Danish Meteorological Institute

Ministry for Climate and Energy

Danish Climate Centre Report 12-03

Terrestrial albedo models: Comparing results from CMSAF EO data and a GCM

Peter Thejll

Colophone

Serial title:
Danish Climate Centre Report 12-03

Title:
Terrestrial albedo models: Comparing results from CMSAF EO data and a GCM

Subtitle:
Authors:
Peter Thejll

Other Contributers:
Responsible Institution:
Danish Meteorological Institute

Language:
English

Keywords:
remote sensing, radiance modelling, satellite data

Url:
www.dmi.dk/dmi/dkc12-03

ISSN:
1399-1957

ISBN:
978-87-7478-615-3 (online)

Version:
1

Website:
www.dmi.dk

Contents

1. Dansk resumé

Jordens klima afhænger af strålingsbalancen - balancen imellem indkommende kortbølget sollys og udstrømmende langbølget varmestråling. Balancen påvirkes af jordens reflektivitet - kaldet albedo - og dermed af jordoverfladens reflektivitet, skymængden og mængden af is og sne. Denne albedo kan måles fra rummet med satelitter, ved at tage billeder - billederne skal derefter behandles for at uddrage fysiske størrelser omkring overladens reflektivitet og skyerne. Dette kræver en model for hvordan overflader og genstande reflekterer lys - såkaldt radians-modellering. Kvaliteten af radiansmodelleringen afgør kvaliteten af de udledte oplysninger.

I det tyske meteorologiske institut's CMSAF projekt arbejdes der på at omdanne observationer fra et antal vejr-satelitter til geofysiske data og dette projekt indkaldte forslag til projekter der kunne hjælpe med evalueringen af udledte datas kvalitet. Vi indsendte et forslag til et Visiting Scientist projekt - og dette er rapporten om resultaterne.

Vi fandt at en bestemt form for anvendelse af de udledte data kunne bruges til at se om forskellige data er indbyrdes konsistente - det skal de jo meget gerne være for at man kan borge for deres kvalitet. Konsistensen skal være sådan at overflader og luften tilsammen reflekterer og spreder lyset, der kommer ind i atmosfæren på en sådan måde at det lys, der reflekteres ud i rummet igen, stemmer med hvad der foregik nede i atmosfæren og på overfladen. Ved at anvende vores metode var det muligt at afsløre kvalitets-problemer i CMSAF data - blandt andet ser der ud til at være problemer i enkelte måneder og der er specifikke problemer med overflade-albedo produktet fra CMSAF i vintersituationer med sne på jorden i skovområder.

Ved at gentage analysen på lignende data fra en klimamodel (kaldet en GCM) var det yderligere muligt at sammenligne kvalitets-niveauet - vi fandt at data fra en klimamodel er mere konsistent end data fra observationer. Det er konsistent med at GCM-data `skal` være indbyrdes konsistente for at modellen kan konvergere. Derudover fandt vi lignende årstidsvariationer i forklaringsgraden i modellen baseret på CMSAF data og en model baseret på GCM-data og dette må betyde at vores metode - baseret på lineær regression - ikke er fuldt ud tilstrækkelig - især til situationer med delvis skydække.

2. Abstract

Earth's climate is governed by the radiative balance, and this in turn is governed by the reflective properties of Earth - its atmosphere (clouds, haze) and its surface (deserts, snow and ice, oceans). The reflectivity - or albedo - of these surfaces can be measured from space using satellites. Pictures from satellites are reduced into geophysical data by radiance modelling.

The DWD's CMSAF group works on the reduction of satellite data into geophysical products and needed a way to test the quality of the derived data, and made a call for proposals for Visiting Scientist projects. We submitted an idea and it was chosen - this report is the result of the VS project.

We had the idea that a certain type of model, based on derived data and observed shortwave fluxes at the top of the atmosphere, could be used to test the consistency of the data products and observations. We based our idea on a simple regression model with CMSAF products for surface albedo, cloudiness etc. as the regressors and shortwave flux at the TOA as regressand.

We were able to show that the tool is useful for detecting data consistency problems - a few months worth of data stood out as being singularly different from the rest, and data quality problems were thus spotted.

We applied the same method to similar data from a general circulation model. We found that the consistency there was higher than in CMSAF data - this is because the data `must` be internally consistent or the GCM would not converge. It also shows that there is room for improvement in the reduction of CMSAF data from observations.

A seasonality in the consistency both for CMSAF data and GCM data shows us that there are limitations to our simple model - particularly in winters there are problems for snow-covered land areas with forest, and when the cloudiness fraction rises the ability of the simple regression model to match the observations falls. This may indicate an over-simplistic treatment of clouds in our model.

Terrestrial albedo models: Comparing results from CMSAF EO data and a GCM

Peter Thejll (pth@dmi.dk)
DKC, Danish Meteorological Institute, Copenhagen, Denmark

October 9, 2012

Contents

Abstract

The fluxes of incoming and outgoing terrestrial radiation must, in equilibrium, be consistent with the radiative properties of the Earth system that influences albedo - e.g. surface albedo, cloud presence, cloud type and so on. We can test this consistency by building simple models of albedo on the basis of the in- and out-going shortwave fluxes as well as system properties. We do this using observation-based data from the CMSAF project [Schulz et al., 2009], and data from the GLIMPSE [Stendel et al., 2006] coupled global climate model. The method allows evaluation of the satellite data quality, and comparison to GCM-data models allows an discussion of the relative properties of the albedo constituents.

The study concludes that albedo-modelling is a powerful way to check CMSAF data quality, as it easily spots data inconsistencies.

1 Introduction

The CMSAF project [Schulz et al., 2009] generates geophysically interesting fields (in time and space) from satellite observations. The fields generated include surface albedo, cloud fraction, cloud optical depth and radiation fields such as the outgoing shortwave radiation, and others.
There is an interest in the community to validate these data in order to make them useful for others. One interest is related to understanding how consistent the radiation fields are with the other derived properties such as surface albedo, cloud-presence and -type, etc. In the EO business in general there is an interest in understanding how correct, e.g. the outgoing shortwave radiation derived from EO data is, compared to reality - i.e. the question of terrestrial albedo - because such EO-based data is subsequently used by the climate modelling community as boundary values for climate modelling efforts. Recently, observations of Earthshine intensity [Pallé et al., 2004] have been used to evaluate the accuracy of EO-based albedo data, and it was shown that discrepancies exist.
We saw an opportunity to use CMSAF products to see if it is possible to understand the strengths and weaknesses of the data. We suggested that a model can be built relating the observed planetary albedo to such factors as land surface albedo, cloud presence and cloud type on a pixel to pixel basis. We also suggested that a similar model could be used on climate-model data for the same quantities and that a comparison of the EO albedo-model and GCM albedo-models could be made and some conclusions drawn about the data quality of both. This report details the investigations performed.

2 The EO Data

CMSAF data are available in many resolutions and with different types of spatial emphasis. We use monthly-mean Full Disk data, labelled "MA". The Full Disk refers to the part of the Earth that can be seen from MSG1, a geostationary satellite placed over Africa, viewing most of the hemisphere centred at lat,lon=0,0.
We are basing our analysis on the incoming and outgoing shortwave fluxes at the top of the atmosphere, the cloud fraction, as well as the surface albedo. These fields are called TRS, TIS, CFC and SAL. We also experiment with cloud optical depth, COT.

2.1 Reviewing the EO data

2.1.1 Tools for reading the CMSAF data

CMSAF data are packaged into `hdf5` and `netcdf` files and can be read using utilities, written in IDL, provided by the CMSAF.
After loading the various libraries needed, an essential piece of code to read and re-grid the data is:

```
; read the data
     data=CM_SAF_read_data(file)
     full_res= *data.data.(0).data

; regrid the data
     regrid=lonlat2reg(full_res, (*data.geolocation.lon), (*data.geolocation.lat),$
     lat0=-60.0,lon0=-60.0,dlon=dlon1,dlat=dlat1,nlat=nlat1,nlon=nlon1, $
     nodata_value=data.data.(0).nodata_value)

; extract the re-gridded data, the re-gridded longitude and the re-gridded latitude
     gridded=regrid.avg
     latgrid=regrid.lat
     longrid=regrid.lon
```

This step is included in the code provided at the end of this report.

2.2 Some EO data quality issues

During this project it quickly became evident that in the monthly mean full-disk data for TRS a daily mean has been substituted at one point. Closer inspection showed that almost all daily data for March 2009 are missing and that the sole existing daily mean is used as the monthly mean. The data for this month are not suitable for further use so March 2009 is omitted from further analysis.
During reading of some TRS fields in the monthly-mean full-disk set it also became clear that for some reason the data field is flipped along a meridian line, as if the longitudes have been reversed. The longitude array itself is the same for all years and months considered, but the data arrays for TRS appear flipped for several months in 2009, starting at February. Furthermore, this problem was not noticed while the software was used at the DWD, but became evident once the same code was used at DMI. Certainly the computer hardware is different at the two institutes, as are the versions of IDL used (7.1 at DWD and 7.0 at DMI), but the rest is the same - the data files having been transferred from DWD to DMI.
For the affected months a simple remedy was used - the data fields were flipped back.

2.3 Pre-analysis treatment of SAL

SAL is given in the monthly-mean full-disk product corresponding to solar zenith angle (SZA) 60 degrees. Of course, only a few pixels have experienced the Sun at this SZA but it was decided to publish the SAL product in this way. We shall call that SAL_{60} from now on. To use SAL_{60} in the way planned we need to convert it to what it would be at the actual SZA. This is done following the procedure described in the ATBD for SAL [Riihelä, 2009]:

$$SAL(\mu) = SAL_{60} \cdot \frac{1+d}{1+2d\mu}, \tag{1}$$

where d is a value, specific to each pixel, that describes the surface type, and $\mu = cos(SZA)$. d is given as a fixed-in-time lookup table by the IGBP (ref here), and $SAL(\mu)$ is the desired surface albedo at $\mu = cos(SZA)$ while SAL_{60} is the published value.

We need to take the published SAL_{60} values and find $SAL(\mu)$. We cannot undo the averaging step without going back to the diurnal-cycle mean products. As this would be quite an elaborate task we instead assume we can make some approximations

$$< SAL(\mu) >=< SAL_{60} \cdot \frac{1+d}{1+2d\mu} >\approx< SAL_{60} > \cdot < \frac{1+d}{1+2d\mu} >, \tag{2}$$

where $<>$ indicates averaging over time. The approximation is not too bad if surface albedo changes slowly over a month, and this will likely be the case for all surfaces that do not get covered by snow. We calculate the term $< \frac{1+d}{1+2d\mu} >$ by looking up the value for d for each pixel and calculating the position of the Sun in the sky, using IDL routines from the GSFC ASTRO library [Landsman, 1995], and averaging over those hours when the Sun is above the horizon for a given pixel for the month in question.

3 Regression model description

In principle, albedo is the ratio of reflected to incoming short-wave radiation. From satellite images the reflected shortwave radiation is determined in the CMSAF data product **TRS**, and is thus available on a pixel-by-pixel basis. At the same time CMSAF data products provide cloud-fraction **(CFC)** as well as cloud-optical depth **(COT)** at visible wavelengths, and the surface albedo product **(SAL)**. Similar data exist from GCM output. Combining the cloud cover, surface albedo and cloud optical depth, with total incoming solar radiation **TIS** we can make regression-based models of the albedo, for instance such as these:

$$\frac{TRS(t)}{TIS(t)} = a + b * CFC(t) + c * SAL(t) + d * COT(t) \quad (3)$$

$$\frac{TRS(t)}{TIS(t)} = a + b * SAL(t) * [1 - CFC(t)] + c * CFC(t) + d * COT(t) \quad (4)$$

$$\frac{TRS(t)}{TIS(t)} = a + b * SAL(t) * [1 - CFC(t)] + c * CFC(t) * (1 - e^{-COT(t)}) \quad (5)$$

Fitting the RHSs to the LHS with standard least squares minimisation, using IDL routine `regress`, we can determine the values of a,b,c and d. We found that the model in equation 4 above was best - In general, using **COT** did not improve model fit anywhere, and was omitted from what follows. We used 35 months of data from the 'full disc view' of the CMSAF data products. We used 36 months of data from the GLIMPSE model.

4 GCM data from the GLIMPSE model

By repeating the above analysis on data taken from a climate model we can understand some strengths and weaknesses of our approach. We have chosen to use the data from a 500-year run of a coupled atmosphere-ocean climate model, called the GLIMPSE model [Stendel et al., 2006].
GLIMPSE was produced at the DKC at DMI and used the ECHAM-4 atmospheric model coupled to the OPYC ocean model to generate 500 year long runs under the influence of time-varying forcings from Sun and volcanoes but constant greenhouse gasses. While this model is no longer considered modern the ease of access to the data made it a good choice for our project.
Using GCM data to build an albedo-model the opportunities and challenges are different:

1. The fluxes are available averaged over all directions - this is simpler than for the EO-data.

2. The surface albedo excludes snow cover - so we constructed a snow-albedo from the model snow depth assigning all grid points with at least some snow the albedo 0.45.

3. The GCM data are known to be internally consistent (otherwise the GCM would not have converged).

Not all the same fields were available from GLIMPSE as were used for the CMSAF data analysis in the previous section, but there was enough overlap that a meaningful comparative regression model could be developed. In all we had access to monthly mean values of net outgoing shortwave radiation at TOA, land surface albedo (time-variable over land only), and cloud fraction. The outgoing shortwave radiation can be calculated from the net if we also provide the incoming shortwave, which was done for each grid position using information on solar angle as a function of season and the solar irradiance used in the model (based on the Lean et al 1994, updated, reconstruction).
The albedo was calculated as the ratio of the outgoing shortwave and the incoming shortwave for all pixels where the Sun was above the horizon. The stored values of the radiation fields are net values - i.e. a summation of the radiance over all angles has been performed.
The regression model was based on the above albedo as regressand and cloud fraction and land surface albedo as regressors. An intercept was allowed.

5 Regression model results

5.1 Full disc view of residuals

In Figure 1 we show regression model residuals (that is, the difference between albedo and the regression model for albedo) for CMSAF data and GCM data. Top row is CMSAF data, bottom row is GCM data, left column is for NH Summer, and right column is for NH Winter. We note that residuals are largest during NH winter months, and smallest during NH summers – both for EO- and GCM-data. Note the particularly large residuals in the northern hemisphere at high latitudes during winter, and tropics, also in winter.

5.2 A closer look at model residuals based on CMSAF data

First, we look at the overall performance of our simple model, as a function of month number. In Figure 2 we see that the fraction of explained albedo variance (R^2) can reach 80% (in GCM data), is somewhat lower at best for CMSAF data, and that data outliers are immediatelys potted (left panel of

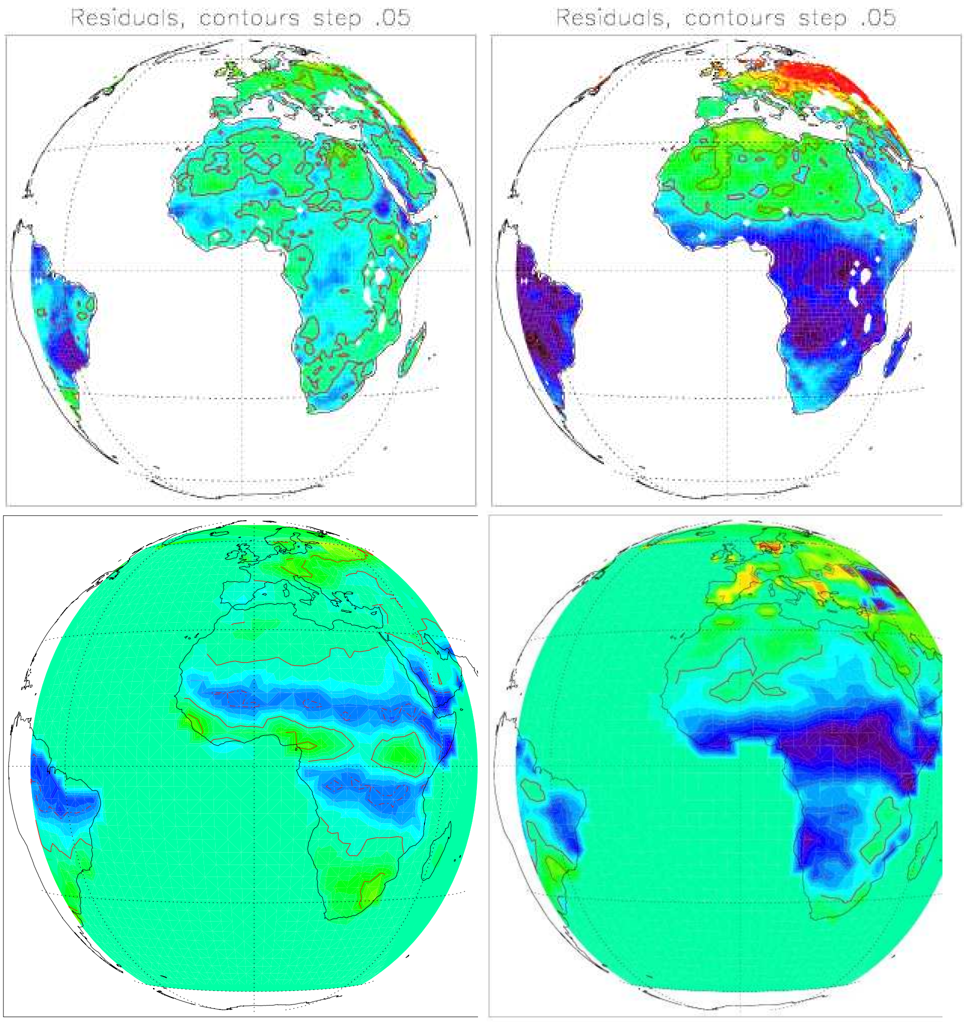

Figure 1: (**Top** left): Full disc view of residuals (observation-model) of the linear model fitted to CMSAF data in Equation 4 for the month of July 2007. (top right): Fit for CMSAF data in February 2008. (**Bottom** left): Full disc view of residuals of the linear model fitted to GCM data in Equation 4 for a model-July month. (bottom right): Fit on GCM data in a model-February. Colors give the magnitude of residuals: Reds, yellows and greens are positive, 0 is between green and light blue, light blue and dark blue and black are negative.

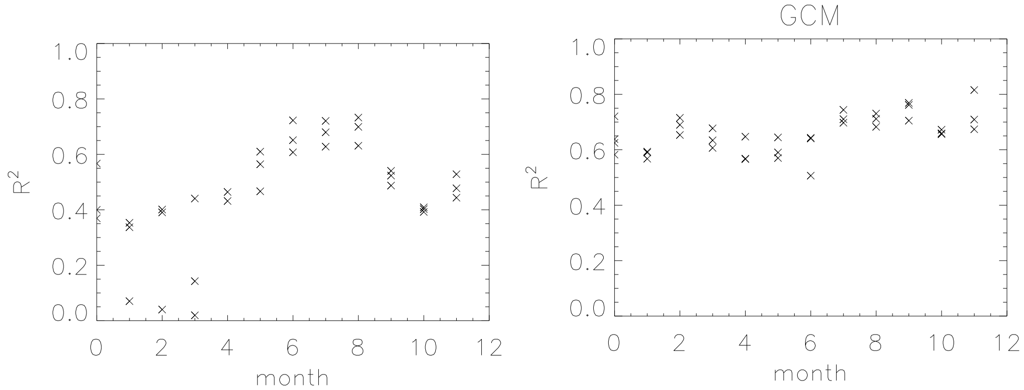

Figure 2: (left) Fraction of variance explained in the Full disc model based on CMSAF data. (right) Fraction of variance explained in the Full disc model based on GCM data. For each of the 35 (CMSAF) or 36 (GCM) months all the points in the monthly-mean field was used in the regression fit, so there is one point for each month - as we have several years of data some months have several points representing that month. Note the evident outliers in the CMSAF case, and the more constant value of R^2 across the months in the GCM case.

figure 2). Figure 5 shows more details of the results from the regression coefficients.

In Figure 3 we look at the role played by CMSAF surface land type on residuals. We see that residual sign is a strong function of surface type. These surface types are used, in CMSAF data processing, to deduce the SAL from radiance data observd by the satellite. Some surface types clearly induce a bias - the residuals are always positive for **EVERGREEN NEEDLE FOREST** and **MIXED FOREST**, and tends to be mainly negative for **CLOSED SHRUB**, **WOODY SAVANNA** and **EVERGREEN BROAD**. Shown as numbers are the counts of pixels involved in each type - and there are about 4770 pixels involved in the model. While land-only points have been the target a few ocean points have been included, near coastlines. Radiance modelling of trees and shrubs is known to be a difficult problem - snow in forested regions is very difficult to model because the albedo depends so much on whether snow is on the trees or on the ground. In the northern hemisphere winter problems seen in figure 3 we may be seeing that problem visualized.

In general, note that the choice of data determines the residuals - had 'forest only' pixels been chosen the residuals would have been different. The choice of 'all land pixels' places all surface types on a comparable basis

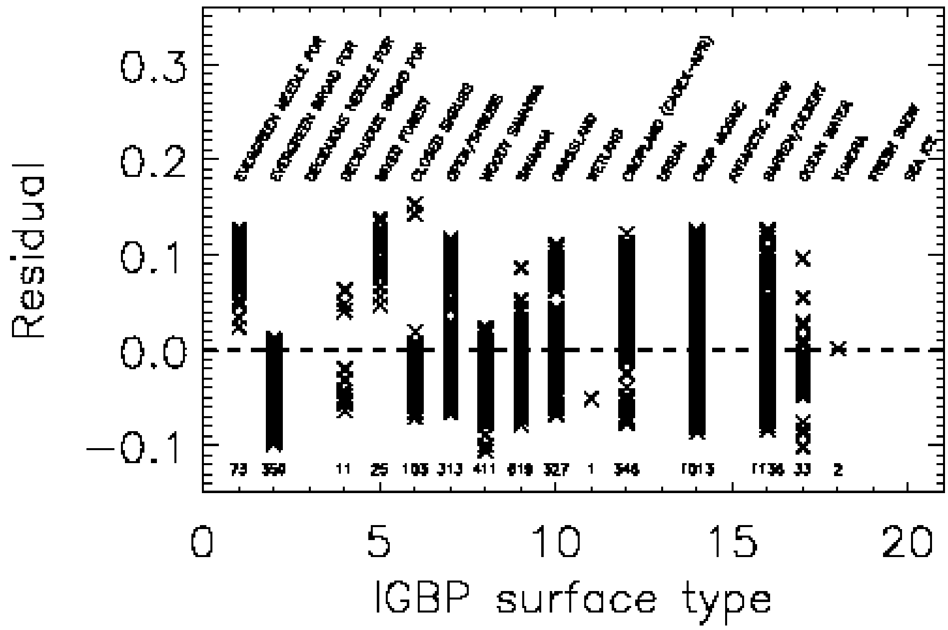

Figure 3: Residuals vs. surface type in absolute terms. The difference between 'observed' and modelled albedo is plotted against the IGBP surface type. The 'observed' albedo is that found from $\frac{TRS}{TIS}$ and the modelled albedo is the linear regression model based on SAL and CFC.

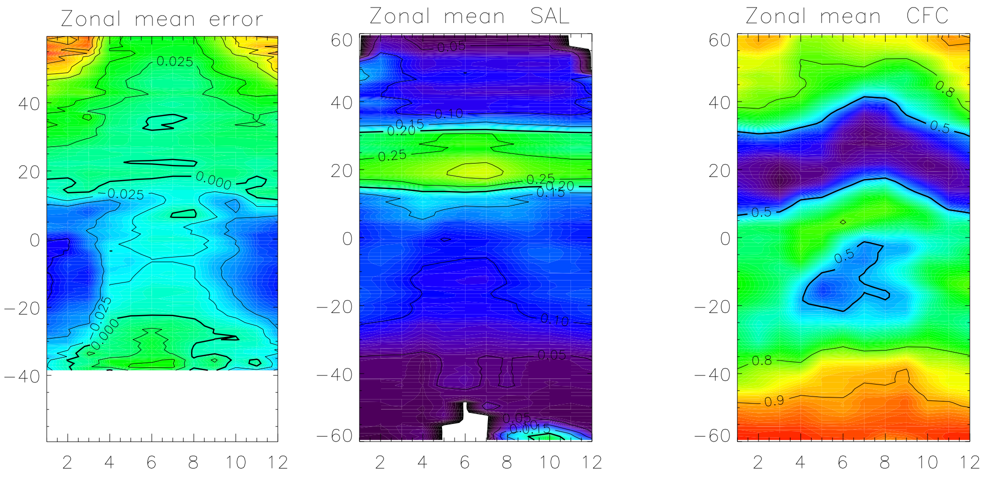

Figure 4: Zonal means of albedo residuals (left) in the CMSAF model, zonal mean of the SAL (middle) and zonal mean of the CFC (right) against month of the year. Averaging was performed over the Full Disc. Note that NH winter indicates a link between regression-model residuals and CFC, not SAL.

in the regression and the results indicate that the radiation field is incompatible with the model - this in turn implies that either the model is too simplistic or that the surface albedo and cloud fraction are not consistent with TRS. Figure 3 suggests that the regression model is to blame to some extent - it is too simplistic - but the problem at high latitudes in the CMSAF data reveal that other factors are needed to explain the residuals. Figure 4 illustrates the problems at high latitudes in winters, and suggest that SAL is not causing the problems as it has low spatial variability compared to the CFC. By restricting the regression in the GCM case to areas where no snow falls (essentially Africa) it is evident (not shown) that problems in the equatorial region persist - i.e. the problem of modelling the cloud contribution to albedo are therefore not a regression artefact of opposite-sign problems elsewhere. Similarly, restricting to latitudes above 20 degrees N shows that the high-latitude winter albedo problem disappears - i.e. that problem could be induced by the regression, as opposed to the equatorial cloud-problem.

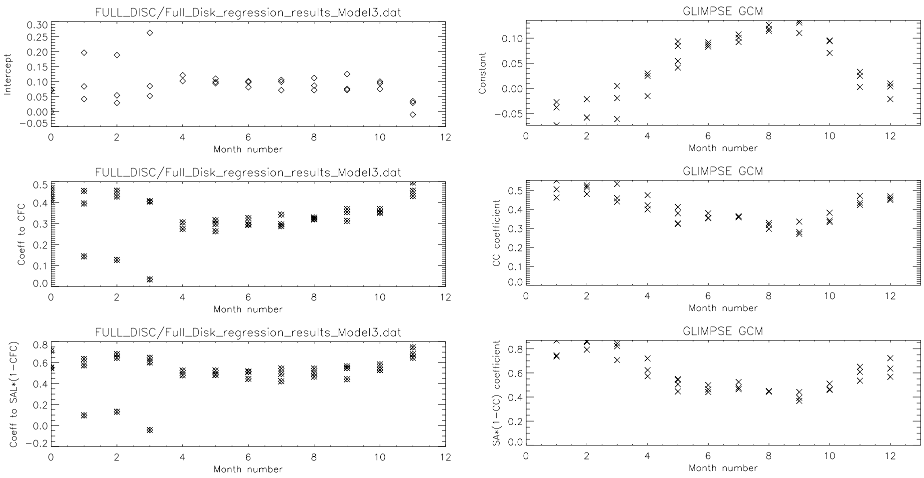

Figure 5: (left) Regression intercept and coefficients using 35 months of CMSAF data. (right) Regression intercept and coefficients using 36 months of GLIMPSE GCM data. Note the different scales due to CMSAF outliers.

6 Further insights from application of a simple model to CMIP model data

If the regression model used above was perfect, we should have R^2=1. We do not. We see seasonal variations, and we see residual outliers, probably due to some data inconsistency. Can we begin to understand why we have seasonal variations in R^2?

It is possible to simplify the radiative transfer problem to a one-layer model, so that surface albedo, atmospheric absorption and atmospheric reflection become the three unknowns in three (nonlinear) equations. With inputs of the upward and downward shortwave fluxes at TOA and at the surface, it is possible to solve for the three unknowns and come to an understanding of their relative importances. The three equations are [Donohoe and Battisti, 2011]:

$$F_{TOA}^{up} = S \cdot (R + SAL\frac{(1-R-A)^2}{1-SAL \cdot R}) \quad (6)$$

$$F_{surf}^{down} = S \cdot (\frac{(1-R-A)}{1-SAL \cdot R}) \quad (7)$$

$$F_{surf}^{up} = SAL \cdot S \cdot (\frac{(1-R-A)}{1-SAL \cdot R}), \quad (8)$$

where SAL is the surface reflectivity or albedo, R is (from now on) the fraction of light reflected by one passage of the atmosphere, and A is the fraction absorbed; F_{TOA}^{up} is the top of atmosphere upward flux, S is the top of atmosphere incoming shortwave flux, and F_{surf}^{down} and F_{surf}^{up} are the surface downward and upward fluxes, respectively. Given a dataset for these fluxes we can solve for SAL, R and A in each gridpoint, and then study the relationship between these quantities.

We have done this using monthly-mean data from three GCM datasets from the CMIP3 project. The CNRM, Bjerknes and Canadian 'control run' 500 year modelsets.

6.1 Results from analysis of CMIP data

We first determine SAL, R and A in each of the twelve months, taking an average over the 500 years for each month, in the European region between 29 and 66 deg N, 0 and 60 degrees E – see Figures 6–8. The derived SAL has been compared (not shown) to the ratio of surface up-welling and down-welling shortwave radiation and is identical to this.

We see that A is well-modelled by an exponential function of the cloud fraction. R is also dependent on cloud fraction but much less so. For cloud fraction = 0 (no clouds) we can extract (Figure 9) the intercepts of the fitted (red) lines for each month so that we gain a picture of what the absorbed and reflected fractions are in the absence of clouds.

6.2 Discussion of CMIP results

It seems that cloud absorption is an exponential function of cloud fraction. The clear-sky absorption seems to be seasonally controlled with more absorption in winter months.

Absorption of shortwave radiation occurs because of molecular absorption bands in water vapour and ozone. Both of these molecules have seasonal cycles - water vapor is mainly present in the region considered in the winter months, while ozone peaks in the Spring months - of these, water vapour is considered to remove more light than ozone, and we therefore expect the A term to be governed by the cycle in water-vapour. This is consistent with what we see.

The increment in reflection as cloud fraction grows from zero to one is small compared to the clear-sky reflection. Clear-sky reflection is almost constant given the inter-model spread.

We are accustomed to expect cloud absorption to be an exponential function of cloud optical depth - not of cloud fraction. A linear growth in absorption with cloud fraction seems to be more in line with expectations - but it is possible that the one-layer model has this as its peculiarity; also, we are not sure how total cloud fraction is determined from model data (the details are hidden in obscure literature). Could it be that cloud fraction is also a proxy for cloud optical depth? If that is so, then we begin to understand the result for A.

We also tend to expect R to be a strong linear function of cloud fraction - the more clouds, the more light is reflected. The relationship is seemingly linear, but the slope we find is small. Also, the clear-sky reflection (see Figure 9) is a large number compared to the contribution that a growing cloud fraction can make - the clear-sky reflection is near 0.2–0.3 and to that a fully cloudy sky can only add something like typically 0.1. Is that in line with expectations? [Donohoe and Battisti, 2011] show values for R in their Figure 2d, and the values are similar to ours. They do not present information on how R should change with cloud fraction, unfortunately.

The clear sky reflection is physically linked to processes like Rayleigh and Mie scattering, and scattering off large dust particles. The model atmosphere may or may not contain dust; but Mie and Rayleigh scattering are probably expressed from canonical formulae in the codes. It is not really conceivable that the formulae for Rayleigh scattering should be different between the GCM codes - and the atmospheric masses in the models are bound to be the same, so Rayleigh scattering should be the same; none of the scattering terms could lessen R but only add to it. Clearly the models behave differently in clear-sky conditions with two having similar minimum clear-sky scattering but the third GC model has almost twice as much minimal clear-sky scattering.

More variable still is each GC model's dependence of R on the cloud fraction - one model increments R by as little as 0.01 when going from clear-sky to fully cloudy conditions. For the same month another model increments R by eight times as much. As clouds are far more complex to model in GCMs, than are clear-sky opacities, we assume that the difference in sensitivity of R on cloud fraction `cl` (called CFC earlier in this text) is due to differences in model implementations.

Astronomers measure the transmission properties of the clear atmosphere frequently and express it as the extinction for a given observation. For 'one airmass' (a passage of a light beam towards or from zenith) the extinction is given as a function of wavelength[1] and is in the range 0.13-0.37 in units

[1] http://www.ls.eso.org/lasilla/Telescopes/2p2/D1p5M/misc/Extinction.html

of magnitudes per airmass (corresponding to 0.11 and 0.29 on a fractional scale) in the blue to visible range. We insert typical values for A and R for clear-sky conditions from our analysis here, into equation 7 and see that at the surface the received shortwave flux is expected to be about 0.7 of what the incoming flux was at the TOA - about 30% of the light being removed from the beam by absorption and scattering - not far from what is given by the extinction tables. With more information on which wavelengths are intended by the 'shortwave' data label used in the GCM datasets we could analyse further if the A term we find is realistic.
Are the numerical solutions for SAL, R and A precise? Inspection (not shown) of the solutions reveal that absolute errors are on the order of $4 \cdot 10^{-4}$ in units of W/m^2 between the left-hand and right-hand side of equations 1–3, above. It seems the solutions are numerically good.

6.3 Summary of CMIP-based results

We have found some interesting things as well as uncovered some yet unsolved problems: On the positive side we have

- The A and R terms together predict an extinction that is close to what is known from independent observation of transmission of light through the clear atmosphere.
- A has a seasonal cycle we seem to understand as being due to water vapour.

On the other hand we have some unsolved problems:

- We are only tentatively understanding why A is an exponential function of cloud fraction - is cloud fraction in the GCM data a proxy for cloud optical depth?
- Why is the R term such a weak function of cloud fraction?

These results suggest that an improvement of the simplistic regression model used above is possible by including an exponential term of the cloud fraction rather than the present linear term.

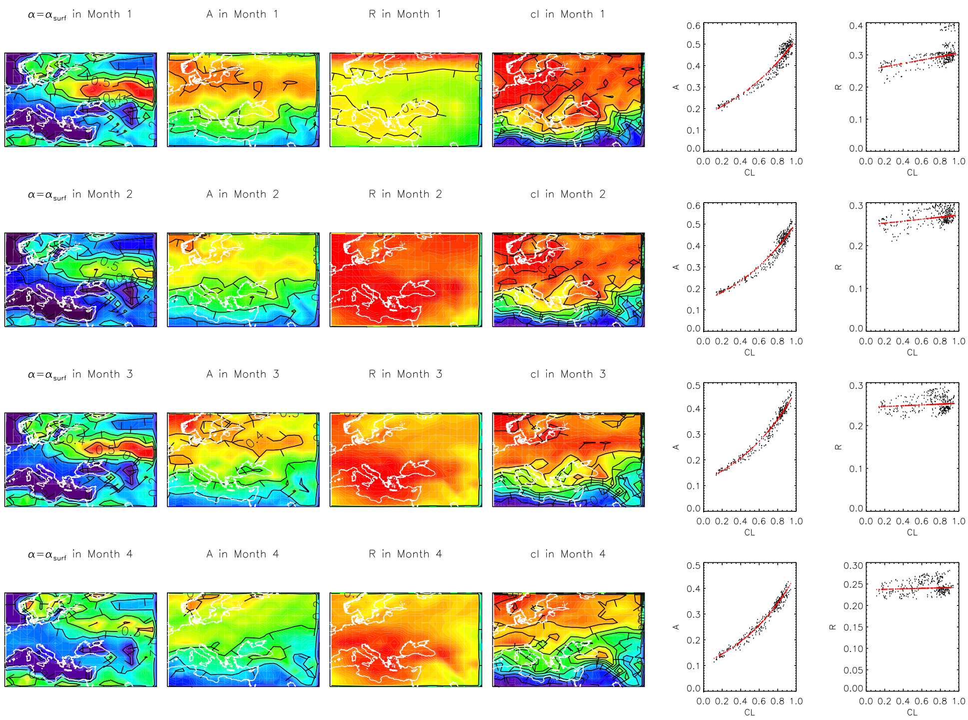

Figure 6: Results for SAL, R and A, from the CNRM GC model, for January to April. 500-year means. One row per month. First panel in a row gives the derived surface albedo. Second panel shows A, third shows R and fourth shows the cloud fraction for reference. Fifth panel in each row is a plot of the value of A as a function of the cloud fraction, over the model grid of points available. Sixth panel is a plot of R against cloud fraction. Overplotted (in red) on the last two panels are robust best-fit exponentials (for A) and straight lines (for R).

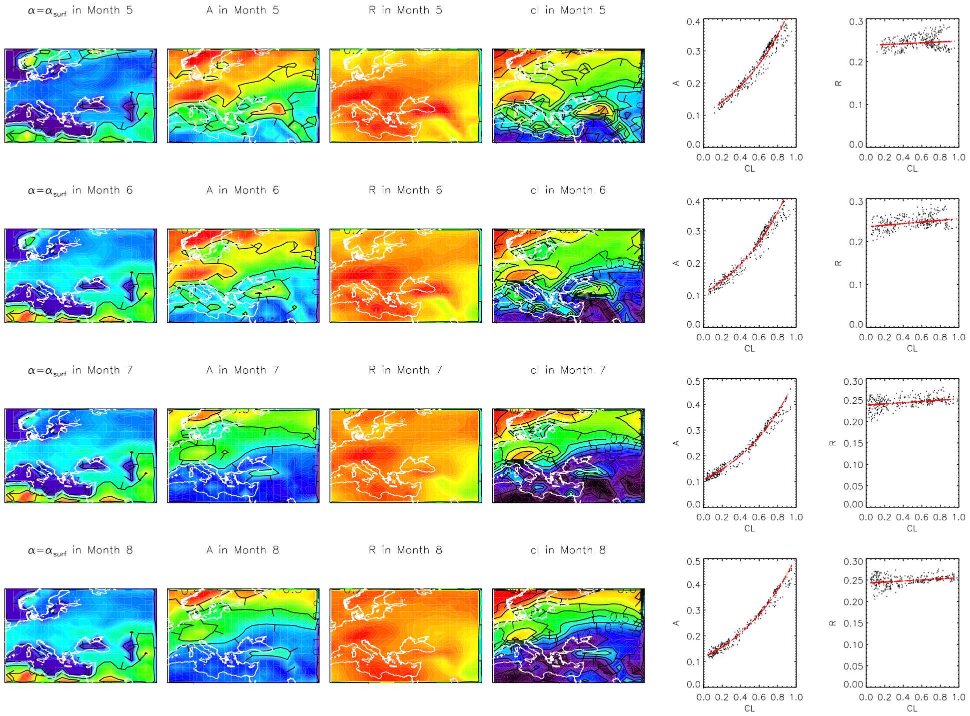

Figure 7: Like Figure 6, but for May to August.

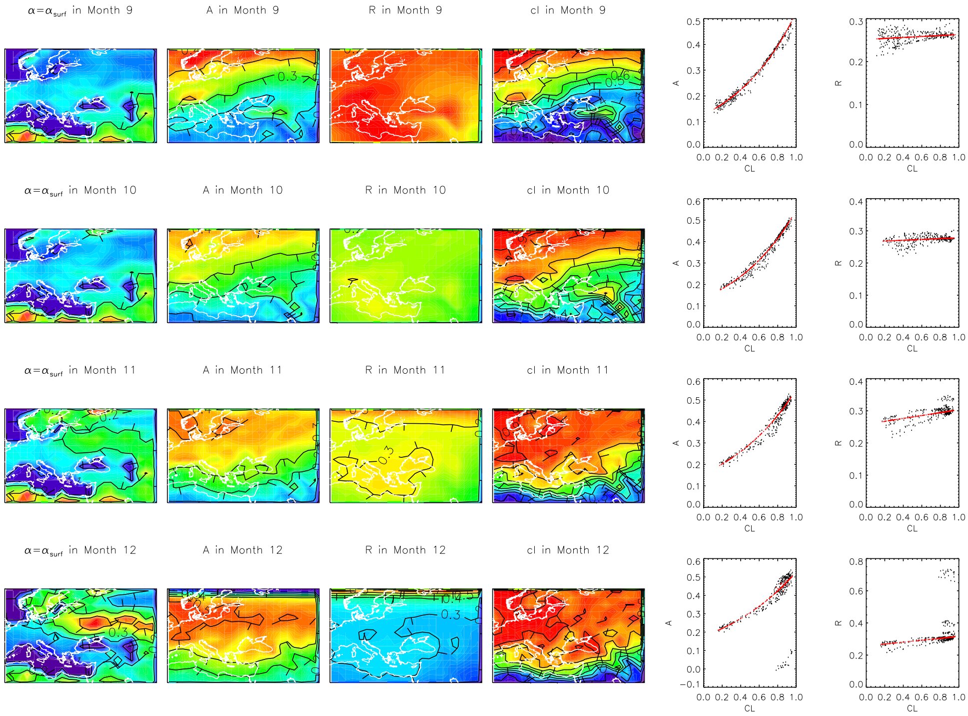

Figure 8: Like Figure 6, but for September to December.

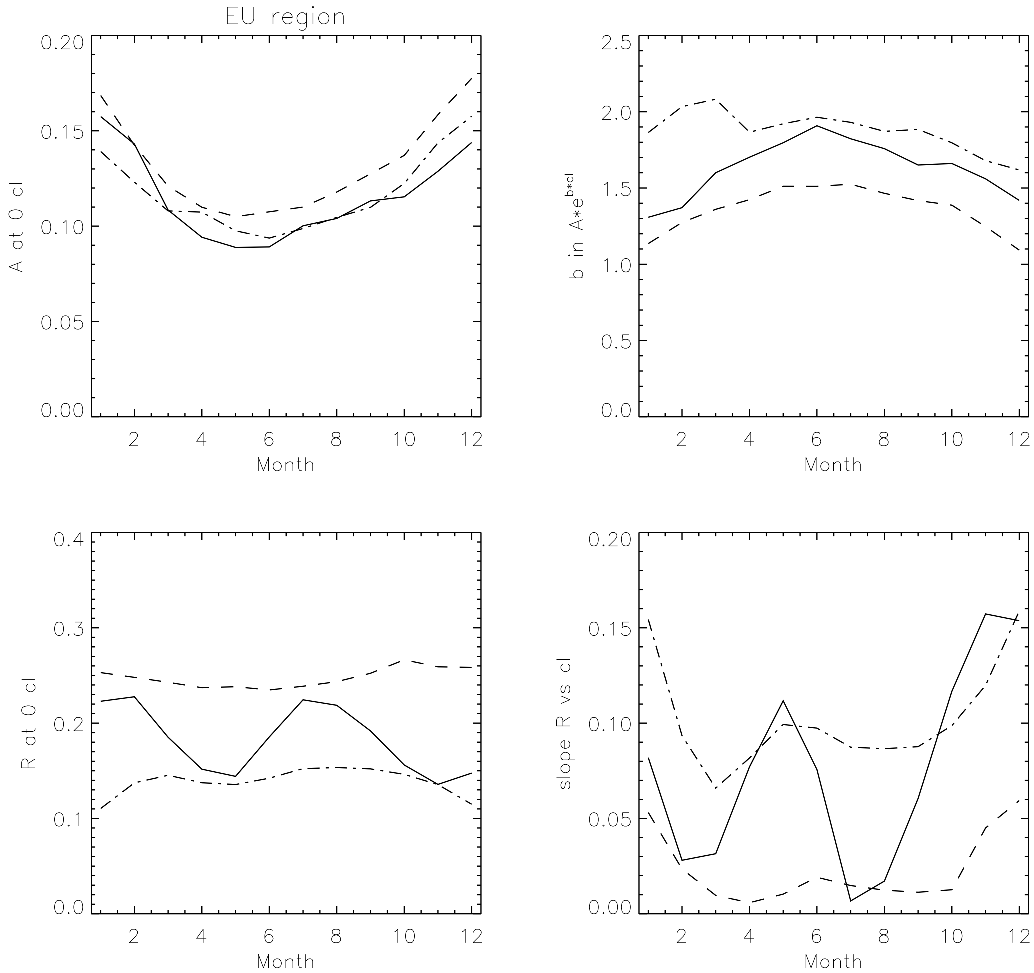

Figure 9: Information from the fits to data from Figures 6– 8. Clockwise from top left the panels are: A at zero cloudiness fraction (`cl`, or CFC) as function of month of year; factor b in the fitted formula $A = a \cdot e^{b \cdot cl}$; slope of the fitted line in the relation between R and `cl`; and R at zero cloudiness. The solid, dashed and dot-dashed lines represent the three GCM models used (see text).

7 Summary of results from analysis of CMSAF, GLIMPSE and CMIP data

- The simple regression model can explain between 40 and 70% of the variance in CMSAF data, for a given month; more in GLIMPSE GCM data.

- Regression model helps spot **data-quality issues in the CMSAF data**: Here, we have found a number of problems - one is in the TRS product where the monthly mean for March 2009 has been replaced with one daily field (March 2009 was omitted from the present analysis at an early stage), while others starting January 2010 are yet unidentified, and note that winter high-latitude areas have a problem, possibly related to snow and/or trees.

- The albedo model fitted to GCM data is better in statistical terms than the CMSAF model; **the seasonality of the regression coefficients are similar**, but the intercept in the GCM fit is less realistic than the intercept in the CMSAF data in that it can become negative.

- Comparing CMSAF model fit to GCM model fit we see similar features: In both cases there is a positive residual at high latitude winters and a negative residual over equatorial areas with clouds. Separation shows that the northern problem may be forced by the regression seeking to minimize residuals, while the equatorial cloud problem seems to be related to the simplicity of the albedo model used.

- The **Main Conclusion we have is that the albedo regression model is a powerful tool** to spot inconsistencies between TRS/TIS, SAL and CFC in the CMSAF data.

- Improvements to the simple linear regression model is suggested – CMIP data analysis indicates that cloud fraction should enter the formulae as an exponential term.

8 Comments, and Future Work

The very simple regression model we use has shown its limitations: Particularly cloud albedo is not well represented by the linear term with CFC. It is well known that clouds at different altitudes have different properties - high thin clouds tend to warm the atmosphere while lower thick clouds tend to cool it: The use of one type of cloud fraction may therefore be overly simplistic, depending on what CFC mainly represents. The failure of the cloud optical depth term COT to have any effect on regression quality may also be a sign of this. In future work we will therefore investigate the use of more detailed representations of clouds in our regression model.

The regression method has shown itself to be a powerful tool to check for CMSAF data consistency, despite the above shortcomings, and it is recommended that a tool, based on the code provided here, is set up and used routinely as part of the data validation procedure.

One potential use for an 'albedo model' like ours is to model the earthshine intensity. In the Earthshine project [Owner-Petersen et al., 2008] we will observe the earthshine intensity very accurately and wish to infer the terrestrial albedo at the time of observation - one way to do this would rely on a forward model of the shortwave flux from Earth in the direction towards the Moon and it is envisaged that spatially detailed models of the distribution of clouds, surface types, ice and so on could model that flux. Previous efforts [Goode et al., 2001] at interpreting observed earthshine intensity relied on representing the earth as uniform light scattering spheres.

9 Appendix 1: Code description and Code listing

9.1 Code description with usage notes

We assume that IDL is installed - at least version 7.1.
Short description: After unpacking the code which is distributed as a tar file named CMSAF_CDOP_STUDY_9.tar, do this:
Go to the directory that results when the tar file is un-tarred. Start IDL, and at the prompt type:

```
idl go_ex18.pro
```

This is a short IDL program that is used to set up a run - it loads various libraries, sets identifying strings, selects the geographical area to work in, and defines the plotting environment. The script runs for a few minutes, *stepping through each month*, and produces screen output as well as some files:

```
Zonal_avg_residuals.ps
Europe_mainlymap_of_annual_SAL_cycle.ps
Europe_mainlyplot_of_residuals.ps
modeltype1\_CMSAF\_nn.txt
```

The files modeltype1_CMSAF_nn.txt appear one for each month and are numbered (nn=1,2,3 etc). These files contain the lon, lat, regressand and then the regressors for each pixel on the map.
The main result of the code is a file containing the regression coefficients and their estimated uncertainties. This is the file with "regression_results_Model" as a substring of its name - prefixed by the identifying string set up in go_ex18.pro.
The regression coefficients can be plotted using

```
idl go_plot_regression_result.pro
```

The output from that is like the plot in Figure 5.

9.1.1 Required input files

- IGBP land surface types file interpolated to the required 1x1 grid.
- 'MA files' for TRS, TIS, SAL, CFC and COT stored in hdf format

- a land-sea mask file with the required grid layout
- `CFCliste.txt` - a file that lists the paths to the CFC files we want to use. It is ordered in a way to match line by line similar files for SAL TRS etc. They are named `SALliste.txt` and so on
- The code `CM_SAF_include.pro` compiles special routines written and available from the CMSAF group at DWD:

```
@CM_SAF_init.pro
@CM_SAF_subroutines.pro
@CM_SAF_defaults.pro
@CM_SAF_read_data.pro
@CM_SAF_display.pro
@CM_SAF_dwd_subroutines.pro
@CM_SAF_compare.pro
@CM_SAF_display_hcp.pro
@CM_SAF_extract.pro
@CM_SAF_statistics.pro
@CM_SAF_overpass
@CM_SAF_display_hcp
@CM_SAF_export
```

- A `netcdf` file - called `CERES_SurfaceAlbedo.nc` - of CERES surface albedos, stored in subdirectory `CERES/`. The first time the code is run it spends a long time extracting information from this CERES file - these extracted data are reused in subsequent, faster, runs.

9.1.2 Code availability

The complete code is made available by `ftp` after arrangement with Peter Thejll (contact pth@dmi.dk), as a (75 Mbyte) `tar` file. It can also be downloaded at the link.

Acknowledgments
We acknowledge the modelling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM) for their roles in making available the WCRP CMIP3 multi-model dataset. Support of this dataset is provided by the Office of Science, U.S. Department of Energy. Martin Stendel at DMI is warmly acknowledged for making GLIMPSE data available, as are the members of the CMSAF group who so readily helped with data-queries and insightful comments during my enjoyable VS stay at DWD.

10 References

References

[Donohoe and Battisti, 2011] Donohoe, A. and Battisti, D. S. (2011). Atmospheric and Surface Contributions to Planetary Albedo. Journal of Climate *24*, 4402–4418.

[Goode et al., 2001] Goode, P., Qiu, J., Yurchyshyn, V., Hickey, J., Chu, M.-C., Kolbe, E., Brown, C. and Koonin, S. (2001). Earthshine Observations of the Earth's Reflectance. Geophysical Research Letters *28*, 1671–1674.

[Landsman, 1995] Landsman, W. B. (1995). The IDL Astronomy User's Library. In Astronomical Data Analysis Software and Systems IV, (Shaw, R. A., Payne, H. E. and Hayes, J. J. E., eds), vol. 77, of Astronomical Society of the Pacific Conference Series p. 437,.

[Owner-Petersen et al., 2008] Owner-Petersen, M., Andersen, T., Ardeberg, A., Thejll, P. and Gleisner, H. (2008). The Earthshine Telescope Project. In Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series vol. 7012, of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Society of Photo-Optical Instrumentation Engineers (SPIE).

[Pallé et al., 2004] Pallé, E., Goode, P. R., Montañés-Rodríguez, P. and Koonin, S. E. (2004). Changes in Earth's Reflectance over the Past Two Decades. Science *304*, 1299–1301.

[Riihelä, 2009] Riihelä, A. (2009). Algorithm Theoretical Basis Document: Surface Broadband Albedo. Doc. No.: SAF/CM/FMI/ATBD/SAL Issue: 1.1. Technical report EUMETSAT Satellite Application Facility on Climate Monitoring.

[Schulz et al., 2009] Schulz, J. P., Albert, H.-D., Behr, D., Caprion, H., Deneke, S., Dewitte, B., Dürr, P., Fuchs, A., Gratzki, P., Hechler, R., Hollmann, S., Johnston, Karlsson, K.-G., Manninen, T., Müller, R., Reuter, M., Riihelä, A., Roebeling, R., Selbach, N., Tetzlaff, A., Thomas, W., Werscheck, M., Wolters, E. and Zelenka, A. (2009). Operational climate monitoring from space: the EUMETSAT satellite application facility on climate monitoring (CM SAF). Atmos. Chem. Phys. *9*, 1687–1709.

[Stendel et al., 2006] Stendel, M., Mogensen, I. A. and Christensen, J. H. (2006). Influence of various forcings on global climate in historical times using a coupled atmosphere ocean general circulation model. Climate Dynamics *26*, 1–15.

3. Previous reports

Previous reports from the Danish Meteorological Institute can be found on:
http://www.dmi.dk/dmi/dmi-publikationer.htm