

Research Article

Climate Cloud Model Forecast Verification-an Engineering PerspectiveKeith D Hutchison ^{1,*}, Barbara Iisager ^{2,†}

1. Center for Space Research, The University of Texas, Austin, TX, USA; E-Mail: keithh@csr.utexas.edu
2. Cloud Systems Research, LLC, Austin, TX, USA; E-Mail: barbara.iisager@gmail.com

† These authors contributed equally to this work.

* **Correspondence:** Keith D Hutchison; E-Mail: keithh@csr.utexas.edu**Academic Editor:** Alfredo Moreira Caseiro Rocha**Special Issue:** [Remote Sensing on Climate Change](#)

Adv Environ Eng Res
2020, volume 1, issue 4
doi:10.21926/aeer.2004004

Received: September 13, 2020
Accepted: November 15, 2020
Published: November 24, 2020

Abstract

The processes relevant to the verification of cloud forecasts generated by climate models are discussed from an engineering perspective. These processes include an assessment of cloud product requirements to be evaluated, the creation of a verification test plan including procedures and data to be analyzed, the development of independent sources of validation or truth datasets, and the quantitative comparisons between the cloud forecast products and the truth data needed to establish model performance. The engineering perspective means minimal effort is focused on assessing the veracity of the physics contained in the cloud forecast model, rather emphasis is upon evaluating the results produced by it. It is postulated that these procedures are critical to improve the reliability of climate model predictions. The World Meteorological Organization has stated accuracy requirements for cloud products created from satellite observations, through the Global Climate Observing System (GSOC) program; however, no similar requirements have been defined for cloud forecast products. A statement of accuracy requirements is urgently needed. Meanwhile, it is assumed herein that cloud observation and cloud forecast requirements are identical. The assessment of model performance exploits high quality, manually-generated cloud truth products created from



© 2020 by the author. This is an open access article distributed under the conditions of the [Creative Commons by Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium or format, provided the original work is correctly cited.

remotely-sensed satellite data which serve as truth data. Results show clouds under-specified in reanalysis cloud datasets created for use to initialize climate models but an over-specification of clouds by the cloud forecast model, in short-range predictions. This system level analysis demonstrates the need to improve the accuracy of cloud forecasts, especially lower-level water clouds which are responsible for most of the uncertainty in climate model predictions.

Keywords

VIIRS; WRF; NAM; cloud forecasting; climate models; model verification

1. Introduction

The critical roles clouds play in climate modeling has also long been recognized since they are the key regulators of the Earth's energy budget [1, 2]. In fact, cloud feedbacks have been identified as the largest internal source of uncertainty in climate change predictions [3, 4] and differences between cloud feedback mechanisms in climate models represent the leading source of spread in estimates of climate sensitivity [5]. While inter-model comparisons have shown that low-level water clouds are responsible for most of this spread in estimates between climate models, there is also a lack in the understanding of the physical processes that control boundary layer clouds and their radiative properties [6]. The establishment of cloud forecast performance standards, e.g. cloud amount or cloud cover fraction (CCf), would be useful to help mitigate these uncertainties in model sensitivities due and to reduce the spread in global temperature predictions. Thus, establishing procedures to verify cloud model forecast performance, from an engineering perspective, could greatly aid in creating these performance standards.

The processes relevant to the verification of cloud forecast products generated by numerical weather prediction and/or global/regional climate models include (1) a definition of cloud product requirements to be evaluated, (2) the articulation of a model evaluation test plan, including the metrics proposed to establish performance, (3) the development of independent source(s) of validation or truth datasets, and (4) the quantitative comparisons between the cloud forecast products and the truth data to establish model performance. From this engineering perspective, the focus is not directed toward assessing the veracity of the physics contained in the cloud forecast model, nor is it on defining the cloud product requirements. These tasks are for the model experts. Instead, focus herein is upon the processes involved in assessing cloud model forecast performance. This process is demonstrated for the CCf product and relies upon high quality, truth data to verify the accuracy of cloud forecast products. These truth data consist of manually-generated cloud truth products created from imagery collected by environmental satellites.

Standards are needed to quantify the accuracy of cloud model performance. Such standards have been established for "observed" cloud products by the World Meteorological Organization (WMO) for essential climate variables (ECVs) created from satellite measurements [7]. For example, the cloud amount ECV accuracy requirement for a 50 km resolution product is $\pm 1-5\%$ with a stability requirement of 0.3-3% across the full range of cloud cover fraction (CCf) values. (Accuracy is defined as the closeness of agreement between product values and truth values and stability is the extent

to which the error in a product remains constant over a long period.) The WMO also states observing accuracy requirements for other cloud ECVs, e.g. cloud top pressure of 15-50 hPa and cloud top temperature of 1-5 K. However, perhaps due to the challenges involved in verifying cloud model forecast model performance, no similar set of requirements has been stated for “forecast” cloud products. Thus, for simplicity, the WMO cloud observing ECV requirements are also assumed to be the cloud model forecast requirements.

Therefore, the engineering processes relevant to the verification of assumed WMO cloud cover forecast products are examined. These procedures are described in a verification test plan that defines the forecast model region of interest, addresses both the forecast model input and output cloud datasets as discussed in the Section 2. In this case, the accuracies of these cloud model products are established using high quality, truth data or CCf_{TRUTH} data created from temporally and spatially coincident, remotely sensed satellite observations as described in Section 3. Since the ability to construct CCf_{TRUTH} datasets is critical to this approach, the creation of these CCf_{TRUTH} data from satellite imagery is presented in Section 4. Examples of cloud observational product accuracies found in model input datasets using these CCf_{TRUTH} data are demonstrated in Section 5 along with the accuracies of cloud forecast products generated from a cloud forecast model. Implications and a summary of this research are contained in the Section 6.

2. Overview of the Verification Test Plan

A test plan is needed to quantitatively assess the accuracy of CCf in data used to initialize the forecast model as well as products generated with the model. For simplicity, this discussion focuses on data previously generated from the WRF (Weather Research and Forecast) forecast model which was initialized with datasets created with the North American Mesoscale (NAM) Forecast System [8, 9]. Both NAM and WRF data sets were analyzed on a 12 km grid but the procedures would be identical to those applicable to any grid size, e.g. 1 km or 50 km. A system-level approach relies upon truth cloud cover or CCf_{TRUTH} fields to first quantify the accuracy of cloud data found within the NAM analyses data fields (CCf_{NAM}) and then to evaluate the accuracy of cloud cover predictions based upon WRF model (CCf_{WRF}).

The quantitative evaluation of these CCf analysis and forecast products requires highly accurate CCf_{TRUTH} data which are derived from manually-generated cloud, no cloud (MG-CNC) analyses of satellite imagery, collocated temporally and spatially with the gridded CCf_{NAM} analyses fields and the CCf_{WRF} gridded cloud forecast fields. The MG-CNC data can be created from imagery collected by a variety of remote-sensing satellite systems. VIIRS (Visible Infrared Imager Radiometry Suite) data, collected by NASA’s NPP (National Aeronautics and Space Administration National Polar-orbiting Program) mission, are used in the analyses that follow. The theoretical basis for the procedures used to create the MG-CNC analyses from VIIRS imagery is discussed in detail below. Those processes have been extensively described in the literature [10-12]. The CCf_{TRUTH} are created by mapping of the MG-CNC analysis onto the different grids of the NAM and WRF cloud products. Comparisons between the WRF and NAM CCf cloud fields and the CCf_{TRUTH} fields result in a quantitative assessment of cloud model performance across the full range of CCf_{TRUTH} values.

3. Verification Test Data Sets

Identifying the region of interest (ROI) requires an assessment of the goals contained in the verification test plan, i.e. availability and quality of truth data, access to initialization test data, forecast model performance characteristics, and occurrences of weather patterns and cloud types of interest across the ROI. Ideally, the climate model ROI would include the entire Earth; however, a smaller ROI was selected for this discussion. The ROI is shown in Figure 1 and includes the Lesser Antilles, portions of Central America, and much of the southeastern USA. In the wintertime, this area is known to contain a variety of weather patterns and cloud types, including lower-level water clouds whose importance to climate modeling was noted above. Additionally, high quality satellite imagery is collected routinely over the ROI from a variety of geostationary and polar-orbiting environmental and meteorological satellites.

3.1 Model Initialization Data

The verification test plan must identify the sources of data to initialize the cloud forecast model. In this case, data contained in the 12 km NAM (ds609.0) dataset, are used to initialize the cloud prediction model. These data are created by NCAR (the National Center for Atmospheric Research) for use as inputs for the WRF Preprocessing System (WPS). NAM data include a total cloud cover fraction (CC_{NAM}) variable that can be compared directly to collocated CC_{TRUTH} data generated from VIIRS data. A portion of the NAM dataset coverage is shown as the lightest gray shade in Figure 1. The complete NAM grid extends from the Gulf of Alaska across Canada into the Nova Scotia area and southward toward the Caribbean and Mexico. The darkest gray shade in Figure 1 shows land-water boundaries over the ROI. The location of VIIRS satellite data, which are temporally and spatially coincident with the NAM data at 1800 UTC on 18 Nov 2014, is shown as the second lightest gray shade. The darkest gray shade within the VIIRS data swath shows individual VIIRS granules that map into the NAM and WRF data contained in the verification test area, which is centered over the southern Florida and Lesser Antilles regions.

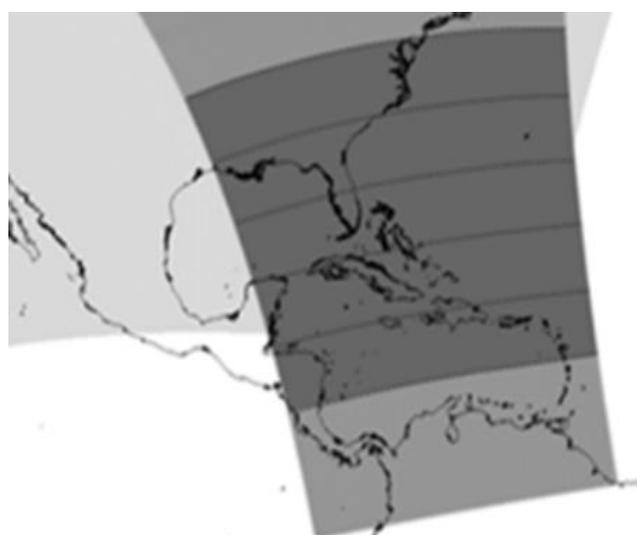


Figure 1 Test area ROI is the overlap of VIIRS (darker gray) and NAM (lighter gray) data on 18-19 Nov 2014.

3.2 Cloud Model Forecasts

The verification test plan must identify the model and settings used to generate the cloud forecasts for the ROI. In this case, forecasts of cloud cover fraction are generated from the NAM datasets using WRF version 3.7.1. Key parameters used to generate these WRF simulations are summarized in Table 1 of Hutchison et al., [8]. No nesting was used. WRF forecasts were examined for the existence of any non-zero cloud liquid water mixing ratios (LWMR) in the WRF variable QCLOUD, which is converted into cloud fractional coverage, at each WRF output level, as described in Eq. 1 of Xu and Randall [13]. Total cloud fractional coverage from these WRF layers can be derived using the vertical integration method described in Boer [14] and Trenberth [15].

3.3 Satellite Data for Creating CCf Truth Data

VIIRS data were collected by the NASA NPP spacecraft which was launched into a sun-synchronous, near-circular polar orbit at an altitude of 824 km at an inclination of 98.74° which creates a period of 101 minutes. This satellite ascends, heading north, across the equator at about 13:30. Orbital characteristics of the VIIRS sensor used to collect data used in this study are shown at: <http://database.eohandbook.com/database/missionsummary.aspx?missionID=413> The VIIRS sensor collects data in 22 spectral bands between about 0.4-12.0 μm. At nadir, the cross-track scanning sensor captures imagery (375 m high resolution) data and radiometric (750 m moderate resolution) data. The resolution of both types of data increases by a factor of 2 as the VIIRS sensor scans from nadir to the edge of its 3000 km data swath. Complete details about the VIIRS instrument design are described in Hutchison and Cracknell [11] and are summarized at: <http://database.eohandbook.com/database/instrumentsummary.aspx?instrumentID=412>.

VIIRS cloud data analyses are created automatically at the NASA NPP ground processing segment. These products rely upon the VIIRS cloud mask (VCM) product [16] along with other ancillary data and algorithms that have been described in the literature [17]. All cloud data products are available via the NOAA (National Oceanic and Atmospheric Administration) CLASS (Comprehensive Large Array-data Storage System) server. The VCM algorithms, which create cloud confidence and cloud phase analyses at a horizontal cell size (HCS) of the VIIRS moderate resolution data [11], form the basis for the downstream automated 3DClouds algorithms. Only the VCM cloud phase product might be used in this analysis. When used, it helps identify water clouds in the CCf_{TRUTH} data. If used, the VCM cloud phase analysis is quality controlled using color imagery composites of VIIRS data, as discussed below.

While not relevant to this study, the pre- and post-launch performance characteristics of the VCM products are well documented [12, 18, 19]. The 3DClouds products are initially created at the VIIRS moderate resolution, but are then aggregated to produce a final product that contains up to four cloud layers, at a 6 km ±1 km HCS across the entire VIIRS swath [20]. The verification of the VIIRS cloud base height product has been described in the literature by Fitch et al., [21] and Hutchison et al., [22]; however, the performance characteristics of other VIIRS 3DClouds products have not been documented so their performance characteristics remain uncertain.

4. Creating Manually-Generated Cloud, No Cloud Datasets

The verification test plan must also clearly define the data proposed to serve as truth and the accuracy of it. In this case, the process is included to convert manually-generated cloud, no cloud analyses into cloud cover fraction truth. Additionally, an understanding of cloud signatures in these multispectral imagery is essential to creating accurate MG-CNC analyses. Therefore, an overview is provided on the theoretical basis of image interpretation and the key parameters influencing signatures seen in satellite imagery. For a more complete discussion, see Hutchison and Lisager [23] and/or Hutchison and Cracknell [11].

4.1 Theoretical Basis for Satellite Data Interpretation

The ability to manually identify a cloud in any given spectral band of imagery is based upon the contrast between the cloud and the surrounding cloud-free background in that band. This contrast is affected by the sensitivity of the radiometer which depends on the ratio of the internally generated signal to that produced by the incoming radiation, e.g. the signal-to-noise ratio or SNR. The optimal sensor design maximizes SNR by increasing the size of the aperture, the field of view and/or the bandwidth or improving detector performance [24]. The SNR specifications for the VIIRS reflective bands are found in Table 4.13 of Hutchison and Cracknell [11]. For simplicity, the sensor's hardware characteristics are ignored and focus is upon the top-of-the-atmosphere (TOA) radiation arriving at the sensor. Thus, this contrast can be expressed by Eq 1 [11], assuming each pixel is either completely cloud-filled or completely cloud-free [25].

$$C = I_{\nu}(0)_{cloud} / I_{\nu}(0)_{background} \quad (1)$$

Depending upon the wavenumber (ν) of the radiation viewed in a given band, the TOA radiance at pressure equal zero, i.e., $I_{\nu}(0)$, may be composed of reflected solar radiation, emitted thermal radiation, or both solar and thermal radiation when observations are made in the 3-5 μm wavelength interval under daytime conditions. For simplicity, consider the case of thermal (infrared) radiation as a narrow (monochromatic) beam of energy emitted from a surface through a cloud-free atmosphere to space. The energy arriving at the sensor is accurately described by Hutchison and Lisager [23] but closely approximated by Eq 2 [25]. Eq 2 states that the vast majority of energy arriving at the satellite sensor during nighttime conditions, under cloud-free conditions, depends primarily upon only three components: the blackbody emission from the Earth's surface, the emissivity of that surface, and the atmospheric transmission from the surface to the TOA. The emissivity of a cloud may differ from its cloud-free background in some bi-spectral band combinations thus improving the cloud-ground contrast [26, 27]. However, in individual spectral bands, small temperature differences often occur between the cloud top and the background surface causing lower-contrast frequently between cloudy and cloud-free pixels, especially in nighttime imagery. A similar analysis would lead to another equation that represents the solar radiation arriving from the cloud top to the sensor.

$$I_{\nu}(0) \sim \epsilon_{\nu} B_{\nu}[T_s] \cdot T_{\nu}(P_s) \quad (2)$$

where the wavenumber dependent value of B_v is the Planck function at the surface temperature T_s , ϵ_v is the surface emissivity and T_v is the transmittance from the surface pressure (p_s) to the TOA.

Figure 2 is color coded, as seen by the scale across the bottom of the image, to depict the spectral signatures of cloud particles and various surface backgrounds in the 5.0-15.0 micron range [11]. (Emissivities of vegetated land = green, water = dark blue, snow = white, bare soil or sand = yellow. Atmospheric transmission = black.) VIIRS band (i.e., M14, M15, M16, I5) centers and widths are shown in medium blue lines with “M” labels at the top of the figure representing “moderate” resolution (750 m) bands as are the “I” imagery bands (i.e., I5) at 375 m resolution. Solid turquoise lines show the absorptive part of the index of refraction for water (K_{Water}) droplets while dashed turquoise lines show that for ice particles (K_{Ice}).

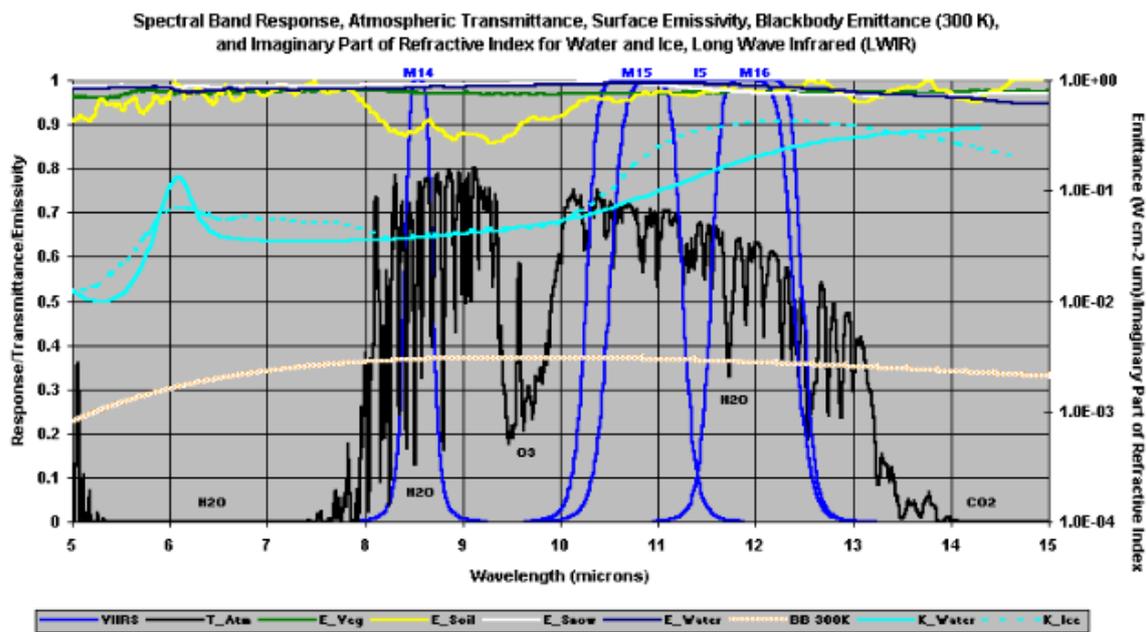


Figure 2 VIIRS bandpasses, atmospheric transmittances, surface emissivities, and the absorptive parts of the indices of refraction for ice/water across the 5000-15000 nm range [11].

A correct interpretation of scene content is crucial for creating an accurate truth analysis and this understanding is facilitated through the use of false color composite images [28]. These color images are created by placing up to three spectral bands into a single RGB display. Bands selected for display exploit differences in cloud and surface reflectance characteristics as well as atmospheric transmittance in the bands. False color composites provide a robust approach to accurately interpret all the features in most scenes and the use of these composites is a fundamental step in creating accurate MG-CNC analyses. Many examples of these composites with VIIRS-type imagery are shown by Hutchison and Cracknell [11].

Once the contents of a scene are understood, it is possible to create highly accurate MG-CNC analysis from multispectral satellite imagery [29]. These MG-CNC analyses can be created by making a binary cloud, no cloud (CNC) mask of a single spectral band, in simplistic scenes such as water clouds over ocean backgrounds, or may require multiple bands of imagery in complex cloud scenes over heterogeneous background conditions. The software used to make these manual analyses and

the preferred spectral bands needed to construct them over different cloud and background conditions is highlighted below in Section 4.3.

4.2 Phenomenology for Feature Extraction in Multispectral Imagery

Figure 2 provided an overview of the phenomenology exploited in the image segmentation process used to construct an MG-CNC analysis. It contains the surface emissivities of different surfaces along with the atmospheric transmittances and indices of refraction of cloud particles in the 5.0-15.0 micron range. Similar figures are available in Hutchison and Cracknell (Figures 4.8-4.11) [11] for all VIIRS imagery (I) and radiometric (M) bands that collect energy from the near-UV to the IR bands, i.e., 412 nm-12013 nm range. Examples of spectral bands used to maximize cloud contrast over different background surfaces has been documented [23]. The analyst uses the ground truth software to segment the scene into sub-regions where the cloud to background contrast is maximized. An example of that contrast maximization process is now shown in Figure 3. Once all the clouds have been identified in separate VIIRS channels, the MG-CNC is created with the truth software.

Figure 3a shows a false color composite of VIIRS imagery collected on 29 March 2013 at 1819 UTC over the southern part of Argentina and Chile. The color image was created with Adobe Photoshop by assigning the VIIRS M1 (centered at 412 nm) band to red, the M10 (1610 nm) band to green and the M16 (12013 nm) band to blue in a Red-Green-Blue (RGB) image. This particular RGB configuration was chosen to show snow/ice as red because the energy contribution from snow/ice is strong in the M1 band compared to the M10 and M16 bands. Densely vegetated surfaces appear dark green since the strongest energy contribution comes from the M10 band while significant energy also comes from the M1 and M16 bands. Water clouds are yellow since they are highly reflective in the M1 and M10 bands but warm in the M16 band while ice clouds have a purplish hue since the maximum energy contributions come from the M1 and M16 bands while the M10 band contributes much less energy. Highly reflective, colder clouds appear gray.

Figure 3b contains the M1 image of the scene while Figure 3c shows the scene in the M5 band (centered at 672 nm). These images show the contrast between cloud features and cloud-free land surfaces is stronger in the M1 band than in the M5 band making the former critically important to creating MG-CNC analyses, especially over bare soil and desert land surfaces [31].

Snow appears black in both the Figure 3d (centered at 1610 nm) and Figure 3e, which shows a brightness temperature difference image (BTD) of the M12 (3700 nm) minus M13 (4050 nm) or (M12-M13) BTD VIIRS image. Contrast between cirrus and water clouds versus land surfaces is more distinct in the M12-M13 BTD image than in the M10 image [32]. In addition, the contrast between water cloud features surrounding the snow/ice is strong in both M10 and M12-M13 BTD images. However, the contrast between snow/ice surfaces and ice clouds is weak in the M10 band but much stronger in the BTD M12-M13 image as seen through the inspection of the clouds over the land in the left-center of the scene.

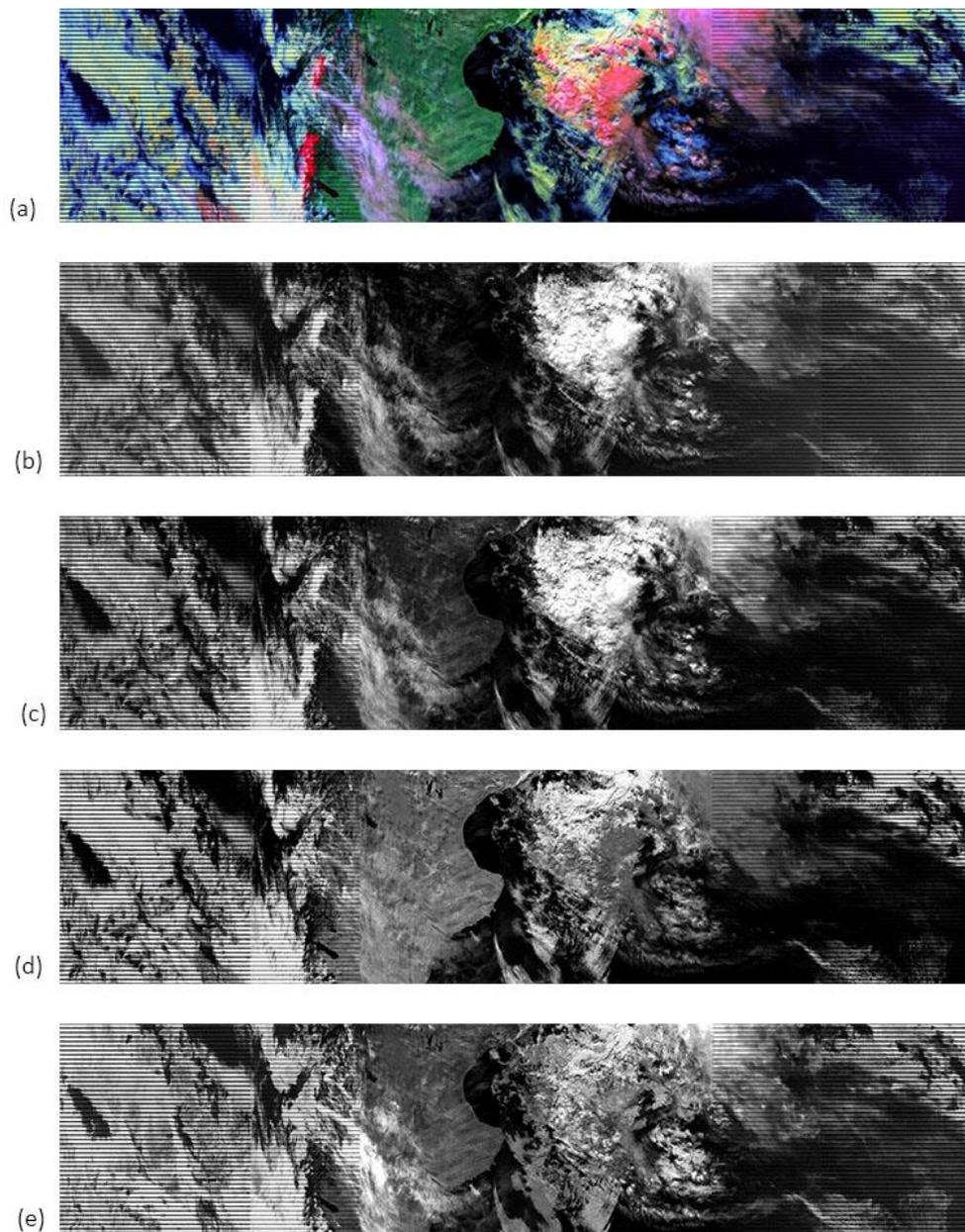


Figure 3 VIIRS imagery collected 29 March 2013 at 1819 UTC over southern Argentina. Figure 3a shows vegetated regions in green since these regions reflect poorly in the M1 (412 nm) band, seen in Figure 3b, and the M5 (672 nm) band seen in Figure 3c. Figure 3d shows cirrus clouds barely visible over land in the VIIRS M10 (1610 nm) imagery but their contrast is much stronger in the M12-M13 BTD image seen in Figure 3e.

4.3 Software to Construct Manually-Generated Cloud, No Cloud Analyses

Once the scene contents have been identified through the use of color composites, the binary MG-CNC analysis is created in each important spectral band with cloud truth software [30]. This software operates on 8-bit gray-scaled imagery to enable the analyst to produce a binary cloud, no cloud image. The analyst can segment the image and identify clouds in each sub-region by making all pixels cloudy that have values that exceed a user-defined threshold. The cloud truth software also allows the analyst to combine the MG-CNC analysis from each spectral band used in the analysis

into a composite MG-CNC analysis of the scene. The utility and accuracy of these MG-CNC analyses were documented during the NASA NPP VIIRS Cloud Mask Calibration/Validation Program [12, 19]. A complete description of the initial version of the software is available through the United States Patent Office [30]. This original version of the software was written in the C programming language for Unix and requires Motif. A more recently version of this software has been migrated to Linux and runs under older versions of the Fedora operating systems. It is currently being migrated to more current versions of Linux, e.g. CentOS.

5. Results

5.1. Creating Manually-Generated Cloud, No Cloud and NAM(WRF) Match-Up Data

The verification test plan must also address the process used in generating matchup datasets between CCf_{TRUTH} data and cloud analyses and cloud forecasts. For example, collocation times differences might be smaller if faster-moving (cirrus) clouds are studied as compared to slower-moving low-level water clouds. In addition, depending upon the grid characteristics of the cloud forecast and analysis datasets, the spatial matchup requirements might differ between CCf_{TRUTH} and these two types of datasets. Initial results from this research focused on comparisons between water clouds in CCf_{NAM} data and CCf_{TRUTH} since these water clouds are responsible for most of this spread in cloud feedback values between climate models [9]. Next, we examined the impact of these NAM data on WRF cloud forecasts [8]. So, the following procedures were applied first with the NAM profile datasets to quantify the accuracy of water clouds in these data before analyzing the accuracy of all cloud fields in the WRF cloud forecasts. Therefore temporal matchup time constraints were minimized.

VIIRS moderate resolution data (and thus MG-CNC analyses) were temporally and spatially collocated with the NAM (WRF) gridded data to identify regions in the MG-CNC analysis that contain either cloud-free or cloudy pixels [8, 9]. After an extended bow-tie trim was applied to reduce any overlapping pixels in the MG-CNC data, the NAM (WRF) data were mapped to a Lambert Conformal projection. An $M \times N$ grid was established with the NAM (WRF) data at the center of each grid. VIIRS data were then matched to the NAM (WRF) grid points by mapping the VIIRS data into the same projection and calculating the NAM (WRF) grid index into which it falls. VIIRS data outside the bounds of the NAM (WRF) grid and NAM (WRF) grid cells not containing VIIRS data were eliminated for further analyses. Next, a temporal restriction was applied to the VIIRS data, i.e. VIIRS scan times were checked to include only those that fell within ± 30 minutes of NAM analysis (or WRF forecast valid times). Distances from the grid center were calculated, through the Vincenty formula, to identify every VIIRS cloud product that was within 6.5 km of each 12 km NAM (WRF) grid center as a spatial restriction. Thus, cloudy and cloud-free pixels were mapped from the binary, MG-CNC mask product into the NAM (WRF) grid. The final 12 km resolution CCf_{TRUTH} for the NAM (WRF) data was then determined by calculating the mean of the MG-CNC data matched to each cell. The VCM cloud phase products were also used to identify occurrences where only cloud-free and/or single-layered water clouds existed in the NAM match-up data. Grid locations found to contain any ice clouds or mixed phase clouds were excluded from further analysis with only pixels containing clear pixels, low-level water clouds or their combination remaining. The CCf_{TRUTH} data are slightly different for CCf_{NAM} and CCf_{WRF} since the NAM and WRF grids are not precisely spatially coincident.

5.2. Generating Cloud Cover Fraction Truth Data

Figure 4 demonstrates the process of creating CCf_{TRUTH} data for the ROI, shown in Figure 1, at 1800 UTC on 19 Nov 2014. Color assignments for the scene shown in Figure 4a are similar to those described in Figure 3 with color codes corresponding identically to the different cloud types.

A MG-CNC analysis was created in an offline process, using the truthing software described in Section 4.3. Figure 4b contains the binary MG-CNC mask for the imagery shown in Figure 4a, where white is cloud and black is cloud-free. The MG-CNC mask is then converted into a mean cloud cover fraction truth (CCf_{TRUTH}) dataset by aggregating it into the 12 km NAM grid projected as discussed in Section 5.1 and shown in Figure 4c. The cloud-filled grids are white while cloud-free grids are black. Grids with CCf_{TRUTH} values in the 10-90% range are assigned the middle shade of gray to depict values containing fractional cloud cover. Truth values in the 10-90% range are seen to contain water clouds through comparisons with the clouds found in the false color image shown in Figure 4a. In order to limit the CCf_{TRUTH} data to only water clouds, the cloud phase restriction is enforced using results similar to those shown in Figure 4d, where water clouds are green and partly cloudy (light blue) while ice clouds are red, brown, and orange. Mixed phase clouds are yellow in Figure 4d and are also excluded from the CCf_{TRUTH} data. The specification of water clouds in the VCM cloud phase analyses, shown in Figure 4d, are seen to be in good agreement qualitatively with the water clouds shown in color composite shown in Figure 4a.

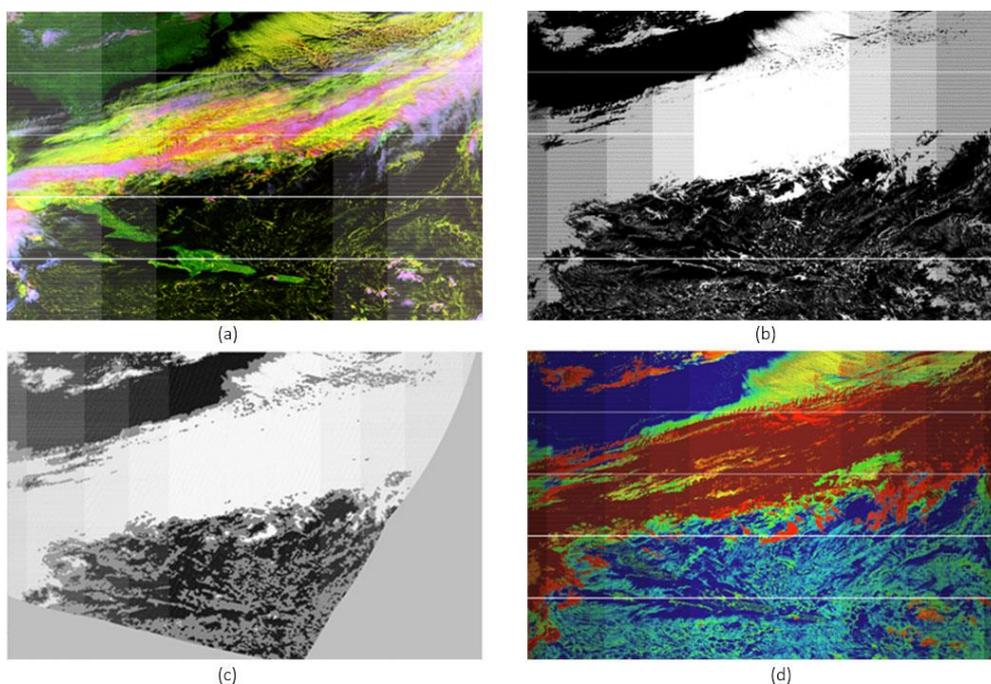


Figure 4 Figure 4a shows the color composite of VIIRS imagery at 1757-1803 UTC on 19 Nov 2014 with land as green, ocean black, water clouds yellow, and ice clouds pinkish-blue. Figure 4b shows the MG-CNC for the VIIRS imagery with clouds as white and no-clouds black. Figure 4c shows the MG-CNC mapped to the NAM grid, i.e. the mean CCf_{TRUTH} . The VCM cloud phase product is shown in Figure 4d, where mixed phase clouds are yellow, water clouds green, thin ice clouds red, multilayered ice over water clouds brown, opaque ice clouds orange, and water clouds green. Subpixel water clouds are light blue [16].

5.3. Evaluating Clouds in the Model Initialization Datasets

Figure 5 shows a comparison between all clouds contained in the NAM data valid on 19 Nov 2014 at 1800 UTC and the mean CCf_{TRUTH} data for all clouds created from VIIRS images collected at the same time. (The VIIRS data were collected between 1757-1803 UTC.) Coastlines are the darkest gray shade in these images followed by backgrounds and cloudless (0-10%), partly cloudy (10-90%) with cloudy grids (90-100%) white. Qualitative comparisons show NAM under-specifies lower-level water clouds (i.e. middle gray shade) especially in areas identified as A, B, and C. Table 1 verifies results quantitatively for three bins: completely cloudy, completely cloud-free and partly cloudy conditions. The performance metrics, shown in column 3, include the total number of VIIRS-NAM match-ups for each bin (counts) contained in the CCf_{TRUTH} data, along with the mean and standard deviation. Column 4 shows the performance statistics for the CCf_{NAM} data while column 5 shows similar results for the CCf_{TRUTH} data. For partly cloudy conditions, the CCf_{NAM} mean values are far less than the CCf_{TRUTH} mean values, i.e. by 70%. In addition, the large standard deviations for the CCf_{NAM} results suggest the NAM data are composed largely of cloudy and cloud-free CCf values. For example, when the CCf_{TRUTH} equals 100%, the CCf_{NAM} is 63.1% and the standard deviation is 43.5%. Thus, it is seen that NAM data under-specify the clouds found to be present in the VIIRS imagery, with water clouds in particular being under represented.

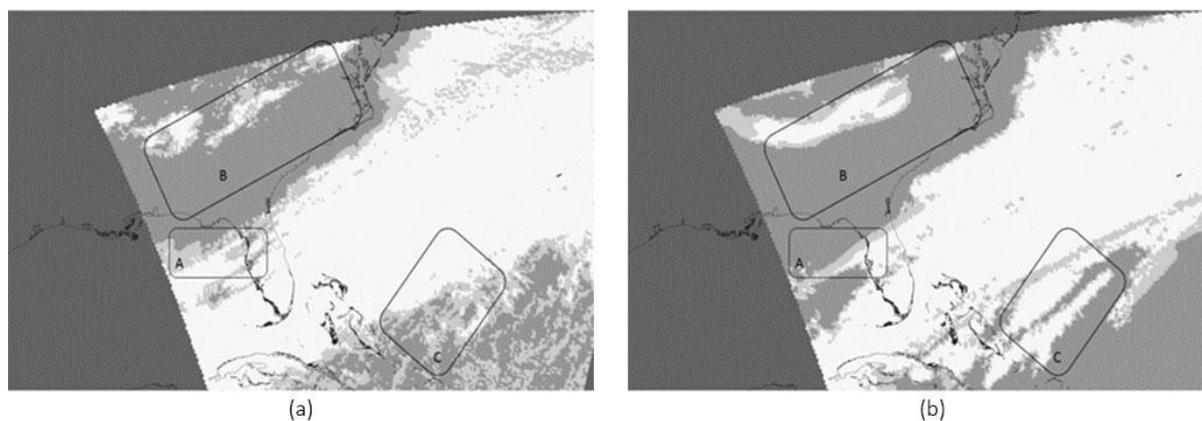


Figure 5 Figure 5a shows the mean CCf_{TRUTH} for all clouds in the VIIRS data at 1800 UTC on 19 Nov 2014. Figure 5b shows CCf_{NAM} for all clouds in the NAM dataset.

Table 1 CCf_{NAM} versus CCf_{TRUTH} for water clouds at 1800 UTC on 19 Nov 2014.

Bin Number	Truth CCf Range (%)	Performance Measure	NAM CCf	Truth CCf
1	$0 \leq CCf_{TRUTH} < 10$	Counts	12689	
		mean (%)	8.0	1.6
		standard deviation (%)	24.6	2.7
2-10	$10 \leq CCf_{TRUTH} < 100$	Counts	5453	
		mean (%)	11.8	17.1
		standard deviation (%)	27.9	2.8
11	$CCf_{TRUTH} = 100$	Counts	71	
		mean (%)	63.1	100.0
		standard deviation (%)	43.5	0.0

5.4 Evaluating Clouds in the Forecast Model Datasets

In the WPS, predicted values of QCLOUD (i.e. LWMR) are converted into cloud fractional coverage at each WRF eta level using Equation 4 of Xu and Randall [13]. These individual layers can then be converted into total cloud cover at each WRF grid [14, 15]. However, since cloud cover in the NAM datasets was found in Table 1 to be essentially binary, and for simplicity of processing, WRF total cloud cover was assumed to equal the single, maximum cloud cover fraction ($CLDFRA_{max}$) at all eta levels across each WRF grid, i.e. cloud layering was ignored.

Simulations of 24 hour cloud forecasts were generated from NAM data valid at 1800 UTC on 18 Nov 2014 using WRF settings described in Section 3.2. The CCf_{NAM} for that dataset is shown in Figure 6a and CCf_{TRUTH} for this scene is in Figure 6b. The CCf_{WRF} WRF 24 hr cloud forecast results valid at 1800 UTC on 19 Nov 2014 are in Figure 6c. The CCf_{WRF} results are summarized in Table 2.

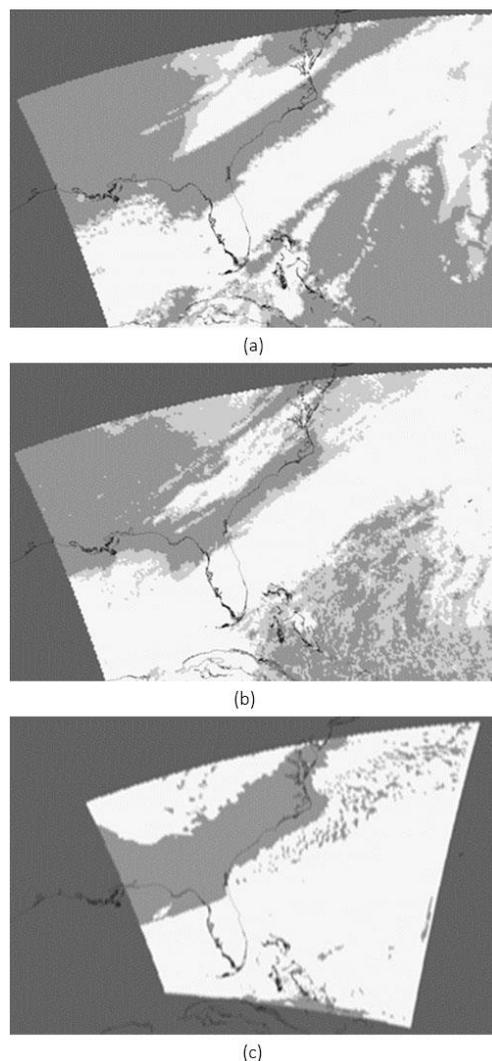


Figure 6 Figure 6a shows CCf_{NAM} data for 1800 UTC 18 Nov 2014 and Figure 6b has the mean truth CCf. Figure 6c shows CCf_{WRF} based upon the $CLDFRA_{max}$ results for a 24h forecast from these NAM data which is valid at the same time as CCf_{TRUTH} shown in Figure 5a.

Table 2 Cloud cover fraction (CCf) frequency from CLDFRA_{max} for all cloud in the WRF 24h forecast versus the occurrences of CCf_{TRUTH}.

CCf _{TRUTH} Interval (%)	CCf _{WRF} Counts	CCf _{TRUTH} Counts
$0 \leq \text{CCf}_{\text{TRUTH}} < 10$	9496	11011
$10 < \text{CCf}_{\text{TRUTH}} \leq 90$	159	7234
$90 < \text{CCf}_{\text{TRUTH}} \leq 100$	21672	13082

The images in Figure 6 show two key deficiencies in these cloud analyses and predictions. First, the CCf_{NAM} results show large areas where clouds are missing compared to the CCf_{TRUTH} image. This is especially evident in the water clouds present in the lower-right quadrant of the image. In addition, the CCf_{NAM} image shows cloud cover to be more binary than found in the CCf_{TRUTH} image. This is confirmed in quantitative comparisons shown between these two datasets in Table 1 of Hutchison et al., [9]. Thus, the WRF simulations are generated from NAM datasets that again under-specify clouds compared to the CCf_{TRUTH} data. However, Figure 6c shows qualitatively that, despite the fact that CCf_{NAM} data under-specifies clouds in the initialization dataset, the CCf_{WRF} cloud forecast results greatly over-predict cloud cover in the 24h forecast. This over-prediction of clouds in the WRF forecast is evident qualitatively by comparing Figure 5a (CCf_{TRUTH}) with Figure 6c. The over-clouding is seen in the areas highlighted in Figure 5a, including the stratocumulus field (Area A) over the Gulf of Mexico, the southeastern-eastern US landmass (Area B), and over the open Atlantic Ocean area northeast of Cuba (Area C). The binary nature of these CCf_{WRF} predictions is shown quantitatively Table 2. Binary cloud fields account for 99.5% of the forecasts in the CCf_{WRF} forecast product while they account for only 76.9% of the grids in the CCf_{TRUTH} data. The results show a lack of skill in the specification of fractional cloud cover amount in the 12 km NAM datasets and even less forecast skill in the prediction of fractional cloud cover amount by the WRF model.

These preliminary results suggest the lack of skill in the WPS modeling system. One cause could center on the conversion of meteorological parameters, e.g. temperature, pressure, moisture, etc., into cloud amount based upon the empirical relationship described by Eq 4 of Xu and Randall [13]. However, additional research would be needed to address the reason for this over-clouding. The specification of clouds in the NAM datasets represents another concern. If major cloud fields are not adequately described on a 12 km grid, it seems less likely that improvement could be expected in climate models which typically use a larger grid. Thus, fidelity in the forecast model initialization data suggests the need to make predictions at higher resolutions using cloud-resolving models. The integration of satellite imagery into the reanalysis system used to generate NAM data should be considered.

6. Conclusions

Clouds, especially lower-level water clouds, are critically important to mechanisms that impact climate studies. However, the apparent difficulty in developing truth measurements, for the verification of NWP and climate model forecasts, appears to result in an omission of forecast model accuracy requirements as well as model performance statistics. Therefore, the processes needed to

develop highly accurate cloud cover fraction truth data for verifying cloud forecast model performance have been addressed. The key process uses manually-generated cloud, no-cloud (MG-CNC) analyses of multispectral satellite imagery that are spatially and temporally collocated with the forecast products. These manually-generated cloud masks are created using special software that allows the analyst to exploit the phenomenological features in a variety of spectral bands to maximize the cloud-background contrast in each segment of the scene. The CNC analysis of each spectral band is then composited to produce a final MG-CNC analysis which becomes CC_{TRUTH} when mapped to the forecast model grid.

MG-CNC products, created from VIIRS data, were used in case studies to examine the cloud cover amount in NAM reanalysis fields as well as the accuracy of WRF cloud forecasts based upon these NAM data. The results showed clouds are under-specified in NAM reanalysis datasets but become over-specified in short-range (24h) WRF cloud forecasts based upon these NAM data. Additionally, both NAM and WRF cloud datasets were strongly binary in nature, i.e. the tendency for grids to be either completely cloudy or completely cloud-free was evident in both cloud products. Few NAM or WRF grids contained fractional cloud cover amounts in the 10-90% range compared to the truth data.

The deficiency of clouds in the NAM datasets might be mitigated by an increased reliance on satellite data products in the reanalysis system. Additionally, the over-clouding in the short range WRF predictions could involve the empirical conversion of meteorological parameters into cloud cover fraction as described by Xu and Randall [13]. Further investigations would be needed to confirm these hypotheses. At this time, it is certain that cloud cover analyses in NAM and cloud forecasts from WRF are binary in nature and show little skill in the accurate prediction of lower-level water clouds. It is concluded that this system level quantitative analysis demonstrates the critical need to improve both the specification of water clouds in NAM reanalysis fields and the accuracy of WRF cloud forecasts in order to reduce the uncertainty in climate model predictions.

Author Contributions

The authors share equally in the work which was based upon prior research published together.

Competing Interests

The authors have declared that no competing interests exist.

References

1. Liou KN, Freeman KP, Sasamori T. Cloud and aerosol effects on the solar heating rate of the atmosphere. *Tellus*. 1978; 30: 62-70.
2. Lohmann U, Feichter J. Global indirect aerosol effects: A review. *Atmos Chem Phys*. 2005; 5: 715-737.
3. Cess RD, Potter GL, Blanchet JP, Boer GJ, Del Genio AD, Déqué M, et al. Intercomparison and interpretation of cloud-climate feedback processes in nineteen atmospheric general circulation models. *J Geophys Res*. 1990; 95: 16601-16615.
4. Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden P, Dai X, et al. *Climate change 2001: The scientific basis*. Cambridge: Cambridge University Press; 2001.

5. Bader D, Covey C, Gutowski W, Held I, Kunkel K, Miller R, et al. Climate models: An assessment of strengths and limitations. Washington D. C: Climate Change Science Program (U.S.); 2008.
6. Bony S, Colman RA, Kattsov VM, Allan RP, Bretherton CS, Dufresne JL, et al. How well do we understand and evaluate climate change feedback processes? J Clim. 2006; 19: 3446-3482.
7. World Meteorological Organization. Systematic observation requirements for satellite-based products for climate. Geneva: WMO; 2001. pp. 128. Available from: https://library.wmo.int/doc_num.php?explnum_id=3710.
8. Hutchison KD, Isager BD, Sudhakar D, Jiang X, Quaas J, Markwardt R. Evaluating WRF cloud forecasts with VIIRS imagery and derived cloud products. Atmosphere. 2019; 10: 521.
9. Hutchison KD, Isager BD, Jiang X. Quantitatively assessing cloud cover fraction in numerical weather prediction and climate models. Remote Sens Lett. 2017; 8: 723-732.
10. Hutchison KD, Hardy KR, Gao BC. Improved detection of optically-thin cirrus clouds in nighttime multispectral meteorological satellite imagery using total integrated water vapor information. J Appl Meteorol. 1995; 34: 1161-1168.
11. Hutchison KD, Cracknell AP. VIIRS-a new operational cloud imager. London: CRC Press of Taylor and Francis Ltd; 2006.
12. Hutchison KD, Heidinger AK, Kopp TJ, Isager BD, Frey R. Comparisons between VIIRS cloud mask performance results from manually-generated cloud masks of VIIRS imagery and CALIOP-VIIRS match-ups. Int J Remote Sens. 2014; 35: 4905-4922.
13. Xu KN, Randall DA. A semiempirical cloudiness parameterization for use in climate models. J Atmos Sci. 1996; 53: 3084-3102.
14. Boer GJ. Diagnostic equations in isobaric coordinates. Mon Wea Rev. 1982; 110: 1801-1820.
15. Trenberth KE. Climate diagnostics from global analyses: Conservation of mass in ECMWF analyses. J Clim. 1991; 4: 707-722.
16. Hutchison KD, Roskovensky JK, Jackson JM, Heidinger AK, Kopp TJ, Pavolonis MJ, et al. Automated cloud detection and typing of data collected by the Visible Infrared Imager Radiometer Suite (VIIRS). Int J Remote Sens. 2005; 20: 4681-4706.
17. Wong E, Hutchison KD, Ou SC, Liou KN. Cloud top temperatures of cirrus clouds retrieved from radiances in the MODIS 8.55 μm and 12.0 μm bandpasses. Appl Opt. 2007; 46: 1316-1325.
18. Hutchison KD, Isager BD, Hauss BI. The use of global synthetic data for pre-launch tuning of the VIIRS cloud mask algorithm. Int J Remote Sens. 2012; 33: 1400-1423.
19. Kopp TJ, Thomas W, Heidinger AK, Botambekov D, Frey AR, Hutchison KD, et al. The VIIRS Cloud Mask: Progress in the first year of S-NPP towards a common cloud detection scheme. J Geophys Res Atmos. 2014; 119: 2441-2456.
20. Baker N. VIIRS Cloud Cover/Layers Algorithm Theoretical Basis Document. Data gov; 2011. pp. 92, Available from: <https://catalog.data.gov/dataset/noaa-jpss-visible-infrared-imaging-radiometer-suite-viirs-cloud-cover-layer-ccl-environmental-d>.
21. Fitch KE. Evaluation of the Visible Infrared Imaging Radiometer Suite (VIIRS) Cloud Base Height (CBH) pixel-level retrieval algorithm for single-layer water. Montgomery: Air University press; 2016.
22. Hutchison KD, Wong E, Ou SC. Cloud base heights retrieved during nighttime conditions with MODIS data. Int J Remote Sens. 2006; 27: 2847-2862.
23. Hutchison KD, Isager BD. Creating truth data to quantify the accuracy of cloud forecasts from numerical weather prediction and climate models. Atmosphere. 2019; 10: 177.

24. Stewart RH. Methods of satellite oceanography. Berkeley: University of California Press; 1985.
25. Liou, KN. An introduction to atmospheric radiation. International geophysics series. Cambridge: Academic Press; 2002.
26. Bell EJ, Wong MC. The near-infrared radiation received by satellites from clouds. Mon Wea Rev. 1981; 109: 2158-2163.
27. Inoue T. On the temperature and effective emissivity determination of semi-transparent cirrus clouds by bi-spectral measurements in the 10 μm window region. J. Meteorol. Soc. Japan. 1985; 63: 88-99.
28. d'Entremont RP, Thomason LW. Interpreting meteorological satellite images using a color-composite technique. Bull Am Meteorol Soc. 1987; 68: 762-768.
29. Hutchison KD, Hardy KR. Threshold functions for automated cloud analyses of global meteorological satellite imagery. Int J Remote Sens. 1995; 16: 3665-3680.
30. Hutchison KD, Topping PC, Wilheit TT. Cloud base height and weather characterization, visualization and prediction based on satellite meteorological observations. US Patent No. 6,035,710. 2000. Available from: <https://patents.google.com/patent/US6035710>.
31. Hutchison KD, Jackson JM. Cloud detection over desert regions using the 412 nanometer MODIS channel. Geophys Res Lett. 2003; 30: 2187-2191.
32. Hutchison KD, Mahoney RL, Lisager BD. Discriminating sea ice from low-level water clouds in split window, mid-wavelength IR imagery. Int J Remote Sens. 2013; 34: 7131-7144.



Enjoy *AEER* by:

1. [Submitting a manuscript](#)
2. [Joining in volunteer reviewer bank](#)
3. [Joining Editorial Board](#)
4. [Guest editing a special issue](#)

For more details, please visit:

<http://www.lidsen.com/journals/aeer>