Geoscientific
Model Development

Multi-sensor cloud retrieval simulator and remote sensing from model parameters – Part 1: Synthetic sensor radiance formulation

G. Wind[1,2], **A. M. da Silva**[1], **P. M. Norris**[1,3], **and S. Platnick**[1]

[1]NASA Goddard Space Flight Center, 8800 Greenbelt Rd. Greenbelt, Maryland 20771, USA
[2]SSAI, Inc., 10210 Greenbelt Road, Suite 600, Lanham, Maryland 20706, USA
[3]Universities Space Research Association, 10211 Wincopin Circle #500, Columbia, Maryland 21044, USA

Correspondence to: G. Wind (gala.wind@nasa.gov)

Abstract. In this paper we describe a general procedure for calculating synthetic sensor radiances from variable output from a global atmospheric forecast model. In order to take proper account of the discrepancies between model resolution and sensor footprint, the algorithm takes explicit account of the model subgrid variability, in particular its description of the probability density function of total water (vapor and cloud condensate.) The simulated sensor radiances are then substituted into an operational remote sensing algorithm processing chain to produce a variety of remote sensing products that would normally be produced from actual sensor output. This output can then be used for a wide variety of purposes such as model parameter verification, remote sensing algorithm validation, testing of new retrieval methods and future sensor studies. We show a specific implementation using the GEOS-5 model, the MODIS instrument and the MODIS Adaptive Processing System (MODAPS) Data Collection 5.1 operational remote sensing cloud algorithm processing chain (including the cloud mask, cloud top properties and cloud optical and microphysical properties products). We focus on clouds because they are very important to model development and improvement.

1 Introduction

Accurate knowledge of cloud cover and cloud properties is important in model studies that involve earth's radiative budget, climate prediction and numerical weather prediction. High thin clouds are observed to have a warming effect on the atmosphere because of their low albedo and low temperature.

Low clouds have a net cooling effect due to their high albedo and relatively small temperature contrast with the surface. (Lee et al., 2009) Clouds and their interactions with aerosols are significant sources of uncertainty in climate prediction studies (IPCC, 2007). In addition, clouds continue to be the main source of climate feedback uncertainty and hence climate sensitivity (e.g., Bony et al., 2006).

The Goddard Earth Observing System Version 5 (GEOS-5) earth system model is maintained by the Global Modeling and Assimilation Office (GMAO) at NASA Goddard Space Flight Center (GSFC). GEOS-5 contains components for atmospheric circulation and composition (including atmospheric data assimilation), ocean circulation and biogeochemistry, and land surface processes. Components and individual parameterizations within components are coupled under the Earth System Modeling Framework (ESMF, Hill et al., 2004). In addition to traditional meteorological parameters (winds, temperatures, etc.; Rienecker et al., 2008), GEOS-5 includes modules representing the atmospheric composition, most notably aerosols (Colarco et al., 2010) and tropospheric/stratospheric chemical constituents (Pawson et al., 2008), and the impact of these constituents on the radiative processes of the atmosphere. GEOS-5 has a mature atmospheric data assimilation system that builds upon the Gridpoint Statistical Interpolation (GSI) algorithm jointly developed with NCEP (Wu et al., 2002; Derber et al., 2003; Rienecker et al., 2008). The GSI solver was originally developed at NCEP as a unified 3D-Var analysis system for supporting global and regional models. GSI includes all the in situ and remotely sensed data used for operational weather prediction at NCEP. GEOS-5 also includes assimilation of

aerosol optical depth (AOD) observations from the MODerate resolution Imaging Spectroradiometer (MODIS) imager on the NASA Earth Observing System (EOS) Terra and Aqua spacecraft; an algorithm for assimilating cloud property information from measurements in the visible and infrared portions of the spectrum is currently under development (Norris and da Silva, 2013). While the GEOS-5 meteorological assimilation includes a wide variety of spaceborne sensor data, traditionally samples containing clouds are carefully screened out. The near real-time GEOS-5 data assimilation and forecasting system runs at a nominal horizontal resolution of 25 km with 72 vertical layers (Rienecker et al., 2008; Molod et al., 2012).

The MODIS instrument (Barnes et al., 1998) is a passive imager, producing a wide variety of remotely sensed data products for land, ocean and atmosphere disciplines from 36 spectral channels. Data Collection 5.1 processing includes algorithms for retrieving cloud cover amount (Ackerman et al., 2006; Frey et al., 2008), cloud top properties such as cloud top pressure and temperature (Menzel et al., 2008) and cloud optical and microphysical properties such as cloud optical thickness, cloud effective radius and cloud water path (Platnick et al., 2003; Wind et al., 2010; Zhang and Platnick, 2011; King et al., 2013).

In this paper we present a technique that brings together remote sensing methods and model-generated fields. We use MODIS geolocation data to sample GEOS-5 fields as if the MODIS instrument were flying over the model fields instead of the earth's surface. Once the sampling is complete, we generate synthetic sensor radiance data for the MODIS footprint. We then replace the contents of the 1 km, 500 m and 250 m resolution MODIS Level-1B (Xiong et al., 2006) radiance files with these simulated radiances and insert the resulting alternate data stream into the start of the MODIS Adaptive Processing System (MODAPS) operational algorithm processing chain for the atmosphere discipline cloud products mentioned above (product designation MOD06 and MYD06 for Terra and Aqua MODIS, respectively). The data stream is fully transparent to the system so that pixel-level (Level-2) retrievals can be aggregated through the same gridded (Level-3) $1° \times 1°$ code (Hubanks et al., 2008; King et al., 2003, 2013) used in MODAPS production (MOD08 and MYD08 for Terra and Aqua, respectively). There are many potential uses for the resulting Level-2 and Level-3 data. Simulated Level-2 granules and Level-3 aggregations can both be compared to the respective archived MODIS data as a means of model validation and study of model biases that could exist. The simulated retrievals can also be compared with GEOS-5 source fields directly to study some aspects of retrieval algorithm behavior and sensitivities since all retrievals are performed with known (prescribed) *truth*.

The synthetic sensor data framework has been developed with instrument flexibility in mind, so that by simple substitution of spectral response functions and data reader, the MODIS instrument can be replaced by other spaceborne or airborne sensors, currently in operation or part of a future concept, and a different sensor data stream can be produced. Thus, identical products from different sensors or different retrieval algorithms for the same sensor can be compared and analyzed in a controlled environment, which can provide insight and lead to improvements in remote sensing algorithms.

This flexibility extends to model data as well. Any climate or weather prediction model fields can be used as long as a means of ingesting the necessary parameters is provided. Thus synthetic retrievals based on multiple models can be compared and analyzed using the same sensor interface in a controlled environment, leading to a consistent diagnostic toolset. Furthermore, this detailed simulation capability can function as a test bed for very fast simulators such as the Cloud Feedback Model Intercomparison Project (CFMIP) Observation Simulator Package (COSP) (Bodas-Salcedo et al., 2011) or the hyperspectral simulator of Feldman et al. (2011).

There have been a number of previous studies of high-resolution simulation of cloud fields and cloud-affected radiances from weather/climate models. We highlight some key papers here (Otkin et al., 2007; Bugliaro et al., 2011; Jonkheid et al., 2012), and discuss their differences with our approach.

The recent paper by Jonkheid et al. (2012) is discussed first because it contains a very useful figure (their Fig. 1) classifying the various types of comparisons that can be made between different cloud property data sets. According to this classification system, their paper is "type III". Namely, it generates synthetic 5 km scale SEVIRI pixels by subcolumn generation from a 25 km resolution Numerical Weather Prediction (NWP) source model, and compares the cloud property retrievals based on the simulated radiances against the source model cloud properties. This is what we refer to here as a "synthetic retrieval validation". Using this approach, they produce a very helpful and thorough analysis of EUMETSAT's Climate Monitoring Science Application Facility (CM SAF)'s cloud physical properties (CPPs) retrievals of cloud water path (CWP). There are at least two main differences between their study and our current MODIS study (apart from the instrument): (a) our predominant focus in this first paper is to describe our simulator and to discuss several anecdotal examples of model validation ("type IV"), although we do intend to use our simulator for type III comparisons in future work. In type IV, cloud property retrievals from real satellite observations are compared with simulated retrievals of the same properties based on simulated radiances from an NWP forecast or analysis valid at the time of the observations. Our type IV comparisons largely eliminate retrieval issues by applying the exact same MODIS retrieval algorithm to both the observed and simulated radiances; (b) Jonkheid et al. (2012) use a rather simplified forward model (at most two cloud layers, one ice and one liquid, in fixed altitude bands) based on cloud fields generated by the "generalized cloud overlap" generator of Räisänen et

al. (2004), together with a look-up table (LUT) approach to SEVIRI radiance generation. They do this for good reason, namely, to have a fast, reasonably accurate simulation tool that can be applied efficiently to the large spatial and frequent temporal coverage of SEVIRI. By contrast, our tool is computationally expensive because it applies a complex subcolumn generation approach and very detailed radiance computations to simulated 1 km MODIS pixels. As such, it is currently most suited to detailed comparison work on case study MODIS granules.

Note that in our type IV comparisons, the attribution of the retrieval space differences back to the model is not necessarily straightforward, and must consider not only model physics and analysis errors but also the realism of the forward model scheme. Nevertheless, based on our previous experience, we do prefer the apples-to-apples retrieval-space comparisons of type IV over type I comparisons of real MODIS retrieval products against GEOS-5 model fields, which are not apples-to-apples comparisons. Invariably, the definition of quantities such as cloud optical thickness (COT), cloud top pressure (CTP), or effective radius (r_e) is not the same for MODIS retrievals and the model. MODIS-retrieved CTP, for example, is a complex product that involves concepts of how far into the cloud the instrument can see in the infrared. This is a particularly acute problem for thin high cirrus overlying lower thicker cloud, in which case MODIS typically retrieves some intermediate CTP. The GEOS-5 model has discrete model layers, each with a known cloud fraction and optical thickness. Various different model-based concepts of CTP can be formulated, such as the pressure at which the top-down cumulative COT reaches some threshold, as in the COSP MODIS simulator of Bodas-Salcedo et al. (2011). Ultimately, the best model-based CTP definition for comparison with MODIS-retrieved CTP will be one that emulates the latter as closely as possible. Similarly, MODIS-retrieved r_e is biased towards the top of a cloud layer, and to make a fair comparison of this quantity with the model, we must understand *how* close to the top of the cloud. Even a comparatively more directly retrievable quantity such as COT has retrieval issues such as detection threshold on the low end and saturation issues (physically or storage based) on the high end. These retrieval issues must be taken into account to make a fair comparison with model COT, typically by applying the same threshold and saturation limits to the model COT prior to comparison with the MODIS retrievals. Ultimately, the point we are making is that type I comparisons are not apples-to-apples, and to make them more so typically involves applying some form of transformation to the model fields to make them more suitable for comparison with the satellite retrievals. This is the whole motivation for the COSP simulator suite. And since such transformations are essentially moving the comparison towards type IV, it is better, in our estimation, to use a full type IV retrieval-space comparison directly, as we do in this paper.

Returning to previous literature, an earlier paper by Otkin et al. (2007) discusses synthetic validation of hyperspectral retrieval algorithms for the now shelved Geosynchronous Imaging Fourier Transform Spectrometer. Otkin et al. (2007) use high-resolution local area mesoscale models to generate temperature, moisture, and other required quantity profiles directly at the spatial resolution ($\sim$ 2–4 km) of the intended GIFTS fields of view. Similar to our approach, the radiative observables are then forward modeled from these profiles and the GIFTS retrieval algorithms applied. But there are several important differences: (a) Otkin et al. (2007) also focus on type III synthetic retrieval validation, by comparing the retrieved temperature and moisture profiles against the source mesoscale profiles, whereas our primary focus here is type IV model validation, by comparing NWP model simulated retrievals against actual in-flight retrievals for MODIS; (b) Otkin et al. (2007) study local regions simulated by mesoscale models and so are able to use direct pixel resolution forecasts. Our goal is to provide a MODIS simulation tool to a global NWP model, with a grid column resolution of around 0.25 degrees, and therefore direct high-resolution pixel simulation at 1 km is not computationally feasible. Rather we use a statistical method to generate an ensemble of pixel-like subcolumns, having realistic subgrid column variability statistics with which to drive a statistical pixel-scale simulation; (c) Otkin et al. (2007) use a simplified single-cloud layer, single cloud-phase forward model. Our forward model also treats mixed-phase and multi-layer clouds.

Bugliaro et al. (2011) discuss validation of SEVIRI retrieval algorithms in a manner similar to that of Jonkheid et al. (2012), but using statistical downscaling, via a spectral extrapolation method, from a 7 km resolution central European NWP forecast. We will discuss spectral extrapolation methods later in the context of the Venema et al. (2010) downscaling method. The Bugliaro et al. (2012) method is similar, essentially using a $k^{-5/3}$ power law to extrapolate the power spectrum of liquid water content from resolved 30 km scales down to scale of 2.33 km, comparable to the SEVIRI pixel scale while removing negative condensate amounts and using a partly randomized small-scale spectral phase to treat vertical correlations between layers.

Finally, we also note the development of the ECSIM simulator framework for the European Space Agency's Earth-CARE mission (Voors et al., 2007; Donovan et al., 2008). ECSIM is a flexible and comprehensive end-to-end simulator package for the active and passive instruments of the upcoming EarthCARE satellite. It is our understanding from Donovan et al. (2008) that the simulator can be driven from cloud resolving model output, much like in Otkin et al. (2007), but that it is the responsibility of the user to generate input cloud fields externally if the simulator is to be driven from a lower resolution NWP model.

The outline of the rest of the paper is as follows: in Sect. 2 we describe the model–sensor interface using the GEOS-

5 model and MODIS imager. Section 3 shows an example of simulation and retrieval of cloud properties on a sample Level-2 data granule. In Sect. 4 we elaborate on future directions for the software suite.

This paper is the first part in a series that will combine the software suite described in detail here with a variety of research applications.

2   Radiance simulations at scales smaller than the model's grid spacing

We start the process by selecting an area and time period of study. It can be as small as a few-pixel subsection of a single MODIS granule or as large as an entire year of MODIS data. The study size is only limited by availability of computing resources. As far as model itself is concerned, there is no need for actual MODIS data to be present, but we specifically want the actual MODIS radiances to be available, so that retrievals from simulated and actual radiances can be compared directly. Similarities and differences in those retrievals can be analyzed and results applied on a variety of levels in order to improve both the model and the sensor retrieval algorithm.

For simplicity's sake in all subsequent references and illustrations, the study area will be taken to be a standard 5 min MODIS data granule (approximate 2000 km in the along track direction by 2300 km). Once the granule is chosen, we proceed to choose model output files that bound the granule time. For example, for the granule at 02:00 UTC, we would select model output at 00:00 and 03:00 UTC. We use a MODIS standard geolocation file (MOD03 product) to define the spatial locations to sample the model fields. Solar and view angle information contained in the same MODIS geolocation file is also used in the simulation. For the examples shown in this paper, we used the GEOS model v.5.7.2 output. A list of specific GEOS-5 fields and products used in this suite is given in Table 1.

2.1   Surface albedo determination

In order to save on computational time, we pre-determine surface albedo for the area of study. The surface albedo data come from a variety of sources. Over ice-free ocean, MODIS geometry and model wind speed are used in a Cox–Munk ocean surface bidirectional reflectance distribution function model (Cox and Munk, 1954) to produce cloud-free ocean surface reflectance. This model reflectance is calculated for four cardinal wind directions and then averaged into a look-up table (LUT), which is a function of wavelength, wind speed, solar and sensor zenith angles and relative azimuth angle. We do the calculation at three wind speeds of 3, 7 and 15 m s^{-1}. The LUT has 33 solar zenith values, 28 view zenith values and 37 relative azimuth values. We linearly interpolate as needed to obtain surface spectral albedo in the selected 15 MODIS channels that have a shortwave reflective

Table 1. GEOS v.5.7.2 fields and products used in simulations.

Field code	Description
U10M	U component of wind speed at 10 m altitude
V10M	V component of wind speed at 10 m altitude
FRSEAICE	Sea ice fraction
FRSNO	Snow fraction
PS	Surface pressure
T2M	Temperature at 2 m altitude
SLP	Mean sea-level pressure
QV2M	Specific humidity at 2 m altitude
O3	Ozone concentration profile
T	Temperature profile
DELP	Level pressure differential profile
RH	Relative humidity profile
CLOUD	Radiative cloud fraction profile
QLLS	Large-scale cloud liquid water mixing ratio
QLAN	Anvil cloud liquid water mixing ratio
QILS	Large-scale cloud ice mixing ratio
QIAN	Anvil cloud ice mixing ratio

component. The MODIS channels used in the simulation and their central wavelengths are listed in Table 2. The ocean reflectance LUT contains data for channels 1–22 and 26 from the table. For the IR (infrared) channels that have no reflective component (27–36), a constant value of 0.015 is used. This value is based on the ocean surface emissivity value suggested by the MODIS cloud top properties algorithm.

Over land several methods are utilized to model the radiances for all MODIS channels. We use the MODIS land surface spectral albedo gap-filled data set (Moody et al., 2005, 2008) that has been updated for MODIS data Collection 6 and is derived from both Aqua and Terra data. In addition to providing 16-day time period averages every 8 days, the gap-filled albedo files are generated for each year separately (instead of aggregating all years together as was done previously). Further, spatial resolution has been improved to 1 km. These files are derived from the Collection 5 MCD42B product (Schaaf et al., 2011). This MODIS land surface albedo product is used directly for channels 1–7 and interpolated linearly to cover other MODIS spectral channels that fall within the 0.47–2.14 μm wavelength range. For wavelengths that are longer than 2.14 μm, we use the surface emissivity data set used in MODIS clear-sky profile retrievals (Seemann et al., 2008). For wavelengths shorter than 0.47 μm, we use a seasonally averaged surface albedo database utilized by the MODIS deep-blue algorithm (Hsu et al., 2004) to obtain the albedo for the 0.41 μm channel and then interpolate for the 0.44 μm channel.

Over snow and sea ice we use the MODIS multiyear average snow/ice albedo data set (Moody et al., 2007). The look-up table is determined by the MODIS pixel latitude, the International Geosphere–Biosphere Programme (IGBP) ecosystem type and snow/ice fractions from GEOS-5 FRSNO and FRSEAICE model fields.

Table 2. MODIS channels used in simulations.

Channel number	Central wavelength (μm)
1	0.65
2	0.86
3	0.47
4	0.55
5	1.24
6	1.63
7	2.13
8	0.41
9	0.44
17	0.91
18	0.94
19	0.94
20	3.7
22	3.9
26	1.38
27	6.2
28	7.3
29	8.5
31	11.0
32	12.0
33	13.2
34	13.4
35	13.8
36	14.2

Table 3. Vertical levels used in simulation

Level number	Level altitude (km)
1	80
2	60
3	50
4	45
5	40
6	35
7	30
8	25
9	20
10	18
11	16
12	15
13	14
14	13
15	12
16	11
17	10
18	9
19	8
20	7
21	6
22	5
23	4
24	3
25	2
26	1
27	0

The resulting surface albedo values are written out to file for each study area so that they can be referenced later for different simulation scenarios involving the same area and time. A good example of such varying scenarios would be repeated experiments with or without the presence of aerosols.

2.2 Water vapor and other gaseous absorbers

After the surface albedo calculations are complete, we proceed to ingest the profiles of temperature, relative humidity, ozone concentration and atmospheric pressure. These profiles are downsampled to 27 atmospheric levels, as listed in Table 3, from the GEOS-5 native 72 vertical levels and sent to an atmospheric transmittance module that uses the correlated-k method (Kratz, 1995) to calculate weights and optical thicknesses for each atmospheric layer due to water vapor and other gaseous absorbers. We are not strictly limited to using the correlated-k method however. Any technique that would produce a profile of optical thickness based on specified atmospheric water vapor and ozone profiles and trace gas amounts will suffice. In cases where the surface is encountered at an altitude higher than 0 km, the profile is trimmed accordingly and the surface level is inserted as the last level to be used. We perform the vertical downsampling in order to save on the computational cost of the synthetic sensor radiance simulation step, with the bulk of

downsampling occurring above the tropopause. We preserve finer vertical resolution in the troposphere. We find this to be permissible as radiance data stored in a MODIS L1B file have an accuracy to only the fifth decimal place, and the uncertainty due to varying the number of vertical levels in the upper atmosphere is less than this data storage accuracy. In our particular setup for MODIS for the channel set of interest, we found deviations in sixth decimal place and less when we decreased the number of levels from 72 to 27. That of course may not hold true for all sensors or studies that a user may attempt, and this setup would need to be re-evaluated as needed. The simulator however is not limited to operating on a specific number of levels.

2.3 Generating cloud subcolumns

Sampling of model cloud-related fields to the MODIS pixel scale is not straightforward because cloud properties typically vary on scales not adequately resolved by the operational 0.25° GEOS-5 resolution. To sample cloud fields, 1 km MODIS pixels for each GEOS-5 grid column are collected, and the same number of pixel-like subcolumns is generated using a statistical model of subgrid column moisture variability. The general approach of Norris et al. (2008) is followed,

namely using a parameterized probability density function (PDF) of total water content for each model layer and a Gaussian copula to correlate these PDFs in the vertical.

In this application, we use the skewed triangle PDF, which allows a simple inclusion of moisture variability skewness, a ubiquitous feature of atmospheric boundary layers. This PDF has a simple scalene form characterized by three parameters: a lower and upper bound and a mode. Under some circumstances, these three parameters can be directly diagnosed from the *layer mean* total water and condensate contents, q_t and q_c, and cloud fraction f, but in many cases some adjustments are necessary to f, and possibly q_c, to achieve consistency. The details of this calculation are beyond the scope of this paper and are described fully in Norris and da Silva (2013). Approximations must also be made in the case of clear or overcast layers, when the triangle is underdetermined.

For the Gaussian copula we use a correlation matrix with a fixed vertical decorrelation scale of 100 hPa, further modified by a multiplicative Riishojgaard (1998) flow-dependent correlation in total water that permits sharper decorrelation across inversion features. Further details are given in Norris and da Silva (2013). Once the correlation matrix is specified, the Gaussian copula correlated ranks of each of the grid column's layers are easily generated (Norris et al., 2008) and then inverted with the cumulative distribution function (CDF) of each layer's skewed triangle distribution. The net result is an ensemble of subcolumns of total moisture content that sample the specified layer PDFs and have the specified vertical correlations and accompanying cloud and condensate overlap properties. The transformation of total moisture content to vapor, liquid water and ice contents assumes the vapor is capped at the GEOS-5 saturation vapor content and that the excess moisture is condensate, split between the phases using an ice fraction linear in temperature over the -35 to $0\,°C$ range. It is these subcolumn condensates, combined with GEOS-5 diagnostic effective radii (a fixed profile across subcolumns based on each grid column's temperature and pressure profiles), that are used to evaluate subcolumn (or "pixel") liquid water and ice optical thicknesses for each layer. These are input to the MODIS radiance simulator code.

Note that the subcolumns generated in this way are horizontally independent (the independent column approximation, or ICA), but are subsequently "clumped", or rearranged, to give horizontal spatial coherence, by using a horizontal Gaussian copula applied to condensed water path. This clumping acts to give the generated clouds a reasonable horizontal structure, such that the cloudy pixels in a grid column are actually grouped into reasonably looking clouds, rather than being randomly distributed. This is important because the MODIS cloud optical and microphysical property retrieval algorithm has some spatial variance tests for potentially partially cloudy pixels, removing cloud edges by the so-called "clear-sky restoral" (Zhang and Platnick, 2011; Pincus et al., 2012). If clumping is not used, then individual pixels generated by ICA stand an exceptionally high chance of being eliminated by the clear sky restoral unless a model grid box has a nearly 100 % cloud fraction.

The clumping algorithm works by applying another Gaussian copula, this time in the horizontal. Copula functions are explained in Norris et al. (2008), and references therein. Basically, an N copula is a joint cumulative distribution function (CDF) in the ranks of a system of N random variables, where, by rank, we mean the cumulative probability, between zero and one, of a variable within its marginal distribution. Thus if the value v of a random variable V_n has a rank of 0.2, it means that 20 % of random samples of V_n from the system will fall at or below v. From this definition, it is clear that each margin of the N copula is a uniform distribution on [0,1]. What the copula does is to correlate these N ranks in a prescribed way (see Norris et al., 2008; Norris and da Silva, 2013), such that, for example, certain variables have a large correlation in their tail properties, and others do not. Furthermore, the Gaussian copula that we use here is easy to sample from, allowing us to generate N tuple ranks with a prescribed correlation structure easily, but such that each variable's rank is individually uniformly distributed on [0,1]. With this background, the clumping algorithm works as follows: a correlation matrix $\mathbf{C}$ is generated between all pixels in a grid column based on the horizontal distance between the actual pixels in the MODIS granule and assuming a nominal 5 km decorrelation length. If there are N pixels, $\mathbf{C}$ is an $N \times N$ matrix. This matrix is used by a Gaussian copula to generate an N *tuple of* correlated ranks, each on [0,1] and each being associated with one of the pixels. Now, as discussed above, each one of these N ranks has a uniform probability density on [0,1], and yet the ranks are correlated by the copula, such that the ranks of nearby pixels are strongly correlated, while the ranks of distant pixels are uncorrelated. Next, by applying each rank to the column condensed water path (CWP) distribution of its associated pixel, we can produce a horizontally correlated CWP field. The CDF of each pixel's CWP can be adequately approximated by the empirical CDF of the CWP of all N simulated pixels within the grid column, since we are assuming a homogeneous statistical process within the grid column (i.e., that the ensemble correlation structure and marginal statistics of CWP are homogeneous on the grid column scale, even if CWP itself is not). In this case, the remainder of the algorithm is simple: by multiplying each rank by N and rounding up non-integers, a new "rank" on $\{1, 2, \ldots, N\}$ is obtained, which is subsequently used to sample (i.e., index, or effectively, to re-order) a list of the N ICA-*simulated* pixels that has previously been sorted by CWP. Because horizontally nearby pixels are more correlated by $\mathbf{C}$, they will have a higher chance of having similar ranks, and therefore similar values of CWP. In this way the pixels are grouped together horizontally into coherent clouds. (Note that this clumping acts on subcolumns as a whole, and independent of the preexisting vertical correlations in the ICA subcolumns, so the clumping will work

better for single cloud layers. For multilayer clouds, the layer that dominates the CWP will dominate the clumping.) Finally, note that this clumping method preserves the simulated grid column cloud fraction. For example, if 20 % of the simulated ICA subcolumns are clear (have zero CWP), then because each pixel's generated rank is uniform on [0,1], each pixel with have a 20 % chance of being assigned a zero CWP.

There are alternative approaches to NWP subcolumn generation. Venema et al. (2010) describe an elegant method for down-scaling coarse-resolution cloud fields (such as NWP models produce) to finer horizontal scales suitable as pixel "subcolumns" for input to 1-D or 3-D radiation calculations. Their method carefully extrapolates the spectrum of variance in the resolved cloud field down to pixel scales using various extrapolation assumptions. Venema et al. (2010) do acknowledge, however, that the accuracy of their downscaling approach depends on the quality achieved by the spectral extrapolation algorithm being used, and ultimately on what assumptions are made regarding the statistical properties of the subresolved scale variance.

We have some reservations about the idea of extrapolating from NWP resolved scales down to pixel scale, particularly because the scales just above the resolution cutoff can be very much affected by numerical diffusion and by the nature of the subgrid cloud and turbulence parameterizations used by the model. Partly for this reason, but also because of the context we describe below, we choose a somewhat simpler approach: of specifying parametrically the nature of the subgrid column variability, both horizontally and vertically, as described above and in detail within Norris and da Silva (2013). Our main reason for this approach is the context of GMAO's ongoing modeling and cloud data assimilation plans. The subcolumn statistical model described, while not currently operational in GEOS-5, is planned for implementation as part of an improved cloud parameterization system, and it is also the statistical model being used by our newly developed cloud data assimilation system. As such, it is natural that we should base our new MODIS simulation tool on this model.

As described above, this subgrid column statistical model includes a simple skewed triangle PDF of total moisture in each grid box, and a modern vertical correlation model, based on ideas advanced by "generalized cloud overlap" (Räisänen et al., 2004; Pincus et al., 2005; Oreopoulos and Norris, 2011) and Gaussian copula overlap (Norris et al., 2008; Norris and da Silva, 2013). Both this vertical correlation model and the simple three parameter intra-layer total water PDF not only have the potential for being diagnosed or prognosticated by our NWP model on physical grounds (cloud microphysics, boundary layer scaling arguments, etc.), but are also ideal for us in the context of the assimilation of high-resolution cloud data from MODIS and other instruments, which contain significant information content on sub-NWP-scale variability. We do not assume that the particularly subgrid-scale model used in this paper is ideal, or

our final destination, but we believe it is a reasonable starting point that is heavily tied to future plans of expansion within GEOS-5.

The clumping algorithm is the only new element of our cloud generation algorithm beyond what is fully described in Norris and da Silva (2013). The clumping algorithm simply re-samples the horizontal position of the ICA-generated grid-column subcolumns or "pixels" to introduce some idea of spatial coherence. This makes negligible difference to the ICA-calculated radiative properties of the subcolumn ensemble, since the subcolumns are not changed in any way, only resampled. So, like the downscaled fields of Venema et al. (2010), our un-clumped ICA-generated subcolumns are still able to give vastly improved mean ID radiative properties for a subcolumn, by averaging over the non-linear radiative calculations on the subcolumn level. Clumping would make a difference if we had been using a 3-D radiation calculation, but this is currently not even close to being possible, because of computational expenses, for the applications we have in mind for our simulation tool. Rather, we do the above "clumping" both for a more realistic visual appearance of the simulated pixels and also because spatial coherence does impact the "clear sky-restoral" within the MODIS optical property retrieval, which affectively weeds out cloud edge pixels. The simple Gaussian copula-based horizontal clumping algorithm we describe has a parametric horizontal decorrelation length scale for condensed water path (CWP), which is currently set at 5 km, but which again can be easily brought within the parameter estimation capabilities of new the cloud data assimilation system, based on the MODIS-observed CWP decorrelation scale.

In summary, the main difference between and the spectral extrapolation downscaling approach of Venema et al. (2010) and of Bugliaro et al. (2011) (discussed in the Sect. 1) and our approach here is one of extrapolating NWP variability to smaller scales versus externally anchoring the subgrid column statistical behavior parametrically. Ultimately the accuracy of our method will depend on the quality of our subgrid column statistical model, and of our choice of its parameters. We anticipate that the quality of this model, and the estimation of its parameters from assimilated high-resolution satellite cloud data will lead to continuously improving accuracy. Similar comments also apply to the Venema et al. (2010) downscaling method, which is also constrained in accuracy by its extrapolation assumptions, and will presumably be improved over time. There *is* one other difference that we acknowledge as a potential problem with our approach, namely that it is grid-column-based and does not deal with discontinuities between grid columns as the extrapolation method does. Our argument is that the error reduction in model radiative properties achieved by subcolumn radiative averaging dominates over other smaller reductions due to grid-column edge effects or 3-D radiative effects, especially for grid columns of 5 km and larger scale (see comments by

Jonkheid et al., 2012). Our target NWP resolution is 25 km, as used in this paper.

The sophisticated subgrid-column cloud generator described above and used in this paper is only one of many possible such generators. A less complicated example, very much akin to the internal GEOS-5 treatment of cloud overlap, would be the following "homogeneous cloud, maximum-random overlap" generator: divide the atmosphere into pressure bands (e.g., low, middle and high bands) with interfaces at 700 and 400 hPa. Say we again wish to generate N subcolumns, $n = 1, \ldots, N$, for the grid column. Then for each pressure band, generate a set of N uniform random numbers $\{r_n\}$ on [0,1], and for each model layer k falling within the band, assign cloudiness to layer k of subcolumn n if $r_n < f_k$, where f_k is that layer's cloud fraction. The fact that the *same* set $\{r_n\}$ is used for each layer k in the band enforces maximum cloud overlap *within* the band. But choosing independent sets of $\{r_n\}$ for each pressure band enforces random overlap *between* the bands. Finally, every subcolumn that is cloudy at layer k shares the same homogeneous in-cloud condensate contents $\frac{q(i,j)k}{f_k}$, where $q_{(i,l)k}$ are the layer mean condensate contents ($i = $ ice, $l = $ liquid water). (Note that this simple generator, as with the more sophisticated generator we use, produces subcolumns of *condensate*. The specification of optical thicknesses from condensate contents proceeds on the subcolumn level for both generators. We emphasize this because the reader should be aware of the traps associated with the alternative strategy of using diagnostic *layer* COT *directly* from global climate model (GCM) (e.g., GEOS-5) output files. When using diagnostic layer COT directly, one must know whether they are in-cloud or "layer mean" and, if in-cloud, for what cloud fraction. One cannot simply interpret the column consequences of model layer cloud diagnostics without a knowledge of the model's cloud overlap.)

Any type IV model validation studies made by comparing real and simulated retrievals can be compromised by a poor forward model (including subcolumn cloud generation and radiative transfer). The more realistic the forward model, the more useful is the model validation, because the real observations use a "perfect forward model". In this paper we have used the sophisticated forward model described (skewed triangle PDFs, copula overlap, copula clumping), rather than the simpler "homogeneous cloud, maximum-random overlap" generator used in the GEOS-5 internal radiation calculations, because we believe the former is a more realistic forward model based on previous studies (Norris et al., 2008; da Silva and Norris, 2013). We have also used the sophisticated DISORT radiation code for satellite radiances rather than trying to adapt the simpler internal GCM radiation codes of GEOS-5 for this purpose. Both these decisions are appropriate – we should use the best forward model possible in order to make the best attribution of differences in retrieval space back to errors in the vapor and condensate distributions

within the model forecast/analysis. Ultimately, in future studies, we must compare model validations made using different forward models in order to gauge the influence of the forward model on the conclusions. But, for now, we simply note that finding a better forward model than the GEOS-5 internal model is a good reason to replace the GEOS-5 internal model. Indeed, the improved forward model we have described is already part of the new GEOS-5 cloud data assimilation system and will also be implemented in GEOS-5 within the near future.

2.4 Radiative transfer calculation

Now that we have collected all the necessary information about atmosphere and cloud layers, we begin the simulation process. The radiative transfer calculations were performed using the discrete ordinate radiative transfer (DISORT) code (Stamnes et al., 1988) with liquid water cloud phase function results from Mie calculations based on gamma distribution water droplet size distributions with an effective variance of 0.1 and bulk ice cloud phase models developed by Baum et al. (2005), both consistent with MOD06. Generally a large number of streams is required to model the forward peak of the phase function and multiple scattering components accurately. The forward peak however can be further truncated, and use of the delta-fit method of Hu et al. (2000) can be considered sufficiently accurate, as described by Ding et al. (2009), for calculations where there is no stored accuracy limit such as the multilayer cloud simulations that use this exact simulation method in Wind et al. (2010). We have applied the delta-fit method and truncation to phase functions before use in the simulator. We have experimented with a different number of computational streams in order to balance speed and desired accuracy. We found that only 16 streams were required to achieve the needed precision. Initial calculations were done with 32 streams; however, the execution time was rather prohibitive. We settled on 16 streams as a balance between execution time and precision as the difference in resulting simulated sensor radiance between 32 and 16 stream simulations was less than 0.5 %. The simulator itself however is not limited as to how many streams the user can specify. The simulator is only limited by available computational resources for particular application. As we pre-calculate the surface spectral albedos, we can save further time by calling DISORT in Lambertian mode with predetermined values. When we encounter cloud subcolumns over the ocean, however, we must adjust the computed Cox–Munk surface albedo to compensate for the diffuse illumination that the presence of the cloud creates. A good value for the diffuse illumination albedo of a water surface is 0.05 (Platnick et al., 2003). We then linearly fit surface albedo as a function of cloud optical thickness, with full diffuse illumination at a total column cloud optical thickness of 3 and full Cox–Munk surface albedo at total column cloud optical thickness of 0.
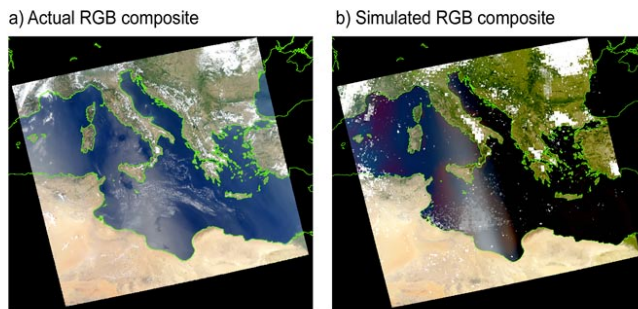
Fig. 1. Synthetic sensor radiance simulation together with an actual MODIS granule that was used as study area. Aqua MODIS granule 2012 day 228 at 12:00 UTC. GEOS-5 temporal fit between 12:00 and 18:00 UTC, 15 August 2012. RGB (0.67 µm, 0.55 µm, 0.47 µm).

3 Example retrievals

In this section we discuss two example results of radiance simulations and subsequent cloud property retrievals. We performed the simulation on the NASA Center for Climate Simulations (NCCS) Discover system using 12 Intel Westmere nodes with 12 cores each. The memory footprint of the software suite is very small, around 80 Mb peak usage, but the process is quite CPU-heavy. A full-resolution 1 km simulation using a full MODIS granule as a study area took about three and a half hours of wall clock time to complete. Figure 1 shows the resulting true-color RGB image of sample MODIS granule 2012 day 228 (15 August) at 12:00 UTC together with the true-color image of the actual MODIS granule before its channel data were replaced. Panel a shows the actual data acquired by Aqua MODIS, and panel b shows the simulation result. GEOS-5 does not assimilate cloudy radiances, and so there should be little expectation of a granule-level feature match. However, in this case the model does remarkably well with cloud placement. Bands of cloud over southern France are present and located properly, as are clouds over the northern Balkans and southern Asia Minor. The orographic clouds over Italy and Greece are also present, as are scattered clouds over the Sahara desert. There are some rather important differences between the model and the actual data, however, when it comes to cloud properties.

Figure 2 shows the results of running the Data Collection 5.1 operational retrieval chain on the resulting L1B file from model fields. Panel a shows the cloud thermodynamic phase, panel b the cloud top pressure, panel c the cloud optical thickness and panel d the cloud effective radius retrieved with the VNSWIR (visible, near- or shortwave infrared) and 2.1 µm channel combination. Figure 3 shows the actual Aqua MODIS retrieval for that same granule using identical panel arrangement.

The cloud field over the central Mediterranean is given improper vertical location by the model. The actual cloud field is retrieved as liquid water and is low cloud, with cloud top
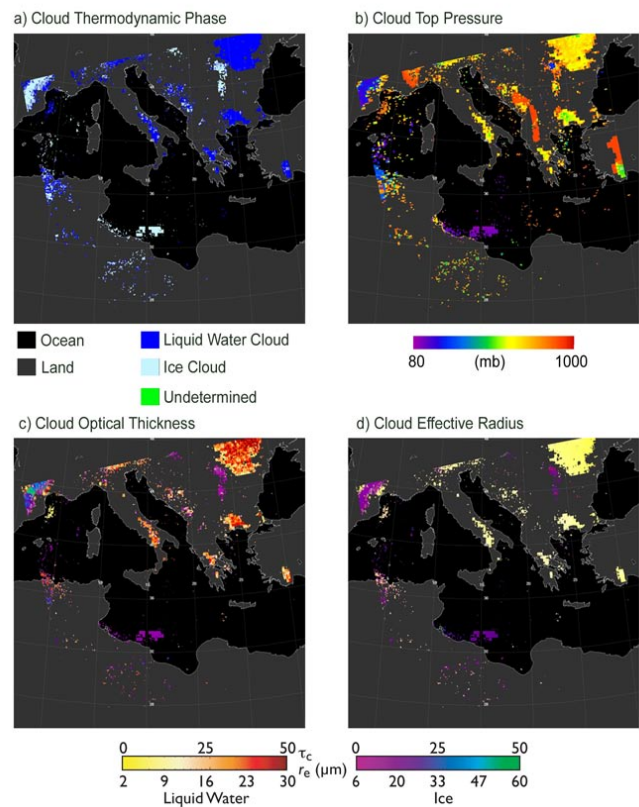

Fig. 2. Example cloud retrieval for simulated granule covered by Aqua MODIS 2012 day 288 12:00 UTC. **(a)** is cloud thermodynamic phase, **(b)** cloud top pressure, **(c)** cloud optical thickness and **(d)** cloud effective radius from the 0.86–2.1 µm band combination.

pressures of 800–900 mb. The model generates a thin cirrus cloud in that location with cloud top pressure of about 100 mb and of course ice thermodynamic phase. We know that this discrepancy is not related to any issue in retrieval method because we are able to examine model fields coming into simulation directly. Incoming model profiles in that area do not contain any cloud in lower troposphere and concentrate all cloud layers in the upper troposphere instead. This has serious implications for outgoing radiation. The cloud field over Romania has a consistent phase, but the model indicates the cloud to be positioned somewhat higher in altitude than the observation and also significantly optically thicker than what is observed. The same is true for the cloud field over NW Turkey.

Figure 4 shows a cloud top pressure/cloud optical thickness joint histogram for the actual granule in panel a and simulated one in panel b. While in this comparison we are not necessarily looking for quantitative evaluation of model parameters, some things do tend to jump out. The actual MODIS granule has mostly low clouds that are moderately thick. The simulated granule on the other hand lacks low clouds almost entirely and instead produces thicker clouds at high altitude. The RGB images look very similar in this
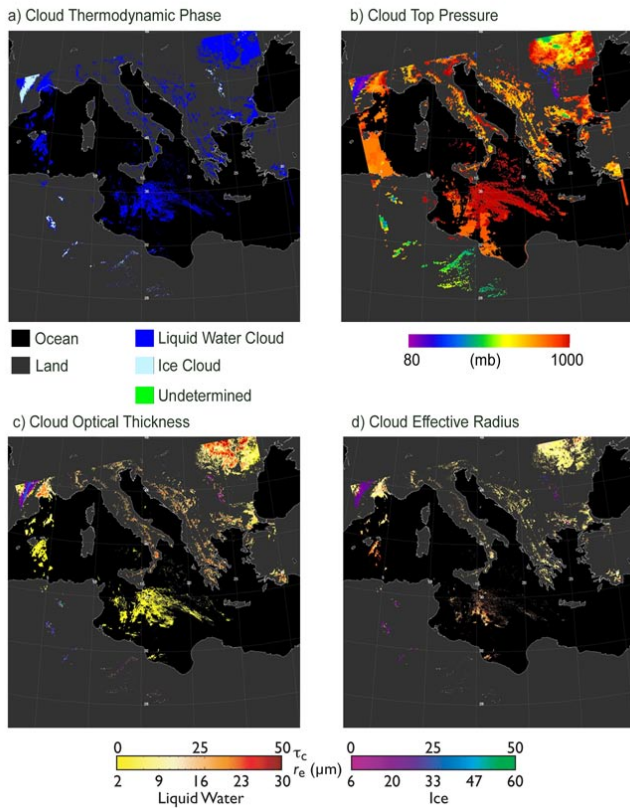
Fig. 3. Actual cloud retrieval for Aqua MODIS 2012 day 228 at 12:00 UTC. **(a)** is cloud thermodynamic phase, **(b)** cloud top pressure, **(c)** cloud optical thickness and **(d)** cloud effective radius from the 0.86–2.1 µm band combination.
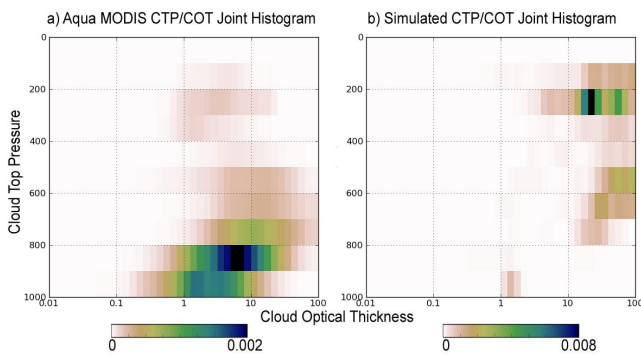

Fig. 4. Joint histograms of cloud optical thickness vs. cloud top pressure for actual **(a)** and model-based **(b)** cloud fields covered by Aqua MODIS 2012 day 228 at 12:00 UTC.

case, so the model is performing well on geographical cloud placement, but fails rather badly when it comes to proper cloud placement in altitude. This kind of disconnect can have some significant implications for earth radiative budget calculations. It is one of our future goals to determine just how frequently such disconnects occur on the global scale.
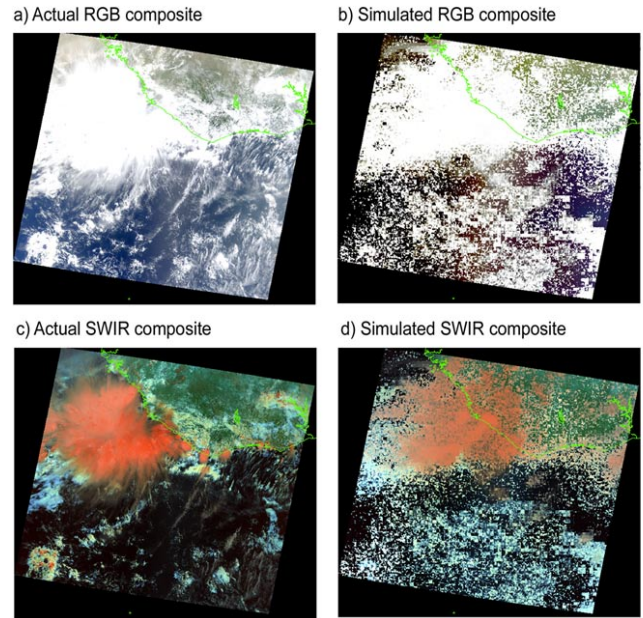

Fig. 5. Synthetic sensor radiance simulation together with an actual MODIS granule that was used as study area. Terra MODIS granule 2013 day 151 at 11:15 UTC. GEOS-5 temporal fit between 06:00 and 12:00 UTC, 31 May 2013. RGB (0.67 µm, 0.55 µm, 0.47 µm), SWIR(0.86 µm, 1.6 µm, 2.1 µm).

Figures 5 through 8 show another simulation example, this time from Terra MODIS 2013 day 151 (31 May 2013) at 11:15 UTC. This granule shows a large convective system over the coast of Liberia and Sierra Leone in western Africa. Figure 5 shows a true-color RGB image for simulated and actual MODIS granule together with the shortwave infrared (SWIR) composite that allows the user to estimate cloud thermodynamic phase visually. Ice clouds appear red in such an image. We cannot show a SWIR composite for Aqua MODIS because of detector issues with the Aqua MODIS 1.6 µm channel that is needed to create the image. Figure 6 shows retrieval results for the simulated granule, and Fig. 7 shows the actual Terra MODIS granule retrievals. Figure 8 shows the joint histograms of cloud top pressure and cloud optical thickness. In this case there is actually reasonable agreement between sensor measurement and model cloud field representation both geographically and vertically. The model could have benefitted from producing somewhat more mid-level clouds, but overall the large convective system that dominates the scene is represented reasonably well, as are the broken clouds around it. The large cloud-free void in retrievals performed on simulated data is due to MODIS operational cloud mask (MOD35) having difficulties. The operational cloud mask uses much more restrictive thresholds in the sun glint region, and as this simulation is performed without aerosols, resulting radiances may not fit well with the cloud mask sun glint thresholds. Data Collection 5 operational MODIS cloud mask however is known to have
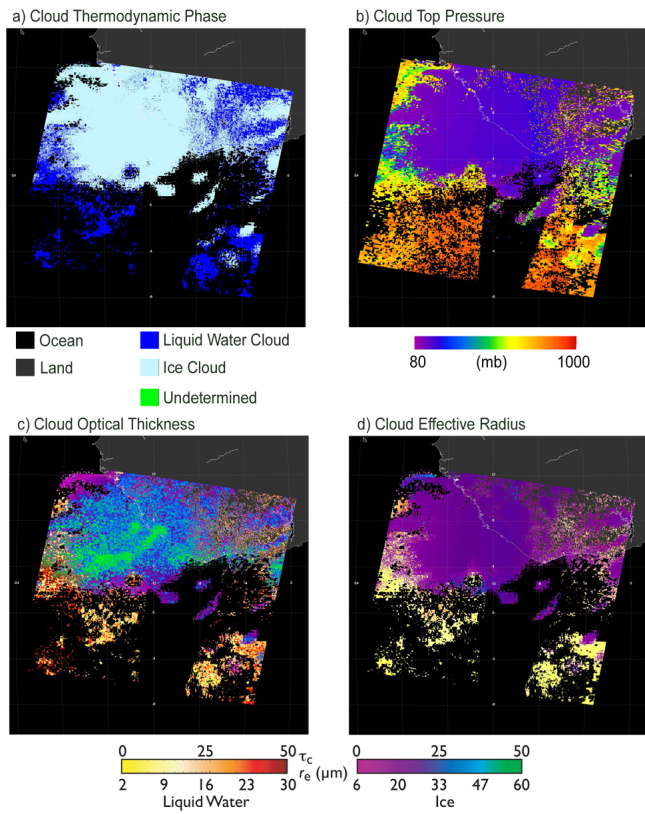
Fig. 6. Example cloud retrieval for simulated granule covered by Terra MODIS 2013 day 151 11:15 UTC. **(a)** is cloud thermodynamic phase, **(b)** cloud top pressure, **(c)** cloud optical thickness and **(d)** cloud effective radius from the 0.86–2.1 μm band combination.

Fig. 7. Actual cloud retrieval for Terra MODIS 2013 day 151 at 11:15 UTC. **(a)** is cloud thermodynamic phase, **(b)** cloud top pressure, **(c)** cloud optical thickness and **(d)** cloud effective radius from the 0.86–2.1 μm band combination.

some retrieval issues in sun glint geometry, and so this particular behavior does not come as a complete surprise. This retrieval may improve when Data Collection 6 algorithms become available, and we would absolutely repeat this simulation with updated cloud mask product code. If the issue persists, then this particular case might be useful as cloud detection diagnostics since the source model data tell us explicitly that the cloud field is indeed present.

The MODIS cloud top pressure retrieval is quite sensitive to ancillary atmospheric profile information (Menzel et al., 2008), and some of differences found in retrievals could be a result of different representations of the atmospheric profile by GEOS-5 and the NCEP Mesoscale Meteorological Model (MM5)-derived model profiles used during the MODIS retrieval.

Situations in which cloud optical thickness retrievals show significant differences tend to be more indicative of significant differences in cloud structure. Unlike cloud top pressure, cloud optical thickness retrievals have very little dependence on atmospheric profile information as there is very little atmospheric absorption in the 0.65 and 0.86 μm channels used to retrieve this quantity.
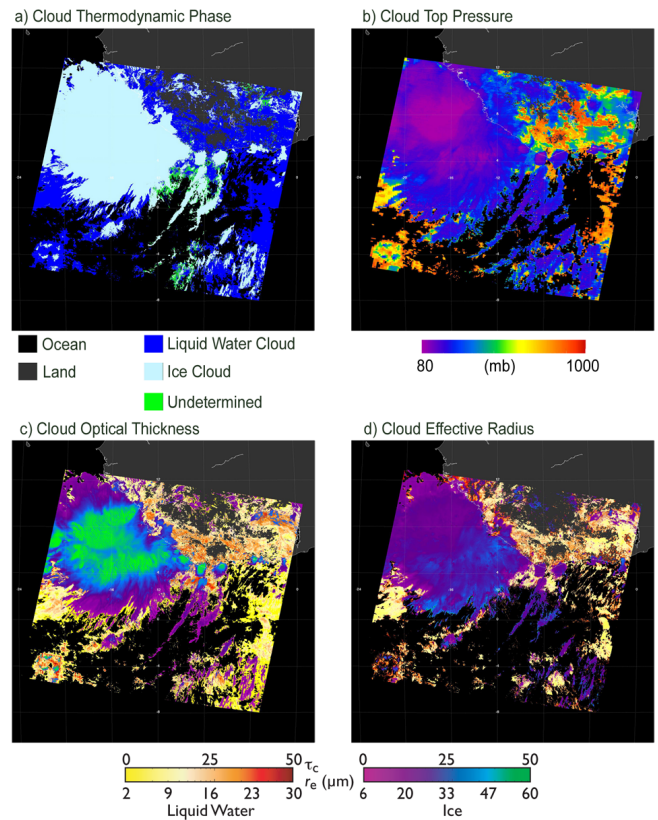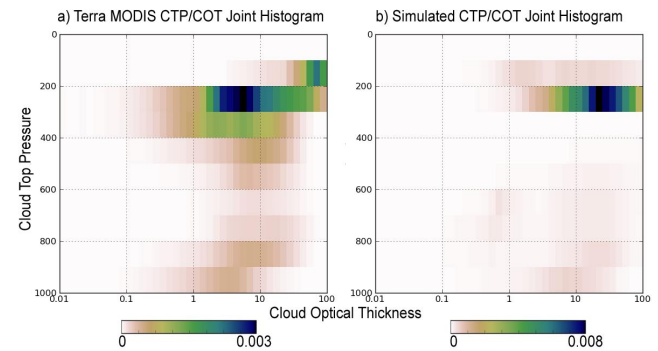

Fig. 8. Joint histograms of cloud optical thickness vs. cloud top pressure for actual **(a)** and model-based **(b)** cloud fields covered by Terra MODIS 2013 day 151 at 11:15 UTC.

Cloud effective radius retrievals from the 2.1 μm channel depend somewhat on the atmospheric profile, but differences in that retrieval are also mainly due to differences in cloud microphysics present in the model and in the actual atmosphere. Retrieved cloud effective radius appears to be somewhat smaller overall for GEOS-5 data than for MODIS. Even though the clumped-ICA cloud formation method allows us

to model some of the scene inhomogeneity normally encountered in actual MODIS data, the present implementation of the simulator does not admit subpixel-based effective radius artifacts such as ones appearing in MODIS (Zhang and Platnick, 2011). Also, GEOS-5 uses a simple diagnostic prescription of cloud effective radius that is loosely based on large-scale observations of aerosol concentrations and their physical connection to cloud droplet size, and with the details being adapted based on consistency with surface radiation budget estimates of shortwave cloud forcing. It is therefore not surprising that, in our preliminary results, there appears to be generally somewhat less variability in retrieved cloud effective radius from GEOS-5 than from real MODIS data.

4 Conclusions and future directions

We have developed a flexible software suite that allows us to interface model fields to operational satellite remote sensing retrieval algorithms. We have presented an example of its operation using the GEOS-5 model and MODIS instrument. We have demonstrated the power of this software in locating and quantifying problems with GEOS-5 cloud optical properties and cloud vertical distributions within the specific geographic and synoptic contexts observed in MODIS cloud granules.

In subsequent papers we will show a number of applications of this software. Our current plans include performing a number of large-scale simulation experiments using GEOS-5 nature run model data with resolutions as high as 7 km. We would like to examine impact of aerosols on cloud retrievals by performing simulations with and without model aerosol fields. Once operational cloud and aerosol retrieval algorithms are applied to such data, we may be able to quantify some aerosol effects on clouds and maybe even find some ways to retrieve aerosols above clouds.

We would also like to examine in detail the performance of COSP, the Cloud Feedback Model Intercomparison Project (CFMIP) Observation Simulator Package, which provides a means of simulating retrieved cloud properties from a variety of instruments, including MODIS, from model fields. The COSP package typically uses approximate, so-called "fast" simulators of cloud properties, whereas we are directly simulating radiances and performing the retrievals using actual operational retrieval codes. We thus have the ability to test the quality of the COSP fast simulators

We also intend to apply this software suite to other remote sensing instruments, such as SEVIRI (Spinning Enhanced Visible Infrared Radiometer Imager) on board the Meteosat Second Generation geostationary satellite series. SEVIRI has a sufficient number of channels that MODIS-style cloud retrieval algorithms can be applied. SEVIRI acquires data in 15 min intervals at 3 km nadir resolution, thus giving fine temporal and spatial resolution for diurnal sampling. Applying the simulated sensor radiance package to SEVIRI would allow us potentially to examine model cloud dynamics at a temporal resolution not offered by polar orbiting satellites.

We are confident that there are many other applications for this software that will be found in the future besides ones outlined above and that it will become a valuable tool for both the remote sensing and modeling communities.

The simulator code is available to users free of charge by contacting the authors and becoming a registered user of this software package so that any updates can be issued directly. There may be additional, wider distribution means in the future if the software shows signs of growing popularity.

Edited by: K. Gierens

References

Ackerman, A., Strabala, K., Menzel, P., Frey, R., Moeller, C., Gumley, L., Baum, B., Seemann, S. W., and Zhang, H.: Discriminating clear-sky from cloud with MODIS Algorithm Theoretical Basis Document (MOD35). ATBD Reference Number: ATBD-MOD-06, available at: http://modis-atmos.gsfc.nasa.gov/reference_atbd.html (last access: 23 July 2013), 2006.

Barnes, W. L., Pagano, T. S., and Salomonson, V. V.: Prelaunch characteristics of the Moderate Resolution Imaging Spectroradiometer (MODIS) on EOS-AM1, IEEE Trans. Geosci. Remote Sens., 36, 1088–1100, 1998.

Baum, B. A., Heymsfield, A. J., Yang, P., and Bedka, S. T.: Bulk scattering models for the remote sensing of ice clouds, Part 1: Microphysical data and models, J. Appl. Meteor., 44, 1885–1895, 2005.

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

Bony, S., Colman, R., Kattsov, V. M., Allan, R. P., Bretherton, C. S., Dufresne, J. L., Hall, A., Hallegatte, S., Holland, M. M., Ingram, W., Randall, D. A., Soden, B. J., Tselioudis, G., and Webb, M. J.: How well do we understand and evaluate climate change feedback processes?, J. Climate, 19, 3445–3482, 2006.

Bugliaro, L., Zinner, T., Keil, C., Mayer, B., Hollmann, R., Reuter, M., and Thomas, W.: Validation of cloud property retrievals with simulated satellite radiances: a case study for SEVIRI, Atmos. Chem. Phys., 11, 5603–5624, doi:10.5194/acp-11-5603-2011, 2011.

Colarco, P., da Silva, A., Chin, M., and Diehl, T.: Online simulations of global aerosol distributions in the NASA GEOS-4 model and comparisons to satellite and ground-based aerosol optical depth, J. Geophys. Res., 115, D14207, doi:10.1029/2009JD012820, 2010.

Cox, C. and Munk, W.: Measurement of the Roughness of the Sea Surface from Photographs of the Sun's Glitter, J. Opt. Soc. Am., 44, 838–850, 1954.

da Silva, A. M. and Norris, P. M.: Monte Carlo Bayesian inference on a statistical model of sub-gridcolumn moisture variability using high-resolution cloud observations. Part II: Sensitivity tests and results, J. Q. R. Meteorol. Soc., submitted, 2013.

Derber, J. C., Purser, R. J., Wu, W.-S., Treadon, R., Pondeca, M., Parrish, D., and Kleist, D.: Flow-dependent Jb in a global gridpoint 3D-Var. Proc. ECMWF annual seminar on recent developments in data assimilation for atmosphere and ocean, Reading, UK, 8–12 Sept. 2003, 2003.

Ding, S., Xie, Y., Yang, P., Weng, F., Liu, Q., Baum, B., and Hu, Y.: Estimates of radiation over clouds and dust aerosols: Optimized number of terms in phase function expansion, J. Quant. Spectrosc. Radiat. Transfer, 110, 1190–1198, 2009.

Donovan, D. P., Voors, R. H., van Zadelhoff, G.-J., and Acarreta, J.-R.: ECSIM Model and Algorithms Document, KNMI Tech. Report: ECSIM_KNMI_TEC_MAD01_R, Version 1.2.2, 24 Oct 2008, Royal Netherlands Meteorological Institute (KNMI), The Netherlands, 2008.

Feldman, D. R., Algieri, C. A., Ong, J. R., and Collins, W. D.: CLARREO shortwave observing system simulation experiments of the twenty-first century: Simulator design and implementation, J. Geophys. Res., 116, D10107, doi:10.1029/2010JD015350, 2011.

Frey, R. A., Ackerman, S. A., Liu, Y., Strabala, K. I., Zhang, H., Key, J., and Wang, X.: Cloud Detection with MODIS, Part I: Recent Improvements in the MODIS Cloud Mask, J. Atmos. Oceanic Technol., 25, 1057–1072, doi:10.1175/2008JTECHA1052.1, 2008.

Hill, C., DeLuca, C., Balaji, V., Suarez, M., and da Silva, A.: The architecture of the Earth System Modeling Framework, Comp. Sci. Engr., 6, 18–28, 2004.

Hsu, N. C., Tsay, S., King, M. D., and Herman, J. R.: Aerosol Properties Over Bright-Reflecting Source Regions, IEEE Trans. Geosci. Rem. Sens., 42, 557–569, 2004.

Hu, Y.-X., Wielicki, B., Lin, B., Gibson, G., Tsay, S.-C., Stamnes, K., and Wong, T.: d-Fit: A fast and accurate treatment of particle scattering phase functions with weighted singular-value decomposition least-squares fitting, J. Quant. Spectrosc. Radiat. Transfer, 65, 681–690, 2000.

Hubanks, P. A., King, M. D., Platnick, S. A., and Pincus, R. A.: MODIS Atmosphere L3 Gridded Product Algorithm Theoretical Basis Document, ATBD Reference Number: ATBD-MOD-30, http://modis-atmos.gsfc.nasa.gov/MOD08_M3/atbd.html (last access: 23 July 2013), 2008.

Intergovernmental Panel on Climate Change (IPCC): Climate Change 2007: The Physical Science Basis, Contribution of Worrking Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge Univ. Press, Cambridge, UK, 2007.

Jonkheid, B. J., Roebeling, R. A., and van Meijgaard, E.: A fast SEVIRI simulator for quantifying retrieval uncertainties in the CM SAF cloud physical property algorithm, Atmos. Chem. Phys., 12, 10957–10969, doi:10.5194/acp-12-10957-2012, 2012.

King, M. D., Menzel, W. P., Kaufman, Y. J., Tanré, D., Gao, B. C., Platnick, S., Ackerman, S. A., Remer, L. A., Pincus, R., and Hubanks, P. A.: Cloud and aerosol properties, precipitable water, and profiles of temperature and humidity from MODIS, IEEE Trans. Geosci. Remote Sens., 41, 442–458, 2003.

King, M. D., Platnick, S., Menzel, W. P., Ackerman, S. A., and Hubanks, P. A.: Spatial and Temporal Distribution of Clouds Observed by MODIS onboard the Terra and Aqua Satellites, IEEE Trans. Geosci. Remote Sens., 57, 3826–3852, 2013.

Kratz, D. P.: The correlated k-distribution technique as applied to the AVHRR channels, J. Quant. Spectrosc. Radiat. Transfer, 53, 501–517, 1995.

Lee, J., Yang, P., Dessler, A. E., Gao, B.-C., and Platnick, S.: Distribution and radiative forcing of tropical thin cirrus clouds, J. Atmos. Sci., 66, 3721–3731, doi:10.1175/2009JAS3183.1, 2009.

Menzel, W. P., Frey, R., Zhang, H., Wylie, D., Moeller, C., Holz, R., Maddux, B., Baum, B. A., Strabala, K., and Gumley, L.: MODIS global cloud-top pressure and amount estimation: Algorithm description and results, J. Appl. Meteor. Climatol., 47, 1175–1198, 2008.

Molod, A., Takacs, L., Suarez, M., Bacmeister, J., Song, I.-S., and Eichmann, A.: The GEOS-5 Atmospheric General Circulation Model: Mean Climate and Development from MERRA to Fortuna, Tech. Rep. S. Gl. Mod. Data Assim., 28, 2012.

Moody, E. G., King, M. D., Platnick, S., Schaaf, C. B., and Gao, F.: Spatially complete global spectral surface albedos: Value-added datasets derived from Terra MODIS land products, IEEE Trans. Geosci. Remote Sens., 43, 144–158, 2005.

Moody, E. G., King, M. D., Schaaf, C. B., Hall, D. K., and Platnick, S.: Northern Hemisphere five-year average (2000–2004) spectral albedos of surfaces in the presence of snow: Statistics computed from Terra MODIS land products, Remote Sens. Environ., 111, 337–345, 2007.

Moody, E. G., King, M. D., Schaaf, C. B., and Platnick, S.: MODIS-derived spatially complete surface albedo products: Spatial and temporal pixel distribution and zonal averages, J. Appl. Meteor. Climatol., 47, 2879–2894, 2008.

Norris, P. M., Oreopoulos, L., Hou, A. Y., Tao, W.-K., and Zeng, X.: Representation of 3D heterogeneous cloud fields using copulas: Theory for water clouds, J. Q. R. Meteorol. Soc., 134, 1843–1864, doi:10.1002/qj.321, 2008.

Norris, P. M. and da Silva, A. M.: Monte Carlo Bayesian inference on a statistical model of sub-gridcolumn moisture variability using high-resolution cloud observations, Part I: Method, J. Q. R. Meteorol. Soc., submitted, 2013.

Oreopoulos, L. and Norris, P. M.: An analysis of cloud overlap at a midlatitude atmospheric observation facility, Atmos. Chem. Phys., 11, 5557–5567, doi:10.5194/acp-11-5557-2011, 2011.

Otkin, J. A., Posselt, D. J., Olson, E. R., Huang, H.-L., Davies, J. E., Li, J., and Velden, C. S.: Mesoscale Numerical Weather Prediction Models Used in Support of Infrared Hyperspectral Measurement Simulation and Product Algorithm Development, J. Atmos. Ocean. Technol., 24, 585–601, doi:10.1175/JTECH1994.1, 2007.

Pawson, S. R., Stolarski, S., Douglass, A. R., Newman, P. A., Nielsen, J. E., Frith, S. M., and Gupta, M. L.: Goddard Earth Observing System chemistry-climate model simulations of stratospheric ozone-temperature coupling between 1950 and 2005, J. Geophys. Res., 113, D12103, doi:10.1029/2007JD009511, 2008.

Pincus, R., Hannay, C., Klein, S. A., Xu, K.-M., and Hemler, R.: Overlap assumptions for assumed-PDF cloud schemes in large-scale models, J. Geophys. Res., 110, D15S09, doi:10.1029/2004JD005100, 2005.

Pincus, R., Platnick, S., Ackerman, S. A., Hemler, R. S., and Hofmann, R. J. P.: Reconciling simulated and observed views of clouds: MODIS, ISCCP, and the limits of instrument simulators, J. Climate, 25, 4699–4720, doi:10.1175/JCLI-D-11-00267.1, 2012.

Platnick, S., King, M. D., Ackerman, S. A., Menzel, W. P., Baum, B. A., Riedi, J. C., and Frey, R. A.: The MODIS cloud products: Algorithms and examples from Terra, IEEE Trans. Geosci. Remote Sens., 41, 459–473, 2003.

Räisänen, P., Barker, H. W., Khairoutdinov, M. F., Li, J., and Randall, D. A.: Stochastic generation of subgrid-scale cloudy columns for large- scale models, Q. J. R. Meteorol. Soc., 130, 2047–2068, 2004.

Rienecker, M. M., Suarez, M. J., Todling, R., Bacmeister, J., Takacs, L., Liu, H.-C., Gu, W., Sienkiewicz, M., Koster, R. D., Gelaro, R., Stajner, I., and Nielsen, J. E.: The GEOS-5 Data Assimilation System – Documentation of Versions 5.0.1, 5.1.0, and 5.2.0. Tech. Rep. S. Gl. Mod. Data Assim., 27, 2008.

Riishojgaard, L. P.: A direct way of specifying flow-dependent background error correlations for meteorological analysis systems, Tellus, 50A: 42–57, 1998.

Schaaf, C. L. B., Liu, J., Gao, F., and Strahler, A. H.: MODIS Albedo and Reflectance Anisotropy Products from Aqua and Terra, In Land Remote Sensing and Global Environmental Change: NASA's Earch Observing System and the Science of ASTER and MODIS, Remote Sensing and Digital Image Processing Series, edited by: Ramachandran, B., Justice, C., and Abrams, M., Vol. 11, Springer-Cerlag, 873 pp., 2011.

Seemann, S. W., Borbas, E. E., Knuteson, R. O., Stephenson, G. R., and Huang, H.-L.: Development of a Global Infrared Land Surface Emissivity Database for Application to Clear Sky Sounding Retrievals from Multispectral Satellite Radiance Measurements, J. Appl. Meteor. Climatol., 47, 108–123, 2008.

Stamnes, K., Tsay, S. C., Wiscombe, W., and Jayaweera, K.: Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media, Appl. Opt., 27, 2502–2509, 1988.

Venema, V., Garcia, S. G., and Simmer, C.: A new algorithm for the downscaling of cloud fields, Q. J. R. Meteor. Soc., 136, 91–106, doi:10.1002/qj.535, 2010.

Voors, R., Donovan, D., Acarreta, J., Eisinger, M., Franco, R., Lajas, D., Moyano, R., Pirondini, F., Ramos, J., and Wehr, T.: ECSIM: The simulator framework for EarthCARE, Sensors, Systems, and Next-Generation Satellites XI, edited by: Meynart, R., Neeck, S. P., Shimoda, H., and Habib, S., Proc. of SPIE, 6744, 67441Y, doi:10.1117/12.737738, 2007.

Wind, G., Platnick, S., King, M. D., Hubanks, P. A., Pavolonis, M. J., Heidinger, A. K., Yang, P., Baum, B.: Multilayer Cloud Detection with the MODIS Near-Infrared Water Vapor Absorption Band, J. App. Met. Clim., 49, 2315–2333, 2010.

Wu, W. S., Purser, R. J., and Parrish, D. F.: Three-dimensional variational analysis with spatially inhomogeneous covariances, Mon. Weather Rev., 130, 2905–2916, 2002.

Xiong, X., Isaacman, A. T., and Barnes, W.: MODIS Level-1B Products. Earth Science Satellite Remote Sensing, Science and Instruments, 1, 33–49, 2006.

Zhang, Z. and Platnick, S.: An assessment of differences between cloud effective particle radius for marine water clouds from three MODIS spectral bands, J. Geophys. Res., 116, D20215, doi:10.1029/2011JD016216, 2011.