CMUG Deliverable

Number:	**D3.1v_1d**
Due date:	**June 2013**
Submission date:	**30 Sept 2013**
Version:	**1.0**

Climate Modelling User Group

Deliverable 3.1v_1d

Technical note on Demonstration of CMF functionality for the assessment of ozone, aerosol and soil moisture

Centres providing input: ECMWF

Version nr.	Date	Status
0.1	06 Jun 2013	Initial Template RD
0.2	25 Jul 2013	Draft distributed to CMUG
0.3	26 Aug 2013	Includes feedbacks from CMUG and O3 CCI
0.4	23 Sep 2013	Addressed ESA comments
1.0	30 Sep 2013	Submit to ESA

Deliverable 3.1v_1d

Technical note

Demonstration of CMF functionality for the assessment of ozone, aerosol and soil moisture

Table of Contents

Demonstration of CMF functionality for the assessment of ozone, aerosol and soil moisture

1. Purpose and scope of the Technical note

ECMWF has developed and put at the disposal of CCI a prototype of an interactive interface for assessing low-frequency (multi-year) variability of statistical averages (typically monthly/regional means) aiming at evaluating the long-term homogeneity and consistency of CDRs.

This document provides a description of this interface, referred to as Climate Monitoring Facility or simply CMF, and a demonstration of its functionality applied to the assessment of the first version of the CCI ozone, aerosol and soil moisture datasets.

2. Brief introduction to CMF and its database

Data quality is a constant preoccupation of any science application. The quality of the conclusions drawn out of any study depends on how good the data records on which the study was based were. In some cases, the concern about data quality has to do with the "there and then". E.g. in Numerical Weather Prediction the usage of a given data set is constantly reviewed based on its quality at that given time and its ability to produce a positive impact on the resulting daily analyses and forecasts. In other applications, such as climate change related studies, not only is data quality at any time critical to provide accurate projections of the future climate, datasets must also be homogeneous over the whole period of availability, and they should be consistent with related variables, e.g. to fully represent biogeochemical cycles. Assessing such long term homogeneity and across-variable consistency is not always possible or easy to do.

The Climate Monitoring Facility (CMF) is an interactive interface that facilitates the evaluation of the multi-year variability of statistical averages computed from a variety of climate data records (CDRs). Thus, the tool is designed precisely to evaluate the long-term homogeneity and perform a consistency analysis of CDRs. An overview of the various ways CMF can aid the long-term data quality assessment of CDRs was described in detail in Part 2 of CMUG (2013b), and thus not repeated in the present document.

Disclaimer:

Like any other tool, the CMF should be used for applications it was designed for (i.e. monitoring and assessing the low-frequency, multi-year variability of regional averages). The comparisons it facilitates are based on pre-calculated statistical regional averages of monthly mean data. As differences may occur in the data coverage of different data streams used to produce those averages, cautioun should be used when assessing their comparison.

The design of this tool was also described in detail by CMUG (2013a). Figure 1, adapted from figure 3 of CMUG (2013a), schematically reminds the reader of the three main parts that constitute CMF. A potential user needs to be familiar with the web interface and be aware of the CMF database (CMFDb) content to be able to use CMF. A detailed description of the CMFDb content was provided by CMUG (2013b) in their Appendix A. Since then, the CMFDb content has been further extended to include additional geophysical datasets and data streams, as listed in tables 1 (for reanalyses) and 2 (for observational datasets) of the present document.

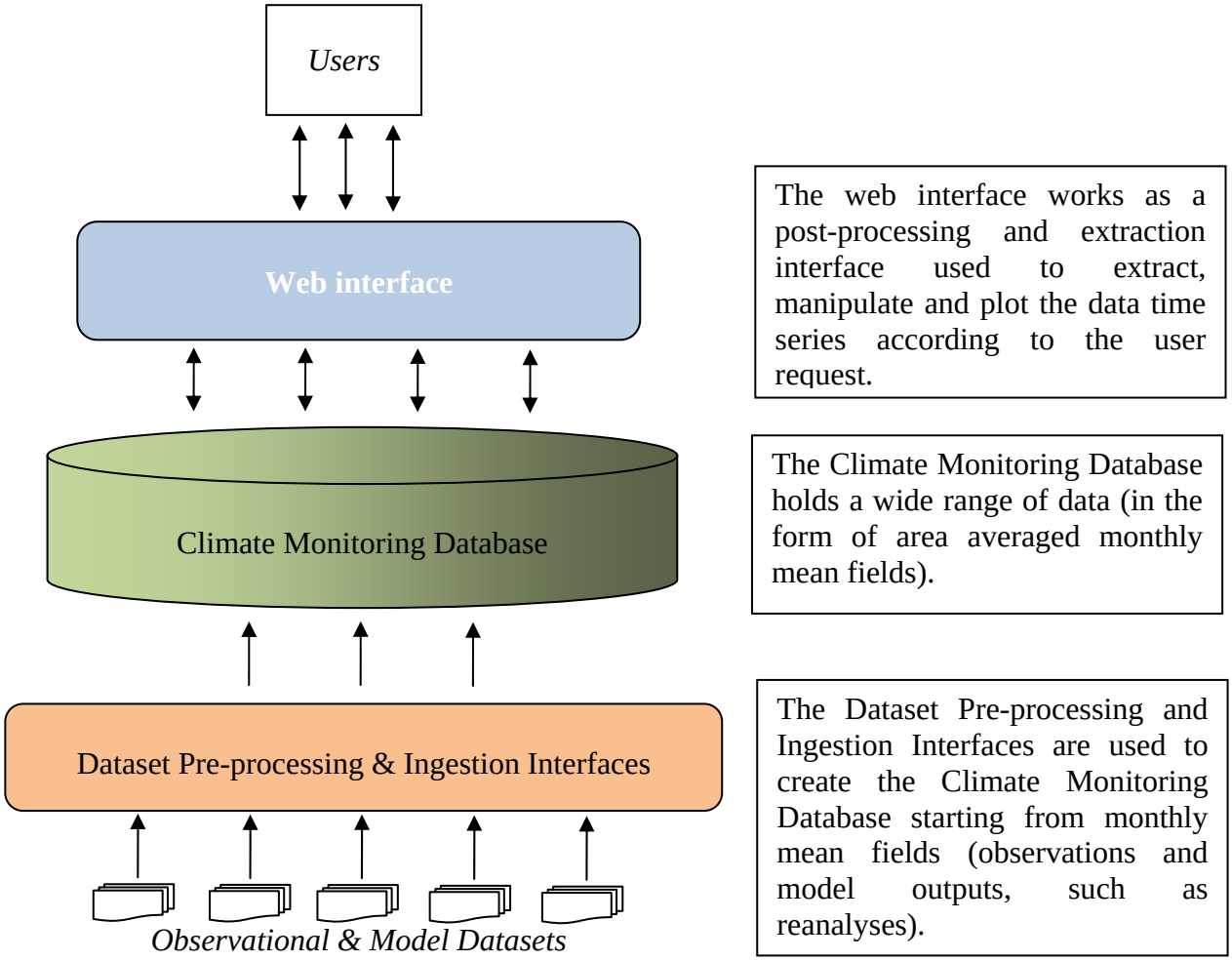

Fig 1: Schematic overview of the main components of the Climate Monitoring Facility adapted from figure 3 of CMUG (2013a).

Name	Data streams in CMFDb	Description
SWVL1	MERRA/Land	Soil Water Vapour Level 1 (0-2cm)
SWVL1-4	ERA-40, ERA-Interim	Soil Water Vapour Level 1 to 4
SWVL1	ERA-Interim/Land	Soil Water Vapour Level 1 (0-7cm)
TCO3	MACC Reanalysis	Total Column Ozone
O3	MACC Reanalysis	Ozone mass mixing ratio
TCCO2	MACC Reanalysis	Total column carbon dioxide
TCCH4	MACC Reanalysis	Total column methane
AOD469	MACC Reanalysis	Total AOD at 469nm
AOD550	MACC Reanalysis	Total AOD at 550nm
AOD670	MACC Reanalysis	Total AOD at 670nm
AOD865	MACC Reanalysis	Total AOD at 865nm
AOD1640	MACC Reanalysis	Total AOD at 1640nm
AFM	MACC Reanalysis	Aerosol Fine Mode
ACM	MACC Reanalysis	Aerosol Coarse Mode

Table 1: Additional geophysical parameters available in the CMFDb since the release of CMUG (2013b) from reanalysis streams.

Name	Data streams in CMFDb	Description
SWVL1	CCI	Soil Water Vapour Level 1 (0-2cm)
TCO3	CCI	Total Column Ozone
O3	CCI	Ozone mass mixing ratio from Nadir profile
O3	LPCCI	Ozone mass mixing ratio from Limb profile
AOD550	SU40, ADV142, ORAC202[#]	Total AOD at 550nm from CCI Aerosol
AOD659	SU40, ADV142	Total AOD at 659nm from CCI Aerosol
AOD865	SU40	Total AOD at 865nm from CCI Aerosol
AOD870	ORAC202	Total AOD at 870nm from CCI Aerosol
AOD1610	SU40, ADV142	Total AOD at 1610nm from CCI Aerosol

Table 2: Like in table 1 but from CCI data. [#]SU40 stands for Swansea University algorithm, version 4.0; ADV142 is the aerosol algorithm available at FMI, version 1.42; ORAC202 is the Oxford-RAL aerosol algorithm, version 2.02.

It is noted that in some cases, e.g. for Aerosol CCI data, different streams had to be created, namely SU40, ADV142 and ORAC202 (see table 2), each referring to a specific aerosol retrieval scheme as a final algorithm decision has not yet been made. Creating multiple streams like in the above example was done mainly to facilitate the present assessment. However, it has to be appreciated that an indiscriminate ingestion of any available dataset cannot be sustainable in the long term. Decisions regarding the data ingestion will need to be made based on a preliminary quality assessment. Possible ways to allow users in the future to compare their interim datasets against either consolidated/previous data versions or different data streams before being ingested in the CMFDb are under study and subject to future development of the current prototype.

In this respect, an important consideration to bear in mind is that in the future (subject to funding) the CMF and its CMFDb are expected to be made publicly available to the international climate community and will constitute an element for promoting the CCI data usage to other international bodies and in other programmes (e.g. ERA-CLIM, and CORE-CLIMAX). Thus, the data ingestion will need to be more carefully regulated.

3. A note on the method

Section 4 discusses the outcome of the comparisons between a selected set of CCI data records with their model equivalent obtained from several reanalysis streams as provided by CMF. The CMFDb was populated with area averaged monthly mean data. In addition, an *area characteristic* (monthly mean) *observation standard deviation* was added as additional parameter in CMFDb. Such a characteristic standard deviation over a given geographical area, *a*, and at time *t*, $\sigma_{a,t}$, was computed as an uncorrected sample standard deviation from the gridded monthly mean standard deviations, i.e.:

$$\sigma_{a,t} = \sqrt{\frac{1}{N_a}\sum_i \sigma_{i,t}^2} \qquad (1)$$

where N_a is the number of grid points over the region, and $\sigma_{i,t}$ is the standard deviation at time *t* of the i-th grid point.

How can the quality of these area characteristic standard deviations be assessed?

A naïve answer to the above question could be to compare the observational standard deviations with the model uncertainties in a similar fashion way the mean values are compared. However, two issues arise in that. First, most reanalysis productions are based on variational data assimilation techniques (Three and Four Dimensional Variational Data Assimilation, 3D-Var and 4D-Var), in which the analysis uncertainties are only implicitly defined and therefore not immediately available as a model output (e.g. Fisher, 2003). The second concerns with the separation between the uncertainty random and systematic components, the latter being essentially related to (model and observational) biases.

An approach that has already been used successfully consists in generating an *ensemble* of data assimilation simulations. The members of the ensemble represent different realizations of the same model initialised from slightly different, but equally probable initial conditions. The spread of such an ensemble can be used as a proxy of the internal climate variability associated with a given variable, more specifically it can be associated to the actual analysis uncertainty (e.g. Houtekamer and Mitchell, 2001; Evensen, 2003). Thus, it can be used to estimate the uncertainties when not available or indeed when available to assess their quality. It is important to note that as the model bias and any other model idiosyncrasy should have similar effects on all members of the ensemble, their contribution to the ensemble spread should be negligible.

As part of the ERA-CLIM project, ECMWF has run a number of pilot reanalyses covering different temporal lengths and at various resolutions. Among those pilot reanalyses, an ensemble of 4D-Var data assimilation was run at low resolution (T159 corresponding to approximately 125km grid resolution) from the beginning of the 20th century onwards. These pilot realizations are mainly used to support the development of the coupled atmosphere-ocean reanalysis system to be used in future reanalysis, and the estimation of the data uncertainty and data homogenization, particularly important in the case of the early 20th century in-situ observations. Among the three variables this document focuses on (i.e. ozone, aerosols, and soil moisture), only ozone data was available from an early 10 member

ensemble of 4D-Var reanalyses. The ensemble spread of the total column and vertically resolved ozone analyses will be used to assess the quality of the ozone data standard deviations and illustrate the method.

In the cases of aerosol and soil moisture, an ensemble of reanalyses was not available. In these cases, either the available reanalysis anomaly[1] (hereafter simply referred to as anomaly) or the residuals between observation and their reanalysis equivalent is compared to the observation standard deviations.

4. Assessment of the CCI ozone, aerosol and soil moisture data

The present assessment is based on the comparisons of the CCI L3 data products with their model equivalent obtained from various available reanalyses (as detailed below). As these reanalyses contain a synthesis of all the good qualities the CCI datasets should have, they can be regarded as precursors of the CCI datasets.

For each ECV and whenever available, the merged dataset was used as obtained from the official data servers as advertised in the CCI teams' web-pages, i.e.:

ECV	Archive	Data access
Ozone	ftp://cci_web@ftp-ae.oma.be/esacci/ozone/	No registration required
Aerosol	http://www.icare.univ-lille1.fr/archive	Registration no longer required
Soil Moisture	ftp.ipf.tuwien.ac.at	Registration required

4.1. Ozone

Ozone_CCI aims at generating three lines of production for Total Column Ozone (TCO_3) and ozone profiles (NP O_3) from nadir UV backscatter sensors, as well as ozone profiles from available limb and occultation sensors (LP O_3).

For each of the three production lines, both L2 and L3 gridded data are produced. In the latter case depending on the production stream, a merged product can be available. As mentioned above the present study focussed on the L3 gridded fields only. Table 3 lists these products and summarises their availability for this study.

It is important to note that the three main ozone reanalyses (ERA-Interim [Dee et al, 2011], MACC [Inness et al 2013] and JRA-25 [Onogi et al, 2005]) were obtained with different ozone constraints, thus their quality can be very different. For example, the Japanese reanalysis (JRA-25) only assimilated ozone retrievals from the Total Ozone Monitoring Sounder (TOMS) on board the Nimbus-7 and Earth Probe satellites. Depending on the period and availability, the ERA-Interim and MACC ozone reanalyses were constrained by a larger set of ozone data, including TOMS, OMI (acronyms not defined in the text are provided in Appendix B), and SCIAMACHY total columns, as well as SBUV, GOME-1, MLS and MIPAS ozone profiles. Details on the data version and data producer, as well as on the exact period during which each ozone product was assimilated can be found in Dragani (2011) for ERA-Interim and in Inness et al (2013) for MACC.

[1] It is argued that for a dataset to be useful it should have an uncertainty smaller than such an anomaly.

Dataset	Product	Version	Acronym	Availability	Period assessed	Reanalysis streams[#]
Total column	Merged	fv0100	TCO_3	Apr 1996-Jun 2011	Apr 1996-Jun 2011	ERA-Interim MACC JRA-25 ERA-20C (10 members)
Nadir profile	N/A	fv0001	NP O_3	1997, 2008	1997, 2008	ERA-Interim MACC ERA-20C (10 members)
Limb profiles	Zonal mean, Merged	fv0002	LP O_3	2007-2008	2007-2008	ERA-Interim MACC ERA-20C (10 members)

Table 3: List of the assessed CCI ozone products and reanalysis streams used for each one. [#]Information about different reanalysis streams were provided in CMUG (2013b). MACC stands for Monitoring Atmospheric Composition and Climate. JRA-25 is the 25 year long Japanese reanalysis. ERA-20C is the reanalysis for the 20th century forced just with surface pressure data that is run as part of the ERA-CLIM project led by ECMWF, and that it will be used to assess the observation uncertainties.

It is important to note that the three main ozone reanalyses (ERA-Interim [Dee et al, 2011], MACC [Inness et al 2013] and JRA-25 [Onogi et al, 2005]) were obtained with different ozone constraints, thus their quality can be very different. For example, the Japanese reanalysis (JRA-25) only assimilated ozone retrievals from the Total Ozone Monitoring Sounder (TOMS) on board the Nimbus-7 and Earth Probe satellites. Depending on the period and availability, the ERA-Interim and MACC ozone reanalyses were constrained by a larger set of ozone data, including TOMS, OMI (acronyms not defined in the text are provided in Appendix B), and SCIAMACHY total columns, as well as SBUV, GOME-1, MLS and MIPAS ozone profiles. Details on the data version and data producer, as well as on the exact period during which each ozone product was assimilated can be found in Dragani (2011) for ERA-Interim and in Inness et al (2013) for MACC.

4.1.1. Differences between the ERA-Interim and MACC ozone reanalyses

Dee et al (2011) and Dragani (2011) discussed the ERA-Interim (1979-present) ozone system and quality of the ozone reanalyses. Flemming et al (2011) and Inness et al (2013) reported on the MACC ozone system and reanalyses (2003-2012). Here, we summarises the differences between the two systems in order to understand and interpret the comparisons with the CCI ozone products.

1) Ozone model:

In ERA-Interim, the ozone model is based on an updated version of the Cariolle and Déqué (1986) scheme with an additional term to parameterize the heterogeneous chemistry at high latitudes (Dragani, 2011). This scheme is normally adequate to describe changes in the

stratospheric ozone, but not for the tropospheric ozone component. The latter in the ERA-Interim reanalyses is essentially constrained indirectly as the residuals between the constraint on the total column ozone and its stratospheric component.

In MACC, the assimilation system for the chemical reactive species, such as ozone, is coupled to a full chemistry transport model (CTM, Flemming et al, 2011; Inness et al, 2013), which provides emissions, deposition, and chemical tendencies for the species included in the system (i.e. O_3, CO, NOx, SO_2 and HCHO). The coupling with a full CTM provides a better characterisation of the tropospheric ozone with impact in the lowermost Stratosphere.

It is anticipated that the difference in the ozone model will affect more the comparisons with the CCI ozone profiles, particularly in the troposphere and lowermost stratosphere and it will be negligible in the comparisons with the CCI total column ozone.

2) Ozone Bias Correction:

An ozone bias correction was applied to the ozone data assimilated in the MACC reanalysis implemented as in the ECMWF IFS system (Dragani, 2009). This ozone bias correction is based on an updated version of the adaptive scheme (VarBC) that was first applied to correct the radiance biases and used to that end in the ERA-Interim reanalysis (Auligné et al, 2007; Dee and Uppala 2009). To avoid long-term drifts in the ozone field, the ozone bias correction scheme requires unbiased reference ozone to use as an anchor according to which all ozone observations can be corrected. Dragani (2009) and Dragani and McNally (2013) showed that the use of the ozone data provided by the Solar Backscatter Ultra Violet (SBUV) 2 instruments to anchor the ozone VarBC was necessary and sufficient to prevent any long-term drift of the ozone field. Dragani (2009) also showed that the quality of the ozone analyses could be improved by using the SBUV data as anchor for VarBC. It is pointed out that these are not directly the 21-level ozone profiles provided by NOAA, but a six ozone layer product obtained by merging together contiguous levels to minimise the potentially negative impact of unaccounted observation error correlations. The six layers are defined as follows: 0.1-1 hPa, 1-2 hPa, 2-4 hPa, 4-8 hPa, 8-16 hPa and 16 hPa-surface.

In contrast, an ozone bias correction became available after the ERA-Interim production had started, and as a reanalysis production is run with a fixed version of the data assimilation system it was no possible to benefit from it. Thus all ozone observations were assimilated in ERA-Interim without any correction.

The regions of the atmosphere and latitudinal bands where the use of VarBC could be more evident are likely the levels above the ozone mixing ration maximum at 10 hPa, particularly in the tropics and at midlatitudes. Ozone data retrieved from Backscatter Ultra Violet (BUV) instruments normally have a larger weight and impact on the ozone analyses at midlatitudes and in tropics than at high latitudes, as a consequence of smaller errors. This is particularly the case in the region of the atmosphere near the ozone maximum and at levels just above it where the BUV instruments are highly sensitive. In the ERA-Interim production, because no bias correction was applied the weight of any ozone observation is determined by the observation error and by the data volume. The SBUV data volume is limited compared with

that of other ozone products that were assimilated at the same time, for example OMI and SCIAMACHY TCO3. In contrast, in the MACC production, despite having the same low data volume, the weight the SBUV ozone data were given in the assimilation was higher than in ERA-Interim in virtue of the fact that it was used to anchor the ozone VarBC.

3) Differences in the assimilated ozone observations:

The ERA-Interim and MACC ozone reanalyses benefitted from the assimilation of almost the same ozone observations with a few small differences. The MACC reanalysis assimilated GOME-1 ozone profiles (retrieved by RAL) from the production start (Jan 2003) to June 2003 when the ERS-2 tape recorder broke down reducing strongly the data availability and coverage. The assimilation of these data in ERA-Interim was carried out only until December 2002. Thus, some differences may be found in the first half of 2003 between the two products.

Both reanalyses benefitted from the assimilation of MLS ozone profiles (Froidevaux *et al.*, 2008). Like any other limb instrument, MLS delivers very accurate ozone profiles that when assimilated can substantially improve the stratospheric distribution of the ozone analyses. The impact of assimilating these observations can be found both in the height resolved ozone field and in the integrated columns. ERA-Interim started the assimilation of the MLS ozone profiles in early 2008, while in the MACC reanalyses they were used from August 2004 onwards.

A change in the ozone bias correction that affected the assimilation of MLS data was implemented on 1 January 2008 in the MACC reanalysis. The ozone VarBC was originally used as described above, anchored to the available SBUV observations. Inness et al (2013) found that in the MACC system, the use of low-vertical resolution SBUV observations as the only anchor to the ozone VarBC was not sufficient to constrain the finer vertical resolution of the MLS ozone products and prevent long-term drifts of the ozone field. For that reason, on 1 January 2008 the MLS ozone profiles started to be used to anchor the ozone VarBC in addition to the SBUV data.

4.1.2. Total Column Ozone

The total column ozone product assessed in the present document consists of the monthly mean gridded field produced by merging observations from several UV nadir instruments, namely GOME, OMI, SCIAMACHY and GOME-2.

One of the requirements, a dataset should fulfil to reach climate quality is the long-term homogeneity that can be limited by algorithm/instrument changes. This is particularly the case of a dataset obtained by merging data from different instruments. Figure 2 shows the time series of the CCI TCO_3 global mean anomaly (residuals from the mean) during the 16 years of data availability.

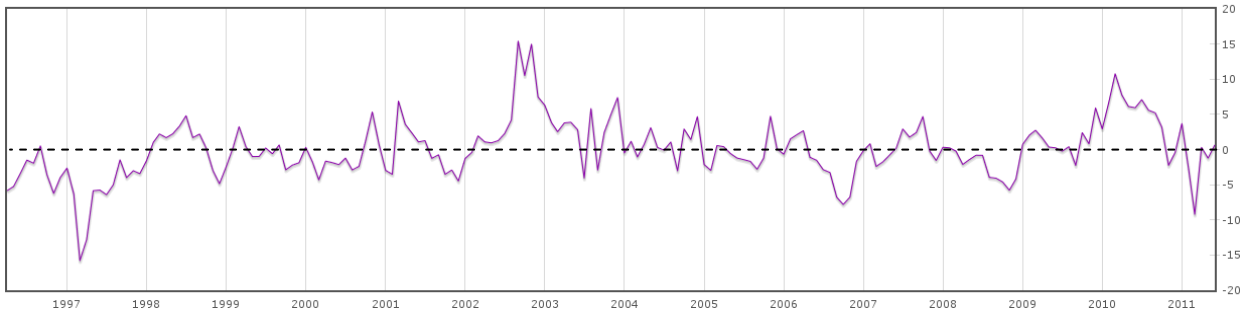

Fig 2: Time series of the global mean (CCI L3) TCO_3 anomaly. Data are in DU.

Figure 2 shows that the global anomaly for this product mostly oscillates around the zero value with variability normally within about ±5DU, which is about 1.5% change compared to a global mean value of about 300DU. During the available 16 years, two are the situations that show a sudden change in the timeseries. One is the negative deep in spring 1997; the second occurred in the second half of 2002. During the first period only the ERS-2 GOME data were used; while the second case occurred in mid-2002 when the SCIAMACHY nadir data were introduced. A confirmation that these two discontinuities were instrument related (e.g. due to an instrument anomaly) could not be found. A further investigation performed by the Ozone CCI team suggested that they instead may be the consequence of anomalous ozone conditions at high latitudes in the NH in March 1997 (where record low ozone values were observed) and in the SH in late summer 2002 (when an anomalous split of the polar vortex occurred) so intense to also affect the global mean values. Although the corresponding ERA-Interim ozone anomaly record does not exhibit these features, indeed the two sharp changes shown above could be a manifestation of a real atmospheric signal. It is noted that by design the CMF tool compares pre-calculated monthly mean area averages from several streams. This means that small differences between various streams can be obtained if and where the data streams present different data coverage, e.g. in this case at high latitudes. This should normally have little consequences on the global mean values unless extreme conditions are experiences over the regions with different data coverage. This could explain the discrepancy between model and observation averages found with the CMF.

Figure 3 presents the comparisons between the global mean CCI TCO3 data and their reanalysis equivalent. It is important to remind the reader that ozone retrievals from some of the instruments used to produce this CCI product were also assimilated in MACC and ERA-Interim. Thus, it is expected that these three datasets are in a reasonably good agreement. This is normally confirmed by figure 3. In contrast, the JRA-25 reanalyses are systematically lower than the other datasets. This comparison already offers an indication of the added value that the assimilation of this dataset could have in the Japanese system. It is argued that if the CCI TCO3 product was assimilated in the Japanese reanalysis, it is likely that the resulting ozone reanalyses exhibited higher values than in the JRA-25 production, and thus they would be more in line with those of ERA-Interim and MACC.

Focussing on the period from 2003 onwards - when both the European reanalyse were available -, it is clear that the ERA-Interim ozone analyses exhibit global mean values that are about 10DU lower than both the MACC and CCI TCO3, although they were all based on similar ozone observations (e.g. OMI and SCIAMACHY). Such a result is important as it

paves the way to further investigate the quality of the CCI TCO3 level 2 data with the view of replacing the SCIAMACHY and OMI TCO3 product that was used in ERA-Interim with the corresponding CCI datasets in future reanalysis productions. These could include for example the ERA-SAT run that will be produced as part of the ERA-CLIM project to replace ERA-Interim.

Figure 3 also shows that the annual cycle presented in the observations is much stronger than that exhibited by ERA-Interim, but in good agreement with the MACC reanalysis from 2003 onwards.

Comparisons at other latitudinal bands show general good agreement between the reanalysis streams and the CCI data with differences typically within the observation errors (not shown), except in the tropics (figure 4). Here, the CCI TCO3 has systematically lower values than ERA-Interim until 2002 and in recent years from 2009 onwards, and higher values than ERA-Interim during in the period 2004-2007. In periods during which limb ozone profiles from MIPAS and MLS were assimilated in ERA-Interim (indicated in figure 4), the level of agreement between the two datasets improves. The agreement with the MACC ozone reanalysis is generally good, with differences smaller than the observation errors.

As the level of agreement between the different datasets depend on the error bars of the Ozone CCI product, we have compared the estimated observational standard deviation with the ensemble spread of a 10-member ensemble as explained in section 3. Figures 5 and 6 present this comparison for the global mean time series and the averages over five latitudinal bands, namely the high and midlatitudes in both hemispheres, and the tropics.

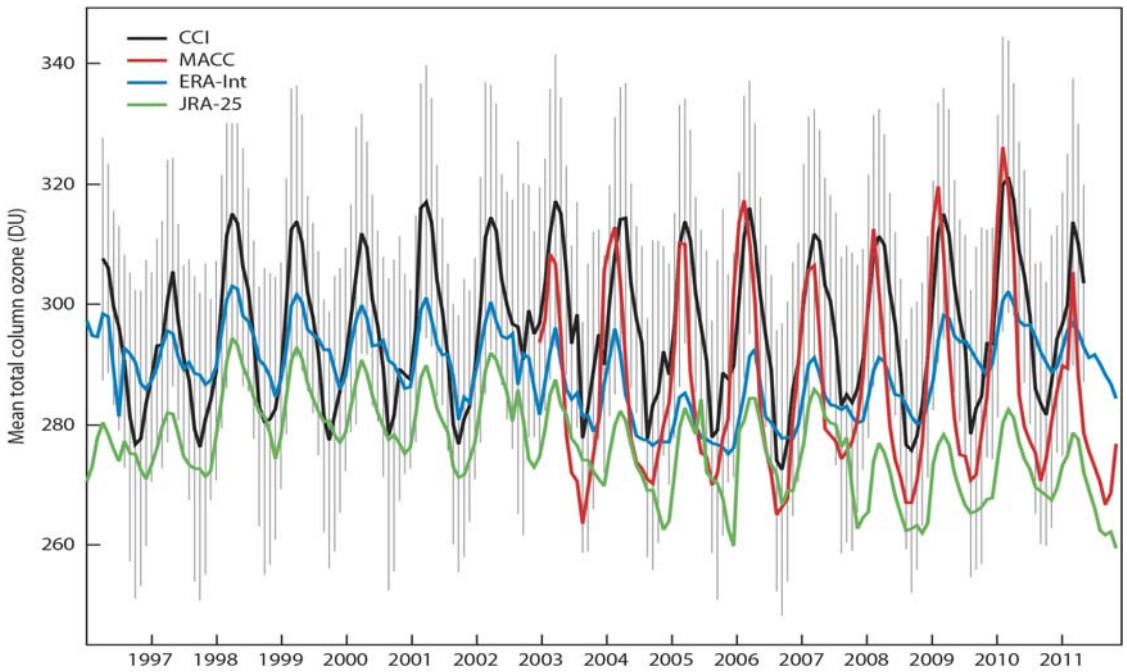

Fig 3: Time series of the global mean monthly mean TCO_3 from Ozone_CCI (black), ERA-Interim (blue), MACC (red) and JRA-25 (green). The vertical bars over-plotted to the CCI data are the standard deviations computed according to Eq (1). Data are in DU.

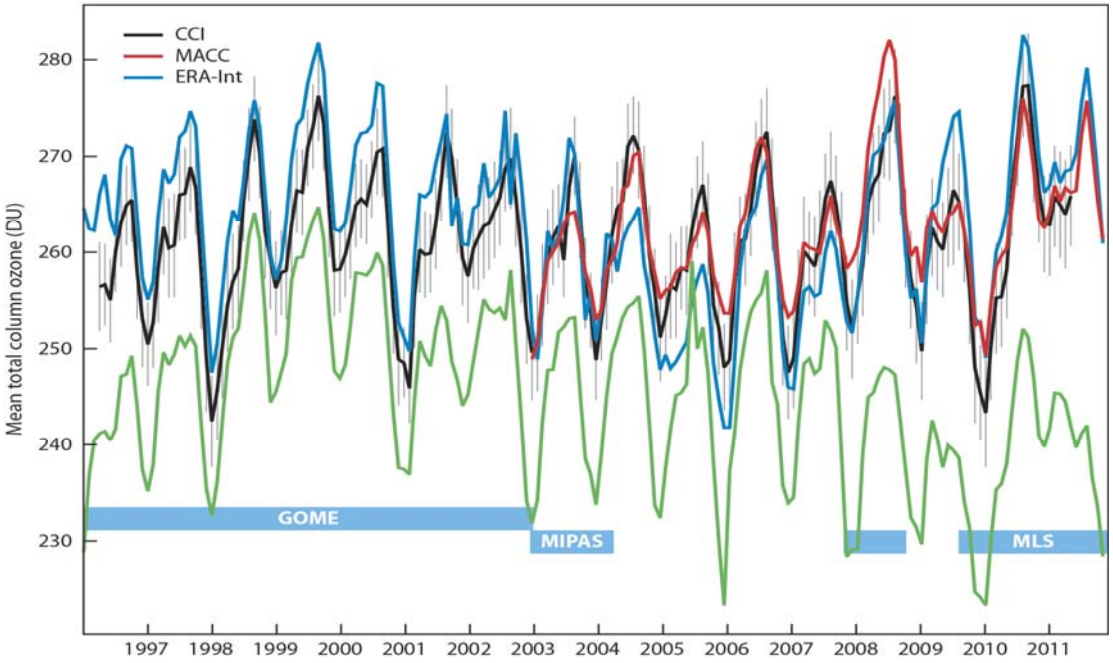

Fig 4: Time series of the monthly mean TCO_3 averaged over the tropics. Over imposed to the TCO3 timeline is the ERA-Interim ozone data usage that were assimilated on top of SBUV/2 ozone data and SCIAMACHY TCO3 from 2002 through December 2008.

In the global mean, the ensemble spread is about half the values of the observation error. By accounting for this overestimation, the level of agreement presented in figure 3 would not be as good as described above.

At mid and high latitudes in both hemispheres latitudes (see panels **b)** and **c)** of figures 5 and 6) where the mean TCO3 time series from the ERA-Interim and the MACC reanalyses well compare with the Ozone CCI product, the standard deviations are comparable with the ensemble spread. Thus, they do not change the conclusions given above.

In contrast, the tropical ozone errors (see panel **a)** of figure 6) appear much smaller and with less variability than estimated by the ensemble spread, which can be as large as 3 to 4 times the observation error. Indeed, if a larger error was considered, the level of agreement in the tropics between the Ozone CCI and the ERA-Interim and MACC reanalyses could be better than shown in figure 4.

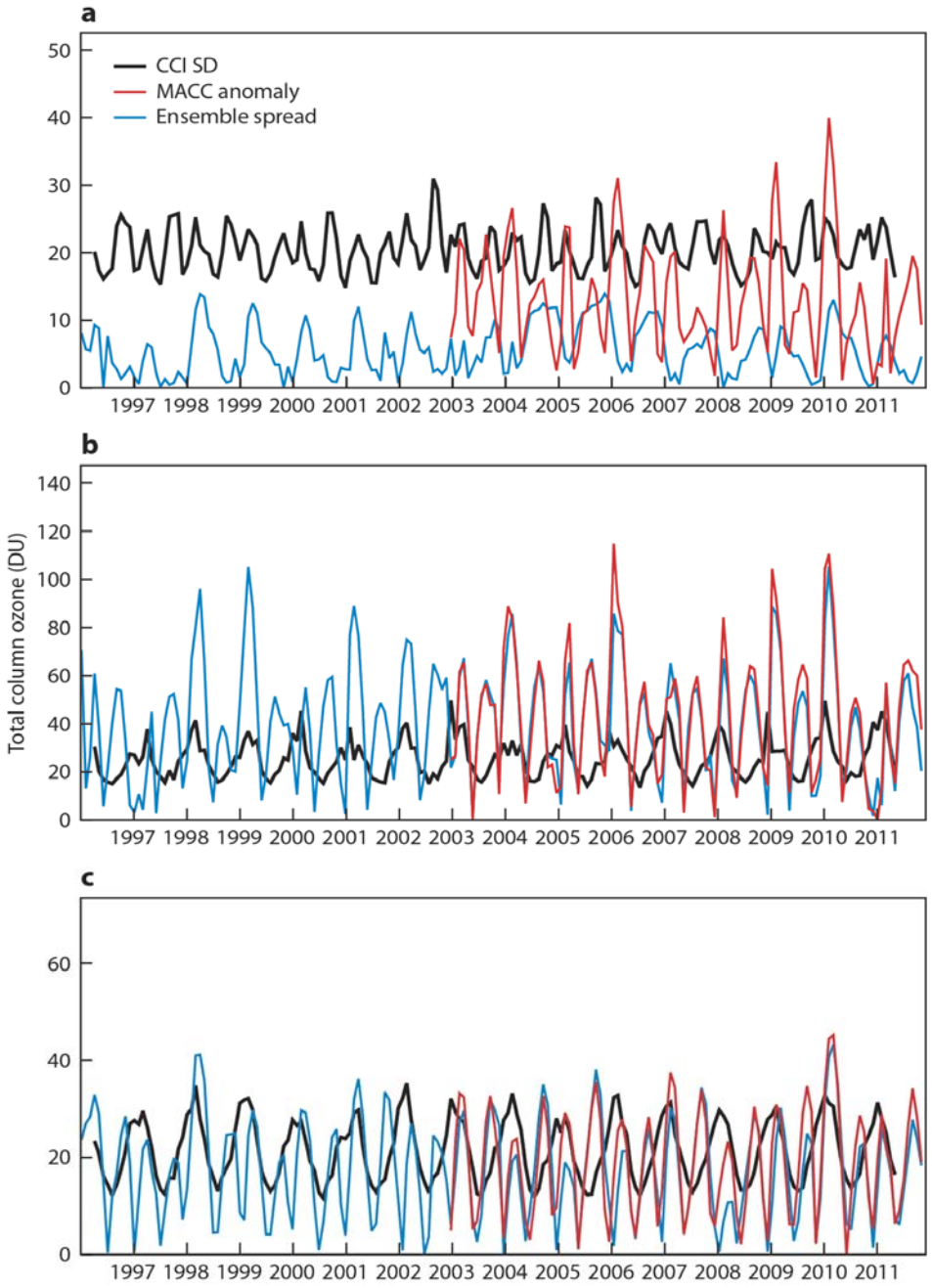

Fig 5: Mean TCO_3 standard deviation averaged over the globe (a), the high (b) and mid (c) latitudes in the NH.

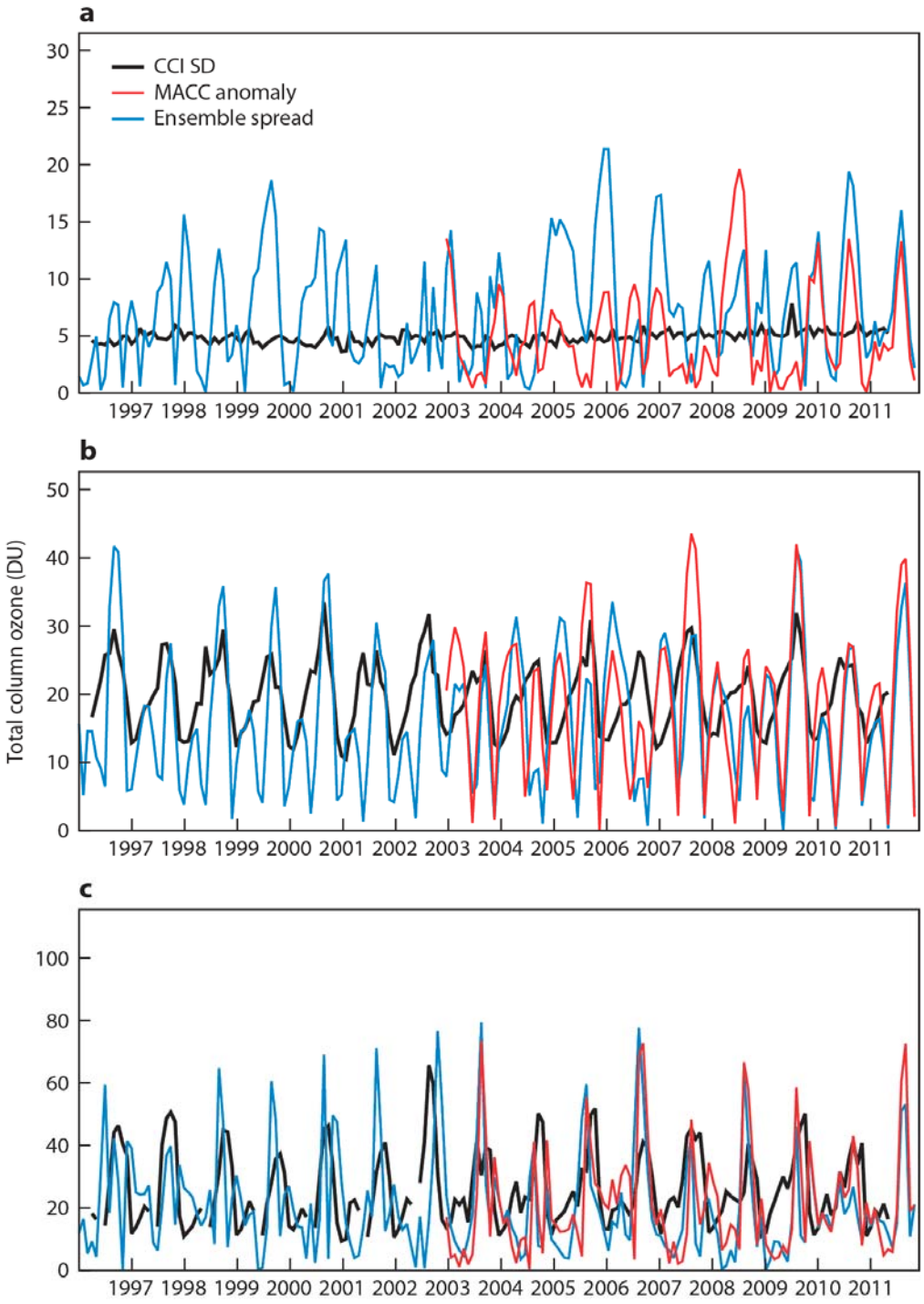

Fig 6: Like in fig 5 but for the tropics (a), the mid (b) and high (c) latitudes in the SH.

A region of particular interest for ozone is represented by the high latitudes, especially in the SH where in spring the ozone hole develops. The comparison between the two European reanalyses and the CCI TCO3 over Antarctica is presented in figure 7 for recent years (2004-2008). This is a period during which the MACC analyses benefitted from the assimilation of MLS ozone profiles, while in the ERA-Interim production its assimilation started at the beginning of 2008. The assimilation of these observations makes a large impact in this region,

particularly during (winter and) spring, when the limited (if any at all) information from UV sensors is often not adequate to properly constrained the ozone analyses. For clarity and to highlight the two situations, the periods of high (summer) and low (spring) ozone values are shown separately in figure 7 (panels a) and b)).

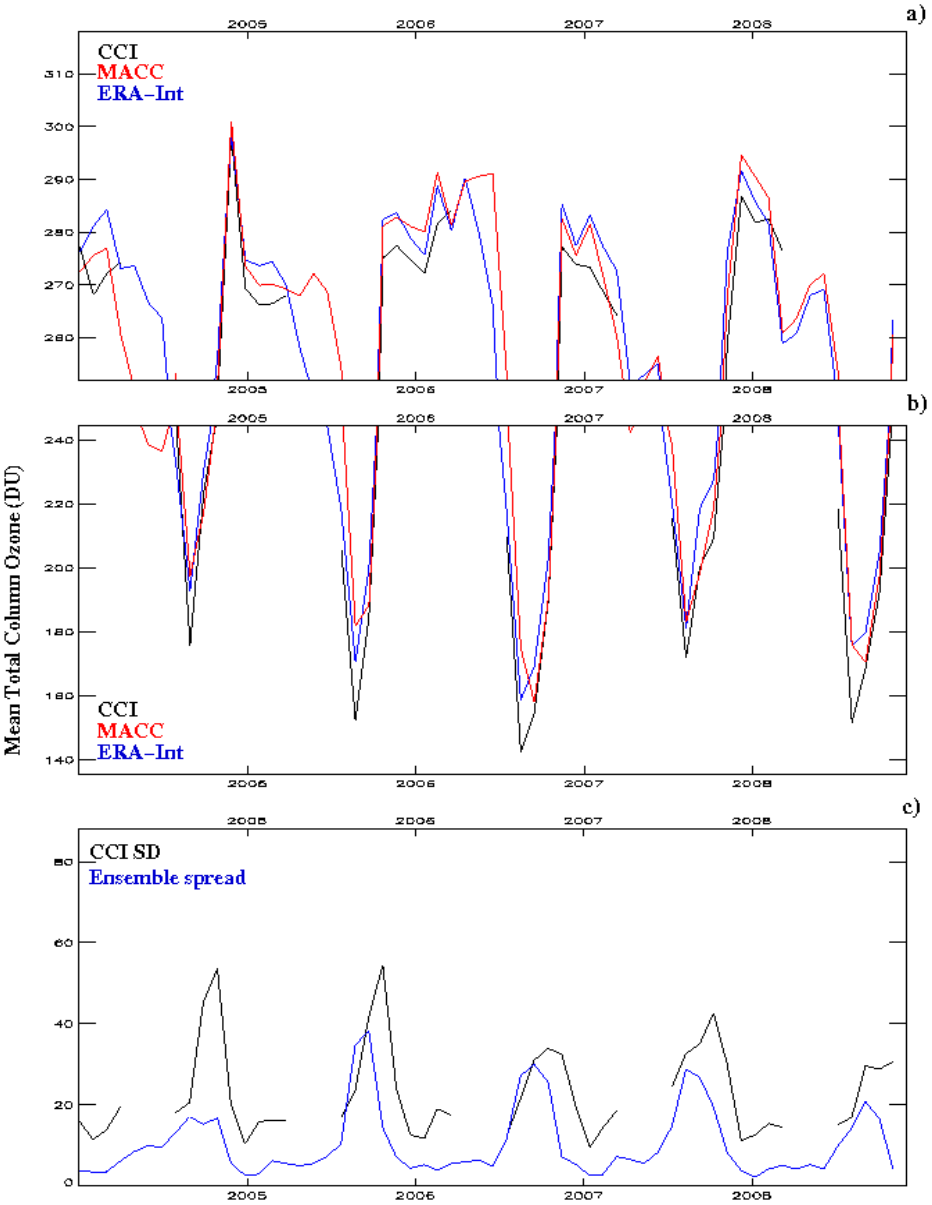

Fig 7: Comparisons between the CCI TCO3 dataset and its equivalent from MACC (red) and ERA-Interim (blue) over Antarctica during the period 2004-2008. Panel (a) highlights the comparison during the summer months; panel (b) focuses on the spring months when the ozone hole develops. Panel (c) shows the comparison between the observation standard deviation and the ensemble spread over Antarctica.

Despite the differences in the assimilated observations and in the ozone model (see section 4.1.1), it appears that the two reanalyses are in very good agreement both during the summer months - when the UV ozone data can provide an appropriate constraint to the ozone analyses - and in spring. In both cases, the differences in the two reanalyses are smaller than those between each of them and the CCI TCO3 dataset. If the MACC TCO3 is, arguably, the best

ozone reanalysis available and the L3 merged dataset can be regarded as a proxy of the L2 product, it may seem that the assimilation of the CCI dataset could produce little improvements in this region, and in some cases potentially degrade the reanalyses. The panel c) of figure 7 shows that over Antarctica the observation errors might be over-estimated.

Additional notes and feedbacks to the Ozone_CCI team:

A preliminary investigation of the data showed some unphysical values in the observation standard deviation of the TCO3 product, consisting of "0" (zeroes) and extremely small values (i.e. $\sigma << 10^{-2}$ DU). The number of grid points affected was relatively low (less than 150 grid points out of a total of 64800), and localized at high latitudes in the NH.

Based on this feedback, the Ozone CCI team found and corrected a bug in the production line responsible for this issue. At the time of writing it is anticipated that forthcoming versions should no longer present this problem.

The TCO3 L3 merged product is not provided with quality information, such as quality flags. These are extremely useful as complementary information to the observation uncertainty. It is recommended to integrate future versions of this product with additional information on data quality in the form of user friendly flags.

4.1.3. Nadir Ozone Profiles

The CCI Nadir Profile Ozone (NPO3) product was retrieved for two non-consecutive years - 1997 and 2008. Although, the NPO3 product will be based on measurements from several instruments (GOME, SCIAMACHY, OMI and GOME-2), only GOME-1 observations were available for 1997. A GOME-1 based ozone profile product (retrieved at RAL and not too dissimilar from the CCI NPO3 product) was assimilated in the ERA-Interim reanalysis, so a good agreement is expected between the two datasets during 1997. Dragani (2010, 2011) showed that the assimilation of this product really improved the quality of the ERA-Interim ozone analyses compared with the previous ECMWF reanalysis (ERA-40, Uppala et al, 2005). OMI and SCIAMACHY retrievals were also assimilated in both ERA-Interim and MACC in the form of integrated columns.

We discuss the comparisons of NPO3 with their model equivalent from the ERA-Interim and MACC reanalyses focussing on four characteristic pressure levels (5, 10, 30, and 100 hPa) among those available in the CMFDb. These levels were selected to be the level where the ozone mixing ratio peaks at 10 hPa, one level just above (5 hPa) and one just below (30 hPa) the ozone peak, as well as one level near the tropopause, at 100 hPa. It is should be noted that the comparisons at 100 hPa are discussed as they provide an indication of the quality of the ozone product in the UTLS, but it is important to remember that nadir UV sensors normally have reduced sensitivity to this region of the atmosphere. Moreover, as anticipated above (section 4.1.1) the comparisons at 100 hPa are also expected to also reflect the differences between the two ozone models used in ERA-Interim and MACC. Figure 8 shows the comparisons between the global mean monthly mean NPO3 and the ERA-Interim and MACC reanalysis equivalent at the four above mentioned pressure levels (**a**), as well as the comparisons between the NPO3 standard deviations and the ensemble spread (**b**) as done above for the TCO3 product. The NPO3 product shows similar features in the two periods.

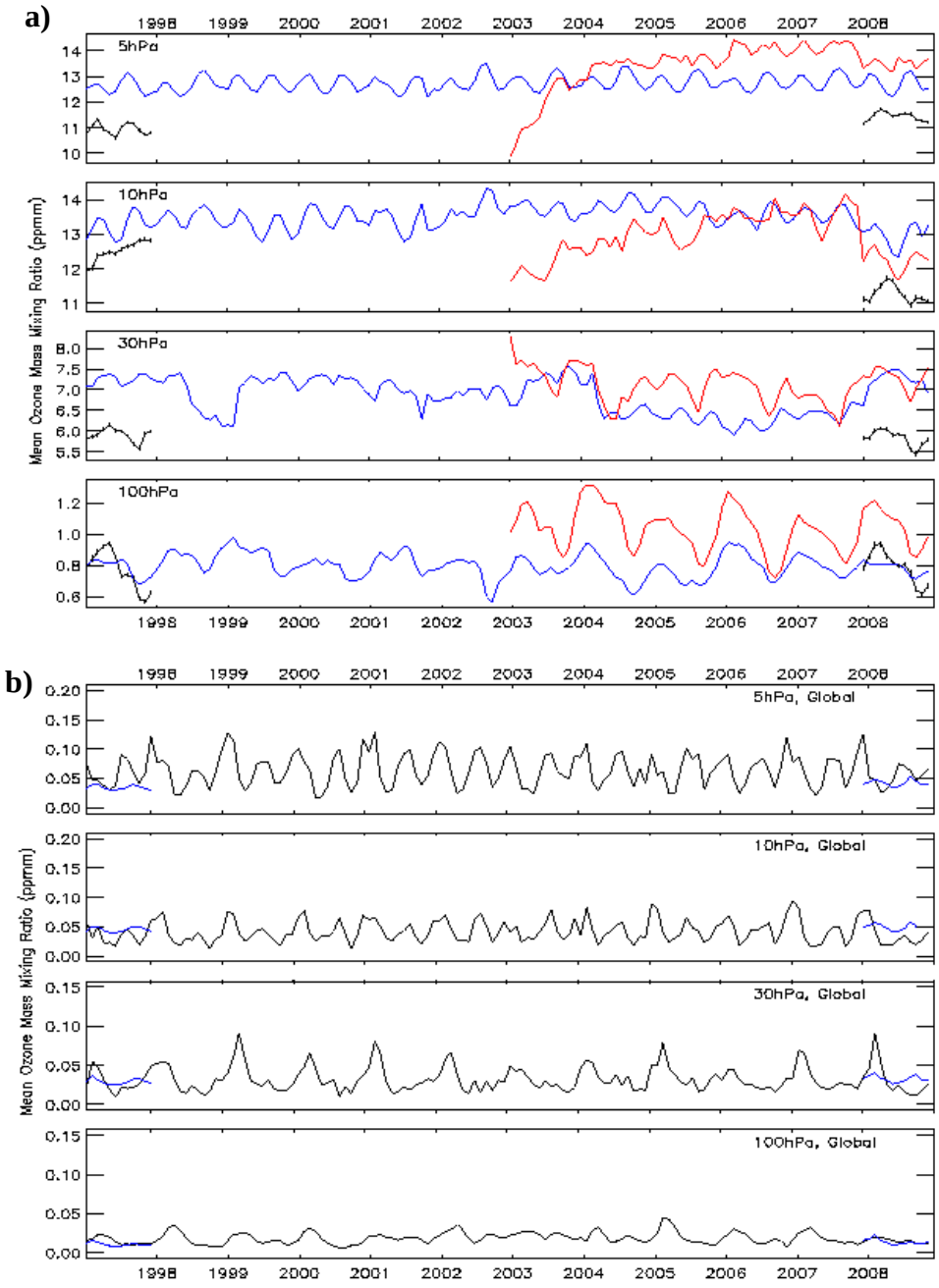

Fig 8: ***a)*** *Time series of the global mean monthly mean (nadir) O_3 mass mixing ratio (ppmm) from Ozone_CCI (black), the ERA-Interim (blue) and MACC reanalyses (red) at four vertical levels (from top to bottom): 5, 10, 30, and 100 hPa.* ***b)*** *Time series of the global mean monthly mean O_3 mass mixing ratio (ppmm) from CCI (blue), and the ensemble spread for O_3 mass mixing ratio obtained from an ERA-Clim 10-member ensemble prototype (black) at the same vertical levels of* ***a)****.*

The NPO3 product exhibits smaller global mean values than their model equivalent in the middle stratosphere where differences are up to 1.5 ppmm. Near the tropopause the agreement of the NPO3 is good with the ERA-Interim ozone reanalyses, but not with the MACC reanalyses which are arguably better characterised thanks to the coupling with a full CTM.

At all vertical levels, and for both years, the residuals between the NPO3 and the reanalysis datasets are much larger than the observation errors. However, the comparisons between the NPO3 errors and the ensemble spread (**b)** panels in figure 8) suggest a good level of agreement, and thus a good characterization of the random component of the observation uncertainty at least in the global mean.

When different latitudinal bands are considered, the level of agreement between the NPO3 product and its model equivalent obtained from ERA-Interim and MACC is generally confirmed (not shown). NPO3 exhibits lower values than the two ozone reanalyses at all levels and both years. The differences are negative and within 0.5 and 1 ppmm. A few exceptions were noticed, for example in the tropical region and at midlatitudes in the SH. In these cases, the observation minus reanalysis residuals were in absolute value larger and up to -3 ppmm in the region of the ozone mixing ration maximum at 10 hPa, particularly against the MACC ozone reanalyses.

As an example, the comparison in the tropical region at 10 hPa is shown in figure 9. In here, the NPO3 retrievals show rather different behaviour in 1997 and 2008, with an average of 1.5 ppmm higher ozone values in 1997 compared with 2008. During 1997, NPO3 and ERA-Interim are in very good agreement, as expected in virtue of the fact that both datasets were based on the same retrievals. This would suggest that by replacing the previous (RAL) GOME-1 retrieval with the CCI product could produce negligible differences. In contrast, in 2008, the NPO3 dataset is, like at other levels and latitudinal bands, about 1 ppmm lower than ERA-Interim, and over 2 ppmm lower than the MACC ozone reanalyses. Independent validation of the ERA-Interim ozone reanalyses (e.g. Dragani 2010, 2011) showed that the latter in the region of the tropical ozone maximum at 10 hPa normally exhibit values that are too low. It is argued that if the L3 dataset is regarded as a proxy of the L2 NPO3 product, then the assimilation of the latter in 2008 could potentially further reduce the values of the tropical ozone maximum and degrade the ozone reanalyses.

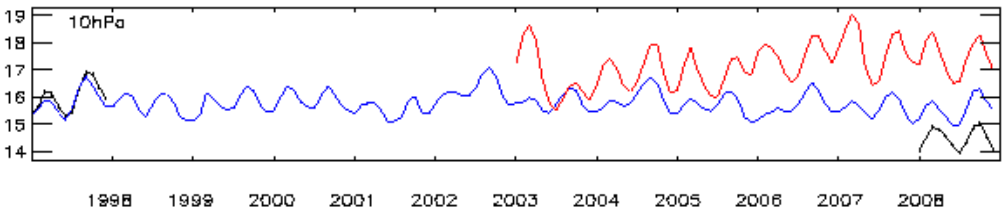

*Fig 9: Like in a**)** of figure 8, but for the tropics at 10hPa.*

In contrast to the comparisons in the mean ozone values, the good level of agreement between the NPO3 errors and the ensemble spread found over the globe is not generally confirmed when focusing on given latitudinal bands. It is noted that the standard deviation over a region was defined according to equation (1) and it is not the mean of the standard deviations.
In general, the NPO3 errors appear to be too small than the predicted natural variability of the ozone field as provided by the ensemble spread at most levels and regions, with values that are normally half the value of the latter. It is also noted that the ensemble spread exhibits a larger variability than that of the NPO3 standard deviation, particularly in the middle

stratosphere. Figure 10 shows the comparisons between the NPO3 standard deviation and the ensemble spread for the tropics as an illustration.

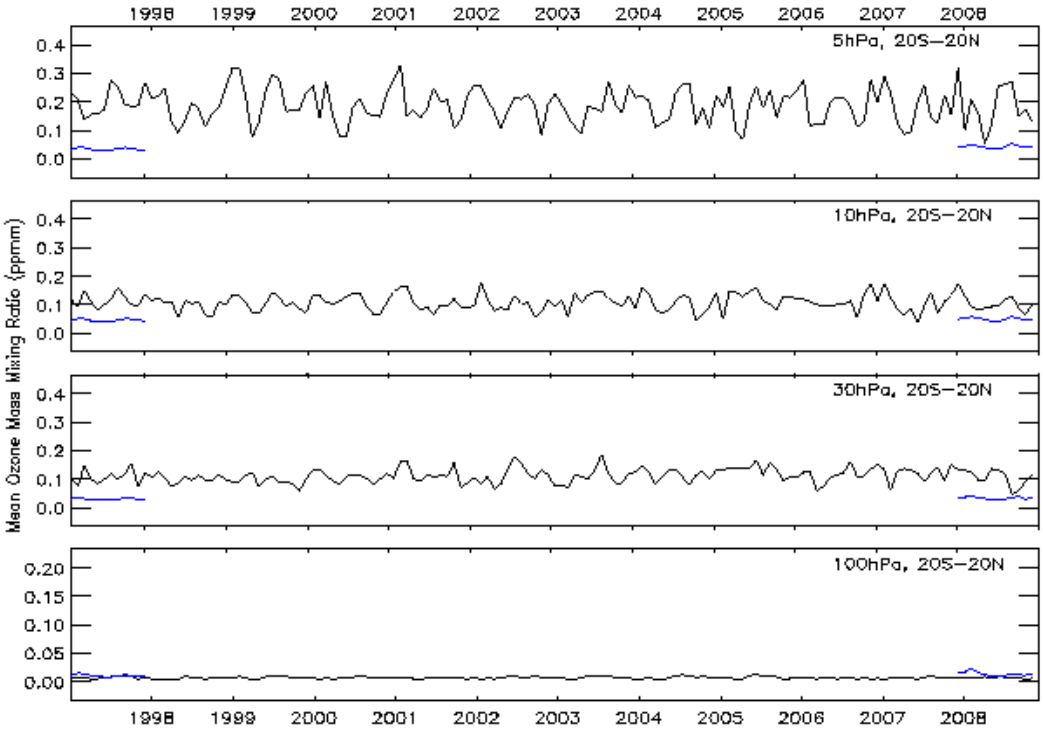

*Fig 10: Like in **b)** of figure 8, but for the tropics.*

It should be pointed out that even if a larger error (e.g. double the size of the current NPO3 standard deviation) was accounted for, it would not be enough to change the conclusions drawn on the level of agreement between the actual product and their reanalysis equivalents.

4.1.4. Limb Profile Ozone

The Ozone CCI team has also created a merged monthly mean zonal mean ozone profile dataset obtained from limb measurements of the ENVISAT (GOMOS, MIPAS, and SCIAMACHY) and ESA Third Party Mission (OSIRIS, SMR, and ACE-FTS) instruments. This product (LPO3) is currently available for two consecutive years only - 2007 and 2008 -, which is too short to provide any feedbacks on the long term homogeneity of this product.

As done for the NPO3, LPO3 was compared with their model equivalent from the ERA-Interim and MACC reanalyses focussing on four characteristic pressure levels (5, 10, 30, and 100 hPa).

Figure 11 shows the two-year long global mean time series for the three datasets at those four characteristic levels. It is clear that although differences exist, the three datasets well compare within the observation error bars at all levels below the ozone mixing ratio maximum at 10 hPa. Above the ozone maximum, the MACC ozone reanalyses exhibit higher values than

LPO3 and ERA-Interim with differences of about 2 ppmm and 1 ppmm, respectively. The difference of 2 ppmm from LPO3 is sufficient to make the MACC timeseries fall outside the observation one standard deviation range.

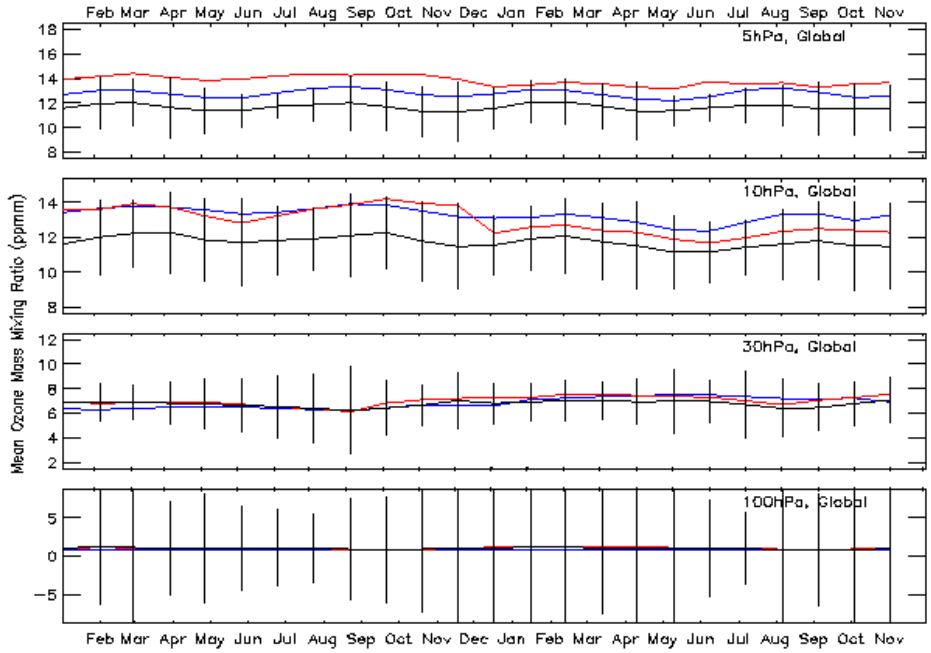

Fig 11: Time series of the global mean monthly mean (limb) O_3 mass mixing ratio (ppmm) from Ozone_CCI (black), the ERA-Interim (blue) and MACC reanalyses (red) at four vertical levels (from top to bottom): 5, 10, 30, and 100 hPa.

When averaging over five latitudinal bands (high latitudes, i.e. 60^o-90^o, midlatitudes, i.e. 20^o-60^o, and tropics, i.e. 20^oS-20^oN), the conclusions drawn from figure 9 are normally confirmed except at 5 hPa at high latitudes (figure 12). At this level, the MACC reanalysis is systematically higher than both LPO3 and ERA-Interim in the latitudinal range between 60^oS and 60^oN, with mean differences of just over 2 ppmm that are larger than the observation errors. The reason for the poor level of agreement at 5 hPa at midlatitudes and tropics between the LPO3 and ERA-Interim datasets on one hand and the MACC reanalysis on the other has to be found in the differences between the two reanalysis productions discussed in section 4.1.1. Among all those differences, the use of an ozone bias correction in MACC but not in ERA-Interim is likely the one difference responsible for the observed disagreement at 5 hPA. The differences can be the result of a different weight given to the SBUV observations in the MACC reanalysis.

At all other levels and latitudinal bands, the level of agreement with the ERA-Interim ozone reanalyses is generally good and typically within the observation errors (not shown). It is noted, however, that the level of agreement between LPO3 the MACC was slightly better in 2008 than during 2007, as a consequence of the change in the MLS ozone bias correction implemented in MACC on 1 January 2008 (as discussed section 4.1.1). The impact of this change was mainly seen at the ozone mixing ratio maximum at 10 hPa (see for example figure 11). The comparison at that level shows differences of about 2 ppmm during 2007, and about 1 ppmm in 2008.

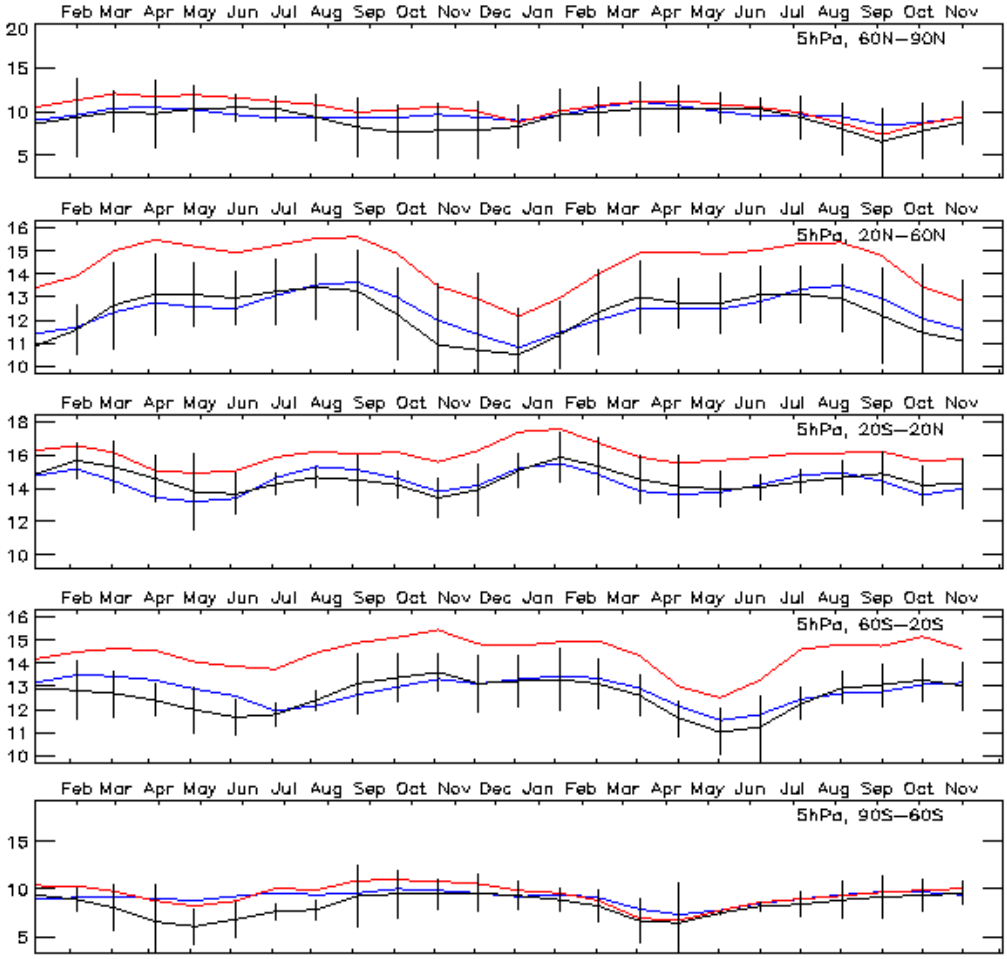

Fig 12: Like in top panel of figure 11, but for averages at different latitudinal bands.

The global ozone anomaly for the three datasets at the four pressure levels analysed above is presented in figure 13. The LPO3 and the ERA-Interim datasets generally show similar seasonal variability at 5 and 10 hPa. At these levels, the MACC ozone anomaly is clearly affected by the bias drift during the whole 2007 as discussed in section 4.1.1, while at the beginning of 2008 the MACC reanalyses might still be affected by spin-up caused by the change in the bias anchor occurred on 1 Jan 2008.

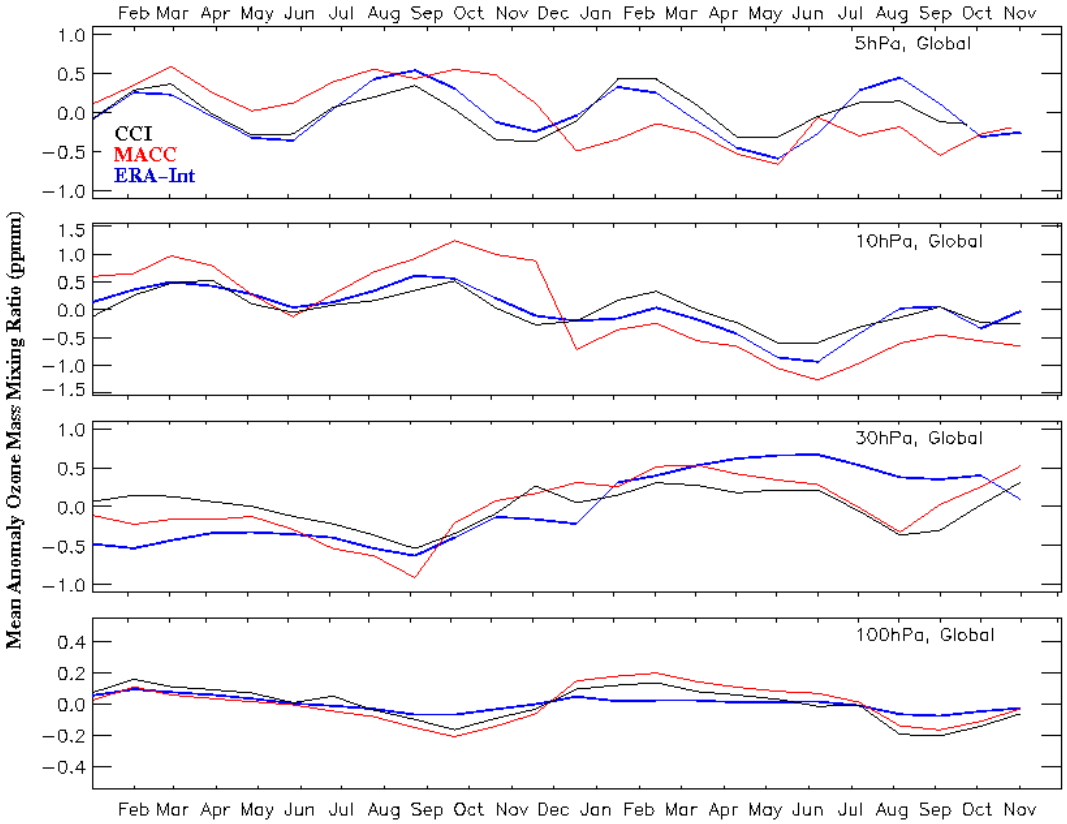

Fig 13: Time series of the global mean monthly mean (limb) O_3 mass mixing ratio anomaly (ppmm) from Ozone_CCI (black), the ERA-Interim (blue) and MACC reanalyses (red) at four vertical levels (from top to bottom): 5, 10, 30, and 100 hPa.

Below the ozone maximum, the level of agreement between the MACC and LPO3 datasets improve substantially, and it is higher than that between ERA-Interim and LPO3. The agreement between LPO3 and ERA-Interim improves during the first half of 2008 when the assimilation of vertically resolved MLS ozone profiles started (and before a period of MLS unavailability).

Closer to the tropopause (100 hPa), the ERA-Interim ozone analyses show little seasonal variations, consistent with the fact that hardly any constraint was available in this region from observations and that the ozone model (based on the Cariolle and Déqué (1986) scheme) is essentially a stratospheric model, and therefore inadequate to provide tropospheric ozone information which is more accurate in the MACC system thanks to the coupling of with CTM.

Comparisons between the observation errors and the ensemble spread at the four vertical levels discussed above are shown in figure 14 for the globally averaged data. The observation error is normally much larger than the ensemble spread (which appears as an almost 0 line). The latter normally varies between 0.05 and 0.15 ppmm across the vertical range discussed (see for example figure 8 - panels b) - and figure 9 from the NPO3 assessment). In contrast, the observation error is on average 2 ppmm (about 13% of the mean value at 10 hPa of 15 ppmm) in the stratosphere and about 9 ppmm near the tropopause (about 9 times a typical

mean ozone mixing ratio of 1 ppmm). The comparisons at the other latitudinal bands confirm the conclusions drawn from figure 14.

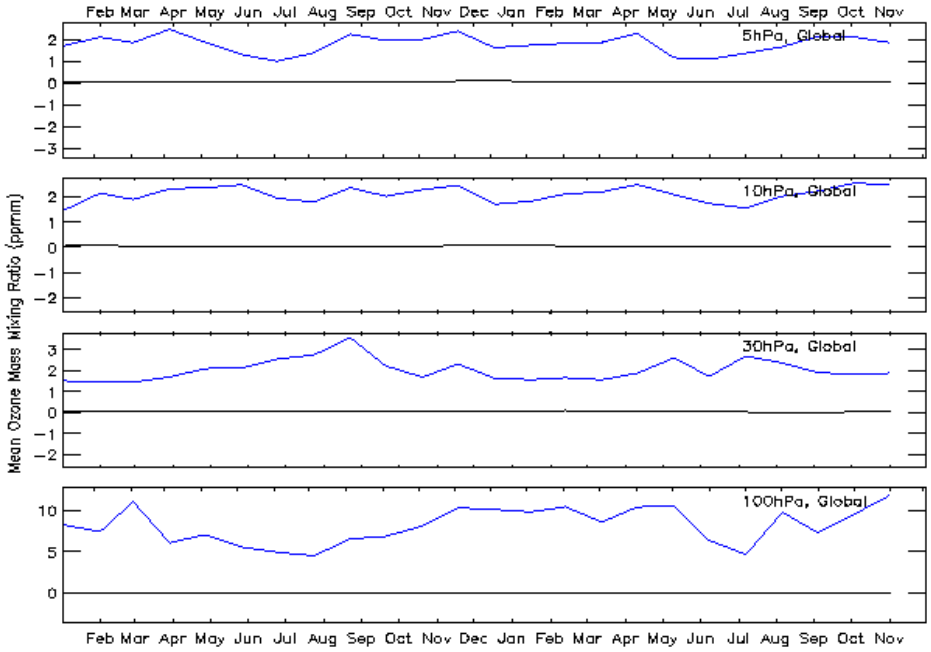

Fig 14: Time series of the global mean monthly mean O_3 mass mixing ratio (ppmm) from Ozone_CCI (blue), and the ensemble spread for O_3 mass mixing ratio obtained from an ERA-Clim 10-member ensemble prototype (black) at the same vertical levels of figure 10.

In summary, the level of agreement between the three datasets in the mid stratosphere can be regarded as being good. A key factor is the fact that both reanalyses, although with some differences, benefitted from the assimilation of the MLS ozone profiles. MLS, like the limb instruments used to produce the LPO3 dataset, provides by design ozone profiles with very accurate vertical resolution that can strongly constrain the modelled ozone field and improve its vertical distribution. Arguably, the orbit-based (Level 2) LPO3 datasets have the potential to positively impact future reanalysis production in virtue of that design, particularly in periods where the MLS data were unavailable. However, a key factor that could prevent these observations being fully exploited within an assimilation system is having over-estimated errors as these provide the weight the observations have when assimilated. Being conservative and preferring "inflated" observation errors is not unusual among data assimilators to account for correlations and representativeness errors. However, the analysis performed here seems to indicate that the LPO3 observation errors in the L3 merged dataset could be too large and it should be understood if this overestimation is a consequence of the L2 ➔ L3 processing step, or it has been inherited by the L2 observation errors themselves.

4.2. Aerosols

Aerosol CCI aims at integrating the major aerosol (European) retrieval schemes in order to generate a well characterised dataset of aerosol properties from the ENVISAT and other European measurements. Using the 2008 as the golden year, a round robin exercise was conducted to compare the ability of the available algorithms. The results highlighted that none of them is able to outperform all the others everywhere, but each has strengths and weaknesses. Among all available algorithms, the three AATSR-based algorithms have been considered in the present study with a focus on the level 3 gridded datasets of multi-spectral aerosol optical depth (AOD). Table 4 provides information on the datasets, data version and provider, and availability for the three products used.

Data set	Version	Provider	Product	Period assessed	Acronym
AATSR_ADV	1.42	FMI	AOD	2008	ADV
AATSR_ORAC	2.02	Uni Oxford/ RAL	AOD	2008	ORAC
AATSR_SU	4.0	Uni Swansea	AOD	2008	SU

Table 4: List of the assessed CCI aerosol datasets.

These three products were compared with reanalyses of AOD produced within the MACC project (Morcrette *et al.* 2009; Benedetti *et al.* 2009; Mangold *et al.* 2011). This aerosol system, that was initially developed in the framework of the Global and regional Earth-system Monitoring using Satellite and in-situ data (GEMS) project, is the first aerosol model to be fully coupled to a numerical weather prediction model with integrated data assimilation. It uses a bin-model for aerosol that includes desert dust, sea salt, organic matter, black carbon and sulphates (Benedetti *et al.* 2009). The aerosol analyses are based on the assimilation of AOD data retrieved from the MODIS measurements on board of Terra and Aqua, selected for its reliability. Remer *et al.* (2005) showed that the MODIS AOD over ocean is more accurate than the land retrievals that have higher uncertainties due to the impact of the surface reflectance that can be confused for aerosol signal over highly reflective surfaces. Benedetti *et al.* (2009) also found differences between the MODIS land retrievals and AERONET (Holben et al. 1998) observations as large as 41% at 550 nm with MODIS showing a positive bias. Several other factors affect the accuracy of the retrievals both over land and ocean (e.g. cloud contamination, assumptions on the aerosol types and size, instrumental uncertainties, etc…) as discussed by Zhang and Reid (2006).

Each of the four datasets (the three described in table 4 and the MACC reanalysis) provides AOD at various wavelengths (summarised in table 5). These are not always exactly the same across the four data records, not even within the three AATSR algorithms. It is understood that this inconsistency will be addressed in future versions. Because of the slightly different definition of the wavelengths in the four datasets (as presented in table 5), the comparisons discussed below used whenever possible the two closest wavelengths, e.g. the AODs at 659 nm and 670 nm are assessed together. The assumption is that any differences seen between the data records are mostly related to differences in the algorithms, and that the impact of using slightly different wavelengths is negligible or at worst small.

	550 nm	659 nm	670nm	865 nm	870 nm	1610 nm	1640 nm
ADV	Yes	Yes				Yes	
ORAC	Yes				Yes		
SU	Yes	Yes		Yes		Yes	
MACC	Yes		Yes	Yes			Yes

Tab 5: List of the available wavelengths for each dataset. The shaded boxes encompassing two close wavelengths indicate that the datasets at both wavelengths were assessed together. A blank box indicates that no data was available for that dataset and at that wavelength.

Figure 15 shows the global mean AOD at the four groups of wavelengths. The uncertainties (computed according to equation 1) are over-plotted to each AATSR dataset. In the global mean, the SU algorithm seems to produce the dataset with the best level of agreement with the MACC reanalyses at the lower wavelengths (from 550nm to 865nm). The ADV dataset is also in good agreement with MACC, showing residuals that are smaller than the observation errors. The residuals between the ORAC dataset and MACC are larger than the ORAC errors at 550 nm, and as large as one standard deviation during most of 2008 at 865 nm.

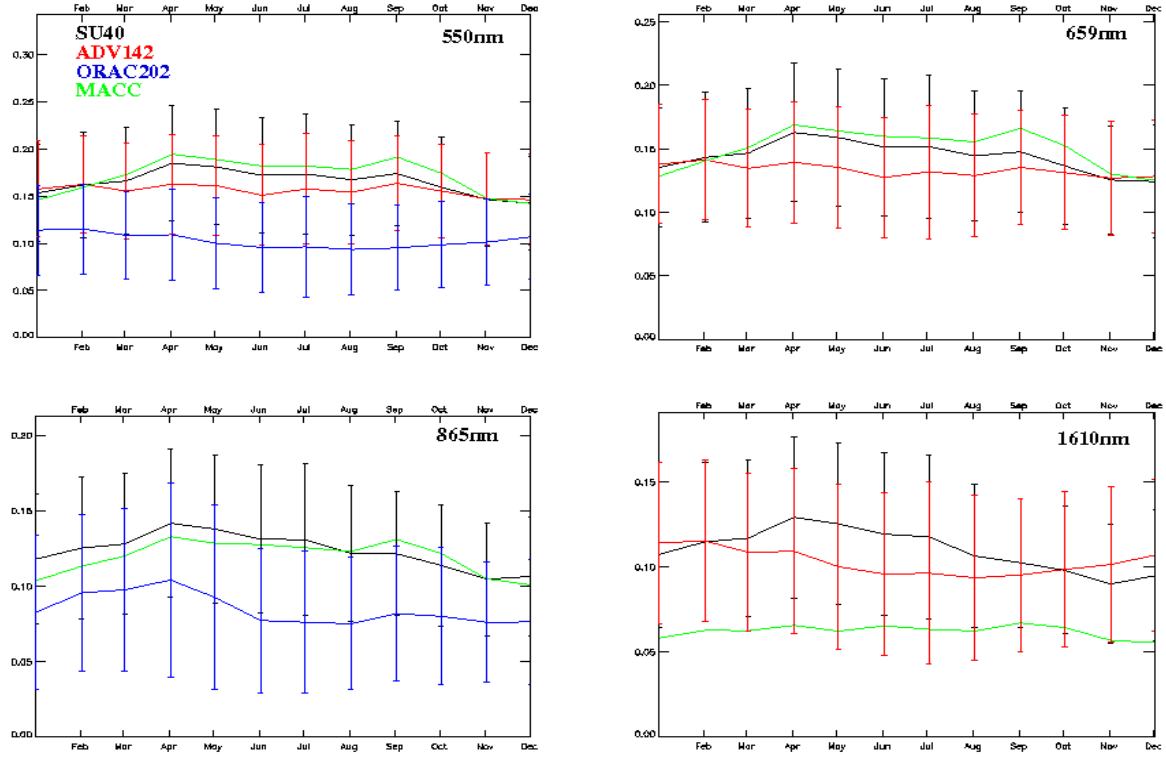

Fig 15: Global AOD at 550,659, 865, 1610nm for the four datasets. The vertical bars represent the observation errors of the three AATSR datasets.

At the longest wavelength (1610nm), the AATSR datasets show large differences with MACC with values that are double what prescribed by the aerosol reanalyses. In this case when accounting for the observation errors, the ADV dataset outperforms the SU one during most of the year.

To assess the strengths and weaknesses of the three algorithms in different situations the comparison presented in figure 15 was repeated for different geographical areas. In particular, figures 16 and 17 show the same comparison of figure 15 but with the data averaged over land and ocean, respectively.

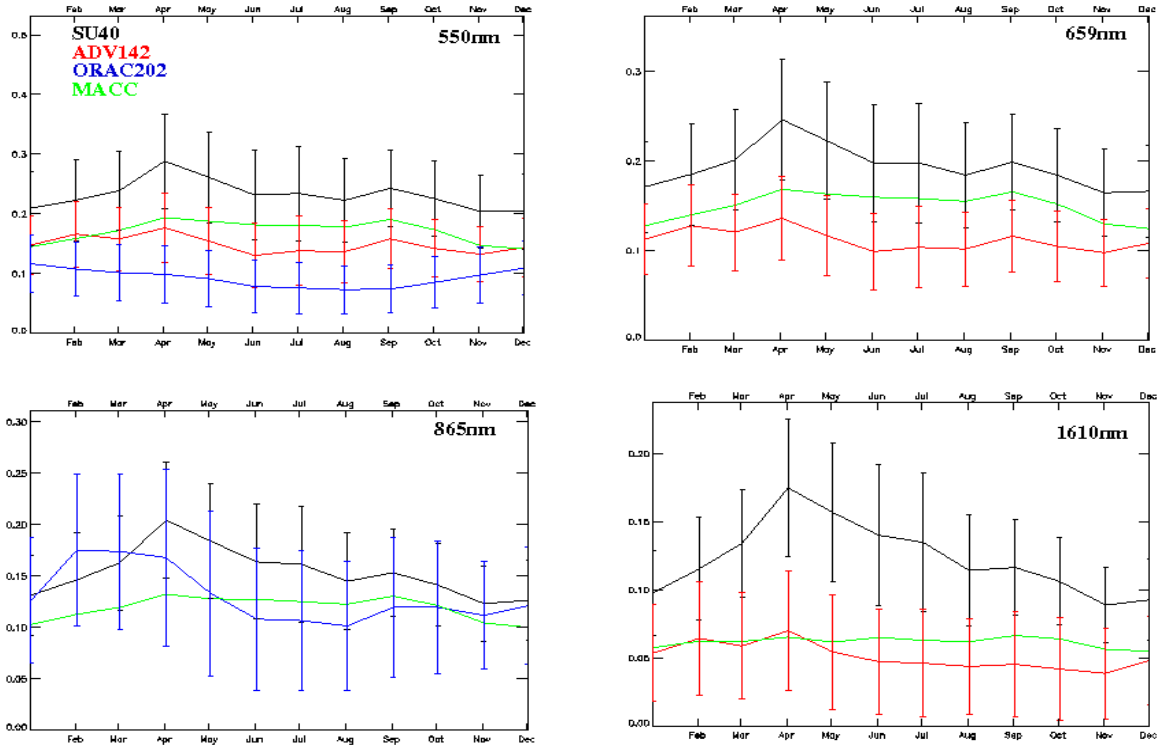

Fig 16: Like figure 15, but over Land.

Over land, the SU dataset shows differences from MACC AODs smaller than the observation errors at wavelengths from 550 to 865 nm, and differences larger than the observation errors at 1610nm. The ADV dataset is in good agreement with the MACC reanalysis at 550 and 659 nm, and in excellent agreement at the longest wavelength (1610 nm), where it outperform the SU algorithm. The ORAC data record shows differences from MACC that are larger than the observation error at 550 nm, while at 865 nm, the level of agreement improves substantially. As mentioned above, Benedetti et al (2009) found that during May 2003 MODIS over-estimates the AOD mixing ratio at 550 nm over land compared with the AERONET data, with differences as large as 41%. They also found that as a result of the assimilation of MODIS data, the AOD analyses showed a positive bias at 550 nm compared with 41 ground stations from the AERONET network, with a mean bias of about 0.012. These results were generally confirmed by a recent study (MACC II, 2013) that found differences between the MACC and AERONET data (at 550 nm) of about 20% during 2008 - i.e. the same year used by Aerosol CCI. If that was taken into account, the level of agreement between the AATSR datasets and the MACC reanalysis would only marginally improve for the ADV and ORAC algorithms, while the residuals of the SU data record from MACC would become larger than shown in figure 16, and - depending on the month - they could exceed the observation error. Provided that an appropriate bias correction could be used to make the assimilation of multiple aerosol datasets viable and efficient, there could be scope to test the assimilation of the ADV dataset

at least over land with a MACC system set-up similar to the one that will be used in future reanalysis.

Over land, the ADV dataset is the one providing overall the best level of agreement with MACC, followed by the SU dataset, and then the ORAC one.

Over oceans (figure 17), the ADV dataset is the one showing the smallest differences from MACC at 550 and 659 nm, followed by the SU dataset. For both of them, the residuals are smaller than the observation errors. At the two longest wavelengths, the SU dataset well agree with MACC within the observation errors, although the residuals at 1610 nm are only marginally smaller than the observation standard deviations. The differences between the ADV and MACC AOD are normally larger than the observation errors at both available wavelengths. The residuals between the ORAC and MACC AODs are normally larger than the observation error at both wavelengths.

Over ocean and accounting for the observation errors, the SU dataset is the one providing overall the best level of agreement with MACC, followed by the ADV dataset.

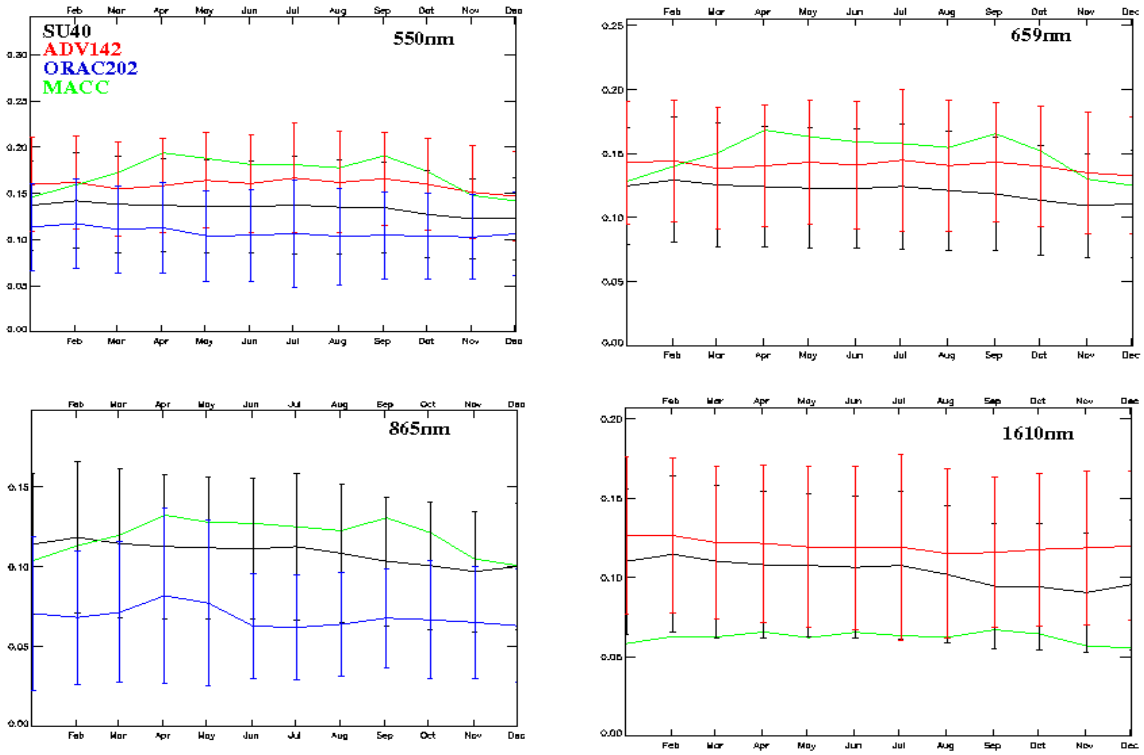

Fig 17: Like figure 15, but over Oceans.

Comparisons over a total of 25 different regions were also performed. Results highly vary depending on the area and wavelength considered. The statistics for each of the three AATSR datasets and at the four groups of wavelengths are summarised in tables A1 and A2 in Appendix A. Tables A1 and A2 seem to indicate that the MACC reanalyses are generally

higher than the three AATSR datasets. Here, we provide a summary of those tables (table 6). For each dataset and at each available wavelength, the number of areas (in absolute and relative terms) where it outperforms the others in the agreement with the MACC AOD is presented.

	550 nm	659 nm	865 nm	1610 nm
SU	12 / 48%	14 / 56%	15 / 60%	9 / 36%
ADV	11 / 44%	11 / 44%	N/A	16 / 64%
ORAC	2 / 8%	N/A	10 / 40%	N/A

Tab 6: Statistics describing the agreement between each AATSR dataset and the corresponding MACC AOD at the four wavelengths. Each box provides two figures representing the total number of areas and their percentage (computed with respect to the number of assessed regions, i.e. 25) where the corresponding dataset outperformed the others in the agreement with MACC.

Overall, the SU dataset has the highest level of agreement with the MACC aerosol reanalysis at the shortest wavelengths, while at the longest wavelength (1610 nm) the ADV dataset agrees with the MACC reanalysis over a much higher number of cases. Clearly, these figures only provide the number of areas where a given AATSR dataset shows the best level of agreement with the MACC reanalysis, without accounting for other factors, e.g. the region extent.

The results presented here and in Appendix A should also be interpreted in the light of the quality of the MACC aerosol reanalyses. Some of the shortcomings of the MACC aerosol model were discussed by Benedetti et al (2009) based on the analysis of assimilation experiments performed for May 2003. More recently MACC II (2013) provided an in-depth analysis of the MACC reanalyses focussing on the 2003-2011 period, and thus encompassing the year considered by Aerosol CCI (i.e. 2008). Some of the issues have already been discussed. It has been mentioned, for instance, that the MACC AOD tend to be over-estimated during the summer months as a consequence of a bias in the MODIS data. This is particularly the case for the 550 nm wavelength. MACC II (2013) found relative bias from the AERONET observation to be as large as +20% in summer at 550 nm (figure 18, left panel).

The right panel of figure 18 shows the geographical distribution of the bias between the MACC AOD reanalyses at 550 nm and the AERONET data. A few regions stand out for having particularly large biases. For example, a large positive bias is found over North America (with relative residuals that range from 60 to 100%). At 550 nm over North America, all AATSR datasets show negative residuals from MACC, implying smaller values than the reanalysis (see table A1) with values that range from -0.034 in the case of ADV, to -0.103 in the case of ORAC. Thus, they all seem to agree with the comparisons with the AERONET data. It is fair to say that any of them if assimilated could potentially improve the level of agreement between the aerosol analyses and AERONET over North America and at 550 nm.

The actual impact is difficult to predict as it will depend on the weight of the observations and their synergy with other aerosol observations assimilated in the system (i.e. the Aqua and Terra MODIS data).

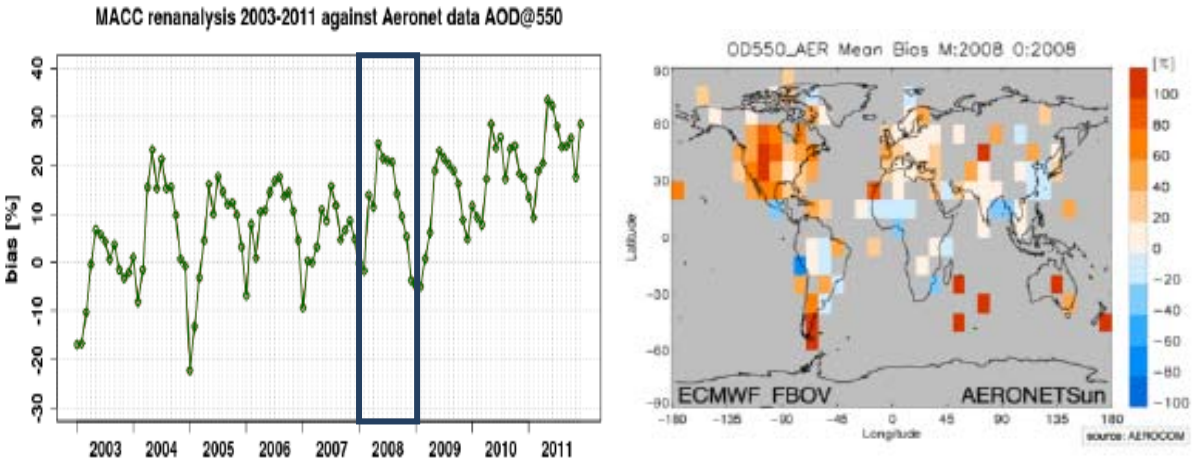

Fig 18: ***Left:*** *Evolution of the mean normalised bias between the MACC aerosol reanalysis and the AERONET data at 550 nm, reproduced from figure S3 in MACC II (2013).* ***Right:*** *Relative bias (MACC-Aeronet)/Aeronet (in %) computed for aggregated monthly data over a 10° by 10° grid boxes. The figure was reproduced from figure 3.5.2 in MACC II (2013).*

Another shortcoming of the model is related to the representation of sea salt, which seems to be overestimated (again at 550 nm) and leads to a high AOD bias in southern oceanic regions. Like in the case of North America, also over the Southern Oceans all datasets exhibit lower values than MACC (see table A1), with similar differences ranging from -0.063 (ADV) to -0.075 (ORAC). Arguably, as noted above the assimilation of any of the three datasets could potentially reduce the bias in the Southern Ocean region.

In addition to the traditional time series, CMF allows users to download the data that were extracted from CMFDb according to their request, so that additional analysis can be performed. Bearing in mind the positive bias of the MACC AODs during summer, we used this option to provide an assessment of the observation uncertainties. Figure 19 shows for example a scatter plot of the observation AOD minus MACC equivalent global mean difference versus the observation uncertainty computed for the three datasets at 550nm. Although the datasets only cover one year and a much longer record is needed to provide statistically significant conclusions, figure 19 does not show issues in the uncertainty associated with the SU40 and ADV142 datasets, as the residuals from the MACC AOD reanalysis are within the observation errors. In contrast, the ORAC202 residuals from MACC AOD are generally larger than the observation standard deviation, suggesting that the dataset might be affected by a systematic bias. Similar outcome were normally found at the other wavelengths and geographical areas (not shown).

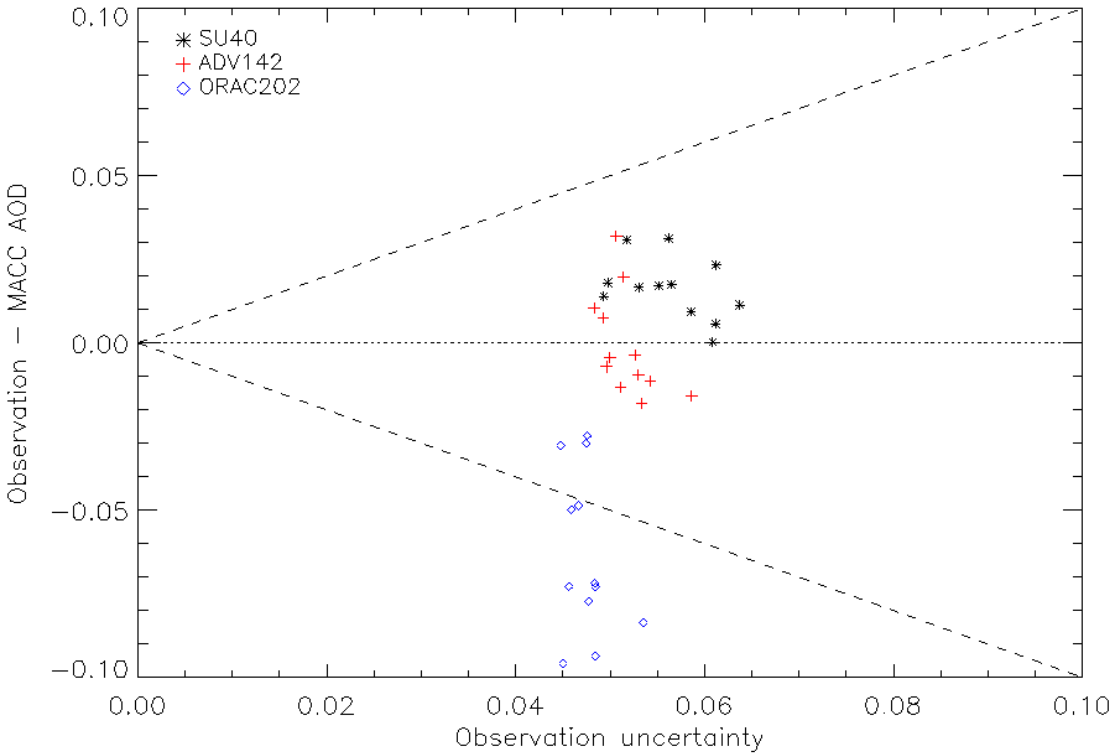

Fig 19: Scatter plot of the observation AOD minus MACC AOD global mean difference versus the observation uncertainty for the three datasets at 550nm.

In general, it is fair to say that the regions of the globe were observations could potentially be most successful in positively impacting the aerosol analyses are those where the aerosol source is mostly of anthropogenic nature (e.g. Eastern Asia, Indian subcontinent), as the aerosol model has limited ability in representing them.

4.3. Soil Moisture

Soil Moisture CCI (SM_CCI) has produced an over 30-year long data record of soil moisture (SM) obtained by combining observations from several active and passive microwave sensors. This report focuses on the quality of the level 3 (gridded) CCI soil moisture data set (Version v0.1) during the entire period of data availability (November 1978 - December 2009) by comparisons against the most recent SM reanalyses available.

The availability of a 30-year long data record permits an assessment of the long-term homogeneity of the data set. Figure 20 shows the temporal evolution of the CCI monthly mean (computed from the daily gridded data on the same grid of the original data) soil moisture anomaly. An indication of the instrument changes for both active and passive instruments has been over-imposed to the timeline.

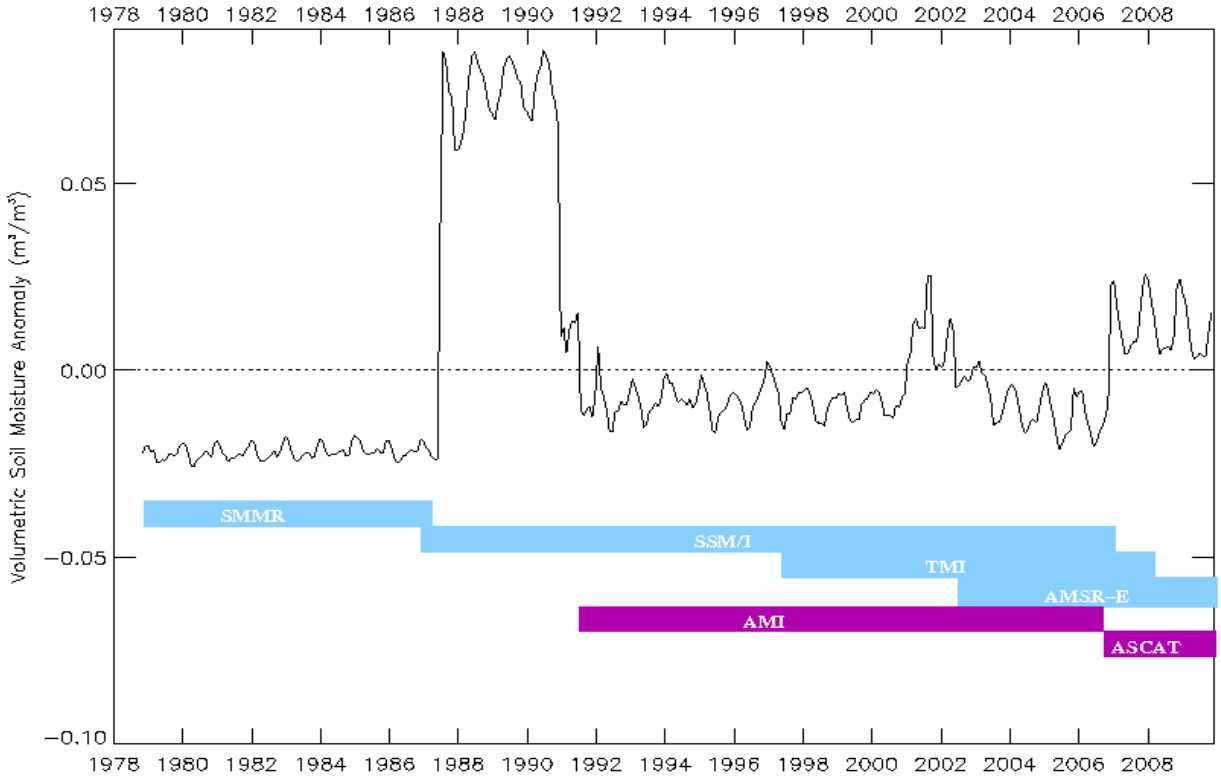

Fig 20: Time series of the CCI monthly mean volumetric soil moisture anomaly averaged over Land. An indication of the changes in the instrument usage is over-plotted to the timeline for both active (magenta) and passive (cyan) sensors.

It is clear that the merged SM product is strongly affected by changes in the observing system used. The change from SMMR to SSM/I appears quite dramatic, and so is the introduction of data from ERS-1 AMI. Other changes can be noticed in 2002 when AMSR-E was introduced and in mid-2006 when AMI data was replaced by ASCAT measurements.

Although studies such as Dorigo *et al* (2012) and Albergel *et al* (2013a) showed some ability of this dataset to provide a reasonable indication of the "sign" of the SM trend estimates (i.e. an indication of whether a given region of the globe is becoming wetter or indeed drier with

time), the changes in the observing system affected the SM time series homogeneity and make the current version unsuitable for providing a reliable trend analysis for SM, and assessing its long-term changes and variability. It is recommended that this aspect is improved in future versions by revising the method of data merging, e.g. by looking at the quality of each data record, and revising the inter-instrumental biases.

Bearing in mind the poor temporal homogeneity, the dataset in its current version has been compared with soil moisture reanalyses from three data streams: ERA-Interim (1979-Present, Dee *et al*, 2011), ERA-Interim/Land (1979-2010, Balsamo *et al*, 2012), and MERRA-Land (1980-2010, Reichle *et al*, 2011). The ERA-Interim/Land (hereafter ERA-Land for simplicity) and MERRA-Land are off-line land-surface simulations associated to the ECMWF ERA-Interim and the NASA MERRA atmospheric reanalysis productions. These off-line simulations provide improved land information compared with the original reanalyses as they permit to make use of a more up-to-date version of the surface-land model (SLM) while forced by the atmospheric reanalysis meteorological fields (temperature, surface pressure, humidity and wind). They are useful for land-model development while also offering an affordable way to improve the land-surface component of the original reanalysis.

It is important to consider that the information provided by the four datasets is not exactly the same. These differences have an impact on their level of agreement and should be taken into account. Satellite instruments are sensitive to the Earth's surface, while models provide the soil moisture content over a soil layer with thickness depending on the model. The MERRA system uses a top layer of soil with 2 cm thickness (i.e. spanning 0-2 cm in the ground). In contrast, the ECMWF SLM provides SM over four soil layers (0-7 cm, 7-28 cm, 28-100 cm, 1-2.89 m) with a much thicker top one.

As the comparisons refer to averages over extended regions, they do not show large differences between the two ERA products, therefore for clarity only the latest dataset (ERA-Land) is shown. For an in-depth assessment of the CCI SM data set against ERA-Interim, the reader can also refer to CMUG (2012).

Figure 21 presents the comparisons between the CCI SM time series and the two land reanalyses averaged over the emerged lands. Both reanalyses present higher SM values than the CCI SM dataset, in agreement with the study performed by Albergel *et al* (2013b) for ERA-Land. The comparison with the ERA-Land dataset shows mean (CCI-Reanalysis) residuals ranging between -0.1 and -0.15 m^3/m^3 (50-75% with respect to the CCI SM), depending on the period. These differences are often larger than the observation error and likely due to the fact that the two datasets - as explained above - do not exactly provide the same information. A higher level of agreement (as it could be expected) is found in the comparisons against MERRA-Land. In this case, the mean (CCI-Reanalysis) residuals are also negative and - depending on the period - with values ranging between 0.05 and 0.075 m^3/m^3 (25-38% with respect to the CCI SM), that are at best comparable with the observation error.

The poor level of homogeneity (discussed above) is here contrasted with that provided by the two reanalyses.

Despite a bias due to the different layer thickness they refer to, the two reanalyses present a clear annual cycle that appears to be in phase. In contrast, a limited annual variability can be found in the CCI SM data set, also also in years when the latter shows variations as large as those in the reanalyses (e.g. during the mid-90s), they appear not to be in phase with the two reanalyses.

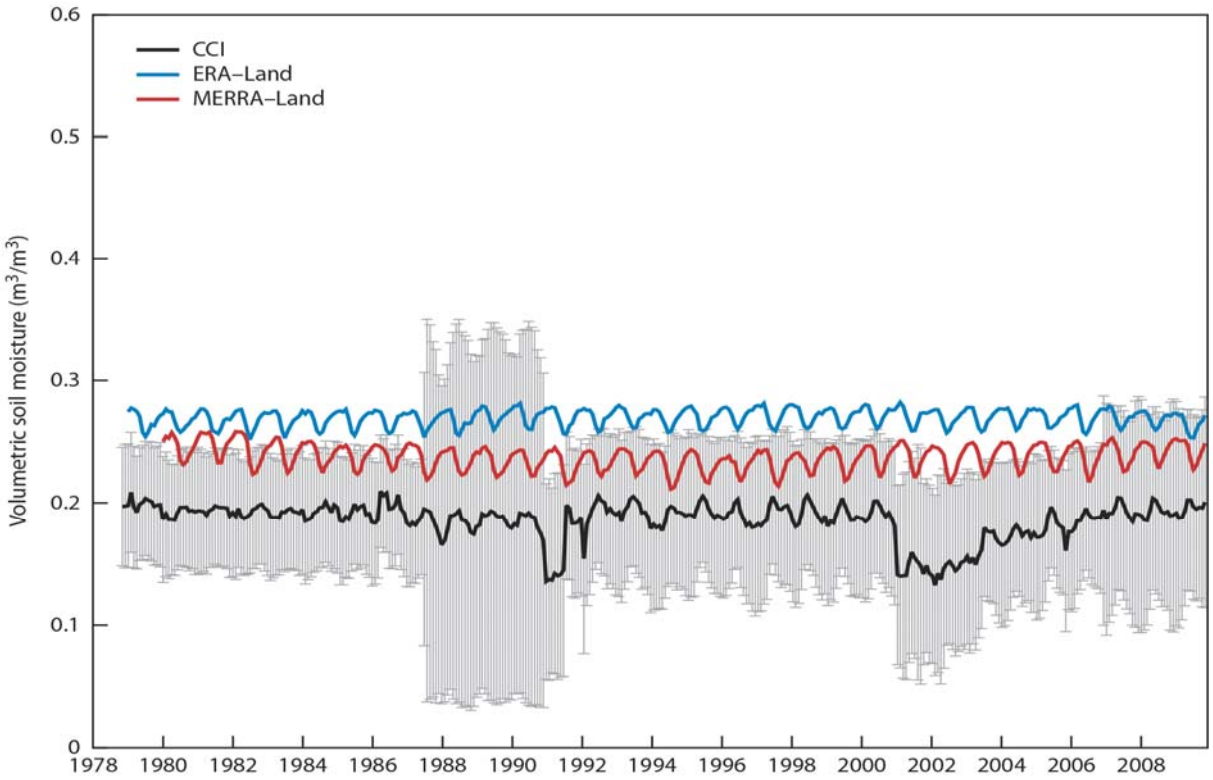

Fig 21: Time series of the monthly mean volumetric soil moisture obtained from CCI SM (black), ERA-Land (blue), and MERRA-Land (red) averaged over Land.

The comparisons performed over other regions of the globe show confirm that the CCI SM normally exhibits lower values than the two reanalysis products, and also that, as expected, the level of agreement is usually higher when compared with MERRA-Land than when it is confronted against ERA-Land, with a few exceptions, e.g. over China and India (figure 22).

A summary of the level of agreement over the whole period of data availability for several geographical areas is given in table 7. The table provides both the temporal mean of CCI SM and the (absolute and relative) residuals from the two reanalyses.

Table 7 gives an overview of the high variability that the SM variable can have from product to product. Advances have certainly been made during the past few years on both the observation and model sides. However, discrepancies between the information that a model can provide and an instrument can measure are a major limitation to properly understand the reported differences. Yet, comparisons with model outputs are important to complement the assessment of remotely sensed soil moisture data that can be obtained with the limited number of in-situ measurements.

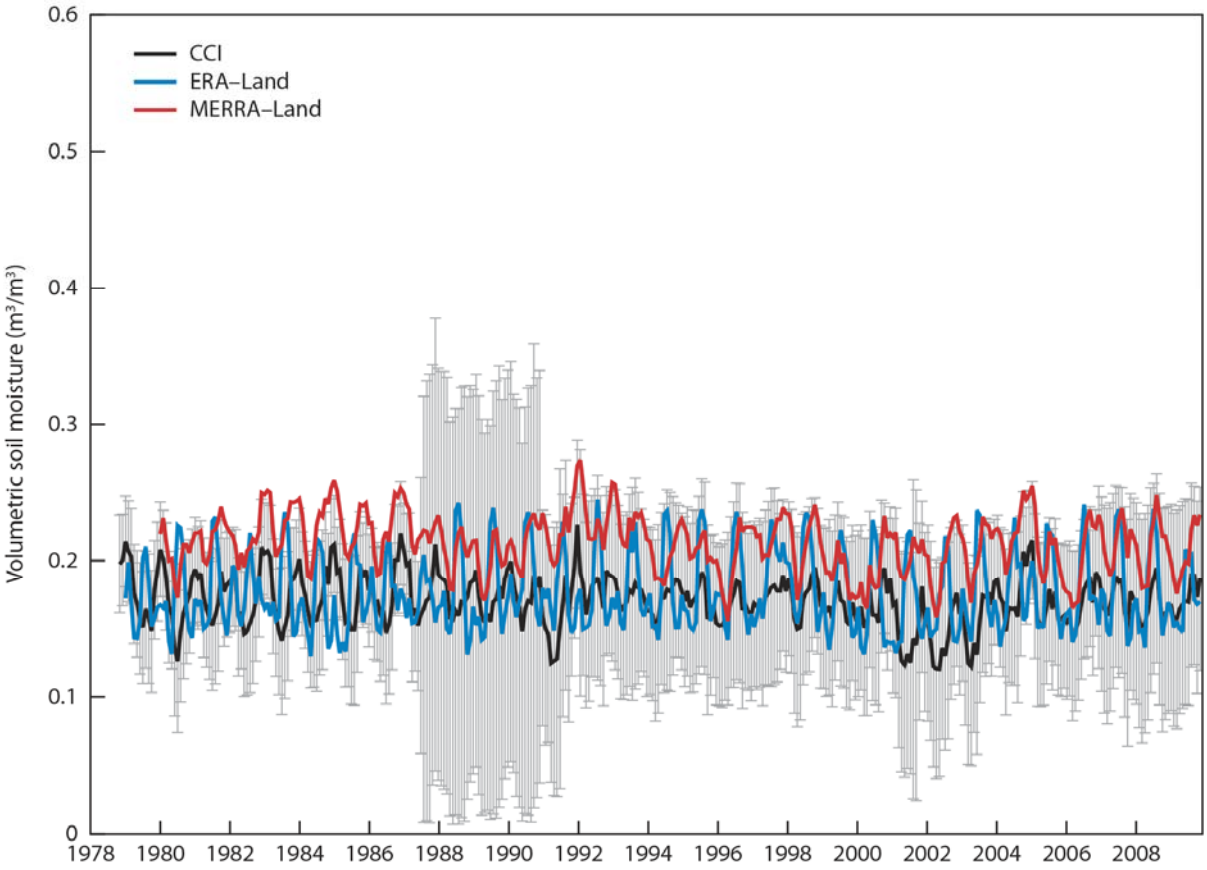

Fig 22: Like in figure 21 but data are averaged over India.

Domain	*CCI SM*	*ERA-Land*		*MERRA-Land*	
		Absolute bias	Relative bias	Absolute bias	Relative bias
Land	0.185	0.084	45.5	0.053	28.5
North Atlantic^{+}	0.212	0.096	43.6	0.045	21.2
North Pacific^{+}	0.198	0.117	58.9	0.031	15.7
Europe	0.222	0.054	24.5	0.043	19.5
Africa	0.259	-0.092	-35.5	0.041	16.0
Asia	0.182	0.118	64.6	0.047	26.0
Australia	0.232	-0.060	-25.8	0.026	11.4
North America	0.181	0.118	65.6	0.035	19.5
South America	0.161	0.181	112.2	0.053	32.7
Scandinavia	0.210	0.105	50.0	0.081	38.3
U.S.A	0.172	0.103	59.7	-0.009	-5.4
Indonesia	0.249	0.133	53.6	0.083	33.5
China	0.250	0.021	8.6	0.047	18.8
India	0.170	0.006	3.7	0.043	25.2
Siberia	0.239	0.102	42.6	0.011	4.6
Euro-Russia	0.234	0.077	32.9	0.000	-0.0
Amazon	0.214	0.166	77.3	0.067	31.1

Table 7: Mean CCI SM and mean (absolute and relative) residuals between each reanalysis dataset and the CCI data set computed over different regions and for the whole period of data availability. The relative residuals are computed as 100.(Reanalysis-CCI)/CCI, and are given in %. Mean SM values and absolute residuals are in m^3/m^3. $^{+}$The mean over these areas is the mean over the emerged regions included within their borders (North Atlantic [70N,30N]-[270W,30E]; North Pacific [70N,30N]-[120W,270E]).*

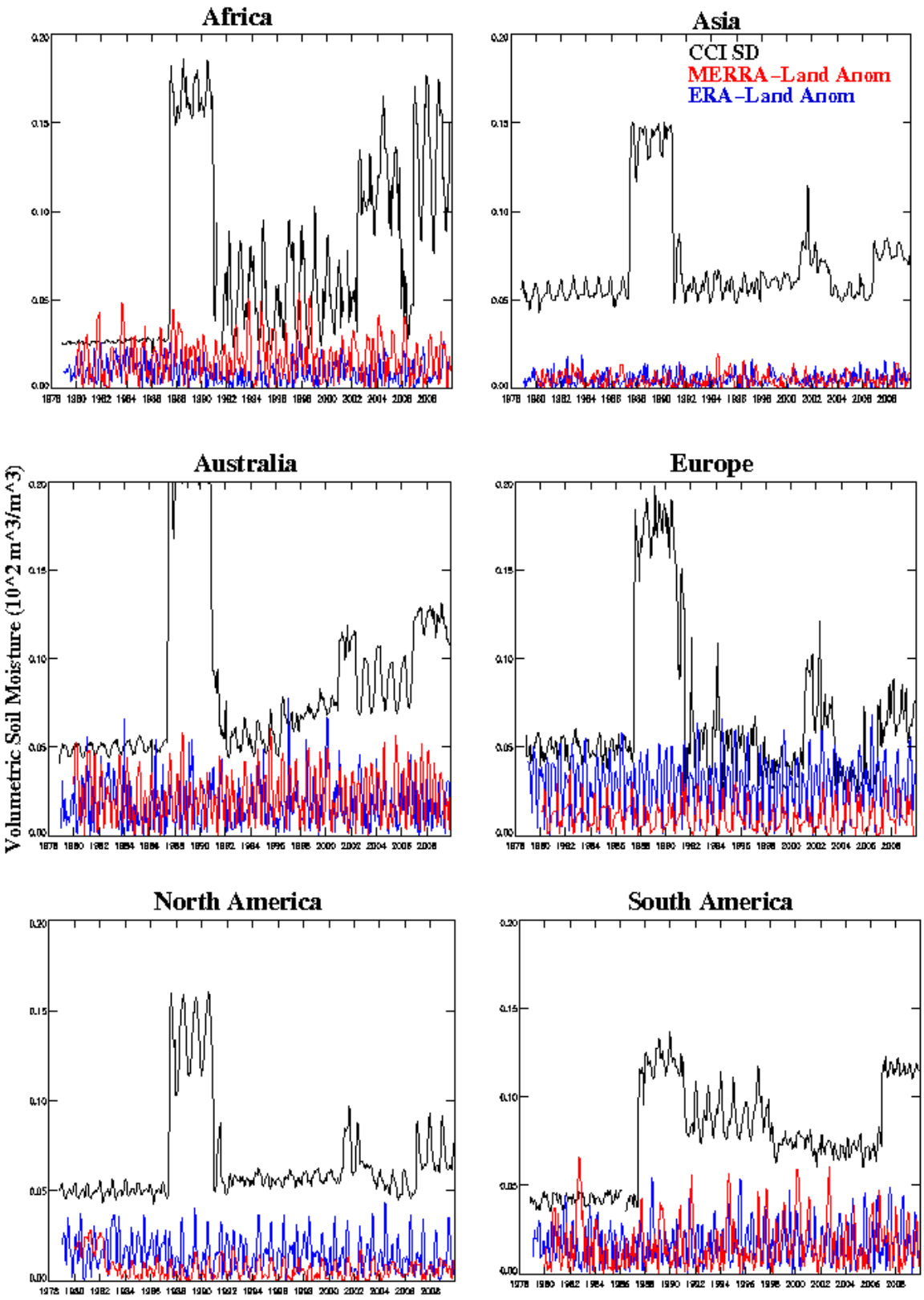

Fig 23: Comparisons between the CCI SM standard deviation (black) and the absolute anomaly from ERA-Land (blue) and MERRA-Land (red) over different geographical regions as given in each panel's title.

Figure 23 shows the comparisons between the CCI SM standard deviation (as given by Eq. 1) and the ERA-Land and MERRA-Land anomalies (as discussed in section 3) computed over different geographical areas. The plots show that the standard deviation has a strong dependence on the instrument. Although this could be expected, it confirms the poor long-term homogeneity of the product as already discussed. Furthermore, the observation standard deviations appear to be larger than both anomalies that arguably can make the dataset of little use to improve model parameterization or constrain soil moisture analyses.

A key element of the CMF is the ability of assessing the consistency between different related variables. It was mentioned in CMUG (2013b) and reminded at the beginning of the present document that the data base that serves the CMF (CMFDb) includes a large number of parameters including many model outputs that are not always well observed if observed at all. It is possible then to explore the level of consistency of each of the CCI ECVs with respect to any of the other variables in CMFDb. As done above, we have extracted the data from CMFDb to perform additional analysis. Figure 24 shows for example a scatter plot of the SM observations versus the ERA-Interim[2] precipitation (top left), evaporation (top right), and their residual, precipitation minus evaporation, (bottom) computed over the U.S.A. for the entire period of data availability of the used datasets. The correlation between each pair of datasets is given on the top right corner of each panel.

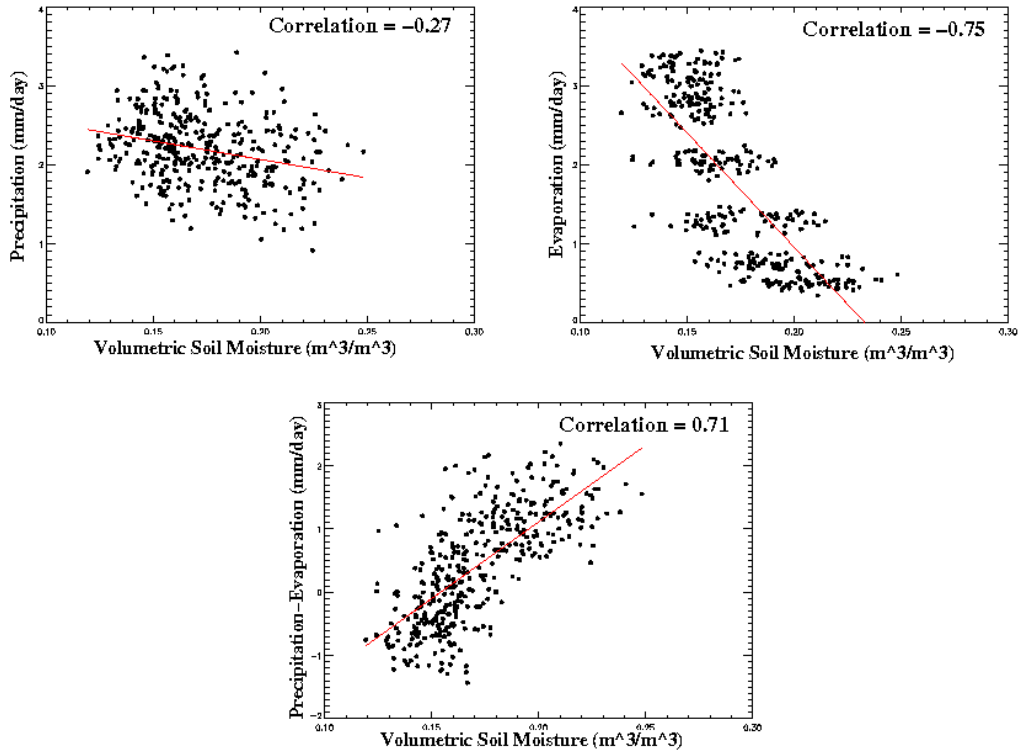

Fig 24: Scatter plot between the SM observation and the ERA-Interim Precipitation (top left), Evaporation (top right), and the Precipitation-Evaporation residual (bottom) over the U.S.A. for the period November 1979 to December 2009. The linear fit is over-plotted (red line). The correlation between each pair of datasets is given on the top right corner. Precipitation, Evaporation and their residual are in mm/day, SM is in m^3/m^3.

[2] Note this is not ERA-Land.

The residual between precipitation and evaporation (P-E) can be regarded as a good proxy of the soil water content. Thus, in first approximation, these two variables should show a good level of correlation, as for instance over the USA where the correlation is 0.71 (figure 24). However, this is not always the case when other geographical areas are considered. Table 8 summarises the correlation between the SM data and these three model variables (P, E, P-E) from the ERA-Interim production over a number of geographical areas. Although other mechanisms besides P and E may also be important in characterizing the soil water content and explain some of the low correlations in table 8 (e.g. the soil composition and saturation), the level of agreement or disagreement between these datasets might also indicate shortcomings and issues in the model land-sea schemes.

	ERA-Interim		
	Precipitation (P)	Evaporation (E)	P-E
Land	0.354	0.049	0.220
20N-60N	0.138	0.174	-0.014
60N-90N	0.238	0.150	0.154
North Pacific[+]	0.520	0.474	-0.280
Europe	-0.012	-0.335	0.345
Africa	-0.098	-0.458	0.121
Asia	0.344	0.264	0.417
North America	0.464	0.323	-0.068
Scandinavia	0.214	0.209	-0.039
U.S.A	-0.267	-0.749	0.710
China	-0.507	-0.528	-0.473
India	-0.260	-0.451	-0.140
Siberia	-0.446	-0.450	0.227
Euro-Russia	-0.266	-0.446	0.419

Table 8: Correlation between the SM observations and the ERA-Interim (note not ERA-Land) model variables for Precipitation (P), Evaporation (E), and their difference (P-E) computed for various geographical areas. [+]The mean over this area is the mean over the emerged regions included within their borders (North Pacific [70N,30N]-[120W,270E]).

This consistency assessment between SM and modelled precipitation/evaporation analyses from ERA-Interim can be regarded as a demonstration of what the CMF can offer. More thorough assessments will be possible in future as the CMFDb expands and more variables from other data streams will become available.

5. Summary

This document aimed at demonstrating the functionality of the Climate Monitoring Facility (CMF) and in doing so at providing an assessment of the first version of the CCI Ozone, Aerosol, and Soil Moisture level 3 (L3) datasets.

CMF, developed by ECMWF and put at the disposal of CCI, allows users to assess the quality and long-term homogeneity of a given data record based on model-observation confrontation. It makes use of a web interface to retrieve, manipulate and plot time-series of datasets from several data streams. At the time of writing, the still evolving database that serves CMF included over 90 different variables (available as monthly mean area averages), and about ten data streams, including various reanalyses and the first version of some of the CCI datasets. Details on the CMF design and database content were given by CMUG (2013a, b).

To facilitate the present assessment, the CMF database (CMFDb) was further extended with monthly mean area mean quantities and area characteristic standard deviations (defined according to Eq. 1 in section 3) computed for the CCI L3 data sets retrieved by the Ozone, Aerosol, and Soil Moisture teams. The monthly mean area mean quantities have been compared with all available reanalyses (e.g. ERA-Interim, MACC etc.), depending on the ECV. The standard deviations for the ozone products were compared with the spread of an ensemble of reanalyses produced as pilot simulation of the ERA-20C reanalysis performed within the ERA-CLIM project. It was discussed that the ensemble spread provides information about the internal climate variability of the ECV under consideration and it can be used to assess the observation uncertainty. The method was illustrated for the ozone ECV, the only variable available in CMFDb from the ERA-20C run among the three ECVs considered in the present assessment. In the case of aerosols and soil moisture, we have used reanalysis anomalies as a proxy for the observation errors, based on the idea that a dataset with errors larger than the field anomaly can unlikely provide useful information.

Summary on the ozone assessment:
Three different L3, merged (when available) products were assessed for ozone (total column ozone, TCO3, and ozone profile from the nadir, NPO3, and limb, LPO3, instruments) over the entire period of data availability.

Two situations (in spring 1997 and in summer 2002) show a sharp change in the TCO3 time series but after further investigation they are likely to be the consequences of anomalous ozone conditions that occurred at high latitudes in the NH in March 1997, and in the SH in mid-2002. The agreement between the two ECMWF reanalyses and the CCI total column ozone (TCO3) was expected to be good and generally within the observation standard deviation as the former were constrained with retrievals from instruments that were also used to produce the merged CCI data record (i.e. GOME, OMI and SCIAMACHY). This was generally confirmed by the comparisons except for ERA-Interim in the tropics where the level of agreement varies with time with some dependence on the observing system assimilated in the reanalysis. Until 2002, ERA-Interim assimilated the GOME ozone profiles, yet the residuals from the CCI TCO3 (consisting of GOME TCO3 data during this period) are larger than the observation error (about 5DU). In periods during which the ERA-Interim reanalysis assimilated limb ozone profiles (from either MIPAS or MLS) the level of agreement between

the two datasets seems to improve with mean residuals smaller than the observation standard deviation. All that said comparisons between the observation errors and the ensemble spread showed that in the tropics the former did not represent the same variability as shown by the ensemble with values being often three to four times smaller. Indeed if that was accounted for, the level of agreement between the two reanalyses and the CCI TCO3 could have been within the observation error also in the tropics. The comparisons with the JRA-25 ozone reanalysis - only constrained by TOMS data - show residuals as large as 25DU that exceed the observation errors at most latitudinal bands. The assimilation of the CCI retrievals would likely provide an added value to future JRA ozone reanalyses and bring them closer to the European ones. Predicting the added value of the CCI TCO3 in the two European systems is less trivial and can only be quantified through assimilation experiments. It is noted, however, that if, on one hand, there are indications that this data record could improve the ERA-Interim replacement if it was used in place of the original retrievals; on the other hand, largely underestimated errors can potentially degrade the resulting analyses if the observations are of poorer quality than the other constrains used.

The L3 ozone profiles from nadir instruments (NPO3) generally exhibited much smaller values than the ERA-Interim and MACC ozone reanalyses across the stratosphere, with stratospheric residuals larger than the observation errors at all latitudes. The observation errors seem to be about half the ensemble spread at most levels and latitudinal bands. Even if larger errors were accounted for, the mean observation minus reanalysis residuals would still exceed the data one standard deviation. The only case where the ERA-Interim and NPO3 agreement is particularly good, and within the observation errors is in the tropics at 10hPa during 1997 when the ERA-Interim assimilated the RAL GOME ozone profiles (which is the actual precursor of the CCI NPO3 product in 1997). While in 2008, it shows too low values.

The merged, zonally averaged limb ozone profiles (LPO3) showed a good agreement with both ERA-Interim and MACC reanalyses at most stratospheric levels below the ozone mixing ratio at 10 hPa, where, however, the observation errors appeared to be largely overestimated. Above the ozone maximum (5 hPa), the agreement with ERA-Interim is also good at all latitudinal bands, while that with MACC is within the observation errors only at high latitudes. This is believed to be a consequence of the use in MACC of an ozone bias correction anchored to SBUV observations. In general, these observations have normally larger weight at the ozone peak and just above it at midlatitudes and in the tropics where the UV instruments have larger sensitivity than for example at high latitudes. In the case of MACC, the weight the SBUV had on the analyses was further increased by fact they anchored the ozone bias correction. The impact was higher in regions of the atmosphere and latitudinal bands where the SBUV instruments are most sensitive and reliable. It is reasonable to believe that, as by design, the L2 LPO3 products can potentially improve the quality of reanalyses if used, particularly in periods where for instance MLS was unavailable. It was noted that having over-estimated uncertainties (also on the L2 retrievals) could potentially limit their beneficial impact on the analyses.

Summary on the aerosol assessment:
Three CCI aerosol datasets (ADV, ORAC and SU) were compared with the MACC aerosol reanalyses at four wavelengths over different geographical regions during 2008. The

comparisons show that the MACC reanalyses are in better agreement with the SU and ADV datasets than with the ORAC one. The actual level of agreement strongly depends on the geographical area and wavelength considered. The indication is that the SU dataset performs better than the other two at the lowest wavelengths (550 to 865 nm) showing in general better agreement with MACC in 50% or more of the cases. In contrast, the ADV dataset is the one that agrees better with MACC at 1610 nm (in 64% of the cases). Across the whole spectral range, the ADV algorithm seems to be closer to MACC over land, while the SU one gives the smallest residuals over oceans. A number of shortcomings in the MACC aerosols reanalyses were discussed and should be taken into account, for instance the 20% positive bias against the AERONET network in the summer months over land at 550 nm caused by a bias in the MODIS data, or the positive bias over the southern oceans. Bearing those shortcomings in mind, an assessment of the data uncertainty was provided. The SU and ADV observation uncertainties are consistent with the observation minus MACC AOD residuals. In contrast, the ORAC uncertainty is normally smaller than the difference between the retrievals and their MACC monthly mean AOD equivalent.

Summary on the soil moisture assessment:
The L3 CCI soil moisture (SM) was compared with reanalyses of SM from ERA-Interim, and two off-line land-surface simulations called ERA-Land and MERRA-Land, associated to the ERA-Interim and the NASA MERRA atmospheric reanalyses, respectively. When averaged over extended geographical regions the ERA-Interim and ERA-Land SM showed negligible to small differences, so only the comparisons against the latter were discussed in detail. The merged CCI SM dataset show poor long-term homogeneity with a clear dependence on changes in the observing system used. This aspect should be improved in future versions as it affects estimate of long-term variability and trend calculations. The comparisons with the land-surface SM simulations show a higher agreement of CCI SM with MERRA-Land than ERA-Land. Depending on the geographical region, differences can be as large as 50% or more from ERA-Land and typically 20-30% from MERRA-Land, with CCI SM typically exhibiting smaller values than the two SM reanalyses. The residuals from ERA-Land were normally larger than the observation errors; the residuals from MERRA-Land were at best as large as the observation errors with a few exceptions. It was noted that the major limitation in assessing SM observations with model outputs consists in matching the information provided by the instruments (that are sensitive to the Earth's surface) with that provided by models, typically a volumetric soil moisture defined over a layer with thickness depending on the model. Yet, comparisons with model outputs are important to complement the assessment of remotely sensed soil moisture data provided by the limited number of in-situ measurements. The SM errors show a strong dependence on the instrument used, confirming the long-term homogeneity issues. Their values were compared with the absolute mean anomaly of ERA-Land and MERRA-Land SM. The former normally showed much larger values than the two latter datasets. An initial consistency assessment between the CCI SM dataset and the ERA-Interim modelled Precipitation, Evaporation, and their residual over various geographical areas was also provided.

The present document clearly demonstrates the ability of the CMF to be a valuable tool for assessing the quality of CDRs. CMF allows one to determine whether changes in the observing system affect the long-term homogeneity of a dataset to make it unsuitable e.g. for

long-term trend analysis. The availability of a large number of variables from a diverse set of data streams makes CMF a very flexible tool to assess whether observed changes in one variable are consistent with those in related ones. Having the statistics pre-calculated over a selected number of areas (over 38 different geographical areas are currently available) facilitates any assessment and makes CMF an ideal tool for providing a fast overview of the data quality over an extended period of time. It was also shown that thanks to the availability of an ensemble of reanalyses in the CMFDb, assessing the observation errors is also possible.

Finally, a word of precaution, like any other analysis tool CMF should be used for applications it was designed for. These include monitoring and assessing the low-frequency, multi-year variability of regional averages. CMF was not designed for example to analyse a single event at one given time for which other tools might be more appropriate, unless that event could produce on monthly mean area averaged fields a detectable signal affecting their long term variability. For example, volcanic eruptions can produce a change in variables such as temperature or atmospheric composition that can be detected in the data time-series and displayed by CMF. However, CMF in its current design cannot show the geographical distribution at a given time of the volcanic plume.

6. Appendix A: Statistics on CCI aerosol AODs vs. MACC AODs

Domain	550 nm			659 nm	
	SU	*ADV*	*ORAC*	*SU*	*ADV*
Global	-0.006±0.007	-0.015±0.016	-0.069±0.023	-0.006±0.007	-0.017±0.015
Oceans	-0.037±0.016	-0.012±0.015	-0.064±0.020	-0.029±0.014	-0.009±0.013
Land	0.061±0.013	-0.022±0.020	-0.078±0.031	0.044±0.014	-0.038±0.017
North Atlantic	-0.064±0.015	-0.048±0.011	-0.095±0.054	-0.058±0.011	-0.045±0.009
North Pacific	-0.080±0.035	-0.043±0.029	-0.138±0.066	-0.068±0.029	-0.040±0.028
Tropical Oceans	-0.029±0.012	-0.031±0.018	-0.092±0.017	-0.025±0.009	-0.030±0.014
Southern Oceans	-0.064±0.018	-0.063±0.010	-0.075±0.015	-0.058±0.016	-0.057± 0.010
Europe	-0.082±0.018	-0.093±0.024	-0.114±0.056	-0.075±0.013	-0.090±0.022
Africa	-0.026±0.024	-0.088±0.038	-0.168±0.059	-0.021±0.024	-0.084± 0.031
Asia	-0.072±0.052	-0.052±0.061	-0.151±0.064	-0.064±0.046	-0.054±0.055
Australia	-0.032±0.011	-0.035±0.009	-0.058±0.014	-0.028± 0.010	-0.032±0.009
North America	-0.060±0.026	-0.034±0.019	-0.103±0.047	-0.051±0.022	-0.029± 0.018
South America	-0.022±0.009	-0.037±0.010	-0.078±0.024	-0.019±0.008	-0.032±0.008
Scandinavia	-0.039±0.037	-0.046±0.021	-0.092±0.035	-0.041±0.029	-0.047±0.019
Britain	-0.114±0.043	-0.072±0.043	-0.046±0.162	-0.105±0.040	-0.072±0.042
Central Europe	-0.068±0.038	-0.094±0.033	-0.073±0.051	-0.065±0.030	-0.096±0.028
Southern Europe	-0.114±0.034	-0.097±0.028	-0.084±0.046	-0.099±0.028	-0.091± 0.026
U.S.A	-0.095±0.029	-0.098±0.031	-0.173±0.044	-0.081±0.023	-0.085±0.027
Indonesia	-0.027±0.015	0.006±0.015	-0.064±0.018	-0.027±0.012	0.002±0.014
China	-0.123±0.063	-0.120±0.080	-0.323±0.108	-0.114±0.056	-0.133±0.073
India	-0.164±0.127	-0.147±0.137	-0.365±0.160	-0.151±0.111	-0.151±0.118
Siberia	-0.047±0.024	-0.037±0.045	-0.084±0.069	-0.043±0.021	-0.037±0.046
Euro-Russia	-0.064±0.017	-0.077±0.037	-0.118±0.071	-0.059±0.015	-0.080±0.037
Congo	-0.064±0.046	-0.065±0.058	-0.216±0.072	-0.059±0.035	-0.067±0.038
Amazon	0.032±0.016	-0.033±0.012	-0.084±0.039	0.023±0.012	-0.033± 0.014
Total	12	11	2	14	11

Tab A1: Mean residual between each AATSR dataset and the corresponding MACC AOD computed over different regions in 2008 for the 550 and 659 nm. The green shaded boxes refer to the dataset that best perform over a given region at a given wavelength. The bottom row provides the number of area over which the three datasets best perform.

Domain	*865 nm*		*1610 nm*	
	SU	ORAC	SU	ADV
Global	0.004±0.007	-0.035±0.012	0.048±0.011	0.041±0.01
Oceans	-0.010±0.011	-0.051±0.010	0.041±0.008	0.058±0.006
Land	0.035±0.016	0.011±0.028	0.061±0.024	-0.011±0.009
North Atlantic	-0.042±0.007	-0.028±0.026	0.007±0.017	0.022±0.014
North Pacific	-0.046±0.019	-0.010±0.046	0.012±0.011	0.033±0.014
Tropical Oceans	-0.005±0.007	-0.062±0.014	0.066±0.012	0.064±0.004
Southern Oceans	-0.047±0.015	-0.088±0.008	-0.016±0.018	-0.014±0.018
Europe	-0.056±0.008	-0.051±0.030	0.010±0.020	-0.005±0.011
Africa	0.005± 0.028	-0.069±0.014	0.103±0.042	0.043±0.010
Asia	-0.040±0.035	-0.056±0.049	0.039±0.016	0.043±0.020
Australia	-0.015± 0.009	-0.063±0.009	0.027±0.010	0.023±0.016
North America	-0.030±0.015	-0.032±0.030	0.025± 0.008	0.046±0.007
South America	-0.006±0.007	-0.051±0.007	0.040±0.007	0.037±0.006
Scandinavia	-0.036±0.021	-0.021± 0.029	-0.005± 0.017	-0.004±0.024
Britain	-0.036±0.021	-0.073 0.038	-0.039±0.050	-0.015± 0.060
Central Europe	-0.052±0.020	-0.026±0.099	-0.004±0.014	-0.035±0.009
Southern Europe	-0.070±0.019	-0.061±0.021	0.005±0.014	0.002±0.009
U.S.A	-0.056±0.015	-0.038±0.043	0.003±0.008	0.004±0.007
Indonesia	-0.014±0.010	-0.052±0.012	0.039±0.008	0.073±0.012
China	-0.081±0.042	-0.122±0.055	0.037±0.015	-0.005±0.019
India	-0.104±0.080	-0.151±0.110	0.056±0.027	0.040±0.021
Siberia	-0.030±0.016	0.051±0.090	0.013±0.015	0.020±0.032
Euro-Russia	-0.043±0.012	-0.002±0.085	0.015±0.018	-0.014±0.015
Congo	-0.031±0.025	-0.068±0.022	0.075±0.016	0.069±0.023
Amazon	0.025±0.009	-0.017±0.009	0.072±0.008	0.045±0.016
Total	15	10	9	16

Tab A2: Like in table A1, but for the two longest wavelengths of 865 and 1610 nm.

7. Appendix B: List of acronyms

AATSR	Advanced Along-Track Scanning Radiometer
ACE-FTS	Atmospheric Chemistry Experiment - Fourier Transform Spectrometer
AMI	Active Microwave Instrument
AMSR-E	Advanced Microwave Scanning Radiometer - Earth observing system
AOD	Aerosol Optical Depth
ASCAT	Advanced SCATterometer
ERS-2	The second European Remote Sensing
GOME	Global Ozone Monitoring Experiment
GOMOS	Global Ozone Monitoring by Occultation of Stars
IFS	Integrated Forecasting System
MERRA	Modern Era Retrospective-analysis for Research and Applications
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding
MLS	Microwave Limb Sounder
MODIS	MODerate resolution Imaging Spectro-radiometer
OMI	Ozone Monitoring Instrument
OSIRIS	Odin Spectrometer and InfraRed Imaging System

SBUV	Solar Backscatter Ultra Violet
SCIAMACHY	SCanning Imaging Absorption SpectroMeter for Atmospheric CHartographY
SMMR	Scanning Multichannel Microwave Radiometer
SMR	Sub Millimetre Receiver
SSM/I	Special Sensor Microwave/Imager
TMI	TRMM Microwave Imager
TRMM	Tropical Rainfall Measuring Mission

8. References

Albergel, C, W Dorigo, R. H. Reichle, G. Balsamo, P. de Rosnay, J. Muñoz Sabater, L. Isaksen, R. de Jeu and W. Wagner (2013a), Skills and global trend analysis of soil moisture from reanalyses and microwave remote sensing, ECMWF Technical Memorandum 695.

Albergel, C, W Dorigo, G. Balsamo, J. Muñoz Sabater, P. de Rosnay, L. Isaksen, L. Brocca, R. de Jeu and W. Wagner (2013b), monitoring multi-decadal satellite earth observation of soil moisture products through land surface reanalyses, ECMWF Technical Memorandum 697.

Auligné T, McNally A, Dee D. (2007). Adaptive bias correction for satellite data in a numerical weather prediction system. *Q. J. R. Meteorol. Soc.*, **133**, 631--642.

Balsamo, G., C. Albergel, A. Beljaars, S. Boussetta, E. Brun, H. Cloke, D. Dee, E. Dutra, F. Pappenberger, P. de Rosnay, J. Muñoz Sabater, T. Stockdale, F. Vitart (2012), ERA-Interim/Land: A global land-surface reanalysis based on ERA-Interim meteorological forcing, ECMWF Technical Memorandum, ERA Report Series n. 13, available from www.ecmwf.int/publications/library/do/references/list/782009.

Benedetti, A., J.-J. Morcrette, O. Boucher, A. Dethof, R. J. Engelen, M. Fisher, H. Flentje, N. Huneeus, L. Jones, J. W. Kaiser, S. Kinne, A. Mangold, M. Razinger, A. J. Simmons, M. Suttie (2009), Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: 2. Data assimilation, J. Geophys. Res., 114, D13205, doi:10.1029/2008JD011115.

Cariolle D, Déqué M. (1986). Southern hemisphere medium-scale waves and total ozone disturbances in a spectral general circulation model. *J. Geophys. Res.*, **91**, 10825–10846.

CMUG, 2012: CMUG ECV Soil Moisture Assessment Report. Technical Note/Deliverable 3.1v_1b, version 0.5, November 2012.

CMUG, 2013a: Status of the ECMWF Climate Monitoring Database Facility. Technical Note/Deliverable 3.5, version 1.1, May 2013.

CMUG, 2013b: The role of Reanalysis in the production and quality assessment of CDRs. Technical Note/Deliverable 3.4, under revision.

Dee, D. P. and S. Uppala (2009). Variational bias correction of satellite radiance data in the ERA-Interim reanalysis, *Q. J. R. Meteorol. Soc*, **135**, 1830-1841.
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. Hólm, L. Isaksen, P. Källberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, and F. Vitart (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc*., **137**, 553-597.

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, DOI: 10.1029/2012GL052988.

Dragani R. (2009). Variational bias correction of satellite ozone data. Technical report R43.8/RD/0934. ECMWF. Available from R. Dragani (rossana.dragani@ecmwf.int).

Dragani, R. (2010). On the quality of the ERA-Interim ozone reanalyses. Part I: Comparisons with in situ ozone measurements, ECMWF, ERA report series 2.

Dragani, R. (2011). On the quality of the ERA-Interim ozone reanalyses: Comparisons with satellite ozone data. *Q. J. R. Meteorol. Soc.*, **137** (658), 1312--1326.

Dragani, R and A.P. McNally (2013). Operational assimilation of ozone-sensitive infrared radiances at ECMWF. Q.J.R. Meteorol. Soc., DOI 10.1002/qj.2106.

Evensen, G. (2003). The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation. *Ocean Dynamics*, **53**, 343–367.

Fisher, M. (2003). Background error covariance modelling. In *Proceedings of the ECMWF Seminar on Recent developments in Data Assimilation for Atmosphere and Ocean*, 8–12 September 2003, Reading, UK, 45–63.

Flemming J, Inness A, Jones L, Eskes HJ, Huijnen V, Schultz MG, Stein O, Cariolle D, Kinnison D, Brasseur G. (2011). Forecasts and assimilation experiments of the Antarctic ozone hole 2008. *Atmospheric Chemistry and Physics*, **11**, 1961--1977.

Froidevaux, L, YB Jiang, A Lambert, NJ Livesey, WG Read, JW Waters, EV Browell, JW Hair, MA Avery, TJ McGee, LW Twigg, GK Sumnicht, KW Jucks, JJ Margitan, B Sen, RA Stachnik, GC Toon, PF Bernath, CD Boone, KA Walker, MJ Filipiak, RS Harwood, RA Fuller, GL Manney, MJ Schwartz, WH Daffer, R.E. Cofield, DT Cuddy, RF Jarnot, BW Knosp, VS Perun, WV Snyder, PC Stek, RP Thurstans, PA Wagner (2008). Validation of Aura Microwave Limb Sounder stratospheric ozone measurement. *J. Geophys. Res*.,**113**.

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, A. Smirnov (1998), AERONET: A

federated instrument network and data archive for aerosol characterization, *Remote Sens. Environ.*, **66**, 1–16.

Houtekamer PL, Mitchell HL. 2001. A sequential ensemble Kalman flter for atmospheric data assimilation. *Mon. Weather Rev.*, **129**, 123–137.

Inness A, Baier F, Benedetti A, Bouarar I, Chabrillat S, Clark H, Clerbaux C, Coheur P, Engelen RJ, Errera Q, Flemming J, George M, Granier C, Hadji-Lazaro J, Huijnen V, Hurtmans D, Jones L, Kaiser J, Kapsomenakis J, Lefever M, Leitão J, Razinger M, A, Schultz MG, Simmons AJ, Suttie M, Stein O, Thépaut JN, Thouret V, Vrekoussis M, Zerefos C, and the MACC team. (2013). The MACC reanalysis: an 8 yr data set of atmospheric composition. *Atmospheric Chemistry and Physics*, **13**, 4073--4109.

MACC II, 2013: Validation report of the MACC reanalysis of global atmospheric composition. Period 2003-2011. Technical Note/Deliverable D_83.4, version 1.1, February 2013.

Mangold, A., H. De Backer, B. De Paepe, S. Dewitte, I. Chiapello, Y. Derimian, M. Kacenelenbogen, J.-F. Léon, N. Huneeus, M. Schulz, D. Ceburnis, C. O'Dowd, H. Flentje, S. Kinne, A. Benedetti, J.-J. Morcrette, O. Boucher (2011), Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: 3. Evaluation by means of case studies, J. Geophys. Res., 116, D03302, doi:10.1029/2010JD014864.

Morcrette, J.-J., O. Boucher, L. Jones, D. Salmond, P. Bechtold, A. Beljaars, A. Benedetti, A. Bonet, J. W. Kaiser, M. Razinger, M. Schulz, S. Serrar, A. J. Simmons, M. Sofiev, M. Suttie, A. M. Tompkins, A. Untch (2009), Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: Forward modeling, J. Geophys. Res., 114, D06206, doi:10.1029/2008JD011235.

Onogi K, Tsutsui J, Koide H, Sakamoto M, Kobayashi S, Hatsushika H, Matsumoto T, Yamazaki N, Kamahori H, Takahashi K, Kadokura S, Wada K, Kato K, Oyama R, Ose T, Mannoji N, Taira R. (2007). The JRA-25 Reanalysis. *J. Meteor. Soc. Japan*, **85**, 369-432.

Reichle, R. H., R. D. Koster, G. J. M. De Lannoy, B. A. Forman, Q. Liu, S. P. P. Mahanama, and A. Toure (2011), Assessment and enhancement of MERRA land surface hydrology estimates, *Journal of Climate*, **24**, 6322-6338, doi:10.1175/JCLI-D-10-05033.1.

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. Holbe (2005), The MODIS aerosol algorithm, products and validation, *J. Atmos. Sci.*, **62**, 947–973.

Tan, D., E. Andersson, M. Fisher, and L. Isaksen (2007), Observing-system impact assessment using a data assimilation ensemble technique: application to the ADM-Aeolus wind profiling mission, *Q. J. R. Meteorol. Soc.*, **133**, 381–390.

Uppala SM, Källberg PW, Simmons AJ, Andrae U, da Costa-Bechtold V, Fiorino M, Gibson JK, Haseler J, Hernandez A, Kelly GA, Li X, Onogi K, Saarinen S, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda MA, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosavac M, Fisher M, Fuentes M, Hagemann S, Holm E, Hoskins BJ, Isaksen L, Janssen PAEM, Jenne R, McNally AP, Mahfouf JF, Morcrette JJ, Rayner NA, Saunders RW, Simon P, Sterl A, Trenberth KE, Untch A, Vasiljevic D, Viterbo P, Woollen J. (2005). The ERA-40 re-analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961-3012.

Zhang, J., and J. S. Reid (2006), MODIS aerosol product analysis for data assimilation: Assessment of over-ocean level 2 aerosol optical thickness retrievals, *J. Geophys. Res.*, **111**, D22207, doi:10.1029/2005JD006898.