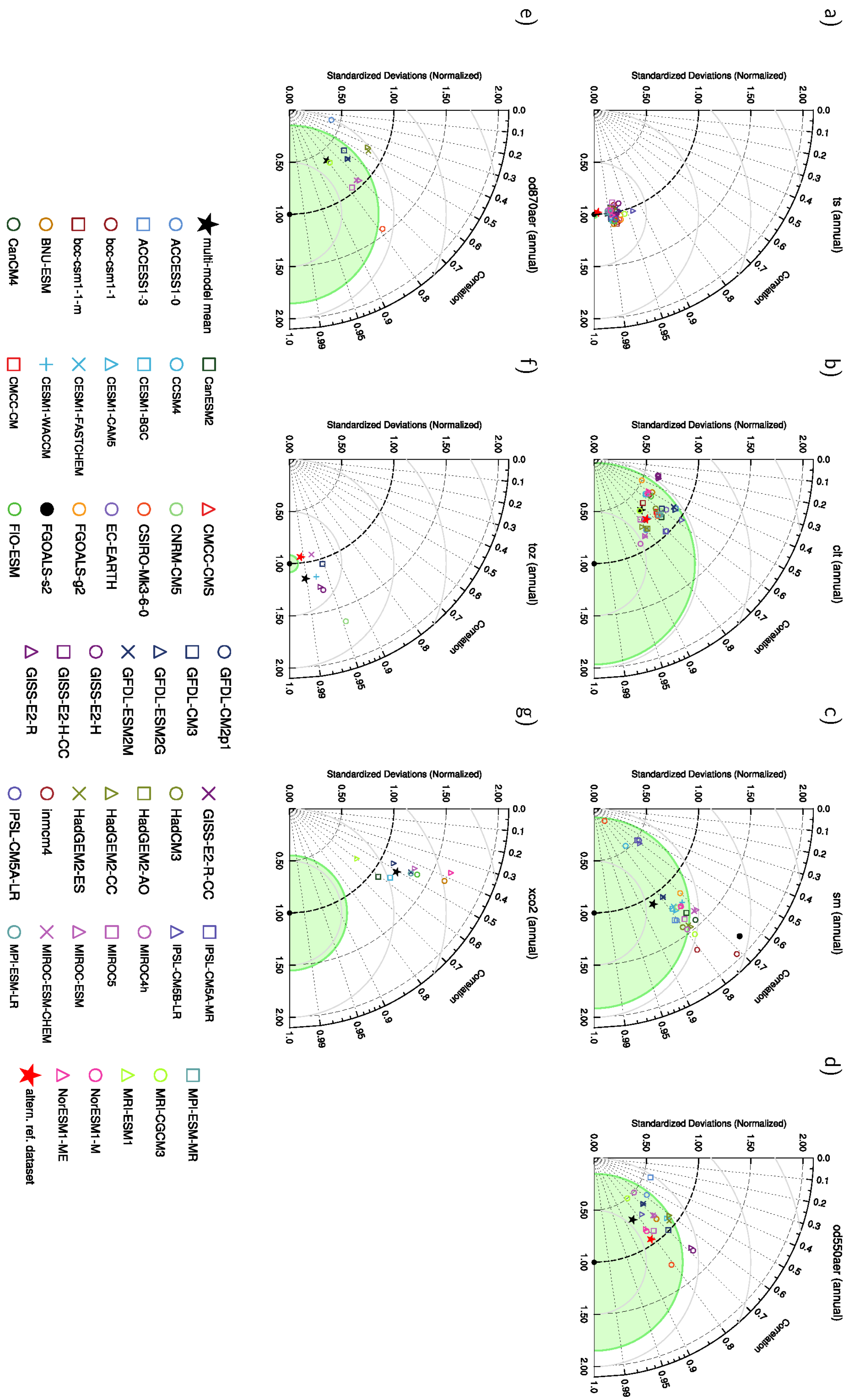

a)
b)
c)
d)
e)
f)
g)
ts (annual)
clt (annual)
sm (annual)
od550aer (annual)
od870aer (annual)
toz (annual)
xco2 (annual)
Standardized Deviations (Normalized)
Correlation
multi-model mean
ACCESS1-0
ACCESS1-3
bcc-csm1-1
bcc-csm1-1-m
BNU-ESM
CanCM4
CanESM2
CCSM4
CESM1-BGC
CESM1-CAM5
CESM1-FASTCHEM
CESM1-WACCM
CMCC-CM
CMCC-CMS
CNRM-CM5
CSIRO-Mk3-6-0
EC-EARTH
FGOALS-g2
FGOALS-s2
FIO-ESM
GFDL-CM2p1
GFDL-CM3
GFDL-ESM2G
GFDL-ESM2M
GISS-E2-H
GISS-E2-H-CC
GISS-E2-R
GISS-E2-R-CC
HadCM3
HadGEM2-AO
HadGEM2-CC
HadGEM2-ES
inmcm4
IPSL-CM5A-LR
IPSL-CM5A-MR
IPSL-CM5B-LR
MIROC4h
MIROC5
MIROC-ESM
MIROC-ESM-CHEM
MPI-ESM-LR
MPI-ESM-MR
MRI-CGCM3
MRI-ESM1
NorESM1-M
NorESM1-ME
altern. ref. dataset

Figure 2. Extended Taylor diagrams showing the multi-year annual average performance of the CMIP5 models in comparison with ESA CCI data for a) SST, b) total cloud cover, c) soil moisture, d) AOD at 550 nm, e) AOD at 870 nm, f) total column ozone, and g) column averaged CO_2 concentration. Panels a) to e) show CMIP5 historical simulations (extended with RCP4.5), panel f) historical simulations (extended with RCP4.5) with interactive ozone chemistry, and panel g) emission driven historical simulations (extended with RCP8.5). The multi-model mean values have been calculated over all models with data available (black stars). Where available alternative observationally based data sets are also shown (red stars, Table 2). The green circles show estimates of the observational uncertainties (RMSE, for details see section 4.2).

In this study, a new extended version of Taylor diagrams is presented that visualizes observational uncertainty: the green circles show estimates of the observational uncertainties (RMSE) that are part of the ESA CCI data sets. Here, the multi-year global average uncertainties given as one sigma of the total standard error normalized by the standard deviation of the observations are shown. The RMSE of a given model compared with the observations is therefore smaller than the 1-sigma uncertainty estimate of the observations if the model lies within the green circle.

SST (Figure 2a): the geographical annual mean patterns of the sea surface temperatures from the models are highly correlated with the ESA CCI data with correlation coefficients ranging between 0.94 and 0.98. However, SST in the subtropical stratocumulus regions as well in the Southern Ocean is overestimated by many models. Another typical model bias found in many simulations is an underestimation of the SST in the equatorial Pacific.

Cloud (Figure 2b): for total cloud cover, the models show a large spread in pattern correlation between 0.25 and 0.88. Most models are, however, not outside of the 1-sigma uncertainty estimate showing that the differences between the models and the observations cannot be solely explained by model deficiencies.

Soil moisture (Figure 2c): can mostly be used for qualitative assessments of the models as the observational uncertainties are larger than the RMSE of many of the individual models.

Aerosol (Figure 2d,e): the integrated aerosol properties AOD at 550 (Figure 2d) and 870 nm (Figure 2e) also show a large inter-model spread. Because of the large observational uncertainties, most models lie within the green circle of the 1-sigma measurement uncertainty making further quantitative assessments difficult. This is also supported by the differences between the ESA CCI data set and the MODIS data for AOD with MODIS being close to 1-sigma of the ESA CCI uncertainty estimate. The linear pattern correlation of most models with the ESA CCI data, however, is smaller than that of the ESA CCI data and MODIS (0.8) showing also differences in the geographical distribution of the simulated AOD (see also section 5.1 and Figure 14).

Ozone (Figure 2f): the correlation coefficients of the modeled total ozone columns with the ESA CCI data are quite high for most models (with interactive chemistry) with values above 0.94 and a ratio of the modeled and the observed spatial standard deviation close to 1. All models are, however, outside of the 1-sigma uncertainty estimate of the observations, which is also the case for the alternative observational data set (NIWA). Differences are found, for instance, in the northern high latitudes where the models tend to overestimate the total ozone columns (see also section 5.7 and Figure 18).

CO_2 (Figure 2g): For the column-averaged CO_2 concentrations, the correlation coefficients of the results from the emission driven simulations with the ESA CCI data are typically quite low and range between 0.4 and 0.6. This is partly caused by a systematical overestimation of XCO_2 concentrations by most CMIP5 models and partly by differences in the geographical patterns such as, for example, in northern Europe or Southeast Asia where the models show distinct local maxima that are not clearly visible in the ESA CCI data.

5 Further insights into the evaluation of CMIP5 models with ESA CCI data

In the following subsections, the evaluation of CMIP5 models using ESA CCI data and comparisons of ESA CCI data with alternative observational data sets are discussed individually for each of the CCI products (sea surface temperature, sea ice, cloud, soil moisture, land cover, aerosol, ozone, and greenhouse gases).

5.1 Sea surface temperature

The implemented diagnostics for sea surface temperature in the ESMValTool include the analysis of the temporal mean fields, their differences as well as a long-term trend analysis and calculation of scalar accuracy skill scores such as, for instance, area weighted RMSDs. All diagnostics can be applied to regional areas of interest defined by the user, e.g., ocean basins.

A major challenge when comparing the ESA CCI SST data to CMIP model results is that the ocean grids used in the various CMIP models differ substantially. Thus, a common target grid needs to be defined for the models and SST observations first. The user can specify the target resolution and target projection in the ESMValTool configuration. For the examples given in Figure 3 and Figure 4, we use a T63 Gaussian grid as a common reference and project all SST data to this grid using an energy conservative approach. In addition, the representativeness of the SST variables largely varies among different models. While the CMIP sea surface temperature variable (tos) corresponds to the temperature in a layer a few centimeters deep in some models, it represents the temperature of a layer of a couple of meters in other models. The ESA CCI SST product used in this study is designed to be representative of the sea surface at a depth of 20 cm. Except under conditions of very low wind stress and strong insolation, the stratification across the upper ~1 m of the ocean tends to be small because of near-surface mixing driven by wind and wave action. Nonetheless, the differing depth definitions need to be considered when interpreting SST differences between different models and between models and observations, particularly for the subset of the comparisons corresponding to situations of likely near surface stratification.

Figure 3 shows an example of a comparison of results from the CMIP5 model MPI-ESM-P with the ESA CCI SST data set. On a global scale, the observed geographical patterns (top) with high temperatures in the equatorial areas and low temperatures close to the poles are well reproduced by the model and so is the global mean SST value of 287 K and its spatial variability. Both the observations and the model show the typical two-armed warm areas in the Niño 3 and Niño 4 areas in the equatorial Pacific. They also show the typical shift of warm water in the eastern northern Atlantic generated by the Gulf Stream, and the colder regions in the Arabian Sea. MPI-ESM-P shows a negative bias in the subtropics and tropics while a positive bias is found in the cold climate zones in both hemispheres (Figure 3, bottom row). A switch in these differences occurs in the temperate zones. The difference plot also shows discrepancies in specific areas such as the underestimation of SST in the central northern Atlantic, from too-zonal behavior of the North Atlantic Drift in the model, or the pattern of overestimation of temperature in ocean upwelling zones on the east of ocean basins (along the Namibian coast, Baja California, etc.). These discrepancies suggest differences in the representation of the wind driven upwelling and western boundary currents.

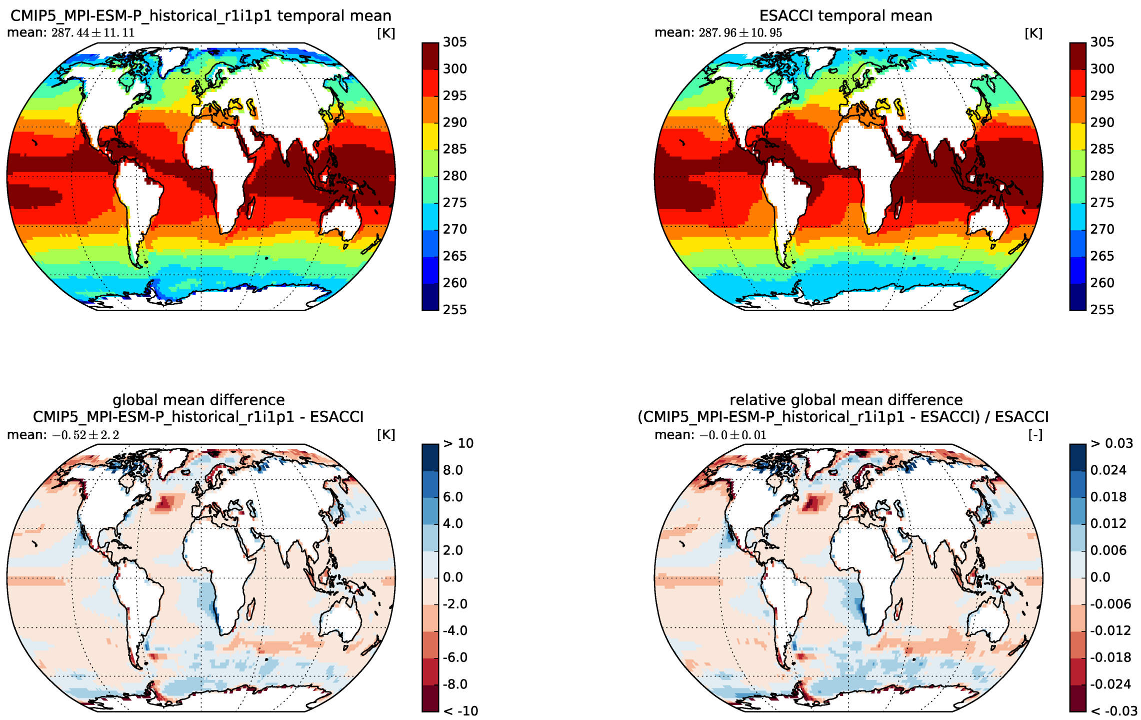


Figure 3. Temporal means of SST in K for the ESA CCI data set (top right) and the CMIP5 model MPI-ESM-P (top left) as well as absolute (bottom left) and relative differences (bottom right).

Discrepancies between 7 exemplary CMIP5 models (GISS-E2-H-CC, GISS-E2-H, IPSL-CM5A-LR, MIROC-ESM-CHEM, MIROC-ESM, MPI-ESM-P, NorESM1-ME) for different ocean basins are shown in Figure 4. Larger basins, like the northern or southern Pacific or Atlantic Ocean, as well as the polar seas show good agreement in SST cycles among the different models and with the ESA CCI data. Differences are larger for smaller basins like the Baltic or Mediterranean Seas or the Niño regions. These larger discrepancies occur due to the size of these smaller regions and their higher sensitivity to small scale fluctuations. Spatial averages of larger basins attenuate such fluctuations. The ESA CCI SST data are sufficient to show the model limitations on such scales, and discriminate, for example, the better seasonal cycle amplitude for the Baltic Sea in MIROC-ESM-CHEM compared to MPI-ESM-P.

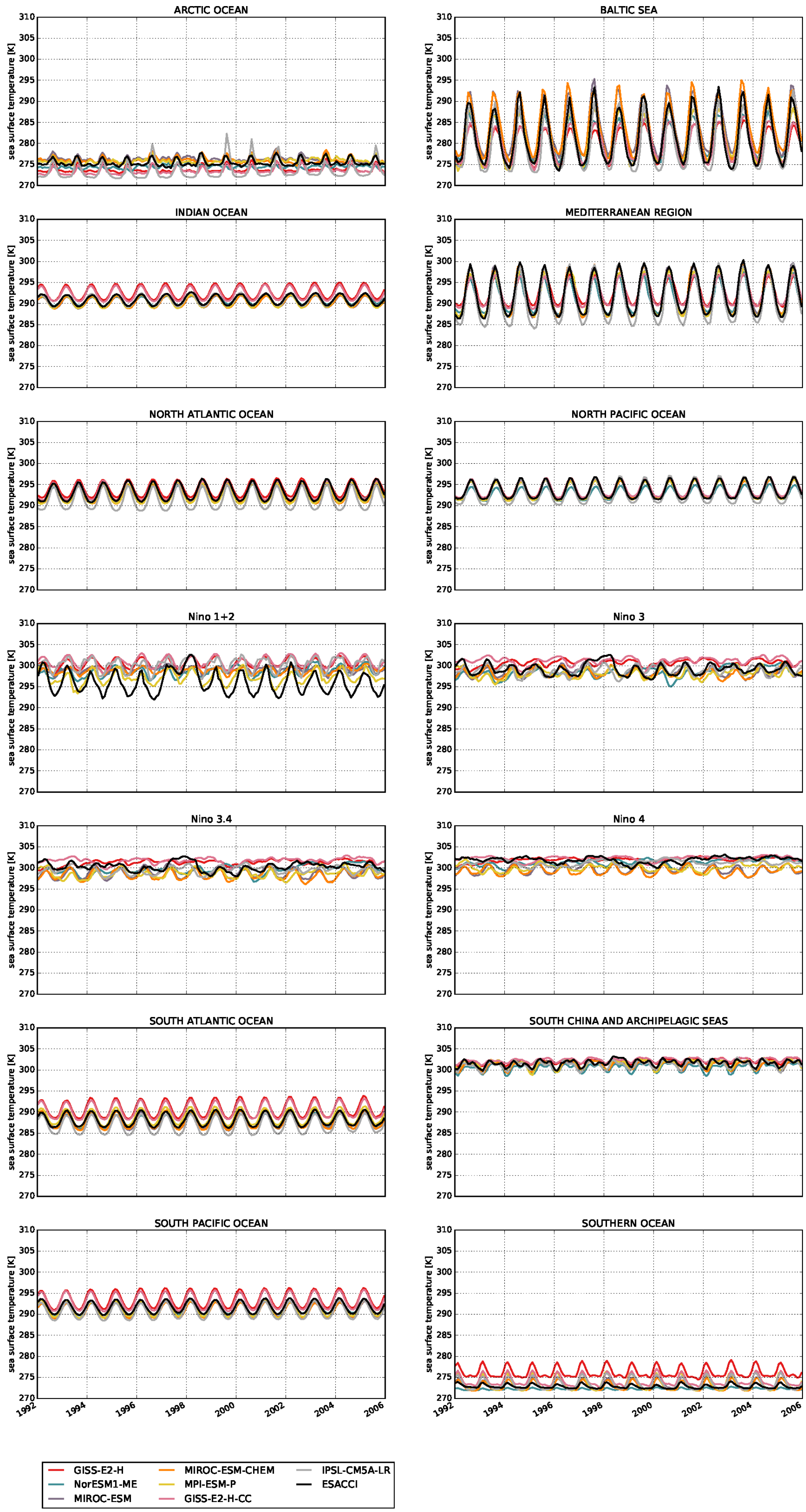

ARCTIC OCEAN
BALTIC SEA
INDIAN OCEAN
MEDITERRANEAN REGION
NORTH ATLANTIC OCEAN
NORTH PACIFIC OCEAN
Nino 1+2
Nino 3
Nino 3.4
Nino 4
SOUTH ATLANTIC OCEAN
SOUTH CHINA AND ARCHIPELAGIC SEAS
SOUTH PACIFIC OCEAN
SOUTHERN OCEAN
sea surface temperature [K]
310
305
300
295
290
285
280
275
270
1992
1994
1996
1998
2000
2002
2004
2006
GISS-E2-H
NorESM1-ME
MIROC-ESM
MIROC-ESM-CHEM
MPI-ESM-P
GISS-E2-H-CC
IPSL-CM5A-LR
ESACCI

Figure 4. Time series of mean SST for different ocean basins from 7 CMIP5 models (see legend) compared with the ESA CCI SST data.

ESA CCI SSTs are relatively unusual in being physics-based (not tuned to drifting buoys) and explicitly aiming to represent the 20-cm depth SST, which should correspond well to model-layer-average SSTs in most circumstances. The new data set therefore provides an independent, accurate (0.1 K), high-stability climate data record.

5.2 Sea ice

The observed rapid decline in Arctic sea ice thickness and extent over the last few decades is one of the most striking indicators for climate change (Stroeve et al., 2012; Lindsay and Schweiger, 2015). The melting of sea ice contributes to the rise of global temperatures through the ice-albedo feedback (Curry et al., 1995). The decline in sea ice extent is a positive feedback where the initial shrinkage in the area of sea ice reduces the albedo and thus reinforces the initial alteration in sea ice area. High-quality observations of sea ice are thus crucial to monitor climate change and to evaluate climate models.

Here, we use data from the National Snow and Ice Data Center (Walsh et al., 2015) as an additional reference data set for the model evaluation and for comparison with the ESA CCI sea ice data. The NSIDC provides two different data sets, each covering the time period from 1979 to present. The main difference between the two data sets is the algorithm used in processing the satellite data: the NASA Team (NSIDC-NT; Cavalieri et al., 1996) and the Bootstrap (NSIDC-BC; Comiso, 2000) algorithm. While the NSIDC-BT algorithm corrects for melt ponds that are treated as open water by synthetically increasing the summer sea ice concentration (sic), such a correction is not included in NSIDC-NT.

The sea ice diagnostics implemented into the ESMValTool include time series of the modeled and observed evolution of sea ice extent (Figure 5) or area as well as polar-stereographic contour plots of sic and sic biases (Figure 6). The sea ice extent has been calculated by adding up the surface area of all grid cells with a sea ice concentration equal or larger than 15%. Satellites in polar orbits do not pass directly over the poles. As a consequence, there is a small area centered around the poles that cannot be observed by these instruments. For the comparison with the model data shown in Figure 5, these pole holes have been filled assuming 100% sea ice cover in this region.

The time series of September Arctic sea ice extent in Figure 5 shows that the spread between the four observational data sets (thick black lines) from ESA CCI and NSIDC is much smaller than the spread among the CMIP5 models (colored lines), which amounts to about 9 million km^2 between CSIRO-Mk3-6-0 (largest positive bias) and GISS-E2-H (largest negative bias). However, the CMIP5 multi-model mean (thick red line) lies most of the time within the observational spread although the RCP4.5 simulation mean does not show the decrease in sea ice extent that has been observed between 2005 and 2013. The sea ice extent from the ESA CCI data sets (thick black lines) is in very good agreement with the NSIDC data sets. ESA CCI SSM/I data show a small positive bias compared with NSIDC-NT of up to 1 million km^2 between 1997 and 2005. ESA CCI AMSR-E data are in very good agreement with both NSIDC data sets. The negative trend over the observed time period from 1990 to 2010 is about 1 million km^2 per decade in all four observational data sets. The magnitude of this trend is, however, underestimated by the CMIP5 multi-model mean.

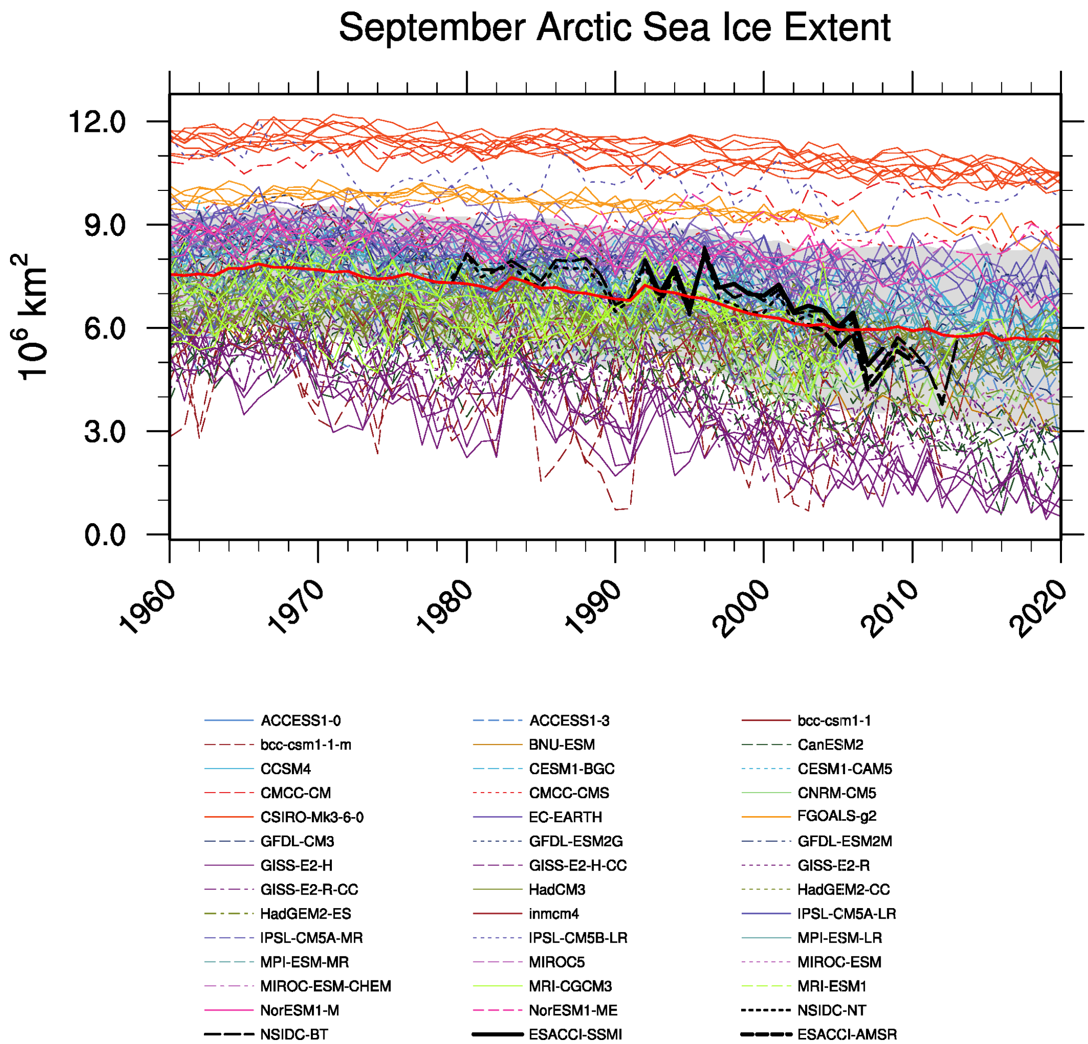


Figure 5. Evolution (1960-2020) of September Arctic sea ice extent in million km^2 from the CMIP5 models (colored lines) and from observations (thick black lines). The pole holes of the satellite data sets have been filled assuming a sea ice concentration of 100%. All available ensemble members from a given model are shown and drawn in the same color as indicated in the legend. The CMIP5 multi-model mean is shown in bold red and the gray shading shows the standard deviation of the CMIP5 ensemble. The observations are from ESA CCI SI and NSIDC. Figure modified from Bräu (2013).

Figure 6 shows polar-stereographic contour maps of Arctic September (upper row) and Antarctic March (lower row) sic, which roughly corresponds to the average annual minimum sea ice extent. As in Figure 5, there is good agreement between the ESA CCI SI SSM/I (left column) and NSIDC-NT (middle column) also in the geographical distribution. The sic from the two data sets differs by less than 0.2 in all grid cells for both Arctic and Antarctic sea ice distributions (not shown). In the Arctic, the CMIP5 multi-model mean slightly underestimates the observed sic in the marginal ice zone of the Central Arctic Ocean and in the East-Siberian and Beaufort Seas by about 0.2 (right column). There is also a small overestimation east of Svalbard. In the Antarctic, the sea ice concentration is underestimated by the CMIP5 multi-model mean in the Weddell Sea as well as in a belt along the coast of the Amundsen, Ross and Somov Seas by up to 0.6.

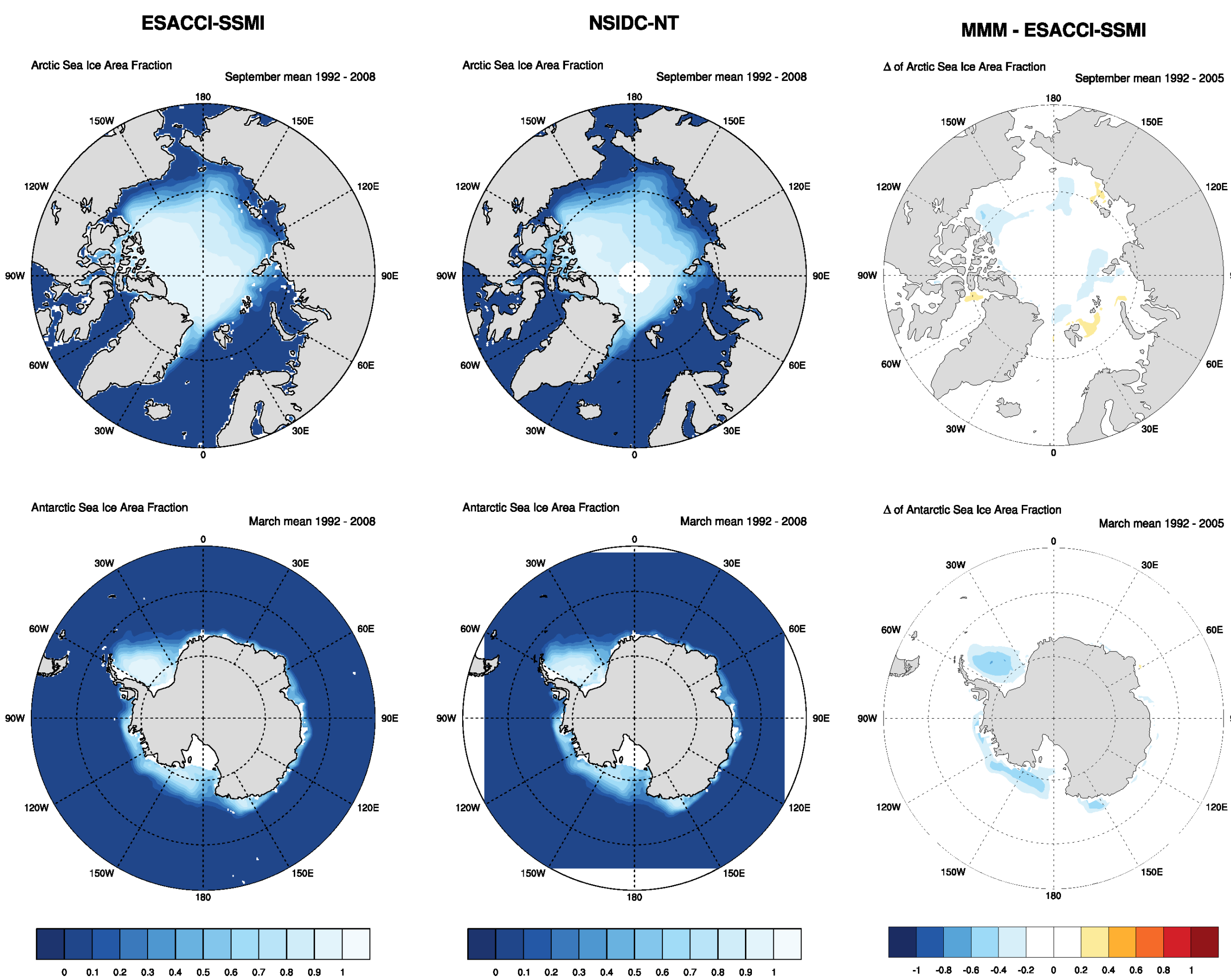


Figure 6. Polar-stereographic map of Arctic September (upper row) and Antarctic March (lower row) sea ice concentration from ESA CCI SI SSM/I (left column) and NSIDC-NT (middle column) observations

averaged over the years 1992-2008. The pole holes of the satellite data sets have been filled assuming a sea ice concentration of 100%. The right column depicts the differences between the CMIP5 multi-model mean and the ESA CCI SI SSM/I observations averaged over the years 1992-2005.

In general, the ESA CCI data show good agreement to the data sets from the NSIDC. For robust assessments of retrospective climate simulations, however, a longer time period is needed and would ideally go from the early 1980s to present. Since the sea ice observational data are no exception in having errors that are inherent to all observations, the daily uncertainty estimates provided by the ESA CCI sea ice team are very useful for a more quantitative model evaluation. These error estimates are based on the extensive algorithm comparison study (Ivanova et al., 2015) and have been underpinned by subsequent validation studies (Kern et al., 2016). The error estimates will be useful for further regional and seasonal assessments of sea ice concentrations.

5.3 Cloud

Clouds strongly affect the Earth's radiative balance and temperature but are challenging to model and observe, leading to large uncertainties in understanding climate variability and change. Model evaluation using long-term, consistent observational data records can help to improve both, the understanding of the present-day climate and the confidence in climate model projections. Modeled clouds and satellite observations are difficult to compare because observed clouds are affected by the satellite instrument's sensitivity, the temporal and spatial sampling and the vertical overlap of the cloud layers, while the clouds in climate models are assumed to be plane-parallel and are of coarse horizontal and vertical resolution. Ideally, a satellite simulator (e.g., Bodas-Salcedo et al., 2011) is used during the model simulation to mimic the satellite viewing geometry, temporal sampling and specific instrument characteristics such as lower cut-off values. Many CMIP5 historical and future scenario simulations, however, have been run without such a satellite simulator. Total cloud cover is the model cloud parameter that most readily can be compared directly to the satellite derived cloud fraction without a simulator, even though models can have substantial cloud cover but very little cloud condensate making those clouds too optically thin to be detected by the satellite instrument.

Here we use the ESMValTool diagnostics mean, bias and interannual variability to compare Cloud_cci AVHRR-PM total cloud cover with other satellite-based cloud data sets and to evaluate CMIP5 models. Figure 7 shows the ESA CCI total cloud cover (clt) in boreal winter (December, January, February) and summer (June, July, August) and the associated total uncertainties derived from comparisons with CALIOP as described in section 2.3. The inherent AVHRR difficulties in detecting clouds during polar night and over high elevation, snow covered areas (North Canada, North East Asia and Himalayas) result in uncertainties of more than 20% in these regions. Comparing the ESA CCI zonal mean cloudiness to other AVHRR cloud data sets such as PATMOS-x (Heidinger et al., 2014) and CLARA-A2 (Karlsson et al., 2013) and the MODIS cloud data set (Platnick et al., 2003) also show the largest observational spread (40-50%) in high latitudes in the winter hemisphere. The ESA CCI uncertainties are also high with values of up to 20% in the subtropical high pressure dry areas. In these regions, the ESA CCI data set has 5-10% less cloud coverage than PATMOS-x and CLARA-A2 (not shown).

The performance metrics results in Figure 1 show that the cloudiness of most CMIP5 models compare well with the ESA CCI data and the alternative reference data set PATMOS-x on a global scale, but there are regional differences as seen in Figure 7. The CMIP5 multi-model mean bias compared to the ESA CCI data shows an underestimation of cloud amount especially in the subtropical stratocumulus regions off the west coasts of North and South America as well as off the coast of Australia as known from many previous studies (e.g., Nam et al., 2012). In contrast, the CMIP5 multi-model mean and most individual models overestimate cloud amounts by 20% over the subtropical high pressure regions with minimum cloud amounts. These biases are smaller (10-15%) if the models are compared to PATMOS-x and CLARA-A2 instead because cloud amounts from these two alternative observational data sets are larger than from the ESA CCI data set in these regions. The CMIP5 models with a normalized RMSD above 0.2 in Figure 1 (CCSM4, CESM1-BGC, HadCM3, MIROC-ESM and MIROC-ESM-CHEM) underestimate cloud amount on a global scale (not shown). The largest inter-model spread (60%) occurs at high latitudes in polar winter, where also the observational data sets have their largest uncertainties as seen in the zonal mean plots in Figure 7. In these cold conditions the amount of cloud condensate is small and the modeled clouds are often thinner than the satellites' detection limit. Here, using a simulator removes part of these model clouds.

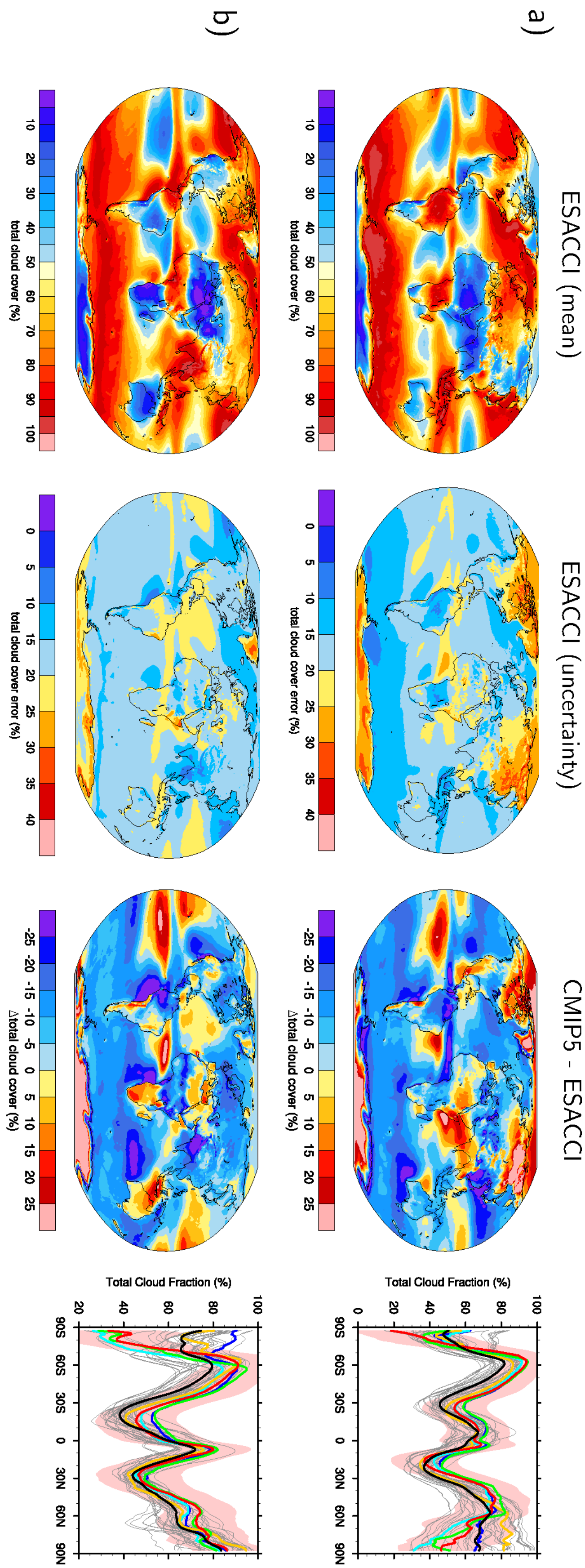

a)
b)
ESACCI (mean)
ESACCI (uncertainty)
CMIP5 - ESACCI
total cloud cover (%)
total cloud cover error (%)
Δtotal cloud cover (%)
Total Cloud Fraction (%)

Figure 7. Maps of the multi-year seasonal mean of total cloud cover and 1-sigma uncertainty from ESA CCI cloud for a) December-January-February (DJF) and b) June-July-August (JJA) averaged over the years 1982-2014. The figure also shows the differences between the ESA CCI data and the CMIP5 multi-model mean as well as zonal means. The zonal mean panels show averages from ESA CCI (red), PATMOS-x (blue), CLARA-A2 (cyan), MODIS (green), ERA-Interim (orange), and the CMIP5 multi-model mean (black). The individual CMIP5 models are shown as thin gray lines and the observational uncertainties of the ESA CCI data (±1-sigma) are shaded in light red. The MODIS data are only available for the years 2003-2014.

Figure 8 shows the interannual variability of total cloud cover for the satellite data sets, the CMIP5 multi-model mean and ERA-Interim. The interannual variability is estimated as the relative temporal standard deviation of the deseasonalized monthly mean time series (Lauer and Hamilton, 2013). All the AVHRR data sets (ESA CCI, CLARA-A2, PAMOS-x) have their largest variability (30-40%) for the dry tropical high pressure regions over the oceans, over north Africa, south Africa and Australia, reflecting the annual shift of the ITCZ and the El Niño/Southern Oscillation (ENSO). MODIS tropical Pacific Ocean variability is smaller than in the AVHRR data sets, since MODIS data are available only for the time period 2003-2014 and thus do not include the strong El Niño events in the 1980s and 1990s, which illustrates the importance of using long-term observational records when evaluating ENSO. The ESA CCI data have larger variability over the tropical Pacific Ocean than the other AVHRR satellite data sets. Time series (not shown) reveal that the ESA CCI clt is of similar magnitude as PATMOS-x and CLARA-A2 for El Niño years when the cloud cover is maximum, while the ESA CCI data have less cloud amount (5-15%) for La Niña years when the cloud cover reaches minima. This results in a larger interannual variability of the ESA CCI data. PATMOS-x data show less variability and higher cloud amounts over the Antarctic than the ESA CCI and CLARA-A2 data. The CMIP5 multi-model mean shows less variability than the observations, especially over the subtropical high pressure regions, where most of the individual CMIP5 models overestimate the total cloud cover. In contrast, the models that underestimate clt in the dry regions (CCSM4, CESM1-BGC, HadCM3, MIROC-ESM, MIROC-ESM-CHEM) show a larger interannual variability.

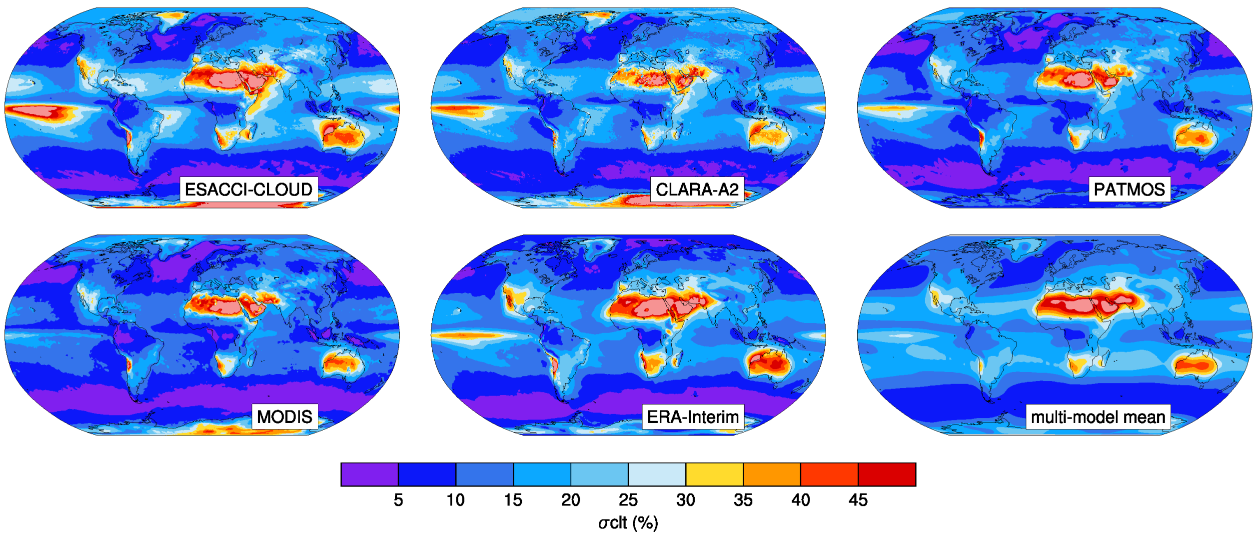


Figure 8. Interannual variability in total cloud cover estimated from the relative temporal standard deviation of the deseasonalized monthly mean time series from 1982 to 2014. Shown are (from top left to bottom right) satellite data (ESA CCI cloud, CLARA-A2, PATMOS-x, MODIS) in comparison with ERA-Interim reanalysis data (lower row, center) and the CMIP5 multi-model mean (lower row, right). The MODIS data are only available for the years 2003-2014.

The Cloud_cci AVHRR-PM total cloud cover data compare well with other existing long-term AVHRR cloud data sets. The ESA CCI pixel-based uncertainties show the user which areas should be interpreted carefully, e.g. polar and high elevation snow covered regions where the passive satellites have problems detecting clouds. The ESA CCI cloud cover data show lower minima than the other AVHRR data sets for the tropical Pacific, which should be investigated further. The other ESA CCI cloud data sets with shorter time records (MODIS, ATSR-2, AATSR and MERIS) can be used for process studies and for narrowing the observational uncertainties. A Cloud_cci satellite simulator has been developed, which can be used in future CMIP simulations and include other cloud variables such as cloud top pressure, optical thickness, effective radius, albedo and liquid/ice water path in the model evaluation. Cloud cover from the CMIP5 models shows the known typical error patterns compared with the ESA CCI data and the other satellite data sets, underestimating clouds in the stratocumulus regions and overestimating clouds in the subtropical dry regions. More detailed analysis of the individual models and the interaction with radiation are needed to understand these biases.

5.4 Soil moisture

The current soil moisture diagnostics implemented in the ESMValTool comprise metrics for the evaluation of soil moisture from regional to global scale and are largely based on Loew et al. (2013) using version 2.2 of the ESA CCI soil moisture data set. These include the comparison of temporal mean fields of soil moisture, as well as the analysis of the co-variability of soil moisture anomalies with precipitation anomalies and the similarity of the spatial patterns of the percentile distributions of the model and observations. The latter is a measure for the similarity of the spatio-temporal dynamics of the soil moisture field (see Loew et al., 2013 for details). Another diagnostic analyzes the long-term trend in soil moisture for both the ESA CCI data set and CMIP models. The non-parametric Mann-Kendall regression is used to assess the statistical significance of long-term soil moisture trends, similar to Dorigo et al. (2012) for the time period 1988-2008. This time period was chosen because the ESA CCI soil moisture data have a poorer temporal sampling prior to this period (Loew et al., 2013). All diagnostics can be applied at the global scale as well as for user-defined regions.

A general challenge when comparing satellite soil moisture with model results is that the observations represent a rather different quantity than the one simulated by the models. CMIP models provide the soil moisture as storage terms for soil layers at specific depths. As the different CMIP models are based on different soil model implementations, these are not necessarily directly comparable as they might differ in their depth and therefore in their temporal dynamics. Currently, the official CMIP5 output comprises two soil moisture variables, which are supposed to represent a 10-cm surface layer (mrsos) or the entire soil column (mrso). Here, we use only data from models that provided the surface layer soil moisture for comparison. The surface layer soil moisture is converted into the volumetric soil moisture content by dividing mrsos by the thickness of the represented layer and by the density of water, which is assumed to be 998.2 kg m^{-3} (20°C). The variable for volumetric soil moisture content compared with the ESA CCI data is called sm (see Table 3).

Satellite soil moisture data typically represent the volumetric soil moisture content (m^3 m^{-3}) of a shallow surface layer, which is also the case for the ESA CCI data set. The soil moisture diagnostics implemented in the ESMValTool compare the volumetric soil moisture content calculated from the model output with observations. All data are aggregated to similar temporal and spatial scales before further analysis.

Figure 9 shows an example of the ESA CCI volumetric soil moisture data compared with soil moisture from the CNRM-CM5 model. The model shows comparable soil moisture patterns in large parts of the globe. A wet bias is observed in the northern latitudes, which might be related to an overestimation of soil moisture due to missing processes in the model (e.g., freeze-thaw dynamics). The model bias can also be related to a dry bias in the ESA CCI observations in these regions as no soil moisture is observed during wintertime and under frozen soil conditions. Relative differences are largest in the desert regions (Sahara, Arabian Peninsula), which is, however, of minor importance due to the overall small absolute soil moisture content in these regions. The wet region along the southern border of the Himalayas is clearly visible in the model but not in the ESA CCI soil moisture. This is most likely due to the complexity in mountainous terrain with large terrain slopes (for both the models and in the satellite soil moisture retrieval algorithms).

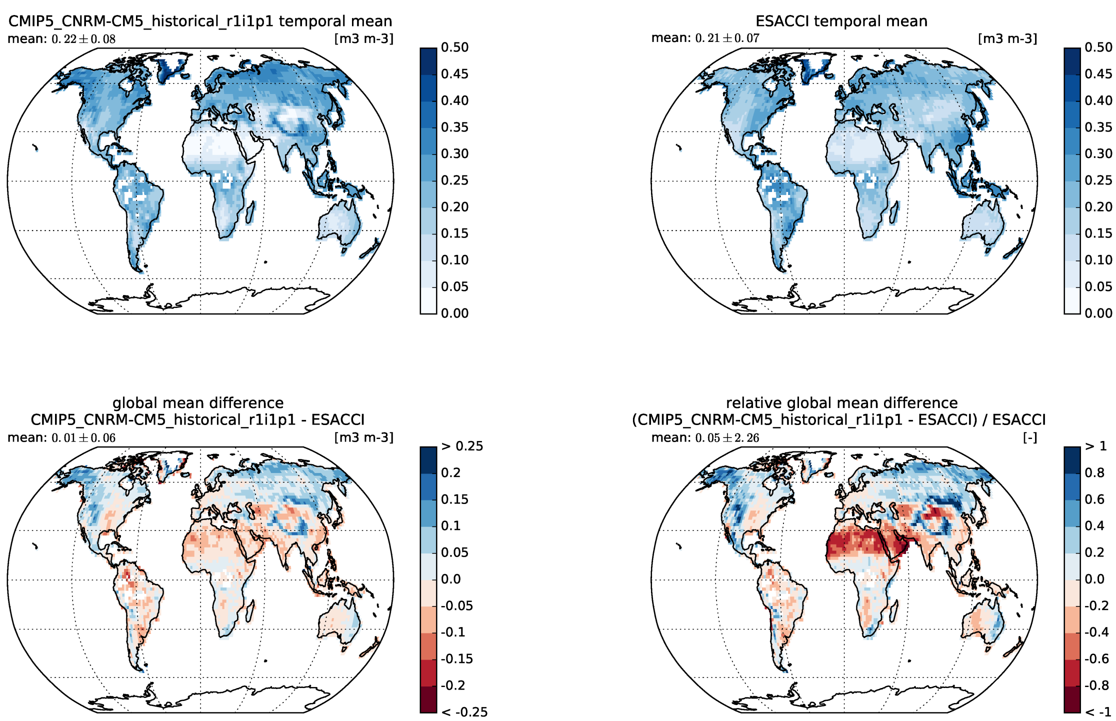


Figure 9. Temporal mean fields of volumetric soil moisture from the CNRM-CM5 model (top left), the ESA CCI soil moisture data set (top right) as well as their absolute (bottom left) and relative differences (bottom right).

The long-term trends in soil moisture during the time period 1988-2008 are compared in Figure 10. The figure illustrates only statistically significant trends ($p < 0.05$). The ESA CCI soil moisture shows decreasing soil moisture in large parts of the globe. Strongest decline of soil moisture is observed in southern Russia, while positive trends are observed in the tropical parts of Africa. Trends in the CNRM-CM5 model are rather different to those obtained from the CCI data set. A significant decline in soil moisture is observed over Europe, while a significant increase of soil moisture is simulated throughout large parts of the northern hemisphere.

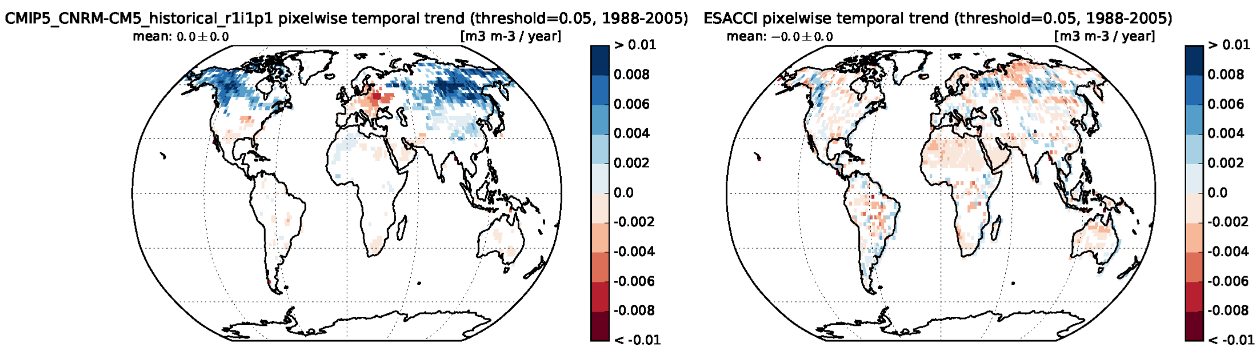


Figure 10. Temporal trend in soil moisture over the period 1988-2008 as derived from the CNRM-CM5 model (left) and the ESA CCI soil moisture data sets (right). Masked areas represent grid cells where the Mann-Kendall correlation coefficient was not statistically significant at the 95% confidence level.

The percentiles of the observed and simulated soil moisture fields are rather similar, which illustrates that both data sets show similar spatial patterns of the soil moisture dynamics. As an example, Figure 11 shows the percentile maps for the 5%, 50% and 95% percentiles for the observed and simulated (CNRM-CM5) soil moisture fields. For each of the percentiles, the spatial autocorrelation coefficient results in very high correlation values ($\rho > 0.9$), which indicates a strong similarity of the spatial patterns.

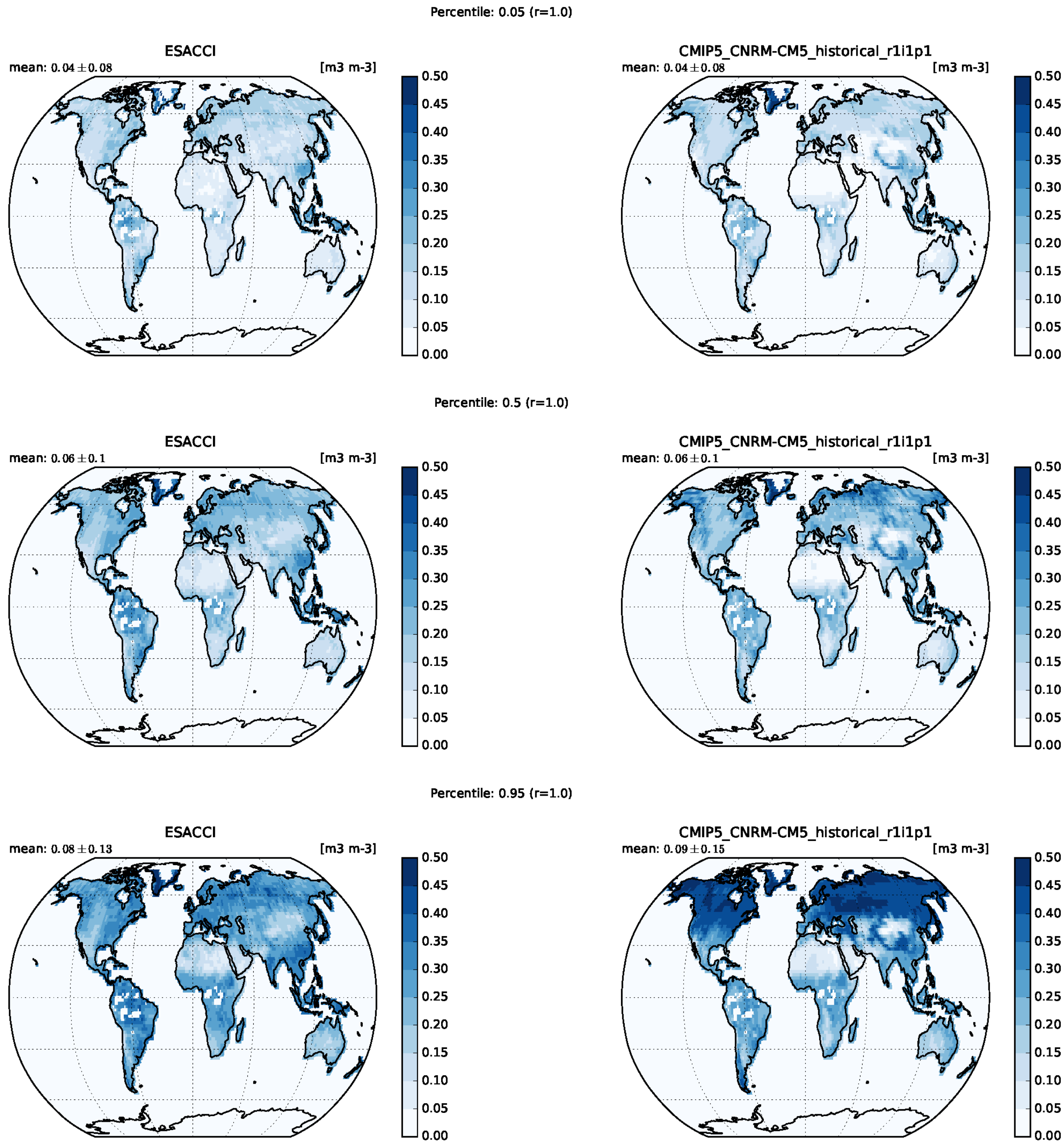


Figure 11. Percentile maps for ESA CCI soil moisture (left column) and soil moisture from CNRM-CM5 (right column). The (from top to bottom) 5%, 50% and 95% percentiles are shown and the spatial correlation coefficient between the model and the observations is provided in the title of each plot.

There is increasing evidence on the quality and consistency of the trends in the ESA CCI soil moisture data set. For example, in a recent special issue in the International Journal of Applied Earth Observation and Geoinformation (JAG) (vol. 48, June 16) several trend papers are presented and reveal reliable trends over many parts of the globe where they were compared with other water related observations including runoff, precipitation, and reanalysis data (see e.g., Wang et al., 2016; Su et al., 2016; Du et al., 2016; Qiu et al., 2016, all in the special issue in JAG). These results give more confidence in the ESA CCI soil moisture trends, especially over the sparsely to moderately vegetated regions. This is also highly relevant to assessing soil moisture variability and change in the context of a changing climate, which has been a great challenge so far.

5.5 Land cover

Benchmarking climate models with land cover information is not straight forward due to the different concepts of representation of terrestrial vegetation in global Dynamic Vegetation Models (DGVM), which are typically based on the concept of PFTs that are supposed to represent groups of land cover with similar functional behavior. Thus, an important first step is to map the ESA CCI land cover classes to PFTs like the ones used in CMIP models (Figure 12) (Poulter et al., 2015). As the PFTs in CMIP models differ, the current ESMValTool diagnostics analyzes only broad surface types (bare soil, grass, shrubs, forests), which is similar to the approach chosen by Brovkin et al. (2013). Land cover is either prescribed in the CMIP models or simulated using a DGVM. In particular for the latter case, an independent assessment of the accuracy of the simulated spatial distributions of major land cover types is desirable in order to evaluate the DGVM accuracy for present climate conditions. The diagnostic currently implemented into the ESMValTool considers the land cover to be static for present climate conditions in the CMIP models. The PFT distribution is then compared against satellite observations from a similar time period.

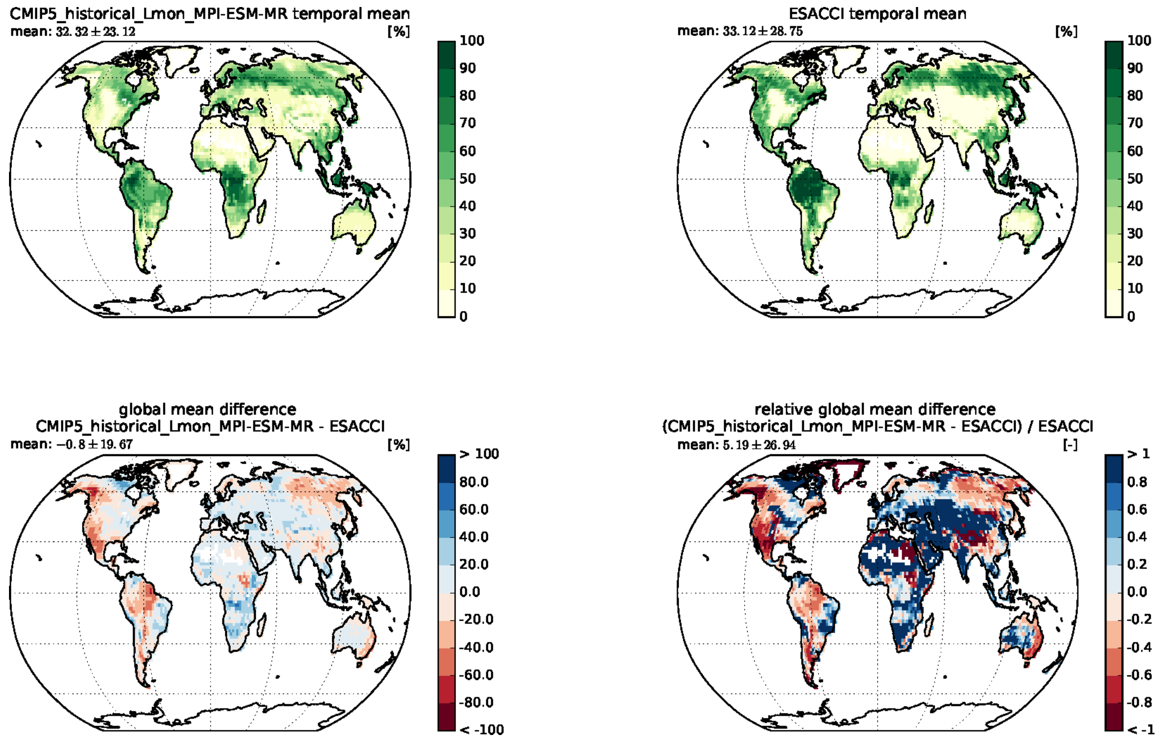


Figure 12. Area fraction (%) of forest and shrub cover in the MPI-ESM-MR model (top left) and the ESA CCI land cover data set (top right) and absolute (bottom left) and relative differences (bottom right). The ESA CCI 2005 epoch was used for the analysis.

Figure 12 and Figure 13 show differences in the area cover fraction for forest type land covers as well as grassland and cropland areas between the ESA CCI land cover product and the MPI-ESM-MR model, which is based on the DGVM JSBACH for the terrestrial component (Brovkin et al., 2009; Brovkin et al., 2013). The tree cover in MPI-ESM-MR is underestimated compared to the ESA CCI data set in the Amazon and along the west coast of North America, while grass and cropland is overestimated in many parts of the globe. Similar analysis results are obtained when using the ESA CCI epoch for the year 2000 instead of 2005.

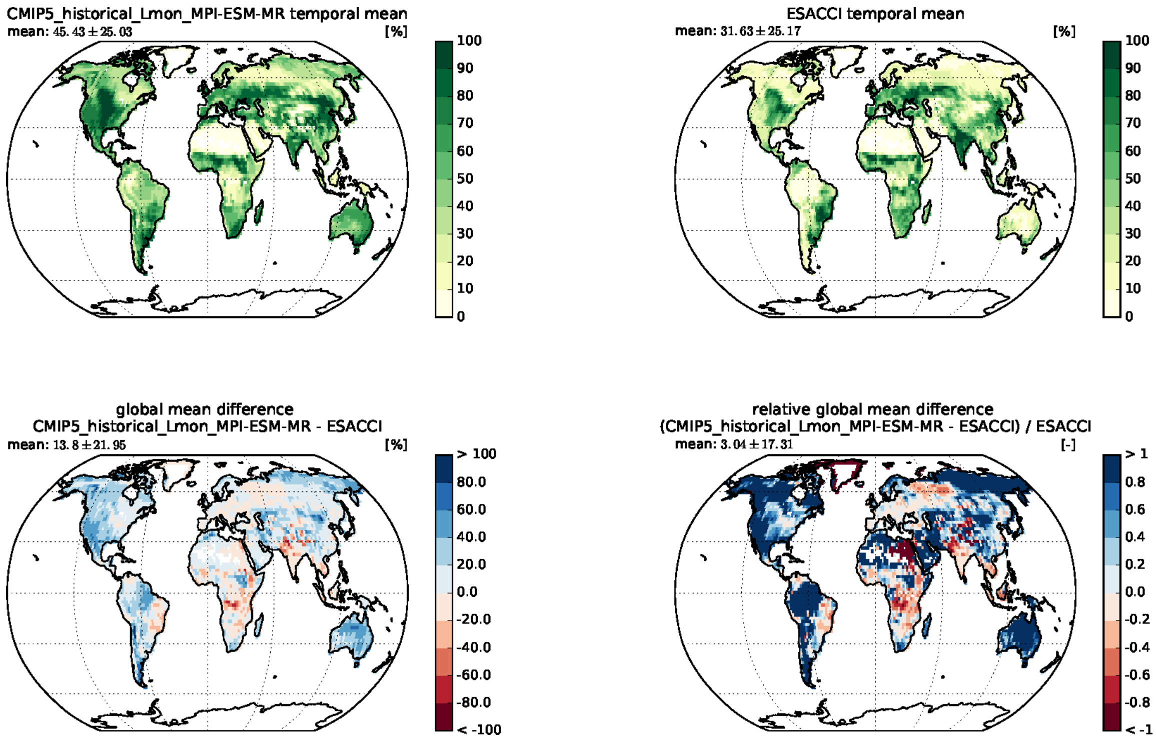
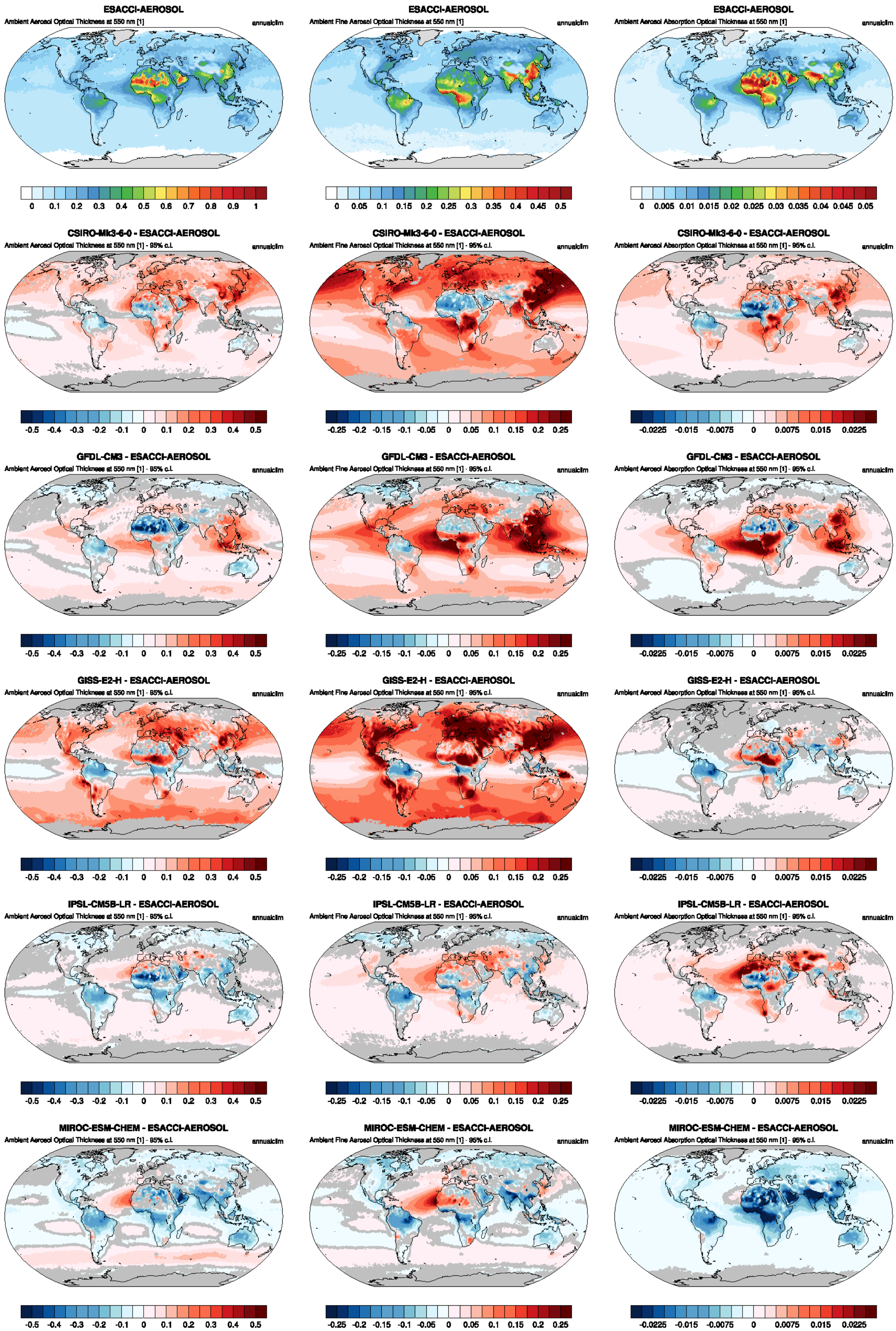


Figure 13. Area fraction (%) of grass and cropland cover in the MPI-ESM-MR model (top left) and the ESA CCI land cover data set (top right) and absolute (bottom left) and relative differences (bottom right). The ESA CCI 2005 epoch was used for the analysis.

The ESA CCI land cover data set provides the first consistent series of high-resolution (300 m) global land cover products derived by combining a whole suite of different sensors including information on PFTs. This has become important in particular for evaluation of ESMs that start to include more complex land cover dynamics in projections of future climate.

5.6 Aerosol

The geographical distribution of the multi-year averages of od550aer, o550lt1aero, and abs550aer, as well as the differences between the ESA CCI data and some exemplary CMIP5 models (CSIRO-Mk3-6-0, GFDL-CM3, GISS-E2-H, IPSL-CM5B-LR, MIROC-ESM-CHEM) are shown in Figure 14. Here, we consider only the CMIP5 models with interactive aerosols and exclude multiple versions of the same model. In general, the models' performance is better over the oceans than over the continents although the SU algorithm used to process the CCI data may underestimate AOD over the oceans. Large model biases are found over the Sahara where some models (especially GFDL-CM3 and IPSL-CM5B-LR) underestimate the aerosol optical depth (left column). This could be caused by an incorrect representation of dust which is consistent with a much better performance of these models for the fine mode optical depth (middle column) in the same region. In addition, the underestimation of AOD over the Sahara might also be partly amplified by an overestimation of AOD in the ESA CCI aerosol product (SU), which is a known problem in this region. In contrast, a substantial positive bias is found over Europe and East Asia (in particular CSIRO-Mk3-6-0 and GISS-E2-H) with similar biases both in the total and in the fine mode optical depth. Significant deviations from the observations are also visible in the modeled absorption optical depth (right column), especially in tropical regions. The contribution of absorption to the aerosol optical depth is, however, quite small. We also note that the satellite uncertainty for abs550aer is larger than for od550aer and od550ltaer.


ESACCI-AEROSOL
Ambient Aerosol Optical Thickness at 550 nm [1]
Ambient Fine Aerosol Optical Thickness at 550 nm [1]
Ambient Aerosol Absorption Optical Thickness at 550 nm [1]
annualclim
CSIRO-Mk3-6-0 - ESACCI-AEROSOL
GFDL-CM3 - ESACCI-AEROSOL
GISS-E2-H - ESACCI-AEROSOL
IPSL-CM5B-LR - ESACCI-AEROSOL
MIROC-ESM-CHEM - ESACCI-AEROSOL

909

Figure 14. Climatological mean AOD (left column), fine mode optical depth (middle) and absorption optical depth (right column) at 550 nm averaged over the period 1997-2011. The first row shows the the observations (ESA CCI ATSR SU v4.21), the other rows the differences between selected CMIP5 models with interactive aerosols and the ESA CCI data. Differences that are not statistically significant at the 95% confidence level are masked out in gray.

As can also be seen in Figure 1, the two satellite data sets used as observational references result in different model performance grades, mainly because of measurement uncertainties inherent to the data sets. To further explore the reason for these differences, the two satellite data sets are compared with ground-based measurements from the AErosol RObotic NETwork (AERONET; Holben et al., 1998) (Figure 15). AERONET data are widely accepted as a reliable reference for aerosol optical depth and are often used for validating satellite products. AERONET data, however, do not provide global coverage with very few measurements particularly over the ocean. The few AERONET sites that are measuring AOD over the ocean are typically near shallow-water areas such as on islands and the coastlines of continents, and thus not representative of open ocean conditions. The Marine Aerosol Network MAN has therefore been established to provide AOD measured with hand-held sun photometers, predominantly on research ships, starting from 2004 (Smirnov et al., 2009). However, in spite of the many cruises included, the data are still sparse making global satellite data sets very valuable for evaluation of ESMs. For consistency, we only consider years that are covered by both, the MODIS and the ESA CCI data sets (2003-2011). Similarly to the models, the largest differences between the two satellite data sets are found over the continents (top row of Figure 15). This is not surprising given that satellite retrievals over the dark ocean surfaces are less sensitive to the assumptions in the retrieval algorithms. The ESA CCI product shows a considerably higher optical depth than

MODIS over the Sahara and seems to be in slightly better agreement with AERONET in this region (however only a few stations are available around the Sahara). Another striking difference between the two data sets is found over Southeast Asia where od550aer from MODIS is higher than the values from the ESA CCI resulting in a slightly better performance when compared to AERONET. The overall performance of the two data sets is quite similar but the MODIS data show a higher correlation (R^2 = 0.85) with AERONET than the ESA CCI data (R^2 = 0.76) as can be seen in the scatter plots in the bottom row of Figure 15.

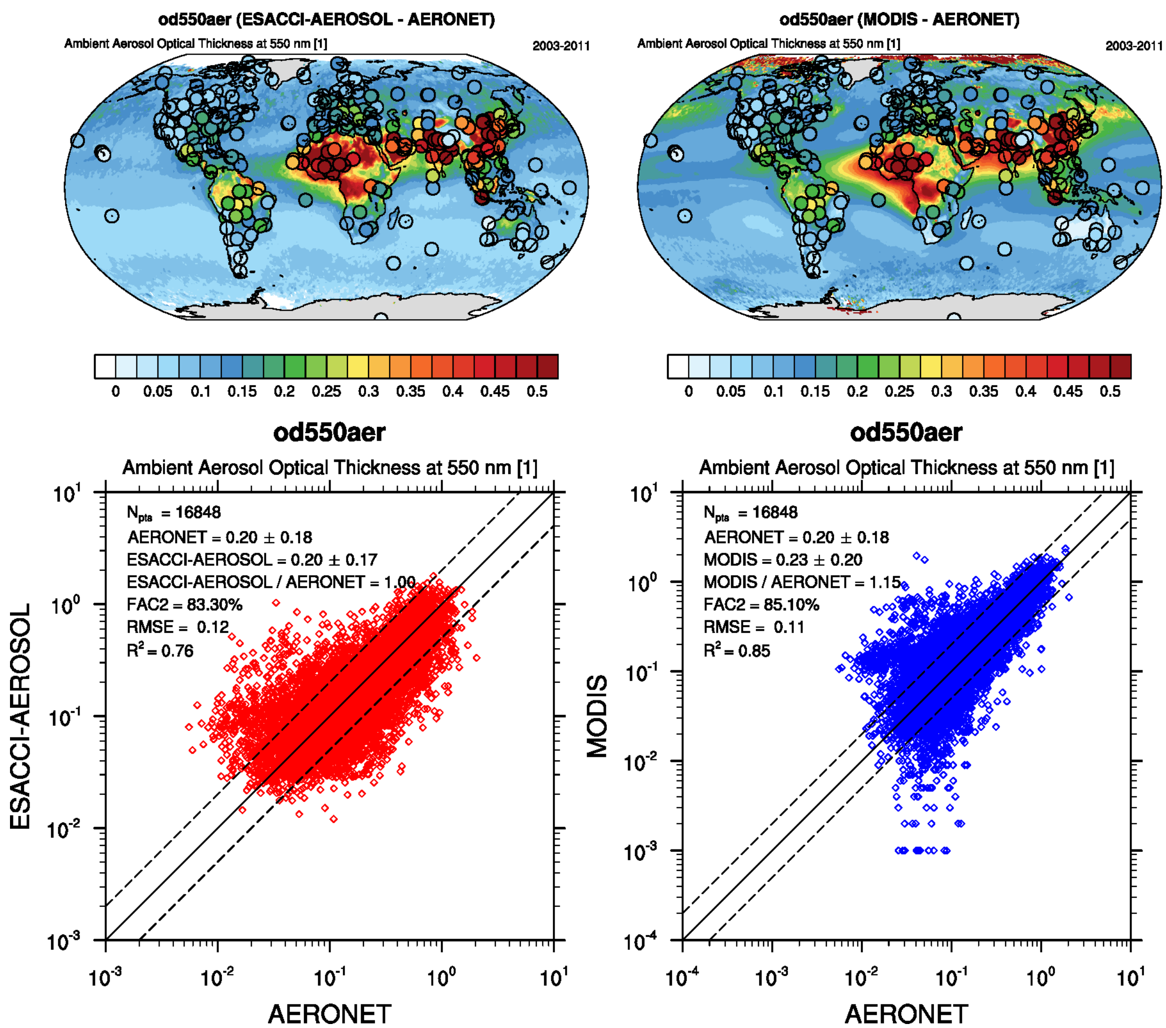


Figure 15. Comparison of AOD at 550 nm from the ESA CCI ATSR SU v4.21 and the MODIS Terra C6 satellite products against the AERONET ground-based measurements for the period 2003-2011. The top

row shows the AERONET values as open circles plotted on top of the satellite data averaged over the same time period. The bottom row shows scatter plots of spatially and temporally collocated measurements on a monthly-mean basis.

With the ESA CCI aerosol and the MODIS data, two independent, long-term satellite data sets are available for model evaluation. This is particularly helpful when there is doubt about the reliability of the comparison with model results by adding the possibility to provide an independent check whether the satellite data are correct. Furthermore, in some areas, the ESA CCI aerosol products provide better correlation with AERONET than MODIS and the addition of ATSR to MODIS data can improve the overall results when used for data assimilation as there are more data available to constrain the model.

5.7 Ozone

For the first time in CMIP, a subset of the models included interactive chemistry in CMIP5. Also in contrast to previous CMIP phases, the models that prescribed ozone in CMIP5 included a time-varying stratospheric ozone climatology (Cionni et al., 2011) rather than a constant forcing. Detailed information on the treatment of ozone in CMIP5 models as well as an evaluation of their performance compared to observations is given in Eyring et al. (2013). Here we repeat some of this analysis by adding the newly available ESA CCI ozone data.

Eyring et al. (2013) divided the CMIP5 models into three classes: (a) CMIP5 models with interactive chemistry, (b) CMIP5 models with semi-interactive chemistry including those models that prescribed ozone data based on results from the underlying CMIP5 chemistry-climate model, and (c) CMIP5 models that prescribed ozone IGAC/SPARC ozone database (Cionni et al., 2011). Here, we focus on the models with interactive ozone chemistry only. The performance of the individual CMIP5 models with interactive chemistry for total ozone columns is similar with respect to both observational data sets (ESA CCI and NIWA) as can be seen in the time series from 1960 through 2010 shown in Figure 16. Differences in both data sets are therefore mostly a result of different statistical methods used to combine the different satellite data sets.

Most CMIP5 models with interactive chemistry overestimate the annual global mean total column ozone compared with the ESA CCI data (Figure 16a) but capture the trend of ozone depletion starting in the 1980s quite well. The October mean total column ozone in the Antarctic (90°S-60°S) is well captured by the CMIP5 models in terms of both, magnitude and trend (Figure 16b).

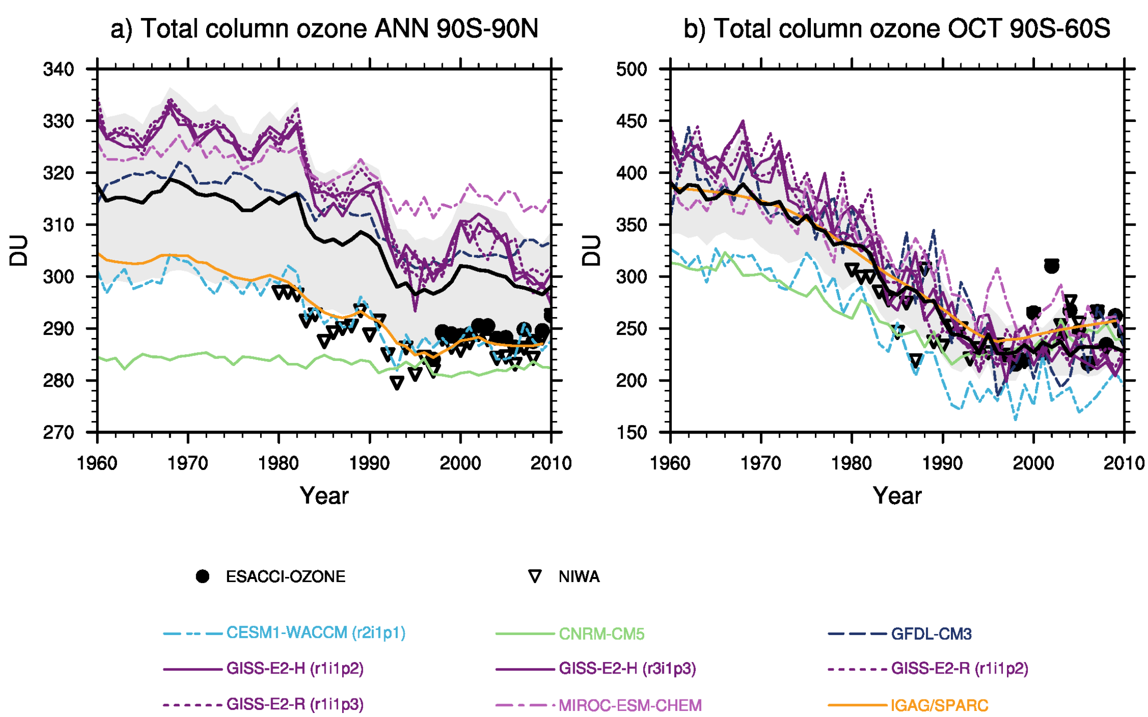


Figure 16. Time series of area-weighted total column ozone from 1960 to 2010 for a) global annual mean (90°S-90°N) and b) Antarctic October mean (60°S-90°S). The figure shows the multi-model mean (black line) and standard deviation (gray shading) as well as individual CMIP5 models with interactive chemistry (colored lines) compared with ESA CCI (filled circles) and NIWA (open triangles) data. The

IGAG/SPARC ozone database (Cionni et al., 2011) is also shown as a reference (orange line). All data sets have been interpolated to the same grid as the ESA CCI observations. During the periods covered by observations, only grid cells in the time series with valid observational data available have been taken into account for calculating the (area-weighted) averages.

Figure 17 shows the climatological vertical profiles of the ozone mixing ratio for different latitude bands and months. Some models simulate ozone only up to 10 hPa, which is just below the layer of maximum ozone concentrations in the stratosphere. Although most models capture the trend and magnitude of total column ozone in Antarctica well, the spread of ozone at 10 hPa in the CMIP5 models is quite large for the same region (80°S).

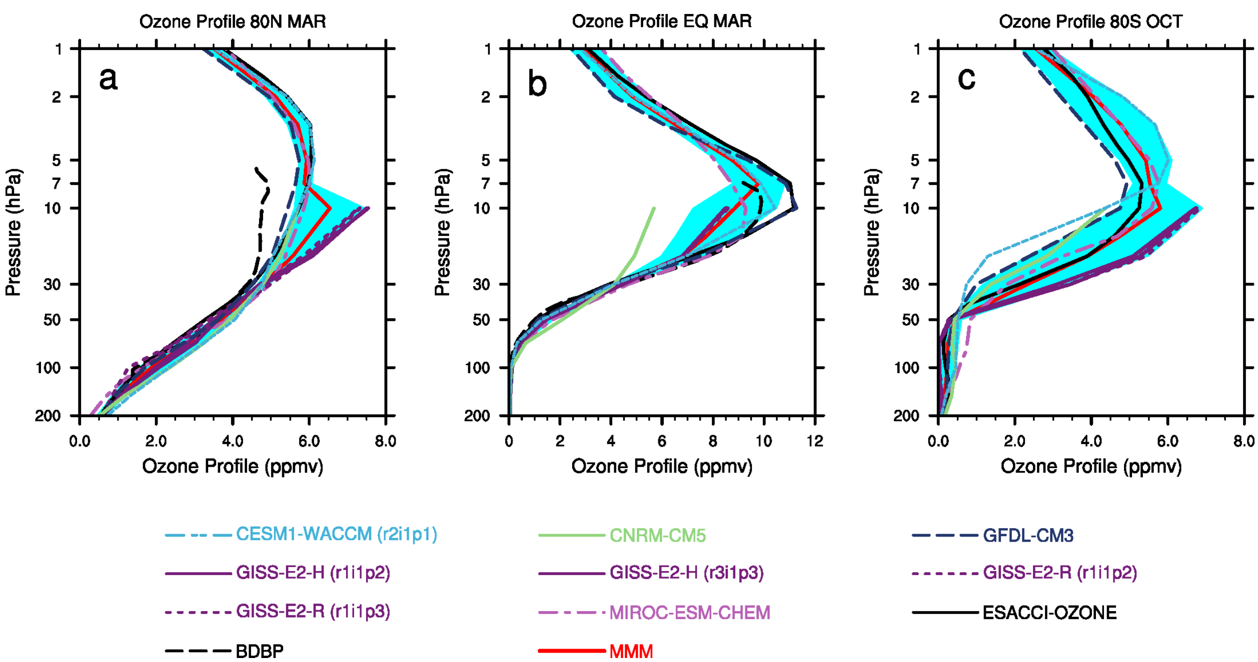


Figure 17. Vertical ozone profile climatologies (2007-2008) at a) 80°N in March, b) the equator in March, and c) at 80°S in October from individual CMIP5 models with interactive chemistry (colored lines) and the ESA CCI ozone data set (solid black line). The multi-model mean (MMM) is shown as a red solid line with one standard deviation of the inter-model spread shown as the light-blue shaded area. For

comparison, also balloon measurements from the Binary Data Base of Profiles (BDBP; Hassler et al., 2008, 2009) are shown for the respective latitudes (2006-2007).

Figure 18 shows the zonally averaged climatological seasonal cycle of total column ozone for the CMIP5 multi-model mean, the two satellite-based reference data sets ESA CCI (Figure 18, upper row) and NIWA (Figure 18, lower row), and the differences of the multi-model mean and the two reference data sets. All data sets (models and observations) have been interpolated linearly to the grid of the observations and all grid cells with no observational data have been excluded from the model data sets. The seasonal cycle is calculated from monthly means averaged over the years 1997 to 2010. As expected, the zonal mean seasonal cycle of total column ozone does not differ much between ESA CCI and NIWA for the above mentioned reason. Only the magnitude of ozone is a few DU higher in northern winter in the ESA CCI data set, which can probably also be attributed to the different merging algorithms used to produce the two data sets. The CMIP5 multi-model mean is able to capture the phase and amplitude of total column ozone but tends to slightly overestimate ozone at the equator throughout the year and underestimate total ozone in Antarctica during summer (November through January). The occurrence of very low ozone values in CMIP5 multi-model mean is delayed by about 1 month compared with the observational data sets and lasts a few weeks longer than shown by the observations.

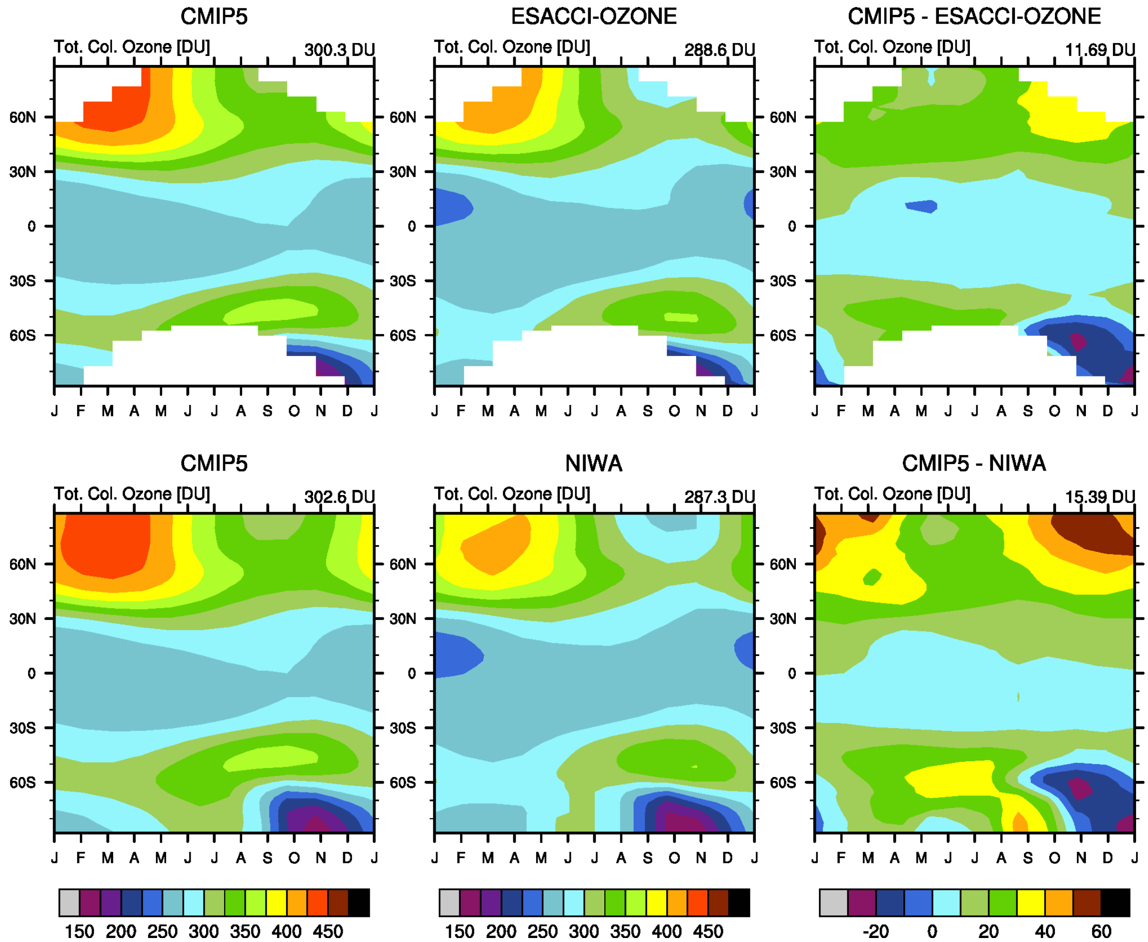


Figure 18. Total column ozone climatologies (1997-2010) for (upper row, from left to right) the multi-model mean of CMIP5 models with interactive chemistry (see Table 1), the ESA CCI ozone data set, and the differences between the CMIP5 multi-model mean and the ESA CCI ozone data. The lower row shows the same plots but for the NIWA combined total column ozone data. The model data have been interpolated to the same grid as the observations. In order to calculate the (area-weighted) global annual averages shown above the individual plots, grid cells in the time series without valid observational data have not been taken into account.

The ESA CCI ozone data sets combine all currently available backscatter nadir spectral UV-Vis sensors, i.e. GOME, SCIAMACHY, GOME-2 and OMI (Lerot et al., 2014) resulting in a harmonized product suitable for analyses of long-term ozone trends (WMO, 2014). The reprocessed ozone profiles from 20 years of observations by GOME, SCIAMACHY and GOME-2 result in a data set of unprecedented accuracy and consistency (Miles et al., 2015; Keppens et al., 2015) well suited for the evaluation of global coupled climate models with interactive chemistry.

5.8 Greenhouse Gases: XCO_2

In order to compare the ESA CCI XCO_2 data set with CMIP5 simulations, only the emission driven simulations (esmHistorical) are used. These simulations were extended until 2014 with results from simulations of the RCP8.5 (esmrcp85). The differences in modeled CO_2 concentrations in the year 2014 between the different emission scenarios (RCP2.6, RCP4.5, RCP8.5) are rather negligible and are therefore not further discussed in the analysis presented here. Here, we focus on those models of the CMIP5 ensemble that provide all necessary data to compare with the ESA CCI GHG data for the full time period (2003-2014): BNU-ESM, CanESM2, CESM1-BGC, FIO-ESM, GFDL-ESM2G, GFDL-ESM2M, MIROC-ESM, MPI-ESM-LR, MRI-ESM1, and NorESM1-ME (see Table 1). These models include an interactive carbon cycle and performed emission driven simulations in which the emissions rather than the concentrations of the greenhouse gases are prescribed (Taylor et al., 2012). This allows the carbon cycle in the models to react to changes in climate by adjusting their carbon fluxes to the new climate conditions and providing the atmospheric CO_2 concentration as an output (Friedlingstein et al., 2006).

For comparison of model and satellite data shown in Figure 19 and Figure 20, the model data were interpolated to the grid of the ESA CCI data set (5°x5°) using local area averaging. Grid cells with missing values in the satellite data were also flagged as missing in the model fields. An important characteristic of the ESA CCI data set is that between 2003-2008 measurements are only over land whereas from 2009-2014 the record contains measurements over land and ocean. The models have been sampled accordingly.

Figure 19 shows the monthly mean time series of XCO_2 comparing ESA CCI data with CMIP5 simulations in four different latitude bands. For all four latitude bands two main features of the time series are very prominent: firstly, the increase in XCO_2 between 2003 and 2014. The increase of about 2 ppm per year is consistent with other observations (Ciais, 2013; Jones and Cox, 2005; Tans and Keeling, 2015) although the absolute values are not directly comparable since the ESA CCI product is an average of the total atmospheric column of atmospheric CO_2 with the concentration at higher altitudes increasing more slowly than at the surface due to mixing (Shia et al., 2006). The CMIP5 multi-model mean shows a positive bias compared with the ESA CCI data of about 5-10 ppm in all four domains. Particularly the CESM-BGC and the GFDL-ESM2M models simulate an XCO_2 bias of about two times higher than the bias of the multi-model mean. The MRI-ESM1 model has the largest negative bias of the models analyzed here with a bias of about -20 ppm. This agrees with findings by Friedlingstein et al. (2014) and Hoffman et al. (2014), who analyzed CO_2 simulated by ESMs. Secondly, the seasonal variation of XCO_2 is more pronounced in the northern hemisphere (30°N-60°N) because of more vegetation exchanging carbon with the atmosphere. We note again that no ESA CCI XCO_2 data over the ocean are available before 2009 (see also section 2.8). Since the main anthropogenic sources of CO_2 are located over land, the CO_2 concentrations over the oceans are slightly lower than over land. Thus, there is a small discontinuity in the XCO_2 time series shown in Figure 19 in the beginning of the year 2009 when measurements over the ocean become available and are included in the calculation of the averages over the different latitude bands. As a consequence of this artifact, the amplitudes of the seasonal cycle in Figure 19 appear slightly reduced in the beginning of 2009.

The emission driven CMIP5 models simulate a large spread in XCO_2 at all latitude bands mainly falling outside the observational (1-sigma) uncertainty of the ESA CCI data. The MPI-ESM-LR model is in good agreement with the annual average XCO_2 values but overestimates the amplitude of the seasonal cycle compared with the ESA CCI data.

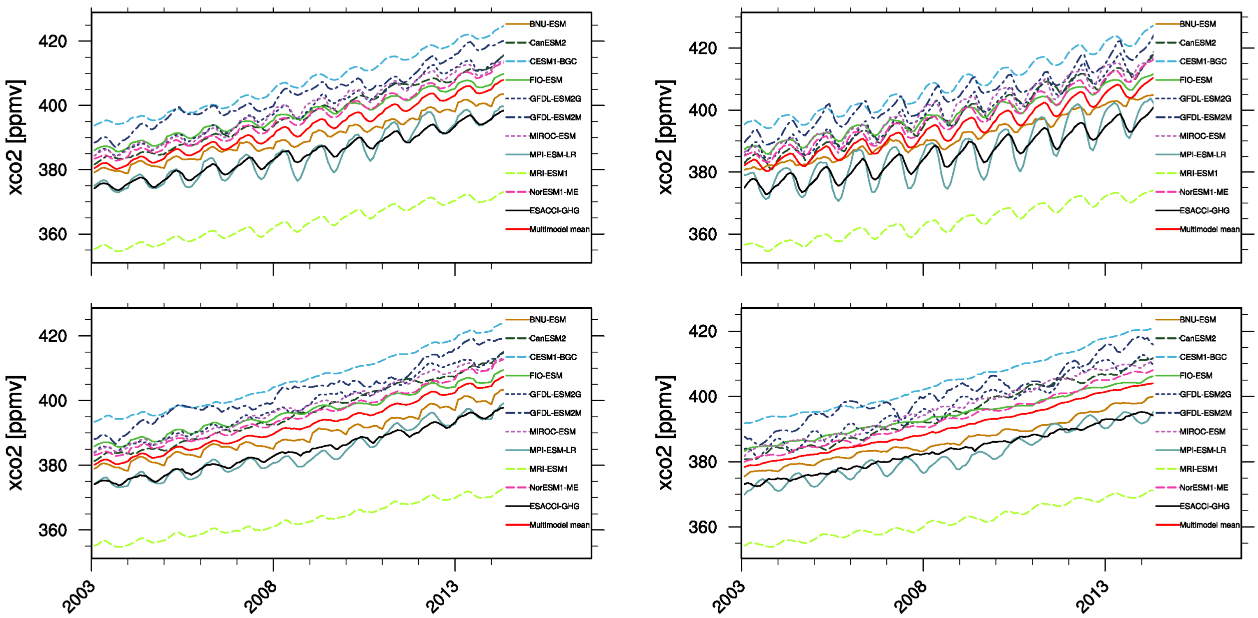


Figure 19. Time series of column averaged carbon dioxide (XCO_2) from 2003 to 2014 from the CMIP5 emission driven simulations for the historical period (2003 to 2005) extended with RCP8.5 simulations (from 2006 to 2014) in comparison with the ESA CCI GHG XCO_2 data. The CMIP5 models are interpolated to the 5°x5° grid of the observations omitting grid cells with no observations. From top left to bottom right: global average, 30°N-60°N, 30°S-30°N, and 60°S-30°S.

The spatial distribution of XCO_2 from the CMIP5 models and the ESA CCI data set is compared by analyzing the deviations from the climatological annual averages (2003-2008 and 2009-2014) shown in Figure 20. Because of the trend in XCO_2, we show the two time periods separately to reduce artifacts caused by XCO_2 data over the ocean only being available in the second half of the ESA CCI record (2009-2014) (Buchwitz et al., 2015). The CMIP5 models have been sampled accordingly averaging only over grid cells with observational data available. Over the continents the ESA CCI data reveal many expected regional features, such as lower XCO_2 concentrations over the tropical rain forests and the boreal forests in the northern high latitudes (Buchwitz et al., 2015). This spatial distribution can be expected because in forest regions and areas with high vegetation more carbon from the atmosphere is taken up by plants via photosynthesis (Keeling et al., 1995). Higher than global average values are found particularly in the northern hemisphere over the United States, Europe, Middle East, India, and China. These basic features are reproduced by the CMIP5 multi-model mean but the annual average XCO_2 values are overestimated by about 6-10 ppm by the models compared with the ESA CCI data in the time period 2003-2008. This bias in the CMIP5 multi-model mean is found to increase slightly to 8-12 ppm in the second half of the ESA CCI XCO_2 record (2009-2014), which could point to possibly slightly too weak carbon sinks in some models (Friedlingstein et al., 2014).

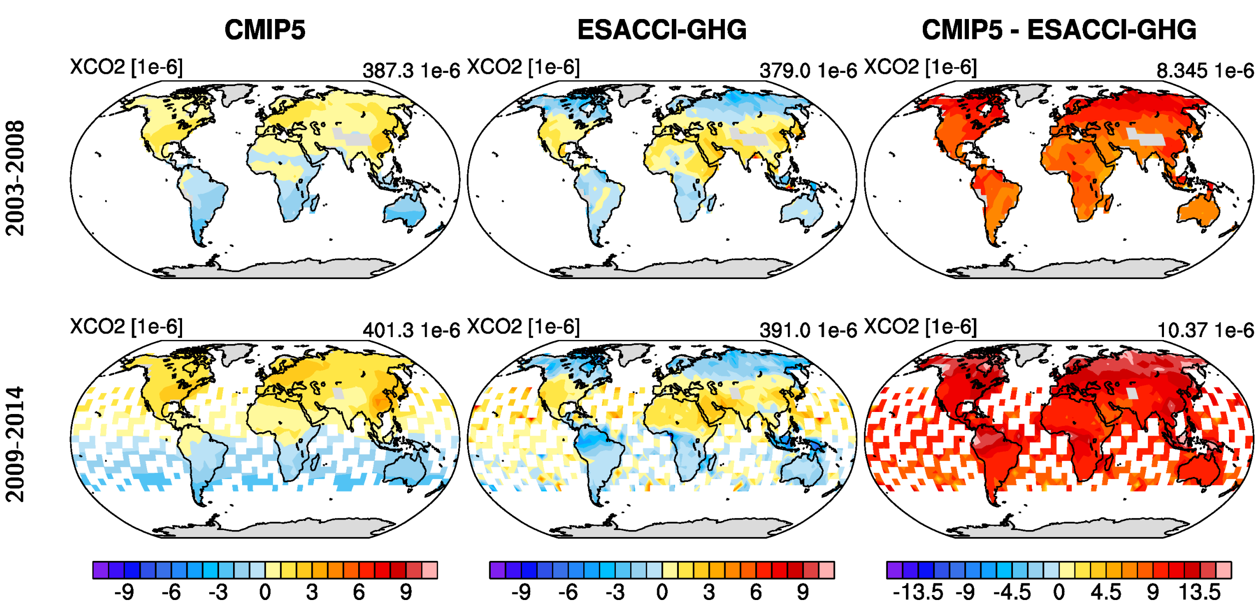


Figure 20. Annual mean XCO_2 climatologies averaged over the years 2003-2008 (top row) and over the years 2009-2014 (bottom row). Shown are deviations from the global annual mean (printed in the right above each panel) for (left) the CMIP5 multi-model mean and (middle) ESA CCI XCO_2. The right panels show the absolute differences between the CMIP5 multi-model mean and ESA CCI XCO_2 data. The CMIP5 results shown are from emission driven historical simulations extended with the respective RCP8.5 scenario.

6 Summary and outlook

Diagnostics for a subset of the ESA CCI Phase 2 data including the CCIs sea surface temperature, sea ice, cloud, soil moisture, land cover, aerosol, ozone, and greenhouse gases have been implemented into the community diagnostics and performance metrics tool ESMValTool. This enhanced version of the ESMValTool has been applied to evaluate a suite of CMIP5 models with the new ESA CCI data sets as well as to compare the new data sets with observations that have already been widely used for model evaluation. The usage of the ESA CCI data in model evaluation has been demonstrated in overview statistics of the models' global average performance using RMSD from the climatological mean seasonal cycle as a metric. The ESA CCI data sets allow for evaluation of new ECVs such as global soil moisture and AOD from fine particles from global coupled (free running) climate models for which consistent and long-term observational data sets have not been previously available. For other variables such as total cloud cover, sea surface temperature, or total ozone columns, the ESA CCI data sets provide the possibility to compare previously available observational data sets in addition to the models. This can help to estimate the uncertainty inherent to model evaluations caused by the choice of a specific observational reference data set for comparison. The error estimates provided as part of the ESA CCI data sets on a per grid basis help to further assess and quantify what a climate model can be realistically expected to reproduce. A new extended version of the Taylor diagram has been presented that includes observational uncertainty estimates and allows to quickly identify models with a RMSE compared to the observations of less than the observational uncertainty (also given as RMSE) by simply gauging the figure. The models cannot be expected to agree perfectly with the observations given the observational uncertainty. In particular for ECVs with large observational uncertainties such as certain cloud properties this helps to avoid over-interpreting model biases that cannot be assessed quantitatively and that might depend significantly on the choice of the reference data set.

In most cases, the ESA CCI data compare well with existing data sets such as, for instance, MODIS AOD, NIWA total ozone, or NSIDC sea ice concentration. The additional value of implementing the ESA CCI data sets into the ESMValTool for these quantities lies particularly in the harmonized and consistently processed data from different platforms and instruments. Such data can now be used by the climate modeling community to evaluate long-term trends and variability of selected modeled ECVs. This is particularly relevant to assessing modeled changes in ECVs related to projected climate change and an important contribution reducing the uncertainties in the projected climate change scenarios.

The ESMs participating in CMIP6 will be more complex than the models of the CMIP5 generation and include new or more detailed processes such as more sophisticated dynamical vegetation models, sea ice treatment or interactive chemistry and carbon cycle. Future releases of the ESMValTool will therefore not only include further ESA CCIs such as ocean color, sea level, ice sheets and fire, but also additional ECVs from already implemented CCIs such as column averaged methane or additional cloud properties such as, for instance, cloud water path, spectral cloud albedo and cloud optical properties.

The aim is to apply the enhanced version of the ESMValTool presented in this paper for routine evaluation of ESMs with observations including the ESA CCI data sets within CMIP6. The CMIP6 results can be analyzed and evaluated together with other evaluation tools and metrics packages such as PMP as soon as the results become available on the ESGF. The application of different analysis/evaluation tools in combination with different and independent observational data sets will help to get a more complete picture of the performance of the quite complex state-of-the-art ESMs, particularly across different ESM domains. This is an important step to identify domains and processes that would particularly benefit from further model improvements and one step further to the ultimate goal of improving our understanding of the climate system and reducing the uncertainties in projections of future climate change.

Code Availability

The enhanced version of the ESMValTool presented in this paper is released under the Apache License, VERSION 2.0. The newly added ESMValTool namelist ‘namelist_lauer16rse.xml’ includes the diagnostics that can be used to reproduce the figures of this paper. This enhanced version will be available from the ESMValTool webpage at http://www.esmvaltool.org/ and from github (https://github.com/ESMValTool-Core/ESMValTool). Users who apply the software resulting in presentations or papers are kindly asked to cite the ESMValTool documentation paper (Eyring et al., 2016b) alongside with the software doi (doi: 10.17874/ac8548f0315) and version number. The wider climate community is encouraged to contribute to this effort and to join the ESMValTool development team for contribution of additional more in-depth diagnostics for ESM evaluation.

Acknowledgments. Financial support for the development of the enhanced version of the ESMValTool presented in this study was provided by ESA’s Climate Change Initiative Climate

Modelling User Group (CMUG). Additionally, funding was received from the European Union's Horizon 2020 research and innovation programme, under grant agreement No. 641816 (CRESCENDO). We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 1 of this paper) for producing and making their model output available.

References

Adler, R. F., G. J. Huffman, A. Chang, R. Ferraro, P.-P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, P. Arkin, and E. Nelkin (2003), The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-Present), J. Hydrometeorol., 4, 1147-1167.

Albergel, C., P. de Rosnay, C. Gruhier, J. Muñoz-Sabater, S. Hasenauer, L. Isaksen, Y. Kerr, and W. Wagner (2012), Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations, Remote Sensing Envir., 118, 215-226.

Arora, V. K., J. F. Scinocca, G. J. Boer, J. R. Christian, K. L. Denman, G. M. Flato, V. V. Kharin, W. G. Lee, and W. J. Merryfield (2011), Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases, Geophys. Res. Lett., 38, L05805, doi: 10.1029/2010GL046270.

Bentsen, M., I. Bethke, J. B. Debernard, T. Iversen, A. Kirkevåg, Ø. Seland, H. Drange, C. Roelandt, I. A. Seierstad, C. Hoose, and J. E. Kristjánsson (2013), The Norwegian Earth System Model, NorESM1-M - Part 1: Description and basic evaluation of the physical climate, Geosci. Model Dev., 6, 687-720, doi: 10.5194/gmd-6-687-2013.

Bi, D., M. Dix, S. Marsland, S. O'Farrell, H. Rashid, P. Uotila, A. Hirst, E. Kowalczyk, M. Golebiewski, A. Sullivan, H. Yan, N. Hannah, C. Franklin, Z. Sun, P. Vohralik, I. Watterson, X. Zhou, R. Fiedler, M. Collier, Y. Ma, J. Noonan, L. Stevens, P. Uhe, H. Zhu, S. Griffies, R. Hill, C. Harris, and K. Puri (2013), The ACCESS coupled model: description, control climate and evaluation, Aust. Met. Oceanogr. J., 63, 41-64.

Bodas-Salcedo, A., M. J. Webb, S. Bony, H. Chepfer, J.-L. Dufresne, S. A. Klein, Y. Zhang, R. Marchand, J. M. Haynes, R. Pincus, and V. O. John (2011), COSP Satellite simulation software for model assessment, Bull. Amer. Meteorol. Soc., 92, 1023-1043, doi: 10.1175/2011BAMS2856.1.

Bodeker, G. E., H. Shiona, and H. Eskes (2005), Indicators of Antarctic Ozone Depletion, Atmos. Chem. Phys., 5, 2603-2615.

Bojinski, S., M. Verstraete, T. C. Peterson, C. Richter, A. Simmons, and M. Zemp (2014), The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy, Bull. Amer. Meteor. Soc., 95, 1431-1443, doi: 10.1175/BAMS-D-13-00047.1.

Bontemps, S., M. Herold, L. Kooistra, A. van Groenestijn, A. Hartley, O. Arino, I. Moreau, and P. Defourny (2012), Revisiting land cover observation to address the needs of the climate modeling community, Biogeosciences, 9, 2145-2157, doi: 10.5194/bg-9-2145-2012.

Bovensmann, H., J. P. Burrows, M. Buchwitz, J. Frerick, S. Noël, V. V. Rozanov, K. V. Chance, and A. P. H. Goede (1999), SCIAMACHY: Mission Objectives and Measurement Modes, J. Atmos. Sci., 56, 127-150.

Bräu, M. (2013), Sea-ice in decadal and long-term simulations with the Max Planck Institute Earth System Model, Bachelor thesis, Ludwig Maximilian University, Munich, Germany.

Brovkin, V., T. Raddatz, C. H. Reick, M. Claussen, and V. Gayler (2009), Global biogeophysical interactions between forest and climate, Geophys. Res. Lett., 36, L07405, doi: 10.1029/2009GL037543.

Brovkin, V., L. Boysen, T. Raddatz, V. Gayler, A. Loew, and M. Claussen (2013), Evaluation of vegetation cover and land-surface albedo in MPI-ESM CMIP5 simulations, J. Adv. Model. Earth Syst., 5, 48-57, doi: 10.1029/2012MS000169.

Buchwitz, M., R. de Beek, J. P. Burrows, H. Bovensmann, T. Warneke, J. Notholt, J. F. Meirink, A. P. H. Goede, P. Bergamaschi, S. Korner, M. Heimann, and A. Schulz (2005), Atmospheric methane and carbon dioxide from SCIAMACHY satellite data: initial comparison with chemistry and transport models, Atmos. Chem. Phys., 5, 941-962.

Buchwitz, M., M. Reuter, O. Schneising, H. Boesch, I. Aben, M. Alexe, R. Armante, P. Bergamaschi, H. Bovensmann, D. Brunner, B. Buchmann, J. P. Burrows, A. Butz, F. Chevallier, A. Chédin, C. D. Crevoisier, S. Gonzi, M. De Mazière, E. De Wachter, R. Detmers, B. Dils, C. Frankenberg, P. Hahne, O. P. Hasekamp, W. Hewson, J. Heymann, S. Houweling, M. Hilker, T. Kaminski, G. Kuhlmann, A. Laeng, T. T. v. Leeuwen, G. Lichtenberg, J. Marshall, S. Noël, J. Notholt, P. Palmer, R. Parker, M. Scholze, G. P. Stiller, T. Warneke, and C. Zehner (2015), The greenhouse gas project of ESA's climate change initiative (GHG-CCI): overview, achievements and future plans, Int. Arch. Photogramm, Remote Sens. Spatial Inf. Sci., XL-7/W3, 165-172.

Buchwitz, M., and M. Reuter (2016), Merged SCIAMACHY/ENVISAT and TANSO-FTS/GOSAT atmospheric column-average dry-air mole fraction of CO2 (XCO2) - XCO2_CRDP3_001, Technical Document, http://esgf-data1.ceda.ac.uk/thredds/fileServer/esg_obs4mips/esacci/ghg/data/obs4mips/crdp_3/CO2/v100/TechNote_SCIAGOSAT_L3_CRDP3_001_XCO2_FINAL.pdf (last accessed 22 September 2016).

Burrows, J. P., E. Holzle, A. P. H. Goede, H. Visser, and W. Fricke (1995), SCIAMACHY - Scanning Imaging Absorption Spectrometer for Atmospheric Chartography, Acta Astronaut, 35, 445-451.

Cavalieri, D.J., C. L. Parkinson, P. Gloersen, and H. Zwally (1996), Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, Arctic, full record, doi: http://dx.doi.org/10.5067/8GQ8LZQVL0VL.

Cho, K., N. Sasaki, H. Shimoda, T. Sakata, and F. Nishio (1996), Evaluation and Improvement of SSM/I Sea Ice Concentration Algorithms for the Sea of Okhotsk, Journal Remote Sensing of Japan, 16(2), 47-58.

Ciais, P., C. Sabine, G. Bala, L. Bopp, V. Brovkin, J. Canadell, A. Chhabra, R. DeFries, J. Galloway, M. Heimann, C. Jones, C. Le Quéré, R. B. Myneni, S. Piao, and P. Thornton (2013), Carbon and Other Biogeochemical Cycles. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Stocker, T. F., D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P. M. Midgley (Ed.), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Cionni, I., V. Eyring, J. F. Lamarque, W. J. Randel, D. S. Stevenson, F. Wu, G. E. Bodeker, T. G. Shepherd, D. T. Shindell, and D. W. Waugh (2011), Ozone database in support of CMIP5 simulations: results and corresponding radiative forcing, Atmos. Chem. Phys., 11, 11267-11292, doi: 10.5194/acp-11-11267-2011.

Clarke, L. E., J. A. Edmonds, H. D. Jacoby, H. Pitcher, J. M. Reilly, and R. Richels (2007), Scenarios of greenhouse gas emissions and atmospheric concentrations, Sub-report 2.1a of Synthesis and Assessment Product 2.1, Climate Change Science Program and the Subcommittee on Global Change Research, Washington DC.

Coldewey-Egbers, M., D. G. Loyola, M. Koukouli, D. Balis, J.-C. Lambert, T. Verhoelst, J. Granville, M. van Roozendael, C. Lerot, R. Spurr, S. M. Frith, and C. Zehner (2015), The GOME-type Total Ozone Essential Climate Variable (GTO-ECV) data record from the ESA Climate Change Initiative, Atmos. Meas. Tech., 8, 3923-3940, doi: 10.5194/amt-8-3923-2015.

Collins, M., S. F. B Tett, and C. Cooper (2001), The Internal Climate Variability of HadCM3, a Version of the Hadley Centre Coupled Model without Flux Adjustments, Clim. Dyn., 17 (1), 61-81.

Collins W. J., N. Bellouin, M. Doutriaux-Boucher, N. Gedney, T. Hinton, C. D. Jones, S. Liddicoat, G. Martin, F. O'Connor, J. Rae, C. Senior, I. Totterdell, S. Woodward, T. Reichler, and J. Kim (2008), Evaluation of the HadGEM2 model, Meteorological Office Hadley Centre, Technical Note 74.

Collins, W. J., N. Bellouin, M. Doutriaux-Boucher, N. Gedney, P. Halloran, T. Hinton, J. Hughes, C. D. Jones, M. Joshi, S. Liddicoat, G. Martin, F. O'Connor, J. Rae, J., C. Senior, S. Sitch, I. Totterdell, A. Wiltshire, and S. Woodward (2011), Development and evaluation of an Earth-System model - HadGEM2, Geosci. Model Dev., 4, 1051-1075, doi: 10.5194/gmd-4-1051-2011.

Comiso, J. C., and F. Nishio (2008), Trends in the Sea Ice Cover Using Enhanced and Compatible AMSR-E, SSM/I, and SMMR Data, J. Geophys. Res., 113, C02S07, doi: 10.1029/2007JC0043257.

Comiso, J. C. (2000), Variability and Trends in Antarctic Surface Temperatures from In Situ and Satellite Infrared Measurements, J. Climate, 13, 1674-1696, doi: 10.1175/1520-0442(2000)013<1674:VATIAS>2.0.CO;2.

Comiso, J. C. (1995), SSM/I Concentrations Using the Bootstrap Algorithm, NASA Reference Publication 1380, 40 pages.

Crow, W. T., A. A. Berg, M. H. Cosh, A. Loew, B. P. Mohanty, R. Panciera, P. de Rosnay, D. Ryu, and J. P. Walker (2012), Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products, Rev. Geophys., 50, RG2002, doi: 10.1029/2011RG000372.

Curry, J. A., J. L. Schramm, and E. E. Ebert (1995), Sea-Ice Albedo Climate Feedback Mechanism, J. Climate, 8, 240-247.

de Leeuw, G., T. Holzer-Popp, S. Bevan, W. Davies, J. Descloitres, R. G. Grainger, J. Griesfeller, A. Heckel, S. Kinne, L. Klüser, P. Kolmonen, P. Litvinov, D. Martynenko, P. J. R. North, B. Ovigneur, N. Pascal, C. Poulsen, D. Ramon, M. Schulz, R. Siddans, L. Sogacheva, D. Tanré, G. E. Thomas, T. H. Virtanen, W. von Hoyningen Huene, M. Vountas, and S. Pinnock (2015), Evaluation of seven European aerosol optical depth retrieval algorithms for climate analysis, Remote Sensing of Environment, 162, 295-315, doi: 10.1016/j.rse.04.023.

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Holm, L. Isaksen, P. Kallberg, M. Kohler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J. J. Morcrette, B. K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J. N. Thepaut, F. and Vitart (2011), The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteor. Soc., 137, 553-597.

Defourny, P., et al. (2015), ESA Land Cover Climate Change Initiative (ESA LC_cci) data: ESACCI-LC-L4-LCCS-Map-300m-P5Y-[2000,2005,2010]-v1.6.1 via Centre for Environmental Data Analysis.

Di Gregorio A. (2005), Land cover classification system: Classification concepts and user manual for software - version 2; Food & Agriculture Organization of the United Nations, R., Ed.; Italy, p. 190.

Dils, B., M. Buchwitz, M. Reuter, O. Schneising, H. Boesch, R. Parker, S. Guerlet, I. Aben, T. Blumenstock, J. P. Burrows, A. Butz, N. M. Deutscher, C. Frankenberg, F. Hase, O. P. Hasekamp, J. Heymann, M. De Maziere, J. Notholt, R. Sussmann, T. Warneke, D. Griffith, V. Sherlock, and D. Wunch (2014), The Greenhouse Gas Climate Change Initiative (GHG-CCI): comparative validation of GHG-CCI SCIAMACHY/ENVISAT and TANSO-FTS/GOSAT CO_2 and CH_4 retrieval algorithm products with measurements from the TCCON, Atmos Meas. Tech., 7, 1723-1744.

Donlon, C. J., M. Martin, J. Starka, J. Roberts-Jones, E. Fiedler, and W. Wimmer (2012), The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system, Remote Sensing of Environment, 116, 140-158.

Donner, L. J., and Coauthors (2011), The Dynamical Core, Physical Parameterizations, and Basic Simulation Characteristics of the Atmospheric Component AM3 of the GFDL Global Coupled Model CM3, J. Climate, 24, 3483-3519.

Dorigo, W., A. Gruber, R. De Jeu, W. Wagner, T. Stacke, T., A. Loew, C. Albergel, L. Brocca, D. Chung, R. Parinussa, and R. Kidd (2014), Evaluation of the ESA CCI soil moisture product using ground-based observations, Remote Sensing of Environment, 162, 380-395.

Dorigo, W., R. de Jeu, D. Chung, R. Parinussa, Y. Liu, W. Wagner, and D. Fernández-Prieto (2012), Evaluating global trends (1988-2010) in harmonized multi-satellite surface soil moisture, Geophys. Res. Lett., 39, L18405.

Dorigo, W. A., W. Wagner, R. Hohensinn, S. Hahn, C. Paulik, A. Xaver, A. Gruber, M. Drusch, S. Mecklenburg, P. van Oevelen, A. Robock, and T. Jackson (2011), The international soil moisture network: A data hosting facility for global in situ soil moisture measurements, Hydrology and Earth System Sciences, 15, 1675-1698, doi: 10.5194/hess-15-1675-2011.

Dorigo, W. A., A. Xaver, M. Vreugdenhil, A. Gruber, A. Hegyiová, A. D. Sanchis-Dufau, D. Zamojski, C. Cordes, W. Wagner, and M. Drusch (2013), Global automated quality control of in situ soil moisture data from the international soil moisture network, Vadose Zone Journal, doi: 10.2136/vzj2012.0097.

Dorigo, W.A, D. Chung, A. Gruber, S. Hahn, T. Mistelbauer, R. M. Parinussa, C. Paulik, C. Reimer, R. van der Schalie, R. A. M. de Jeu, and W. Wagner (2016), [Hydrological Cycle] Soil Moisture [in: “State of the Climate 2015”], Bull. Am. Meteorol. Soc., 97(8), S31-S32.

Du, E. A. Di Vittorio, W. D. Collins (2016), Evaluation of hydrologic components of community land model 4 and bias identification, Int. J. Appl. Earth Observ. Geoinform., 48, 5-16, doi: 10.1016/j.jag.2015.03.013.

Dufresne, J.-L., M.-A. Foujols, S. Denvil, A. Caubel, O. Marti, O. Aumont, Y. Balkanski, S. Bekki, H. Bellenger, R. Benshila, S. Bony, L. Bopp, P. Braconnot, P. Brockmann, P. Cadule, F. Cheruy, F. Codron, A. Cozic, D. Cugnet, N. de Noblet, J.-P. Duvel, C. Ethé, L. Fairhead, T. Fichefet, S. Flavoni, P. Friedlingstein, J.-Y. Grandpeix, L. Guez, E. Guilyardi, D. Hauglustaine, F. Hourdin, A. Idelkadi, J. Ghattas, S. Joussaume, M. Kageyama, G. Krinner, S. Labetoulle, A. Lahellec, M.-P. Lefebvre, F. Lefevre, C. Levy, Z. X. Li, J. Lloyd, F. Lott, G. Madec, M. Mancip, M. Marchand, S. Masson, Y. Meurdesoif, J. Mignot, I. Musat, S. Parouty, J. Polcher, C. Rio, M. Schulz, D. Swingedouw, S. Szopa, C. Talandier, P. Terray, N. Viovy, N. Vuichard (2013), Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5, Clim Dyn, 40, 2123-2165, doi: 10.1007/s00382-012-1636-1.

Embury, O., C. J. Merchant, and M. J. Filipiak (2012a), A reprocessing for climate of sea surface temperature from the Along-Track Scanning Radiometers: basis in radiative transfer, Remote Sensing of Environment, 116, 32-46, doi: 10.1016/j.rse.2010.10.016.

Embury, O., C. J. Merchant, and G. K. Corlett (2012b), A reprocessing for climate of sea surface temperature from the Along-Track Scanning Radiometers: initial validation, accounting for skin and diurnal variability, Remote Sensing of Environment, 116, 62-78, doi: 10.1016/j.rse.2011.02.028.

Eyring, V., J. M. Arblaster, I. Cionni, J. Sedlacek, J. Perlwitz, P. J. Young, S. Bekki, D. Bergmann, P. Cameron-Smith, W. J. Collins, G. Faluvegi, K.-D. Gottschaldt, L. W. Horowitz, D. E. Kinnison, J.-F. Lamarque, D. R. Marsh, D. Saint-Martin, D. T. Shindell, K. Sudo, S. Szopa, and S. Watanabe (2013), Long-Term Ozone Changes and Associated Climate Impacts in CMIP5 Simulations, J. Geophys. Res., 118(10), 5029-60.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor (2016a), Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937-1958, doi: 10.5194/gmd-9-1937-2016.

Eyring, V., M. Righi, A. Lauer, M. Evaldsson, S. Wenzel, C. Jones, A. Anav, O. Andrews, I. Cionni, E. L. Davin, C. Deser, C. Ehbrecht, P. Friedlingstein, P. Gleckler, K.-D. Gottschaldt, S. Hagemann, M. Juckes, S. Kindermann, J. Krasting, D. Kunert, R. Levine, A. Loew, J. Mäkelä, G. Martin, E. Mason, A. S. Phillips, S. Read, C. Rio, R. Roehrig, D. Senftleben, A. Sterl, L. H. van Ulft, J. Walton, S. Wang, and K. D. Williams (2016b), ESMValTool (v1.0) – a community diagnostic and performance metrics tool for routine evaluation of Earth System Models in CMIP, Geosci. Model Dev., 9, 1747-1802, doi: 10.5194/gmd-9-1747-2016.

Eyring, V., P. J. Gleckler, C. Heinze, R. J. Stouffer, K. E. Taylor, V. Balaji, E. Guilyardi, S. Joussaume, S. Kindermann, B. N. Lawrence, G. A. Meehl, M. Righi, and D. N. Williams (2016c), Towards improved and more routine Earth system model evaluation in CMIP, Earth Syst. Dynam., 7, 813-830, doi: 10.5194/esd-7-813-2016.

Ferraro, R., D. E. Waliser, P. Gleckler, K. E. Taylor, and V. Eyring (2015), Evolving obs4MIPs to Support the Sixth Coupled Model Intercomparison Project (CMIP6), B. Am. Meteorol. Soc., doi: 10.1175/BAMS-D-14-00216.1.

Flato, G., J. Marotzke, B. Abiodun, P. Braconnot, S. C. Chou, W. Collins, P. Cox, F. Driouech, S. Emori, V. Eyring, C. Forest, P. Gleckler, E. Guilyardi, C. Jakob, V. Kattsov, C. Reason, and M. Rummukainen (2013), Evaluation of Climate Models. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Stocker, T. F., D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley (Ed.), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Friedlingstein, P., P. Cox, R. Betts, L. Bopp, W. von Bloh, V. Brovkin, P. Cadule, S. Doney, M. Eby, I. Fung, G. Bala, J. John, C. Jones, F. Joos, T. Kato, M. Kawamiya, W. Knorr, K. Lindsay, H. D. Matthews, T. Raddatz, P. Rayner, C. Reick, E. Roeckner, K.-G. Schnitzler, R. Schnur, K. Strassmann, A. J. Weaver, C. Yoshikawa, and N. Zeng (2006), Climate-Carbon Cycle Feedback Analysis: Results from the C4MIP Model Intercomparison, J. Climate, 19, 3337-3353.

Friedlingstein, P., M. Meinshausen, V. K. Arora, C. D. Jones, A. Anav, S. K. Liddicoat, and R. Knutti (2014), Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks, J. Climate, 27, 511-526, doi: 10.1175/JCLI-D-12-00579.1.

Gent, P. R., G. Danabasoglu, L. J. Donner, M. M. Holland, E. C. Hunke, S. Jayne, D. M. Lawrence, R. B. Neale, P. J. Rasch, M. Vertenstein, P. H. Worley, Z.-L. Yang, and M. Zhang (2011), The Community Climate System Model version 4, J. Climate, 24(19), 4973-4991, doi: 10.1175/2011JCLI4083.1.

GCOS (2010), Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC, August 2010.

Gleckler, P. J., K. E. Taylor, and C. Doutriaux (2008), Performance metrics for climate models, J. Geophys. Res., 113, D06104.

Gleckler, P. J., C. Doutriaux, P. J. Durack, K. E. Taylor, Y. Zhang, D. N. Williams, E. Mason, and J. Servonnat (2016), A More Powerful Reality Test for Climate Models, Eos Trans. AGU, doi: 10.1029/2016EO051663.

Good, S, and N. Rayner (2014), Product Specification Document, Project Document SST_CCI-PSD-UKMO-201, http://www.esa-sst-cci.org/PUG/documents (last accessed 13 July 2016).

Hagemann, S., A. Loew, and A. Andersson (2013), Combined evaluation of MPI-ESM land surface water and energy fluxes, J. Adv. Model. Earth Syst., 5, 259-286, doi: 10.1029/2012MS000173.

Hassler, B., G. E. Bodeker, and M. Dameris (2008), Technical Note: A new global database of trace gases and aerosols from multiple sources of high vertical resolution measurements, Atmos. Chem. Phys., 8, 5403-5421.

Hassler, B., G. E. Bodeker, I. Cionni, and M. Dameris (2009), A vertically resolved, monthly mean, ozone database from 1979 to 2100 for constraining global climate model simulations, International Journal of Remote Sensing, 30(15-16), 4009-4018.

Hazeleger, W., C. Severijns, T. Semmler, S. Ştefănescu, S. Yang, X. Wang, K. Wyser, E. Dutra, J. M. Baldasano, R. Bintanja, P. Bougeault, R. Caballero, A. M. L. Ekman, J. H. Christensen, B. van den Hurk, P. Jimenez, C. Jones, P.Kållberg, T. Koenigk, R. Mc Grath, P. Miranda, T. van Noije, T. Palmer, J. A. Parodi, T. Schmith, F. Selten, T. Storelvmo, A. Sterl, H. Tapamo, M. Vancoppenolle, P. Viterbo, and U. Willén (2010), EC-Earth, A Seamless Earth-System Prediction Approach in Action, Bull. Am. Meteor. Soc., 91, 1357-1363, doi: 10.1175/2010BAMS2877.1.

Heidinger, A. K., M. J. Foster, A. Walther, and X. Zhao (2014), The Pathfinder atmospheres-extended AVHRR climate dataset, Bull. Am. Meteor. Soc., 95(6), 909-922, doi: 10.1175/BAMS-D-12-00246.1.

Hoffman, F. M., J. T. Randerson, V. K. Arora, Q. Bao, P. Cadule, D. Ji, C. D. Jones, M. Kawamiya, S. Khatiwala, K. Lindsay, A. Obata, E. Shevliakova, K. D. Six, J. F. Tjiputra, E. M. Volodin, and T. Wu (2014), Causes and implications of persistent atmospheric carbon dioxide biases in Earth System Models, J. Geophys. Res. Biogeosci., 119, 141-162, doi: 10.1002/2013JG002381.

Holben, B. N., T. F. Eck, I. Slutsker, D. Tanré, J. P. Buis, A. Setzer, E. Vermote, J. A. Reagan, Y. J. Kaufman, T. Nakajima, F. Lavenu, I. Jankowiak, and A. Smirnov (1998), AERONET – A Federated Instrument Network and Data Archive for Aerosol Characterization, Remote Sens. Environ., 66, 1-16, doi: 10.1016/s0034-4257(98)00031-5.

Hollmann, R., C. J. Merchant, R. Saunders, C. Downy, M. Buchwitz, A. Cazenave, W. Wagner (2013), The ESA climate change initiative: Satellite data records for essential climate variables, Bull. Am. Meteor. Soc., 94(10), 1541-1552, doi: 10.1175/BAMS-D-11-00254.1.

Hourdin, F., M.-A. Foujols, F. Codron, V. Guemas, J.-L. Dufresne, S. Bony, S. Denvil, L. Guez, F. Lott, J. Ghattas, P. Braconnot, O. Marti, Y. Meurdesoif, L. Bopp (2013), Impact of the LMDZ atmospheric grid configuration on the climate and sensitivity of the IPSL-CM5A coupled model, Clim. Dyn., 40, 2167-2192, doi: 10.1007/s00382-012-1411-3.

Hubert, D., J.-C. Lambert, T. Verhoelst, J. Granville, A. Keppens, J.-L. Baray, A. E. Bourassa, U. Cortesi, D. A. Degenstein, L. Froidevaux, S. Godin-Beekmann, K. W. Hoppel, B. J. Johnson, E. Kyrölä, T. Leblanc, G. Lichtenberg, M. Marchand, C. T. McElroy, D. Murtagh, H. Nakane, T. Portafaix, R. Querel, J. M. Russell III, J. Salvador, H. G. J. Smit, K. Stebel, W. Steinbrecht, K. B. Strawbridge, R. Stübi, D. P. J. Swart, G. Taha, D. W. Tarasick, A. M.

Thompson, J. Urban, J. A. E. van Gijsel, R. Van Malderen, P. von der Gathen, K. A. Walker, E. Wolfram, and J. M. Zawodny (2016), Ground-based assessment of the bias and long-term stability of 14 limb and occultation ozone profile data records, Atmos. Meas. Tech., 9, 2497-2534, doi: 10.5194/amt-9-2497-2016.

Hung, K.C, H. Gu, and Z. A. Zong (2004), A modified superconvergent patch recovery method and its application to large deformation problems, Finite Elements in Analysis and Design, 40(5-6).

Huffman, G. J., and D. T. Bolvin (2013), GPCP Version 2.2 SG Combined Precipitation Data Set Documentation, available online at ftp://precip.gsfc.nasa.gov/pub/gpcp-v2.2/doc/V2.2_doc.pdf (last accessed 13 September 2016), http://precip.gsfc.nasa.gov.

Hurrell, J. W., M. M. Holland, P. R. Gent, J. E. Kay, P. J. Kushner, J.-F. Lamarque, W. G. Large, D. Lawrence, K. Lindsay, W. H. Lipscomb, M. C. Long, N. Mahowald, D. R. Marsh, R. B. Neale, P. Rasch, S. Vavrus, M. Vertenstein, D. Bader, W. D. Collins, J. J. Hack, J. Kiehl, and S. Marshall (2013), The Community Earth System Model: A Framework for Collaborative Research, Bull. Am. Meteor. Soc., 94(9), 1339-1360, doi: 10.1175/BAMS-D-12-00121.1.

Ivanova, N., L. T. Pedersen, R. T. Tonboe, S. Kern, G. Heygster, T. Lavergne, A. Sørensen, R. Saldo, G. Dybkjær, L. Brucker, and M. Shokr (2015), Inter-comparison and evaluation of sea ice algorithms: towards further identification of challenges and optimal approach using passive microwave observations, The Cryosphere, 9, 1797-1817, doi: 10.5194/tc-9-1797-2015.

Ji, D., L. Wang, J. Feng, Q. Wu, H. Cheng, Q. Zhang, J. Yang, W. Dong, Y. Dai, D. Gong, R.-H. Zhang, X. Wang, J. Liu, J. C. Moore, D. Chen, and M. Zhou (2014), Description and basic evaluation of Beijing Normal University Earth System Model (BNU-ESM) version 1, Geosci. Model Dev., 7, 2039-2064, doi: 10.5194/gmd-7-2039-2014.

Jones, C. D., and P. M. Cox (2005), On the significance of atmospheric CO_2 growth rate anomalies in 2002-2003, Geophys. Res. Lett., 32.

Kalnay, E., et al. (1996), The NCEP/NCAR 40-Year Reanalysis Project, Bull. Am. Meteor. Soc., 77, 437-471, doi: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Karlsson, K.-G., A. Riihelä, R. Müller, J. F. Meirink, J. Sedlar, M. Stengel, M. Lockhoff, J. Trentmann, F. Kaspar, R. Hollmann, and E. Wolters (2013), CLARA-A1: a cloud, albedo, and radiation dataset from 28 yr of global AVHRR data, Atmos. Chem. Phys., 13, 5351-5367, doi: 10.5194/acp-13-5351-2013.

Keeling, C. D., T. P. Whorf, M. Wahlen, and J. Vanderplicht (1995), Interannual Extremes in the Rate of Rise of Atmospheric Carbon-Dioxide since 1980, Nature, 375, 666-670.

Keppens, A., J.-C. Lambert, J. Granville, G. Miles, R. Siddans, J. C. A. van Peet, R. J. van der A, D. Hubert, T. Verhoelst, A. Delcloo, S. Godin-Beekmann, R. Kivi, R. Stübi, and C. Zehner

(2015), Round-robin evaluation of nadir ozone profile retrievals: methodology and application to MetOp-A GOME-2, Atmos. Meas. Tech., 8, 2093-2120, doi:10.5194/amt-8-2093-2015.

Kern, S., K. Khvorostovsky, H. Skourup, E. Rinne, Z. S. Parsakhoo, V. Djepa, P. Wadhams, and S. Sandven (2015), The impact of snow depth, snow density and ice density on sea ice thickness retrieval from satellite radar altimetry: results from the ESA-CCI Sea Ice ECV Project Round Robin Exercise, The Cryosphere, 9, 37-52, doi: 10.5194/tc-9-37-2015.

Kern, S., A. Rösel, L. T. Pedersen, N. Ivanova, R. Saldo, and R. T. Tonboe (2016), The impact of melt ponds on summertime microwave brightness temperatures and sea ice concentrations, The Cryosphere Discuss., doi: 10.5194/tc-2015-202.

Khoei, S. A., and A. R. Gharehbaghi (2007), The superconvergent patch recovery technique and data transfer operators in 3d plasticity problems, Finite Elements in Analysis and Design, 43(8).

Koukouli, M., D. Balis, I. Zyrichidou, C. Lerot, M. Van Roozendael, J-C. Lambert, J. Granville, J-P. Pommereau, F. Goutail, G. Labow, S. Frith, D. Loyola, R. Spurr, and C. Zehner (2015), Evaluating a new homogeneous total ozone climate data record from GOME/ERS-2, SCIAMACHY/Envisat, and GOME-2/MetOp-A, J. Geophys. Res., 120(23), 12,296-12,312, doi: 10.1002/2015JD023699.

Kuze, A., H. Suto, M. Nakajima, and T. Hamazaki (2009), Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring, Appl. Opt., 48, 6716-6733.

Kwok R, and G. F. Cunningham (2015), Variability of Arctic sea ice thickness and volume from CryoSat-2, Phil. Trans. R. Soc. A., 373, 20140157, doi: 10.1098/rsta.2014.0157.

Lauer, A., and K. Hamilton (2013), Simulating Clouds with Global Climate Models: A Comparison of CMIP5 Results with CMIP3 and Satellite Data, J. Climate, 26, 3823-3845, doi: 10.1175/JCLI-D-12-00451.1.

Lavergne, T., and E. Rinne (2014), ESA Sea Ice Climate Change Initiative: Phase 1, D3.4 Product User Guide (PUG), doc ref: SICCI-PUG-13-07, version: 2.0, date: 29 August 2014.

Lerot, C., M. Van Roozendael, R. Spurr, et al. (2014), Homogenized total ozone data records from the European sensors GOME/ERS-2, SCIAMACHY/Envisat and GOME-2/Metop-A, J. Geophys. Res., 119, 1-20, doi: 10.1002/2013JD020831.

Levy, R. C., S. Mattoo, L. A. Munchak, L. A. Remer, A. M. Sayer, F. Patadia, and N. C. Hsu (2013), The Collection 6 MODIS aerosol products over land and ocean, Atmos. Meas. Tech., 6, 2989-3034, doi: 10.5194/amt-6-2989-2013.

Li, L., B. Wang, Y. Yu, et al. (2010), Development and evaluation of GAMIL2.0 and FGOALS2.0-g. In: The 5th C20C Workshop, 25-28 October 2010, IAP, Beijing.

Li, W., P. Ciais, N. MacBean, Shushi Penga, P. Defourny, and S. Bontemps (2016), Major forest changes and land cover transitions based on plant functional types derived from the ESA CCI Land Cover product, Int. J. Appl. Earth Observ. Geoinform., 47, 30-39, doi: 10.1016/j.jag.2015.12.006.

Lindsay, R., and A. Schweiger (2015), Arctic sea ice thickness loss determined using subsurface, aircraft, and satellite observations, The Cryosphere, 9, 269-283, doi: 10.5194/tc-9-269-2015.

Liu, Y. Y., R. M. Parinussa, W. A. Dorigo, R. A. M. de Jeu, W. Wagner, A. I. J. M. van Dijk, M. F. McCabe, and J. P. Evans (2011), Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals, Hydrology and Earth System Sciences, 15, 425-436, doi: 10.5194/hess-15-425-2011.

Liu, Y. Y., W. A. Dorigo, R. M. Parinussa, R. A. M. de Jeu, W. Wagner, M. F. McCabe, J. P. Evans, and A. I. J. M. van Dijk (2012), Trend-preserving blending of passive and active microwave soil moisture retrievals, Remote Sensing of Environment, 123, 280-297, doi: 10.1016/j.rse.2012.03.014.

Loeb, N. G., B. A. Wielicki, D. R. Doelling, G. L. Smith, D. F. Keyes, S. Kato, N. Manalo-Smith, T. Wong (2009), Toward optimal closure of the Earth's top-of-atmosphere radiation budget, J. Climate, 22, 748-766, doi: 10.1175/2008JCLI2637.1.

Loeb, N. G., J. M. Lyman, G. C. Johnson, R. P. Allan, D. R. Doelling, T. Wong, B. J. Soden, and G. L. Stephens (2012), Observed changes in top-of-the-atmosphere radiation and upper-ocean heating consistent within uncertainty, Nat. Geosci., 5, doi: 10.1038/NGEO1375.

Loew, A., T. Stacke, W. Dorigo, R. de Jeu, and S. Hagemann (2013), Potential and limitations of multidecadal satellite soil moisture observations for selected climate model evaluation studies, Hydrology and Earth System Sciences, 17, 3523-3542.

Loew, A. (2008), Impact of surface heterogeneity on surface soil moisture retrievals from passive microwave data at the regional scale: The upper danube case, Remote Sensing of Environment, 112, 231-248, doi: 10.1016/j.rse.2007.04.009.

Long, C. M., Lindsay, K., S. Peacock, J. K. Moore, and S. C. Doney (2013), Twentieth-Century Oceanic Carbon Uptake and Storage in CESM1(BGC)*, J. climate, 26, 6775-6800, doi: 10.1175/JCLI-D-12-00184.1.

Loyola, D. G., R. M. Coldewey-Egbers, M. Dameris, H. Garny, A. Stenke, M. Van Roozendael, C. Lerot, D. Balis, and M. Koukouli (2009), Global long-term monitoring of the ozone layer - a prerequisite for predictions, International Journal of Remote Sensing, 30, 15, 4295-4318, doi: 10.1080/01431160902825016.

Marsh, D. R., M. J. Mills, D. E. Kinnison, J.-F. Lamarque, N. Calvo, and L. M. Polvani (2013), Climate Change from 1850 to 2005 Simulated in CESM1(WACCM), J. Climate, 26, 7372-7391, doi: 10.1175/JCLI-D-12-00558.1.

Mcgarragh G., C. Poulsen, G. Thomas, A. Povey, O. Sus, C. Schlundt, S. Stapelberg, S. Proud, M. Christensen, M. Stengel, and R. Grainger (2016), The Community Cloud retrieval for CLimate (CC4CL). Part II: The optimal estimation algorithm, manuscript in preparation.

Merchant, C. J., O. Embury, J. Roberts-Jones, E. Fiedler, C. E. Bulgin, G. K. Corlett, S. Good, A. McLaren, N. Rayner, S. Morak-Bozzo, and C. Donlon (2014a), Sea surface temperature datasets for climate applications from Phase 1 of the European Space Agency Climate Change Initiative (SST CCI), Geoscience Data Journal, 1, 179-191, doi: 10.1002/gdj3.20.

Merchant, C. J., O. Embury, J. Roberts-Jones, E. K. Fiedler, C. E. Bulgin, G. K. Corlett, S. Good, A. McLaren, N. A. Rayner, and C. Donlon (2014b), ESA Sea Surface Temperature Climate Change Initiative (ESA SST CCI): Analysis long term product version 1.0, NERC Earth Observation Data Centre, 24th February 2014, doi: 10.5285/878bef44-d32a-40cd-a02d-49b6286f0ea4.

Miles, G. M., R. Siddans, B. J. Kerridge, B. G. Latter, and N. A. D. Richards (2015), Tropospheric ozone and ozone profiles retrieved from GOME-2 and their validation, Atmos. Meas. Tech., 8, 385-398, doi: 10.5194/amt-8-385-2015.

Nam, C., S. Bony, J. L. Dufresne, and H. Chepfer (2012), The ‘too few, too bright’ tropical low-cloud problem in CMIP5 models, Geophys. Res. Lett., 39, L21801, doi: 10.1029/2012GL053421.

Parinussa, R. M., A. G. C. A. Meesters, Y. Y. Liu, W. Dorigo, W. Wagner, and R. A. de Jeu (2011), Error estimates for near-real-time satellite soil moisture as derived from the land parameter retrieval model, IEEE Geoscience and Remote Sensing Letters, 8, 779-783, doi: 10.1109/LGRS.2011.2114872.

Phillips, A. S., C. Deser, and J. Fasullo (2014), Evaluating Modes of Variability in Climate Models, Eos Trans. AGU, 95(49), 453-455.

Platnick, S., M. D. King, S. A. Ackerman, W. P. Menzel, B. A. Baum, J. C. Riedi, and R. A. Frey (2003), The MODIS cloud products: algorithms and examples from Terra, IEEE Transactions on Geoscience and Remote Sensing, 41, 2, 459-473, doi: 10.1109/TGRS.2002.808301.

Popp, T., G. de Leeuw, C. Bingen, C. Brühl, V. Capelle, A. Chedin, L. Clarisse, O. Dubovik, R. Grainger, J. Griesfeller, A. Heckel, S. Kinne, L. Klüser, M. Kosmale, P. Kolmonen, L. Lelli, P. Litvinov, L. Mei, P. North, S. Pinnock, A. Povey, C. Robert, M. Schulz, L. Sogacheva, K. S., D. S. Zweers, G. Thomas, L. Gijsbert Tilstra, S. Vandenbussche, P. Veefkind, M. Vountas, and Y. Xue (2016), Development, Production and Evaluation of Aerosol Climate Data Records from European Satellite Observations (Aerosol_cci), Remote Sensing, 8, 421; doi: 10.3390/rs8050421.

Poulter, B., N. MacBean, A. Hartley, I. Khlystova, O. Arino, R. Betts, S. Bontemps, M. Boettcher, C. Brockmann, P. Defourny, S. Hagemann, M. Herold, G. Kirches, C. Lamarche,

D. Lederer, C. Ottlé, M. Peters, and P. Peylin (2015), Plant functional type classification for earth system models: results from the European Space Agency's Land Cover Climate Change Initiative, Geosci. Model Dev., 8, 2315-2328, doi: 10.5194/gmd-8-2315-2015.

Qiao, F., Z. Song, Y. Bao, Y. Song, Q. Shu, C. Huang, and W. Zhao (2013), Development and evaluation of an Earth System Model with surface gravity waves, J. Geophys. Res. Oceans, 118, 4514-4524, doi: 10.1002/jgrc.20327.

Qiu, J., Q. Gao, S. Wang, and Z. Su (2016), Comparison of temporal trends from multiple soil moisture data sets and precipitation: The implication of irrigation on regional soil moisture trend, Int. J. Appl. Earth Observ. Geoinform., 48, 17-27, doi: 10.1016/j.jag.2015.11.012.

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, 4407.

Rayner, N, S. Good, and T. Block (2015), SST CCI Product User Guide, Project Document SST_CCI-PUG-UKMO-201, http://www.esa-sst-cci.org/PUG/documents, 2015 (last accessed 13 July 2016).

Remer, L. A., Y. J. Kaufman, D. Tanré, S. Mattoo, D. A. Chu, J. V. Martins, R. R. Li, C. Ichoku, R. C. Levy, R. G. Kleidman, T. F. Eck, E. Vermote, and B. N. Holben (2005), The MODIS Aerosol Algorithm, Products, and Validation, J. Atmos. Sci., 62, 947-973, doi: 10.1175/jas3385.1.

Reuter, M., M. Buchwitz, O. Schneising, F. Hase, J. Heymann, S. Guerlet, A. J. Cogan, H. Bovensmann, and J. P. Burrows (2012), A simple empirical model estimating atmospheric CO2 background concentrations, Atmos. Meas. Tech., 5, 1349-1357.

Reynolds, R. W., D. B. Chelton, J. Roberts-Jones, M. J. Martin, D. Menemenlis, and C. J. Merchant (2013), Objective Determination of Feature Resolution in Two Sea Surface Temperature Analyses, J. Climate, 26, 2514-2533, doi: 10.1175/JCLI-D-12-00787.1.

Ricker, R., S. Hendricks, V. Helm, H. Skourup, and M. Davidson (2014), Sensitivity of CryoSat-2 Arctic sea-ice freeboard and thickness on radar-waveform interpretation, The Cryosphere, 8, 1607-1622, doi: 10.5194/tc-8-1607-2014.

Roberts-Jones, J., E. K. Fiedler, and M. J. Martin (2012), Daily, Global, High-Resolution SST and Sea Ice Reanalysis for 1985-2007 Using the OSTIA System, J. Climate, 25, 6215-6232, doi: 10.1175/JCLI-D-11-00648.1.

Robert-Jones, J., K. Bovis, M. J. Martin, and A. McLaren (2016), Estimating background error covariance parameters and assessing their impact in the OSTIA system, Remote Sensing of Environment, 176, 117-138, doi: 10.1016/j.rse.2015.12.006.

Rodgers, C. D. (2000), Inverse Methods for Atmospheric Sounding: Theory and Practice, World Scientific Publishing.

Roeckner, E., R. Brokopf, M. Esch, M. Giorgetta, S. Hagemann, L. Kornblueh, E. Manzini, U. Schlese, and U. Schulzweida (2006), Sensitivity of simulated climate to horizontal and vertical resolution in the ECHAM5 atmosphere model, J. Climate, 19, 3771-3791.

Rotstayn, L., M. Collier, M. Dix, Y. Feng, Y., H. Gordon, S. O'Farrell, I. Smith, J. and Syktus (2010), Improved simulation of Australian climate and ENSO-related climate variability in a GCM with an interactive aerosol treatment, Int. J. Climatology, 30(7), 1067-1088, doi: 10.1002/joc.1952.

Sakamoto and Coauthors (2012), MIROC4h - a new high-resolution atmosphere-ocean coupled general circulation model, Journal of the Meteorological Society of Japan, 90(3), 325-359, doi: 10.2151/jmsj.2012-301.

Sandven, S., et al. (2015), ESA Sea Ice Climate Change Initiative (ESA Seaice_cci) data: ESA CCI SIC v1.11, via Centre for Environmental Data Analysis, 16 Feb 2016.

Santoro, M., U. Wegmüller (2014), Multi-temporal Synthetic Aperture Radar Metrics Applied to Map Open Water Bodies, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 2014, 7, 3225-3238.

Schmidt, G. A., and Coauthors (2006), Present day atmospheric simulations using GISS ModelE: Comparison to in-situ, satellite and reanalysis data, J. Climate, 19, 153-192.

Scoccimarro E., S. Gualdi, A. Bellucci, A. Sanna, P.G. Fogli, E. Manzini, M. Vichi, P. Oddo, and A. Navarra (2011), Effects of Tropical Cyclones on Ocean Heat Transport in a High Resolution Coupled General Circulation Model, J. Climate, 24, 4368-4384.

Shia, R.-L., M.-C. Liang, C. E. Miller, and Y. L. Yung (2006), CO_2 in the upper troposphere: Influence of stratosphere-troposphere exchange, Geophys. Res. Lett., 33, L14814, doi: 10.1029/2006GL026141.

Smirnov, A., B. N. Holben, I. Slutsker, D. M. Giles, C. R. McClain, T. F. Eck, S. M. Sakerin, A. Macke, P. Croot, G. Zibordi, P. K. Quinn, J. Sciare, S. Kinne, M. Harvey, T. J. Smyth, S. Piketh, T. Zielinski, A. Proshutinsky, J. I. Goes, N. B. Nelson, P. Larouche, V. F. Radionov, P. Goloub, K. Krishna Moorthy, R. Matarrese, E. J. Robertson, and F. Jourdin (2009), Maritime Aerosol Network as a component of Aerosol Robotic Network, J. Geophys. Res., 114, D06204, doi: 10.1029/2008JD011257.

Smith, S. J., and T. M. L. Wigley (2006), MultiGas forcing stabilization with minicam, The Energy Journal Special issue, 3, 373-392.

Sofieva, V. F., N. Rahpoe, J. Tamminen, E. Kyrölä, N. Kalakoski, M. Weber, A. Rozanov, C. von Savigny, A. Laeng, T. von Clarmann, G. Stiller, S. Lossow, D. Degenstein, A. Bourassa, C. Adams, C. Roth, N. Lloyd, P. Bernath, R. J. Hargreaves, J. Urban, D. Murtagh, A. Hauchecorne, F. Dalaudier, M. van Roozendael, N. Kalb, and C. Zehner (2013), Harmonized dataset of ozone profiles from satellite limb and occultation measurements, Earth Syst. Sci. Data, 5, 349-363, doi: 10.5194/essd-5-349-2013.

Stengel M., C. Poulsen, G. Thomas, M. Christiansen, C. Carbajal Henken, R. Preusker, J. Fischer, O. Sus, S. Stapelberg, C. Schlundt, A. Devasthale, U. Willén, K.-G. Karlsson, A. Povey, D. Grainger, S. Proud, J. F. Meirink, A. Feofilov, and R. Hollmann (2016a), Cloud property datasets retrieved from AVHRR, MODIS, AATSR and MERIS in the framework of the Cloud_cci project, manuscript in preparation.

Stengel M., O. Sus, S. Stapelberg, C. Schlundt, C. Poulsen, and R. Hollmann (2016b), ESA Cloud Climate Change Initiative (ESA Cloud_cci) data: AVHRR-PM CLD_PRODUCTS v2.0 via Deutscher Wetterdienst.

Stevens, B., M. Giorgetta, M. Esch, T. Mauritsen, T. Crueger, S. Rast, M. Salzmann, H. Schmidt, J. Bader, K. Block, R. Brokopf, I. Fast, S. Kinne, L. Kornblueh, U. Lohmann, R. Pincus, T. Reichler, and E. Roeckner (2013), Atmospheric component of the MPI-M Earth System Model: ECHAM6, Journal of Advances in Modeling Earth Systems, 5, 146-172, doi: 10.1002/jame.20015.

Stroeve, J. C., M. C. Serreze, M. M. Holland, J. E. Kay, J. Malanik, and A. P. Barrett (2012), The Arctic's rapidly shrinking sea ice cover: A research synthesis, Clim Change, 110, 1005-1027, doi: 10.1007/s10584-011-0101-1.

Su, B., A. Wang, G. Wang, Y. Wang, T. Jiang (2016), Spatiotemporal variations of soil moisture in the Tarim River basin, China, Int. J. Appl. Earth Observ. Geoinform., 48, 122-130, doi: 10.1016/j.jag.2015.06.012.

Sus, O., M. Jerg, C. Poulsen, G. Thomas, S. Stapelberg, G. Mcgarragh, A. Povey, C. Schlundt, M. Stengel, and R. Hollmann (2016), The Community Cloud retrieval for CLimate (CC4CL). Part I: A framework applied to multiple satellite imaging sensors, manuscript in reparation.

Susskind, J., C. Barnet, J. Blaisdell, L. Iredell, F. Keita, L. Kouvaris, G. Molnar, and M. Chahine (2006), Accuracy of geophysical parameters derived from Atmospheric Infrared Sounder/Advanced Microwave Sounding Unit as a function of fractional cloud cover, J. Geophys. Res., 111, D9, doi: 10.1029/2005JD006272.

Tans, P., and R. Keeling (2015), http://www.esrl.noaa.gov/gmd/ccgg/trends/ (last accessed 31 August 2016).

Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res., 106 (D7), 7183-7192.

Taylor, K. E., R. Stouffer, and G. A. Meehl (2012), An overview of CMIP5 and the experiment design, Bull. Amer. Meteor. Soc., 93, 485-498.

Teixeira, J., D. Waliser, R. Ferraro, P. Gleckler, T. Lee, and G. Potter (2014), Satellite Observations for CMIP5: The Genesis of Obs4MIPs, Bull. Amer. Meteor. Soc, 95, 1329-1334.

Tian, B., E. J. Fetzer, B. H. Kahn, J. Teixeira, E. Manning, and T. Hearty (2013), Evaluating CMIP5 models using AIRS tropospheric air temperature and specific humidity climatology, J. Geophys. Res., 118, D50117, 114-134.

Tsendbazar, N.-E., S. de Bruin, S. Fritz, and M. Herold (2015), Spatial Accuracy Assessment and Integration of Global Land Cover Datasets, Remote Sens, 7, 15804-15821.

Van Roozendael, M., et al. (2015), ESA Ozone Climate Change Initiative (ESA Ozone_cci) data: L3S-TC-MERGED-DLR v1.0 via Centre for Environmental Data Analysis, 2 August 2013.

Voldoire, A., E. Sanchez-Gomez, D. Salas y Mélia, B. Decharme, C. Cassou, S.Sénési, S. Valcke, I. Beau, A. Alias, M. Chevallier, M. Déqué, J. Deshayes, H. Douville, E. Fernandez, G. Madec, E. Maisonnave , M.-P. Moine, S. Planton, D.Saint-Martin, S. Szopa, S. Tyteca, R. Alkama, S. Belamari, A. Braun, L. Coquart, and F. Chauvin (2013), The CNRM-CM5.1 global climate model: description and basic evaluation, Clim. Dyn., 40, 2091, doi: 10.1007/s00382-011-1259-y.

Volodin, E. M., N. A. Dianskii, and A. V. Gusev (2010), Climate model INMCM4.0, Izvestia RAS, Atmospheric and Oceanic Physics, 46(4), 448-466, doi: 10.1134/S000143381004002X.

Wagner, W., W. Dorigo, R. de Jeu, D. Fernandez, J. Benveniste, E. Haas, M. Ertl (2012), Fusion of active and passive microwave observations to create an Essential Climate Variable data record on soil moisture, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume I-7, 2012, XXII ISPRS Congress, 25 August - 01 September 2012, 315-231, Melbourne, Australia.

Walsh, J. E., W. L. Chapman, and F. Fetterer (2015), Gridded Monthly Sea Ice Extent and Concentration, 1850 Onward, Version 1, Boulder, Colorado, USA.

Wang, S., X. Mo, S. Liu, Z. Lin, and S. Hu (2016), Validation and trend analysis of ECV soil moisture data on cropland in North China Plain during 1981-2010, Int. J. Appl. Earth Observ. Geoinform., 48, 110-121, doi: 10.1016/j.jag.2015.10.010.

Watanabe, M., T. Suzuki, R. O'ishi, Y. Komuro, S. Watanabe, S. Emori, T. Takemura, M. Chikira, T. Ogura, M. Sekiguchi, K. Takata, D. Yamazaki, T. Yokohata, T. Nozawa, H. Hasumi, H. Tatebe, and M. Kimoto (2010), Improved climate simulation by MIROC5: Mean states, variability, and climate sensitivity, J. Climate, 23, 6312-6335.

Watanabe, S., T. Hajima, K. Sudo, T. Nagashima, T. Takemura, H. Okajima, T. Nozawa, H. Kawase, M. Abe, T. Yokohata, T. Ise, H. Sato, E. Kato, K. Takata, S. Emori, and M. Kawamiya (2011), MIROC-ESM 2010: model description and basic results of CMIP5-20c3m experiments, Geosci. Model Dev., 4, 845-872.

Willet, K. M., A. J. Dolman, D. F. Hurst, J. Rennie, and P. W. Thorne (2014), [Global Climate] Overview [in: "State of the Climate 2013'], Bull. Am. Meteor. Soc., 95(7), S5-S8.

Wise, M., K. Calvin, A. Thomson, L. Clarke, B. Bond-Lamberty, R. Sands, S. J. Smith, A. Janetos, and J. Edmonds (2009), Implications of limiting CO_2 concentrations for land use and energy, Science, 324, 1183-1186.

World Meteorological Organization (WMO) (2014), Scientific Assessment of Ozone Depletion: 2014, World Meteorological Organization, Global Ozone Research and Monitoring Project - Report No. 55, 416 pp., Geneva, Switzerland.

Wu, T.-W. (2012), A mass-flux cumulus parameterization scheme for large-scale models: Description and test with observations, Clim. Dyn., 38, 725-744.

Wu, T., R. Yu, F. Zhang, Z. Wang, M. Dong, L. Wang, X. Jin, D. Chen, L. Li (2010), The Beijing Climate Center for Atmospheric General Circulation Model (BCC-AGCM2.0.1): Description and its performance for the present-day climate, Clim. Dyn., 34, 123-147.

Wunch, D., G. Toon, J. Blavier, R. Waschenfelder, J. Notholt, B. Connor, D. Griffith, and V. SherlockSherlock (2011), The Total Carbon Column Observing Network, Philos. T. Roy. Soc. A, 369, 2087-2112, 2011, doi: 10.1098/rsta.2010.0240.

Wunch, D., G. C. Toon, V. Sherlock, N. M. Deutscher, X. Liu, D. G. Feist, and P. O. Wennberg (2015), The Total Carbon Column Observing Network's GGG2014 Data Version. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, available at: doi: 10.14291/tccon.ggg2014.documentation.R0/1221662 (last accessed 27 November 2015).

Yukimoto, S., H. Yoshimura, M. Hosaka, T. Sakami, H. Tsujino, M. Hirabara, T. Y. Tanaka, M. Deushi, A. Obata, H. Nakano, Y. Adachi, E. Shindo, S. Yabu, T. Ose, and A. Kitoh (2011), Meteorological Research Institute-Earth System Model Version 1 (MRI-ESM1) - Model Description, Technical Report of the Meteorological Research Institute, 64, 83pp.

Appendix – list of abbreviations and acronyms

AATSR	Advanced Along-Track Scanning Radiometer
ACE	Atmospheric Chemistry Experiment
ADV	Advanced along-track scanning radiometer (AATSR) Dual-View
AEROCOM	Aerosol Comparisons between Observations and Models
AERONET	AErosol RObotic NETwork
AIRS	Atmospheric Infrared Sounder
AMSR-E	Advanced Microwave Scanning Radiometer - Earth Observing System
ana4MIPs	analyses for Model Intercomparison Projects
AOD	Aerosol Optical Depth
ATBD	Algorithm Theoretical Basis Documents
ATSR(-2)	Along-Track Scanning Radiometers (2)
AVHRR	Advanced Very High Resolution Radiometer
BDBP	Binary Data Base of Profiles
CALIOP	Cloud-Aerosol Lidar with Orthogonal Polarization
CC4CL	Community Cloud retrieval for CLimate
CCI	Climate Change Initiative
CERES	Clouds and the Earth's Radiant Energy System
CLARA-A2	CLoud, Albedo and RAdiation dataset, AVHRR-based
CMIP5/6	Coupled Model Intercomparison Project Phase 5/6
CMOR	Climate Model Output Rewriter
CMUG	Climate Modelling User Group
CO_2	carbon dioxide
CRESCENDO	Coordinated Research in Earth Systems and Climate: Experiments, kNowledge, Dissemination and Outreach
CVDP	Climate Variability Diagnostics Package
DECK	Diagnostic, Evaluation and Characterization of Klima
DGVM	Dynamic Global Vegetation Model
DJF	December, January, February
DU	Dobson Unit
EBAF	Energy Balanced And Filled
ECV	Essential Climate Variable
ENVISAT	Environmental Satellite
ENSO	El Niño Southern Oscillation
ERS-2	European Remote Sensing Satellite 2
ESA	European Space Agency
ESGF	Earth System Grid Federation
ESM	Earth System Model
ESMValTool	Earth System Model Evaluation Tool
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites

FMI	Finnish Meteorological Institute
FTS	Fourier Transform Spectrometer
GCOS	Global Climate Observing System
GHG	greenhouse gases
GOME/-2	Global Ozone Monitoring Experiment / 2
GOMOS	Global Ozone Monitoring by Occultation of Stars
GOSAT	Greenhouse gases observing satellite
GPCP	Global Precipitation Climatology Project
GRACE	Gravity Recovery and Climate Experiment
HadISST	Hadley Centre Sea Ice and Sea Surface Temperature data set
IASI	Infrared Atmospheric Sounding Interferometer
IGAC	International Global Atmospheric Chemistry
ITCZ	Inter-Tropical Convergence Zone
JAG	International Journal of Applied Earth Observation and Geoinformation
JJA	June, July, August
L2/3/4	Level 2/3/4
MAN	Marine Aerosol Network
MERIS	MEdium Resolution Imaging Spectrometer
Metop	Meteorological Operational Satellite
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding
MMM	Multi-Model Mean
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NDVI	Normalized Differenced Vegetation Index
NH	Northern Hemisphere
NIR	near-infrared
NIWA	National Institute of Water and Atmospheric Research
NOAA	National Oceanic and Atmospheric Administration
NSIDC	National Snow and Ice Data Center
obs4MIPs	observations for Model Intercomparison Projects
OMI	Ozone Monitoring Instrument
ORAC	Oxford-Rutherford Appleton Laboratory (RAL) Aerosol and Clouds
OSI SAF	Satellite Application Facility on Ocean and Sea Ice
OSIRIS	Optical Spectrograph and InfraRed Imaging System
OSTIA	Operational Sea surface Temperature and sea-Ice Analysis
PATMOS-x	Pathfinder Atmospheres Extended
PCMDI	Program for Climate Model Diagnostics and Intercomparison
PFT	Plant Functional Type

1798	POLDER	POLarization and Directionality of the Earth's Reflectances
1799	PMP	PCMDI metrics package
1800	RAL	Rutherford Appleton Laboratory
1801	RCP	Representative Concentration Pathways
1802	RMSD	relative space-time Root-Mean-Square Deviation
1803	RMSE	Root-Mean-Square Error
1804	SCIAMACHY	Scanning Imaging Absorption Spectrometer for Atmospheric CHartographY
1805	SECM	Simple Empirical CO_2 Model
1806	SH	Southern Hemisphere
1807	SI	Sea Ice
1808	SMR	Sub-Millimetre Radiometer
1809	SPARC	Stratospheric Processes and their Role in Climate
1810	SPOT	Satellite Pour l'Observation de la Terre
1811	SSM/I	Special Sensor Microwave Imager
1812	SWIR	short-wave infrared
1813	SST	Sea Surface Temperature
1814	SU	Swansea University retrieval algorithm
1815	TANSO	Thermal And Near-infrared Sensor for carbon Observation
1816	TCCON	Total Carbon Column Observation Network
1817	UV	ultraviolet
1818	Vis	visible spectral range
1819	WCRP	World Climate Research Programme
1820	WMO	World Meteorological Organization