Supplementary Materials

Adam M. Wilson, Walter Jetz

June 13, 2014

Contents

1 Background

Table SM1 describes the existing satellite-derived cloud climatologies along with their spatial and temporal grain and extent.

2 Methods

2.1 MOD09 Cloud Detection Algorithm

The MOD09 surface reflectance product includes an internal cloud mask in the PGE11 program which relies on two reflective and one thermal test (Petitcolin and Vermote, 2002; Roger and Vermote, 1998; Vermote et al., 2001). The reflective tests include the shortwave and middle infrared data combined in

Name	Description	Spatial Domain	Spatial Grain	Temporal Domain	Temporal Grain	Reference
GEWEX / ICCP	Compiled from 12 satellite products for comparison study	Global	1^o (≈110km)	1983–2009	Monthly	Stubenrauch et al. (2013)
HIRS	Cloud frequency from NOAA/HIRS/2	Global	≈20km	1979–2001	Daily	Wylie et al. (2005)
AVHRR PATMOS-x	Cloud product derived from NOAA's Advanced Very High Resolution Radiometer (AVHRR)	Global	0.1^o (≈11km)	1981–2010	Daily	Foster and Heidinger (2012)
GridSat	IR, water vapor and visible bands combined from multiple calibrated geostationary satellites. Not currently available.	Global, with missing data early in the record	0.07^o (≈8km)	1980–present	3-hour	Knapp et al. (2011)
Tropical MODIS Cloud Climatology	Optical and IR data from MODIS MOD35 algorithm	40^oS – 40^oN	1km	2000–2006	monthly, diurnal	Mulligan (2006)
MODIS Cloud Climatology	Derived from thresholded RGB images from MODIS data.	Scattered regions mostly in tropics	250m	2003–present	Monthly climatologies	Douglas (2013)

Table SM1: Existing satellite-derived cloud-related products with their spatial and temporal grain and extent.

the 'middle infrared anomaly' index (MIRA= $\rho_{20,21} - 0.82\rho_7 + 0.32\rho_6$, where ρ indicates MODIS band number). The second test uses reflectance at 1.38 microns (1.38mic=ρ_{26}). The MIRA and the 1.38mic reflectance are designed to be complementary, with MIRA efficiently detecting low or high reflective clouds (Petitcolin and Vermote 2002), while 1.38mic effectively detects high (and potentially not very reflective) clouds. Additionally, a thermal test is used to identify pixels with high infrared reflectance anomalies (e.g. fires, sun-glint, and high albedo surfaces) with respect to near surface (2m) air temperature computed by the NCEP reanalysis model (Kalnay et al. 1996). The daily cloud flags were extracted from bit 10 of the daily surface reflectance product "state_1km" Scientific Data Set (SDS) from both the Terra and Aqua satellites (MYD09GA and MOD09GA). Combining cloud observations from both products was necessary to minimize scan line-artifacts due to satellite orbits. Terra daytime imagery is collected at approximately 10:30am local time, while Aqua is from approximately 1:30pm, so the mean combined product represents mean mid-day cloud frequency. The daily 2000-2013 archive (approximately 260TB of data) were processed to calculate the mean and standard deviation of monthly cloud frequency using the Google Earth Engine API `http://earthengine.google.org/` and projected to geographic coordinates at 30-arc-second spatial resolution (≈1km). Due to the algorithms use of tests based on reflectance data, the flag is only available for daytime scenes and thus high latitudes have missing data during winter months. These data are referred to below as the MODIS cloud frequency (MODCF) dataset.

2.2 Removal of Orbital Artifacts

The MODIS orbit results in systematic gaps in the daily global coverage near the equator (Gregg and Casey, 2007) that results in nearly longitudinal artifacts (15^o for Terra and 345^o for Aqua) in the long-term cloud frequencies. To remove these features, we used the Variational Stationary Noise Remover (VSNR, Fehrenbach et al., 2012; Fehrenbach and Weiss, 2013, available at `http://www.math.univ-toulouse.fr/~weiss/Codes/VSNR/VNSR_VariationalStationaryNoiseRemover.html`), a Bayesian image restoration technique implemented in Matlab. The VSNR is well suited to remove these artifacts because it allows specification of the shape and scale of known artifacts. We explored various filter shapes and evaluated longitudinal profiles before and after correction and selected parameters that minimized the artifacts (see Figure SM1). We used a gabor filter with y=200, x=5, and θ=15 for Terra and θ=-15 for Aqua.

2.3 Calculation of Seasonal Metrics

2.3.1 Inter and Intra-annual Variability

Let m index months ($m \in 1 : 12$) and y index years ($y \in 2000 : 2014$). The timeseries of monthly cloud frequencies $CF_{m,y}$ (proportion of days with cloud flag equal to 1) was calculated separately from the daily MOD09GA and MYD09GA. These were then summarized to the 'climatological' cloud frequency mean and standard deviation: μ_m =mean($CF_{m,y}$) and σ_m =SD($CF_{m,y}$). The inter-annual variability was then calculated as mean(σ_m) and intra-annual variability (seasonality) as SD(μ_m).

2.3.2 Seasonal Concentration

We also quantified the seasonality of cloud frequencies following Markham Markham1970 and considered mean monthly cloud frequencies to represent vector quantities with both magnatude (cloud frequency) and direction (month). The sum of the twelve vectors then represents a vector incapsulating both the direction (month) and seasonal concentration (magnitude) of the cloud frequency for each pixel. Dividing the magnatude by the mean annual cloud frequency results in an index ranging from 0 (equal cloud cover throughout the year) to 100 (all observed clouds occurred in a single month).

3 Validation

3.1 Station Observations

The monthly CF were validated using a global observational dataset of synoptic weather reports collected at 5388 stations over 1971-2009 (Eastman and Warren, 2012). We extracted the mean "total cloud" amount for each month, which represents the mean proportion of the sky that was covered by all types of cloud during the observations in that month. Comparison of these observations to satellite data must take into account that the sampling radius of these observations (the visible sky) depends on cloud height, cloud thickness, the curvature of the earth, and other factors, but is typically much larger than a single 1km MODIS pixel. We followed Dybbroe, Karlsson, and Thoss (2005) and took the mean monthly MODCF for a circle with 16km radius around each station location. Additionally, this converts the temporal MODCF to mean cloud amount within the sample radius to make it comparable to the station observations.

3.1.1 Monthly Validation

The monthly MODCF (including data from 2000-2013) were compared to station observations using linear models over the full station record (1970-2009) and the MODIS era (2000-2009) to assess accuracy and relevance of the 14-year satellite-derived data for estimating long-term monthly climatologies. For the full record comparison, the station dataset was filtered to include only stations with at least 20 observations per month for at least 20 years, which retained 4679 stations. Several countries (notably the USA, Canada, and New Zealand) converted from human cloud observations to automated laser ceilometers over the past

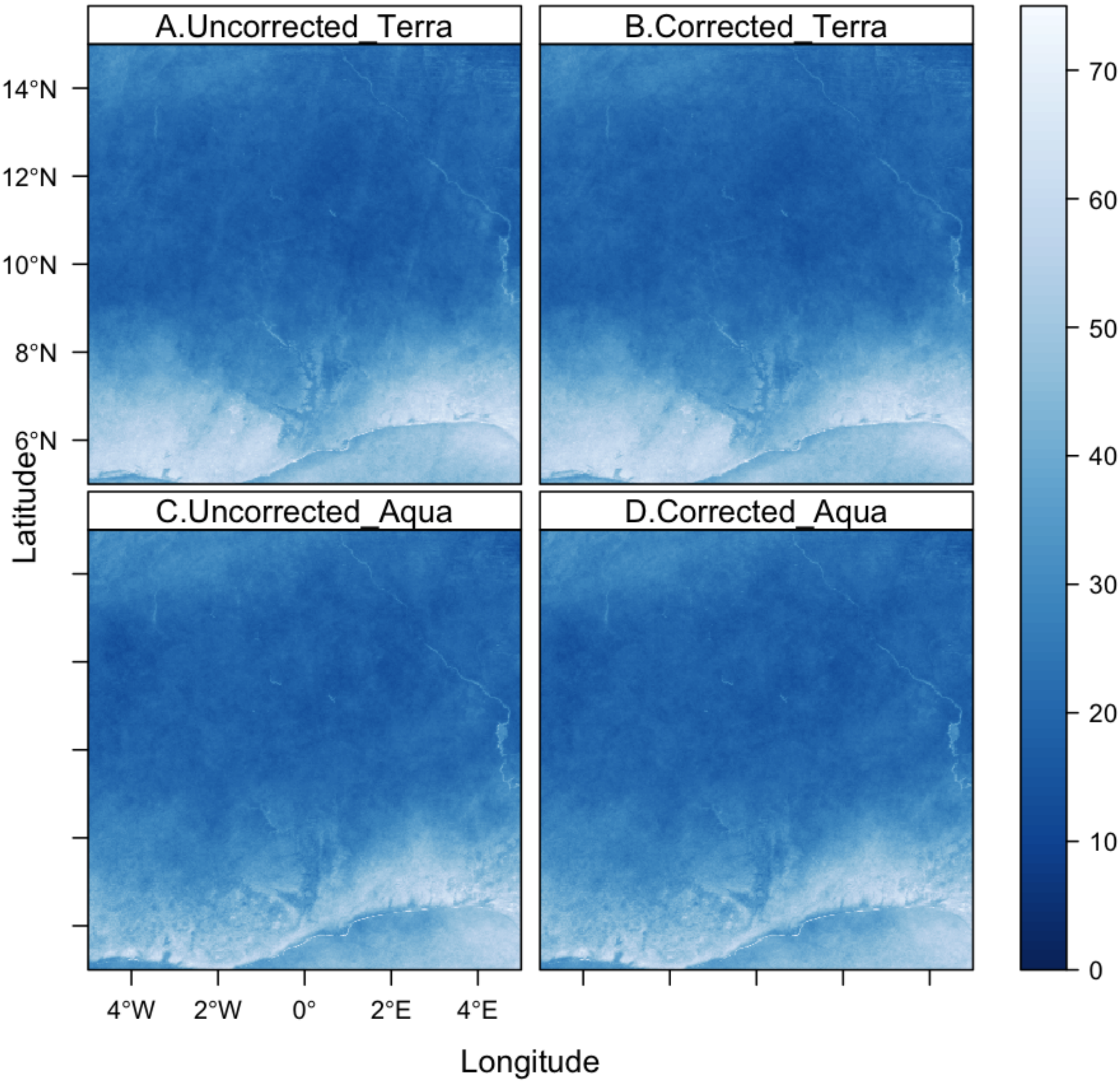


Figure SM1: Comparison of January cloud frequency over the Southwestern Sahara from A) uncorrected Terra and B) corrected Terra, C) Uncorrected Aqua, and D) Corrected Aqua. Note the banding in the uncorrected data resulting from variable observation frequency due to orbital artefacts of the MODIS Satellite.

decade leading to a decline in the number of observations over 1997-2009 (Eastman and Warren, 2012). For the MODIS era comparison, we included only stations with at least 20 observations per month for the full 10-year period (2000-2009), so the number of stations available was reduced to 1558.

Month/Season	Mean	n	R2	RMSE
DJF	58.62	4185	0.81	8.93
MAM	56.39	2879	0.72	8.28
JJA	54.89	5726	0.76	9.64
SON	57.14	4231	0.83	7.67
January	59.02	1403	0.80	9.38
February	57.27	1399	0.80	8.93
March	56.28	1438	0.76	8.22
April	56.51	1441	0.67	8.33
May	56.54	1450	0.66	8.89
June	55.06	1426	0.73	9.99
July	53.78	1428	0.79	10.26
August	54.17	1422	0.82	9.07
September	55.05	1421	0.81	8.07
October	57.03	1419	0.83	7.17
November	59.36	1391	0.84	7.60
December	59.59	1383	0.82	8.44

Table SM2: Summary of validation data by month and season

3.1.2 Seasonal Validation

In addition to monthly validation we also performed the same validation on the seasonal (DFJ,MAM,JJA,SON) mean values for MODCF and the station observations.

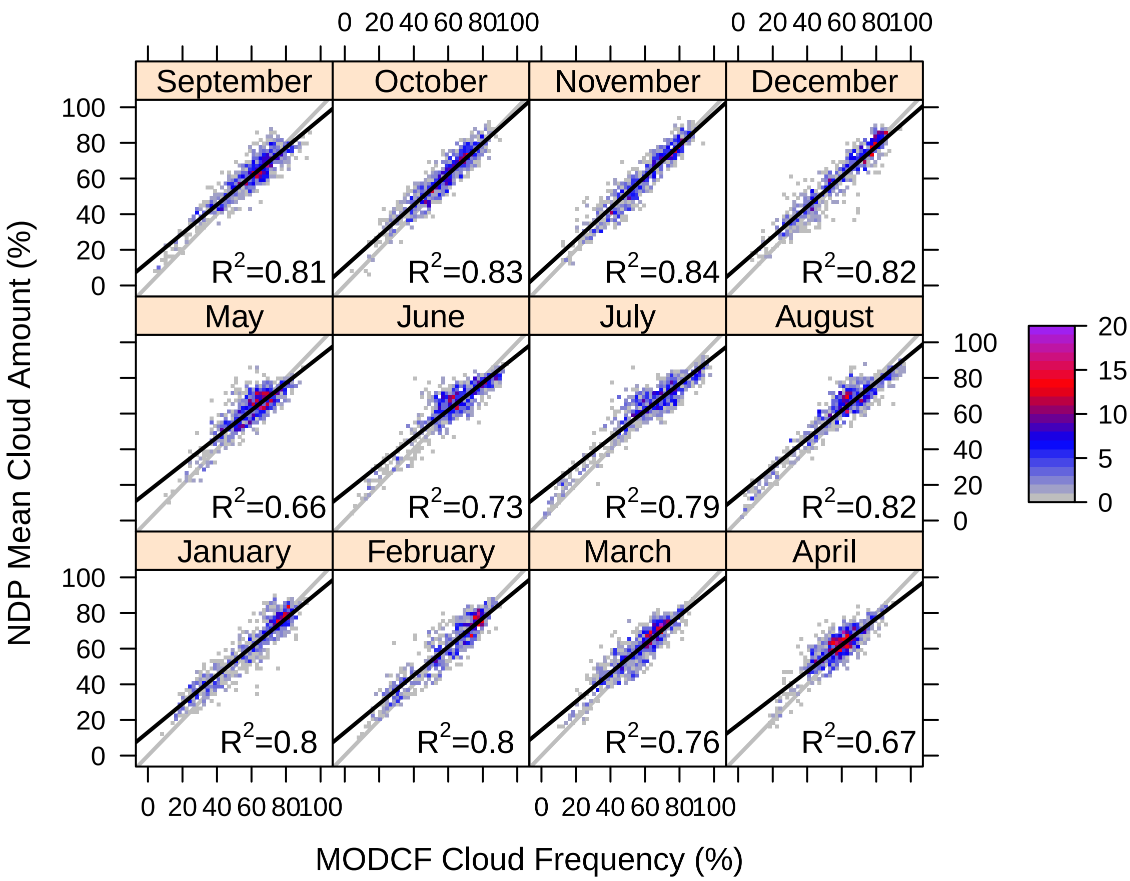


Figure SM2: Mean monthly cloud amount over 2000-2009 from 5388 global stations versus mean 2000-2014 MOD09 cloud frequency by month. Coefficient of determination is shown in each panel. Colors represent the number of station observations within each grid cell of the scatterplot.

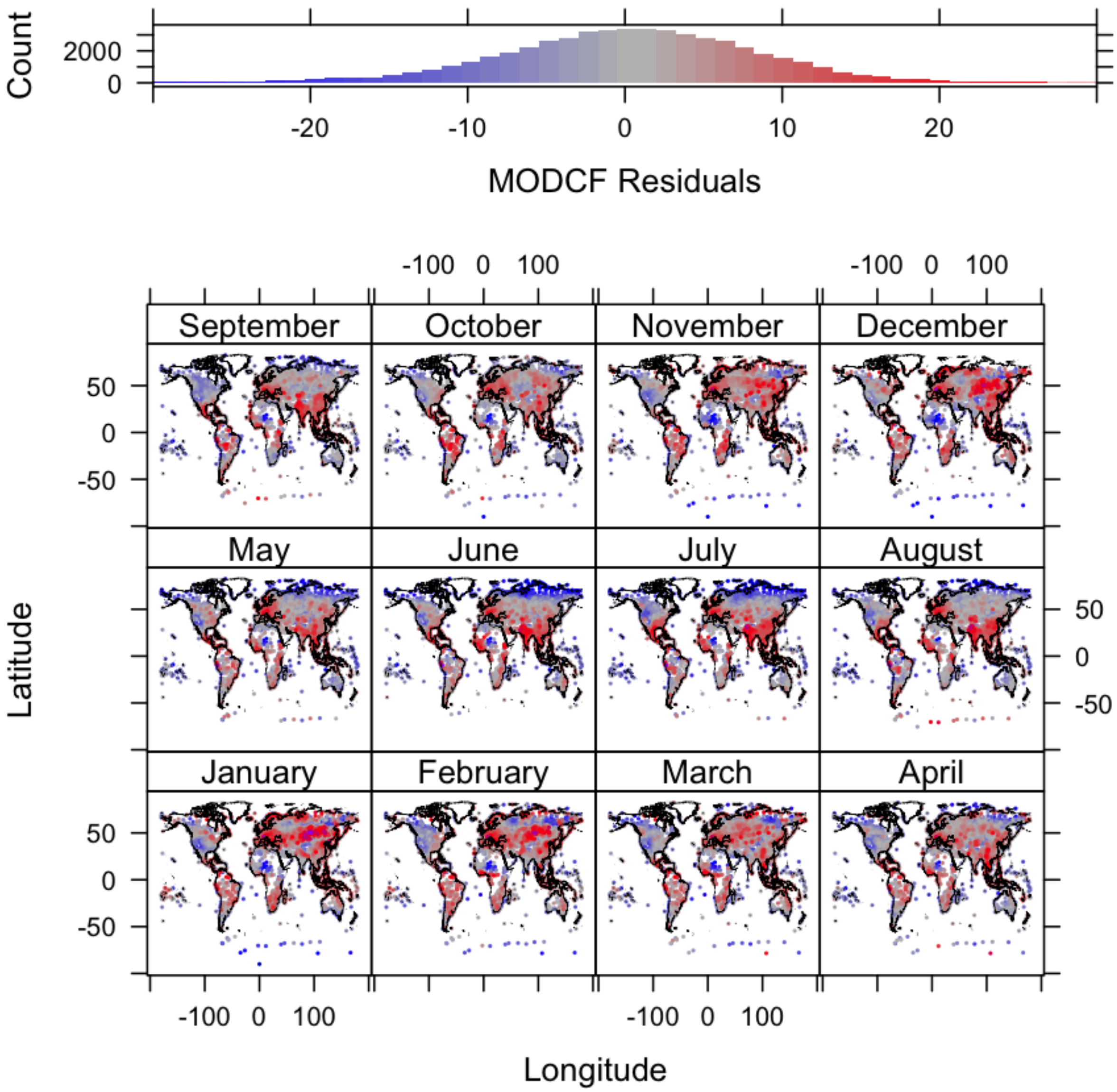


Figure SM3: Histogram and spatial distribution of residuals from linear model between station and satellite cloud amount at station locations. Negative (positive) values indicate locations where MODCF was less than (greater than) expected given the global relationship between MODCF and station observations.

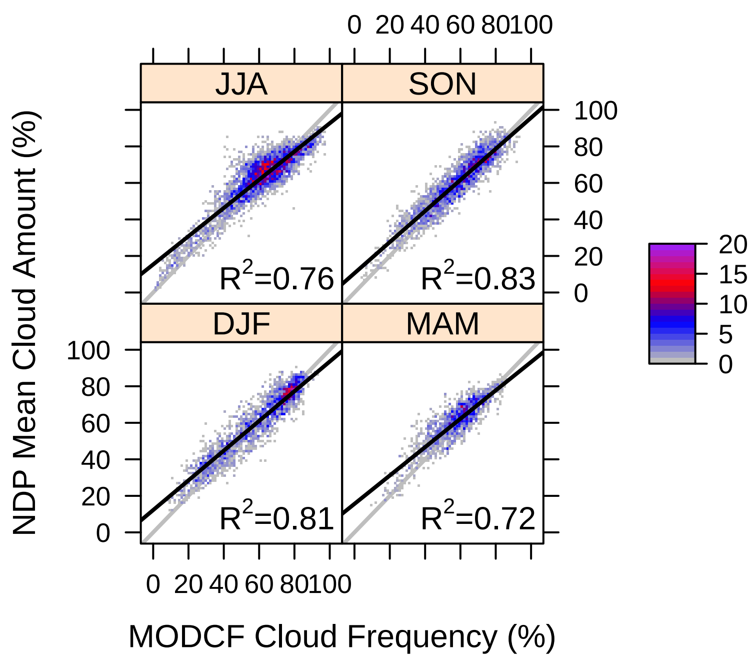
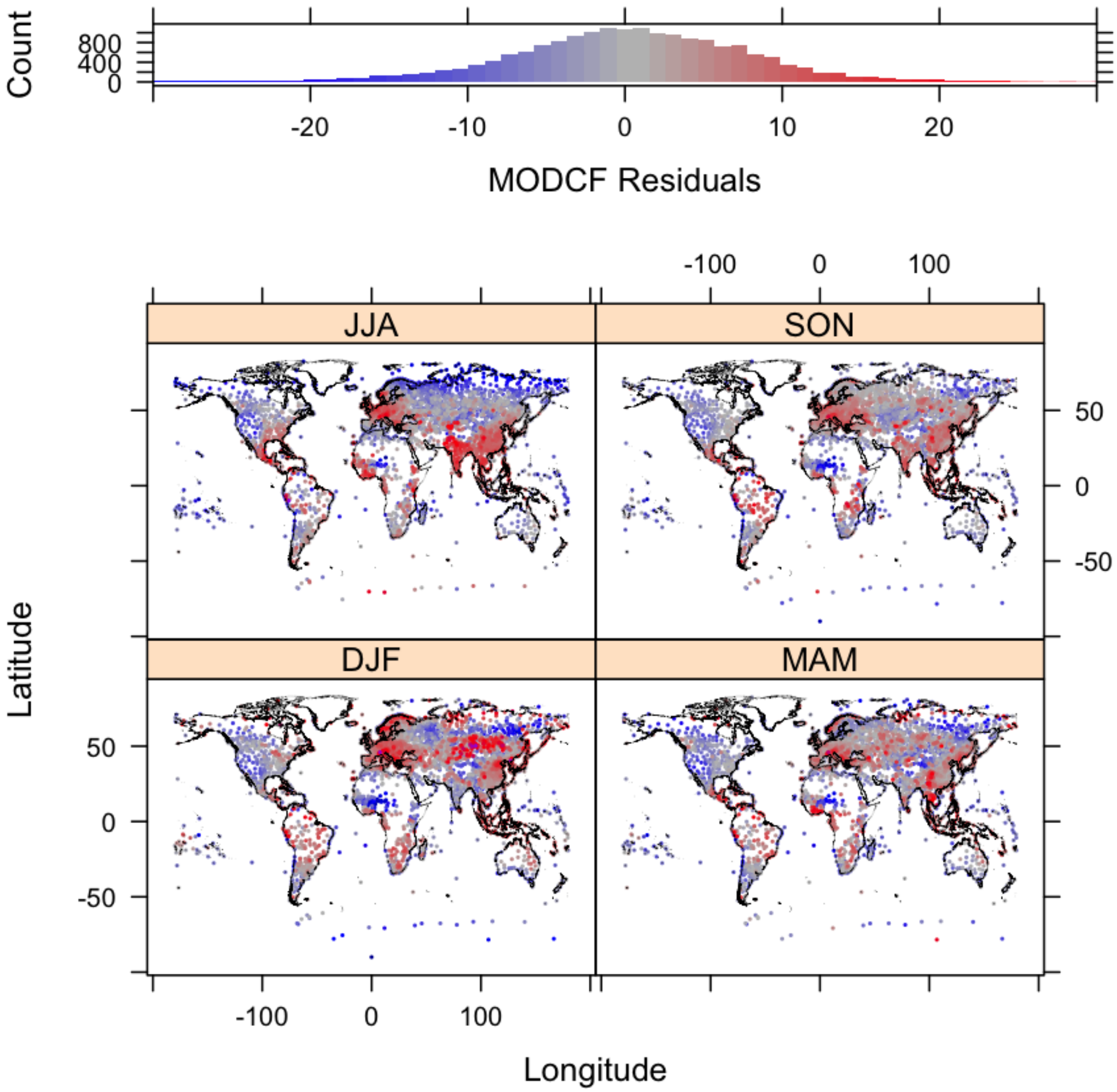


Figure SM4: Mean seasonal cloud amount over 2000-2009 from 5388 global stations versus mean 2000-2014 MOD09 cloud frequency by month. Coefficient of determination is shown in each panel. Colors represent the number of station observations within each grid cell of the scatterplot.



3.2 Temporal Stability

To assess the accuracy of the MODCF product in estimating multi-decadal cloud frequencies, we used linear models between the satellite climatologies (derived using data collected 2000-2014) and station observations divided into two periods: 1) the full station record (1970-2009) and 2) a subset including only the MODIS-era (2000-2009).

[1] "extract"

The MODCF is able to explain 0.78% of the variability in the observed station data across all months over 2000-2009, and 0.74% of the variability over the full record (1970-2009, **??**). The relationship is consistent when separated by month, with R^2 values ranging from 0.69 (May and June) to 0.82 (September and October, Figure 2). The station observations tend to record less cloud than MODCF below 20% (especially during

	1970-2009	2000-2009
Intercept	18.08 (0.11)***	13.41 (0.19)***
MODCF	0.76 (0.00)***	0.80 (0.00)***
R-Squared	0.74	0.78
RMSE	8.08	7.98
n	53678.00	17021.00

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table SM3: Comparison of validation models for full station record (1970-2009) and MODIS era (2000-2009). Stations were included if they had at least 20 years of data for full record or 10 years for MODIS-era record

the boreal summer, Figure 2). This feature is driven primarily by lower cloud frequency observed at high latitude stations (note band of negative values at high latitudes in Figure 3). MODIS CF tends to be higher than station observations in Central Asia and India and lower in the Sahel through much of the year.

4 Results

4.1 Seasonal variability

Caption for seasonal plot:
Near-global 1-km seasonal mean (DJF: December, January, February; MAM: March, April, May; JJA: June, July, August; SON: September, October, November) cloud frequency (proportion of days flagged as cloudy) derived from MODIS MOD09 internal cloud mask algorithm over 2000-2012.

4.2 Biome Summaries

To illustrate and contrast the spatial variability in cloud frequency within and between Earths ecoregions, we summarized MODCF within each of the up to 14 biomes in each geographic 'realm' delineated by the "Terrestrial Ecoregions of the World" dataset (Olson et al. 2001).

Figure 1: Cloud frequency seasonality and variability for each terrestrial biome separated by geographic realm. The grey lines represent the seasonality for 1000 randomly selected locations within each region and the blue line is a thin-plate spline representing the overall seasonality.

Table SM4: Biome and realm codes used in Table SM5.

code	realm	biome
AT_1	Afrotropics	Tropical & Subtropical Moist Broadleaf Forests
AT_2	Afrotropics	Tropical & Subtropical Dry Broadleaf Forests
AT_7	Afrotropics	Tropical & Subtropical Grasslands, Savannas & Shrublands
AT_8	Afrotropics	Temperate Grasslands, Savannas & Shrublands
AT_9	Afrotropics	Flooded Grasslands & Savannas
AT_10	Afrotropics	Montane Grasslands & Shrublands
AT_12	Afrotropics	Mediterranean Forests, Woodlands & Scrub
AT_13	Afrotropics	Deserts & Xeric Shrublands
AT_14	Afrotropics	Mangroves
AT_98	Afrotropics	Lake
AN_11	Antarctic	Tundra
AA_1	Australasia	Tropical & Subtropical Moist Broadleaf Forests
AA_2	Australasia	Tropical & Subtropical Dry Broadleaf Forests
AA_4	Australasia	Temperate Broadleaf & Mixed Forests

AA_7	Australasia	Tropical & Subtropical Grasslands, Savannas & Shrublands
AA_8	Australasia	Temperate Grasslands, Savannas & Shrublands
AA_10	Australasia	Montane Grasslands & Shrublands
AA_11	Australasia	Tundra
AA_12	Australasia	Mediterranean Forests, Woodlands & Scrub
AA_13	Australasia	Deserts & Xeric Shrublands
AA_14	Australasia	Mangroves
IM_1	IndoMalay	Tropical & Subtropical Moist Broadleaf Forests
IM_2	IndoMalay	Tropical & Subtropical Dry Broadleaf Forests
IM_3	IndoMalay	Tropical & Subtropical Coniferous Forests
IM_4	IndoMalay	Temperate Broadleaf & Mixed Forests
IM_5	IndoMalay	Temperate Conifer Forests
IM_7	IndoMalay	Tropical & Subtropical Grasslands, Savannas & Shrublands
IM_9	IndoMalay	Flooded Grasslands & Savannas
IM_10	IndoMalay	Montane Grasslands & Shrublands
IM_13	IndoMalay	Deserts & Xeric Shrublands
IM_14	IndoMalay	Mangroves
NA_2	Nearctic	Tropical & Subtropical Dry Broadleaf Forests
NA_3	Nearctic	Tropical & Subtropical Coniferous Forests
NA_4	Nearctic	Temperate Broadleaf & Mixed Forests
NA_5	Nearctic	Temperate Conifer Forests
NA_6	Nearctic	Boreal Forests/Taiga
NA_7	Nearctic	Tropical & Subtropical Grasslands, Savannas & Shrublands
NA_8	Nearctic	Temperate Grasslands, Savannas & Shrublands
NA_11	Nearctic	Tundra
NA_12	Nearctic	Mediterranean Forests, Woodlands & Scrub
NA_13	Nearctic	Deserts & Xeric Shrublands
NA_98	Nearctic	Lake
NA_99	Nearctic	Rock & Ice
NT_1	Neotropics	Tropical & Subtropical Moist Broadleaf Forests
NT_2	Neotropics	Tropical & Subtropical Dry Broadleaf Forests
NT_3	Neotropics	Tropical & Subtropical Coniferous Forests
NT_4	Neotropics	Temperate Broadleaf & Mixed Forests
NT_7	Neotropics	Tropical & Subtropical Grasslands, Savannas & Shrublands
NT_8	Neotropics	Temperate Grasslands, Savannas & Shrublands
NT_9	Neotropics	Flooded Grasslands & Savannas
NT_10	Neotropics	Montane Grasslands & Shrublands
NT_12	Neotropics	Mediterranean Forests, Woodlands & Scrub
NT_13	Neotropics	Deserts & Xeric Shrublands
NT_14	Neotropics	Mangroves
NT_98	Neotropics	Lake
NT_99	Neotropics	Rock & Ice
OC_1	Oceania	Tropical & Subtropical Moist Broadleaf Forests
OC_2	Oceania	Tropical & Subtropical Dry Broadleaf Forests
OC_7	Oceania	Tropical & Subtropical Grasslands, Savannas & Shrublands
PA_1	Palearctic	Tropical & Subtropical Moist Broadleaf Forests
PA_4	Palearctic	Temperate Broadleaf & Mixed Forests
PA_5	Palearctic	Temperate Conifer Forests
PA_6	Palearctic	Boreal Forests/Taiga
PA_8	Palearctic	Temperate Grasslands, Savannas & Shrublands
PA_9	Palearctic	Flooded Grasslands & Savannas

PA_10	Palearctic	Montane Grasslands & Shrublands
PA_11	Palearctic	Tundra
PA_12	Palearctic	Mediterranean Forests, Woodlands & Scrub
PA_13	Palearctic	Deserts & Xeric Shrublands

Table SM5: Mean (SD) monthly cloud frequency summarized by biome and geographic realm. See Table SM4 for Code descriptions.

Code	January	February	March	April	May	June	July	August	September	October	November	December
AA_1	86.1 (7.3)	84.7 (7.6)	83.7 (8)	79.9 (9.8)	78 (10.4)	77.4 (11.3)	80.2 (11.5)	77.8 (13.5)	77.1 (14.6)	74.6 (13.3)	79.7 (10.6)	82.5 (9.3)
AA_10	66.3 (17.9)	66.3 (17.1)	65.4 (17.2)	65.3 (16.4)	69.9 (13.2)	71.8 (14.2)	73.4 (14.7)	73.5 (13.2)	74.6 (13.3)	72.9 (13.5)	71.5 (14.6)	72.8 (13.9)
AA_11	83.1 (6.2)	83.4 (5.9)	84.4 (5.8)	84.1 (6.4)	81.5 (5.3)	82.5 (6)	81.8 (5.8)	82.4 (4.9)	84.6 (6.3)	81.9 (6.5)	85.3 (7.1)	83.4 (6.5)
AA_12	27.2 (5.7)	35.4 (6.2)	34.3 (7.6)	41.9 (9)	49.1 (10.4)	53.2 (7.5)	53.8 (9)	49.7 (11.3)	45.4 (12.3)	39.1 (11.3)	39 (7.8)	33.2 (7)
AA_13	35.2 (10.4)	39.2 (7.3)	34.9 (7.8)	28.8 (7.4)	27.7 (7)	27.8 (9.8)	22.1 (10.9)	15.2 (9.4)	16.7 (6.3)	23.5 (4.8)	32.9 (6.1)	37.4 (8.6)
AA_14	80 (6.2)	79.5 (5.4)	78.3 (5.8)	73.6 (5.9)	73.9 (7.7)	76.8 (7.8)	79.6 (8.8)	77.5 (10.8)	75.6 (9.7)	69 (8.8)	73 (7)	76.9 (6.7)
AA_2	87.7 (7.4)	84.4 (8.5)	78.8 (8.8)	66.3 (13.2)	63.3 (14)	56.9 (15.9)	54.5 (17.4)	45.7 (18.9)	43.8 (18.3)	53 (16.8)	67.3 (14.2)	86 (8.9)
AA_4	51.5 (12.2)	57.5 (10.2)	55.8 (11)	56.3 (11.2)	55.4 (12.7)	60.1 (8.2)	58.2 (11.8)	56.2 (14.3)	54.7 (15.6)	56.6 (12.7)	59.4 (9.4)	58.7 (11.9)
AA_7	69.8 (11.3)	65.4 (10.4)	58.6 (11.6)	40.6 (12.1)	34.3 (12.5)	25.7 (15.1)	19.5 (15.4)	16.6 (13.4)	21.6 (10.7)	32.2 (10.4)	48.8 (11.4)	61.1 (11.2)
AA_8	40.2 (10.3)	46 (7.4)	38.4 (8.6)	33.9 (9)	36.3 (10.2)	45.1 (9.2)	38.7 (13.8)	33.6 (14.8)	31.3 (13.7)	34.6 (11.2)	45.9 (6.8)	44.2 (9.4)
AN_11	34.4 (17.4)	41.6 (19.4)	53.1 (18.7)	70.4 (20.2)	72.5 (17.5)	89 (9.4)	77 (11.4)	75.3 (13.5)	68.2 (16.4)	55.2 (18.3)	41.8 (17.9)	32.8 (17.8)
AT_1	60.9 (18.7)	68.2 (16.9)	71.9 (15.1)	73.8 (14.5)	69.9 (15.1)	70.1 (16.9)	71.9 (17.7)	75.1 (18.7)	72.6 (18.5)	70.6 (16.8)	66.7 (16.3)	60.6 (18.6)
AT_10	53.7 (20.6)	51.6 (18.6)	53.8 (14.7)	53.1 (14.4)	43.3 (19)	40.8 (25.3)	41 (29.5)	43.7 (27.8)	44.8 (23.4)	53 (15.8)	52.6 (18)	52.1 (20.9)
AT_12	28.3 (13.4)	28.4 (13)	30.3 (10.1)	38.9 (7.6)	44.3 (5.5)	42.6 (5.3)	39 (6)	42.8 (6.9)	40.3 (8.3)	41.3 (10.3)	34.7 (10.8)	33.9 (12.9)
AT_13	35.1 (17.8)	34.3 (17.8)	31.8 (15.4)	29.7 (12.5)	22.1 (13.4)	20 (14.9)	20.6 (17.9)	21.4 (17.6)	18.9 (12.4)	23.6 (12.4)	27.4 (15)	29.1 (16.1)
AT_14	52.4 (17.9)	55.7 (21.6)	59.1 (22.5)	60 (23.8)	61.8 (23.6)	68.7 (25.2)	71.3 (26.5)	71.5 (29)	67.6 (28.6)	63.1 (24.5)	56.5 (20.1)	51.6 (15.5)
AT_2	76.7 (8.7)	69.9 (10.5)	56.5 (12.3)	36.8 (10.6)	23.8 (8.7)	17.5 (10.6)	18.2 (11.7)	20.9 (13.8)	26.1 (11.7)	40.9 (14.6)	58.1 (13.9)	70.9 (11.2)
AT_7	44.2 (27.7)	45.7 (25.7)	50.7 (22.3)	52.5 (20)	46.4 (21.2)	44.4 (24.7)	47.5 (27.2)	51.5 (27.9)	48.4 (23)	49.5 (20.2)	45.8 (25.4)	43.4 (28.8)
AT_8	20.1 (9.1)	13.9 (8.5)	16.7 (8.6)	24.9 (9)	15.3 (10.8)	19.6 (10.1)	31.1 (9)	29.7 (10.1)	17.5 (11.8)	10.2 (10.9)	15.6 (11.2)	18.7 (11.3)
AT_9	48.2 (23.8)	48.8 (18.2)	50.7 (13.2)	50.9 (16.4)	46.3 (23.3)	44.4 (25.4)	44.5 (27.6)	44.9 (29.1)	40.8 (21.9)	46.5 (13.5)	48.6 (19.7)	47.1 (26.4)
AT_98	53.5 (19.1)	56.7 (14.5)	57.5 (10.5)	52.4 (13.7)	42.9 (15.2)	32.5 (13.6)	29 (15.4)	35.9 (18.6)	43 (17.3)	52 (17.4)	55.5 (18.4)	53.7 (18.3)
IM_1	56.1 (28.2)	54.2 (29.5)	56.8 (26.4)	62.3 (21.7)	70.2 (16.8)	80.8 (10.3)	84 (10.5)	82 (11)	77.6 (10.6)	65.3 (19.2)	56.9 (25.2)	55.1 (27.9)
IM_10	92.1 (3.9)	88.8 (5.7)	86.2 (7.6)	84.3 (9.5)	86.3 (7.7)	83.4 (7.4)	86.8 (5.7)	85.4 (6.5)	88.8 (5.8)	90.6 (4.9)	91.1 (5.4)	92.5 (4.3)
IM_13	23.4 (12.1)	22.6 (12.4)	21 (9.4)	25.5 (10.9)	26.6 (14.7)	55.5 (21.1)	78.2 (15.7)	77.7 (15.2)	52.8 (22.5)	25 (25.4)	23 (17.8)	20.2 (13.4)
IM_14	48.9 (30.4)	43.4 (29.3)	47.2 (25.2)	55.3 (20.1)	70.8 (16)	79.4 (11.6)	83.7 (10.4)	82.2 (12.5)	80.1 (11.7)	67.8 (17.9)	58.2 (26.4)	55.6 (30)
IM_2	26.4 (16.2)	24.8 (16.3)	30.8 (20.2)	40.3 (23)	50.9 (23.5)	78.1 (9.8)	90.3 (5.8)	90.2 (5.1)	76.6 (9.8)	49.8 (22.6)	37.7 (21.9)	30 (22)
IM_3	38.9 (16.4)	45.8 (16.5)	43.1 (18.2)	45.1 (20.6)	46.6 (24.6)	60.9 (22.6)	78.4 (18.5)	79.9 (15.3)	61.9 (22.9)	35.3 (27)	29 (22.5)	33.1 (19.9)
IM_4	46.3 (14.6)	57.4 (15.9)	61.2 (19)	67.8 (20.9)	68.2 (22.5)	76.6 (21.8)	83.2 (17.5)	81.3 (15.4)	69 (20.6)	49.3 (23.6)	37.7 (17.5)	39.8 (15.2)
IM_5	45.2 (15.2)	56 (16.6)	58.8 (21.3)	62.6 (23.2)	61.9 (28.4)	68.9 (28.4)	77.7 (24.3)	78.8 (21.6)	66.5 (27.6)	46.7 (28.5)	36.7 (20.3)	38 (16.9)
IM_7	36.2 (9.6)	24.6 (4)	20.8 (5.4)	24.6 (10.1)	35.3 (12.9)	65.6 (11.3)	84 (4.6)	77.3 (5.4)	60.6 (7.8)	28.6 (10.8)	13.1 (4.3)	19.1 (5.1)
IM_9	18.2 (9)	12.2 (6.1)	10.7 (4.7)	15.3 (6.3)	19.2 (11.9)	58.1 (7.2)	86.6 (4.7)	84.4 (6)	50.5 (9.1)	10.2 (5.6)	17.3 (8.9)	17.4 (10.8)
NA_11	65.2 (19.6)	59.9 (19.1)	62.2 (15.7)	56.7 (14.1)	48.7 (11.4)	42.9 (13)	44 (11.9)	59.3 (10.1)	68 (6.4)	70.9 (10.9)	70.8 (15.7)	66.4 (20.2)
NA_12	54.6 (13.6)	55.1 (12.9)	52.1 (11)	51.1 (11.9)	43.6 (12)	28.8 (12.7)	18.2 (11.3)	21.3 (11.4)	34.8 (12.4)	45.7 (10.7)	51.1 (11.7)	54.1 (12.4)
NA_13	33 (18.1)	31.8 (16.2)	31.6 (14.9)	32.8 (12.8)	26.9 (13.2)	21.4 (15.7)	20.1 (17)	18.1 (14.2)	16 (12.2)	18.8 (11.7)	25.9 (15.7)	31.2 (17.7)
NA_2	63.8 (21.6)	61 (23.7)	65.2 (15.9)	71.1 (10.5)	76.2 (6.8)	84.2 (4.9)	84 (6)	78.5 (7.1)	76.7 (6.1)	76.7 (8.2)	63.5 (14.4)	64.6 (18.7)
NA_3	65 (17.7)	64.8 (15.2)	62.6 (10.9)	62.7 (7.9)	62.2 (9.9)	60.7 (14)	59.8 (16.6)	58.3 (16)	59.2 (14.8)	62.6 (15)	67.6 (16.7)	66.2 (17)
NA_4	55.4 (20.9)	57.1 (18.6)	59.3 (16.3)	66.6 (12.3)	68.1 (12.2)	64.9 (15.9)	64.2 (17.7)	60.9 (17.7)	58.7 (16.6)	59.8 (14.5)	59.2 (16.8)	56.9 (19.6)
NA_5	57.5 (18.1)	48.9 (17)	53 (15.2)	55.8 (12.8)	58.2 (10.2)	53.5 (9.7)	52.2 (8.7)	62.9 (6.1)	68.9 (6)	71.7 (10.8)	65.5 (15.1)	63.6 (17.9)
NA_6	56.4 (19)	52 (16.2)	52.4 (12.5)	54.6 (8.2)	50.8 (8.5)	49.4 (12.3)	48.8 (14.5)	43.9 (14.3)	43.3 (14.2)	50.2 (14)	59.9 (16.5)	59.2 (16.6)
NA_7	30.9 (9)	32.2 (9.6)	34.9 (9.2)	47.5 (15.5)	46.3 (19.1)	40 (24.1)	43 (28.3)	39.2 (25.7)	34.2 (19.9)	36.3 (14.3)	38.7 (11.2)	35 (8.4)
NA_8	45.6 (14.8)	54.9 (14.7)	56.9 (14.5)	60.7 (14.2)	61.2 (15.5)	60.8 (19)	62.3 (19.9)	59.4 (19.4)	52.7 (19.4)	44.2 (18.4)	39.1 (17.4)	39.8 (17.3)
NA_98	29.3 (2.7)	30.1 (2.2)	22.4 (2.5)	20.5 (2.7)	11.2 (2.5)	14.9 (3.8)	45.3 (8.4)	42.3 (12.5)	35.5 (9.4)	21.1 (2.5)	21.1 (2.4)	28.7 (2.1)
NA_99	38.5 (6.3)	33.2 (5.4)	30.7 (7)	29.2 (8.9)	31.6 (12.9)	48.1 (14.3)	69.6 (10.7)	63 (11.9)	60.3 (13.3)	37.3 (10.8)	28.5 (7.5)	33.2 (5.7)
NT_1	66.9 (10.6)	66.7 (8)	62.8 (6.1)	60 (6)	64.3 (5)	62.1 (6.3)	60.3 (7.2)	57.9 (6.4)	55.6 (5)	62.1 (9.5)	65.3 (12.1)	70.7 (10.4)
NT_10	39.8 (12.4)	41.6 (11.7)	38.2 (14.2)	36.1 (15.4)	34 (15.8)	29.3 (13.9)	37.9 (16)	35.2 (14.3)	33.5 (15.3)	30.6 (10.9)	34.8 (12.7)	42.1 (12.7)

NT_12	70.3 (12.5)	66.9 (9.8)	62.2 (7.7)	59.4 (6.4)	52 (11.5)	41.5 (13.8)	37.3 (11.2)	36 (13.2)	42.2 (13.8)	56.8 (12)	66.5 (10.7)	69.6 (13.9)
NT_13	7.3 (15.8)	15.1 (21.1)	40.2 (16.7)	62 (33.1)	28.5 (23.8)	21.8 (18.1)	24.4 (16.6)	54.1 (21.8)	49.7 (17.3)	33.1 (16.1)	13.8 (17.9)	6 (12.6)
NT_14	77.4 (10.2)	78.8 (11.3)	77.7 (11)	74.4 (11.3)	70.4 (11.9)	62.1 (15.7)	57.8 (18.2)	55.9 (18.4)	62.9 (13.7)	73.1 (9.9)	75.4 (10)	76.7 (9.8)
NT_2	62.7 (12.3)	63.7 (9.6)	67.5 (13.6)	65.5 (14.2)	65.6 (13.6)	61.8 (15.7)	53.2 (18.9)	52.8 (16.8)	53.2 (14.9)	58 (14.4)	63.5 (14.9)	64.5 (10.9)
NT_3	59.3 (12.3)	52.8 (13.1)	50.3 (13.4)	54.2 (11.4)	58 (11.2)	57.4 (12.7)	60 (10.1)	63.4 (8.2)	65.4 (6.8)	70.7 (7.6)	70.2 (11.5)	65.9 (11.1)
NT_4	57.9 (2.9)	61.9 (4.3)	55.2 (4.8)	54.4 (6.3)	53.5 (7.5)	54.6 (9.8)	64.2 (8.7)	58.7 (9.1)	57 (6.7)	45.8 (5.8)	48.3 (5.2)	59.6 (3.6)
NT_7	54.2 (10.5)	53.3 (6.5)	55.3 (8.4)	53.7 (8.7)	55.9 (9.8)	50.3 (11.7)	40.8 (10.1)	41.9 (9.5)	41.9 (8.4)	49.6 (10.1)	53.2 (11.6)	57.3 (9.8)
NT_8	64.7 (14.7)	57.8 (16)	53.7 (17.7)	51.6 (18.3)	44.5 (17.7)	40.3 (16.1)	45 (15.7)	55.6 (14)	62.5 (14.3)	65.9 (12.5)	68.7 (14.6)	69.1 (13.9)
NT_9	42.6 (6.9)	48.8 (5.9)	40.7 (8)	35.5 (8)	23.4 (7.1)	15 (7.3)	12.8 (10.1)	10.8 (8.8)	12.7 (5.8)	23 (4.7)	35.3 (6)	45.5 (6.3)
NT_98	58.2 (21.4)	56.3 (23.8)	55.4 (23)	53.9 (22.5)	54.4 (21.6)	55 (20.4)	54 (20)	52.1 (20.4)	56.3 (17)	60.6 (16.9)	57.2 (21)	56.3 (21.4)
NT_99	48.5 (16.1)	41.7 (16.6)	43.1 (16.1)	48.6 (19.4)	62.5 (18.8)	75.9 (12.2)	77.6 (11)	76.3 (11.5)	78 (11)	68.9 (13.9)	53.5 (18.2)	45.6 (18)
OC_1	62.4 (25.8)	59.4 (23.1)	64.6 (21.9)	67.7 (18.6)	72.7 (11.6)	78.5 (9.3)	75.2 (10.9)	76.4 (10.8)	75.4 (13.3)	75.9 (14.4)	71.7 (18.5)	68.1 (22.1)
OC_2	67.8 (14.4)	68.7 (11.8)	65.7 (12.8)	61.1 (13.1)	55 (12.8)	47.4 (16.5)	39.9 (17.1)	36.7 (16.5)	46.6 (12.5)	62.9 (13.2)	65.8 (15.4)	68.1 (13.7)
OC_7	38.9 (12.4)	42.5 (11.7)	42 (10)	43.4 (9.8)	54 (6.6)	57.8 (8.5)	54.3 (9.9)	54.7 (10.5)	52.1 (10.6)	50 (10.1)	46.5 (11.3)	43.7 (13)
PA_1	64.4 (15)	63.4 (10.8)	55.7 (10)	49.4 (7.7)	49.8 (7.5)	46 (13.4)	40.5 (13.6)	37.9 (14.7)	46.2 (9.6)	60.7 (10.7)	60.5 (13.1)	63.4 (13.4)
PA_10	96.3 (6.2)	93.8 (8)	94.5 (6.2)	93.9 (6.6)	93.5 (7)	93.7 (6.5)	92.6 (8.4)	93.3 (7.5)	93.8 (6.4)	95.6 (5.5)	96 (5.7)	96.9 (5.2)
PA_11	73.5 (13.9)	73.8 (12.8)	73.8 (12.3)	72 (13.5)	69.4 (14)	68.4 (15)	66.7 (16)	67.7 (17.1)	68.8 (17.2)	72.2 (14.9)	76.1 (14.5)	73.8 (13.8)
PA_12	65.1 (17.3)	68.4 (15)	71.5 (12.3)	68.5 (13.1)	63.8 (13.9)	61.5 (15.9)	59.5 (16)	62.4 (17.5)	66.2 (16.8)	68.9 (12.9)	70.5 (14.5)	66.3 (15.5)
PA_13	33.3 (9.1)	42.2 (10.9)	51 (10.4)	48.6 (13.5)	42.6 (13.2)	39.3 (17.4)	39.9 (16.4)	39.5 (17.1)	45.3 (18.7)	49.5 (14)	47.7 (11.5)	40 (9)
PA_4	56 (24.5)	54 (25.3)	45.4 (28.3)	38.7 (25.5)	36.2 (23.3)	34.9 (26.8)	34.3 (25.6)	33.2 (26)	36.3 (26.3)	40.2 (27.3)	41.8 (26.5)	49.4 (27.6)
PA_5	9.5 (7.8)	10.9 (7.7)	12.6 (11)	20.3 (12.5)	34 (16.8)	40 (20)	35.3 (19.6)	34.4 (19.2)	27.2 (17.6)	24.5 (18.1)	17.4 (14.4)	12.6 (11.9)
PA_6	61 (20.6)	64.1 (22.1)	61.2 (22.5)	58.4 (23.5)	54 (24.5)	49 (23.8)	44.4 (23)	39.7 (22.8)	37 (20.9)	44.7 (20)	51.4 (20.9)	57.4 (20.7)
PA_8	53.9 (19.2)	52 (21.1)	51.8 (20.3)	54.5 (19.3)	59.1 (18.4)	62.4 (16.3)	60.2 (14.7)	57.2 (15.8)	57.2 (16.9)	57.2 (17.7)	54.6 (17.8)	53.4 (17.4)
PA_9	58.8 (10.2)	51.5 (10.3)	42.4 (10.4)	25.5 (6.1)	19.6 (3.6)	12.7 (2.2)	15.2 (2.3)	16.1 (2.1)	17.7 (4.3)	35.9 (6.3)	42.7 (9)	57.7 (11.5)

4.3 Spatial Resolution

Figure 2: Comparison of cloud related products for a region near Parque Nacional Jaua-Sarisariama in Southern Venezuela. Top left: MODCF developed in this paper (1km resolution). Top right: PATMOS-x AVHRR data formatted for the Global Energy and Water cycle Experiment (GEWEX) Cloud Assessment (1 degree, 110km). Lower left: WorldClim mean annual precipitation (mm). Lower right: Elevation (m).

Figure 2 illustrates the increased fine-grain detail available in MODCF compared with easily available coarse-grain cloud data from the GEWEX Cloud Assessment in a region near Parque Nacional Jaua-Sarisariama in Southern Venezuela. The 1-km dataset captures the effects of orographic cloud formation due to the complex topography (compare with elevation in the lower right panel). Figure 2 also shows mean annual precipitation from the WorldClim dataset26. WorldClim is available at the same resolution (30-arc seconds) as the product described here, but was developed from interpolated station data using only latitude, longitude, and elevation as covariates. The artifacts near stations and treatment of precipitation as a simple function of elevation are apparent in the interpolated precipitation, while much finer detail of orographic cloud effects is apparent in the Cloud Frequency. While these products (cloud frequency and precipitation) are not directly related, the possibility of incorporating cloud frequency in the interpolation of precipitation is promising.

5 Limitations and Caveats

The MOD09 cloud algorithm was designed to minimize confusion over snow and ice by taking the surface air temperature into account, however there are possibly inflated cloud frequency over snow-covered areas which are not well represented in our validation data set. Like many cloud masks, the MOD09 detection algorithm has a binary response (cloudy/not cloudy) and does not retain an estimate of confidence in cloud state (i.e. probability that the pixel was actually cloudy given the tests). The other MODIS cloud mask (MOD35) converts the continuous probabilities into four bins (certainly clear, probably clear, probably cloudy, and confidently cloudy), and is available at the satellite swath-level (which would avoid any sampling problems introduced by the orbital parameters and the MODLAND selection criteria). However, due to spatially heterogeneous application of cloud tests (even in the recently reprocessed Collection 6), the MOD35 mask is unsuitable for generating spatially consistent maps of cloud frequencies at 1-km resolution (Wilson et al., 2014). Liu and Liu Liu and Liu (2013) introduced an interesting alternative method of estimating cloud cover based on multi-year timeseries of MOD09 surface reflectance, which is promising but currently based on the frequency of clouds between 8-day MODLAND composites and thus cannot estimate the true daily cloud frequency (e.g. a cloudy observation in a single 8-day MODLAND window indicates 8 cloudy days, but a clear observation could indicated 1-7 clear days). Other approaches have been developed to estimate continuous probabilities rather than binned classifications (e.g. Heidinger et al. 2012), but these have not been applied to MODIS data. However, there the human-observed station data used for validation have known biases at the low and high cloud amounts (ELABORATE and cite). There is evidence of a negative bias in MODCF due to increased frequency of observations at high latitudes and the MODLAND algorithm (Figure SM5).

5.1 Latitudinal Effects

The MODIS polar orbit results in more frequent observations at high latitudes and gaps in daily coverage near the equator. Since the MODLAND daily compositing algorithm chooses the best (least-cloudy) observation for each pixel, the increased number of observations leads to a greater chance of at least one clear pixel (Vermote, Kotchenova, and Ray 2011). This could lead to a negative bias in cloud frequencies derived from MODLAND products at high latitudes (visible in Figure 3). We used a linear model between latitude and the station anomalies to assess the presence of a latitudinal bias in the MODCF product. The MODCF tends to overestimate CF at higher latitudes in winter months, and underestimate it in summer months.

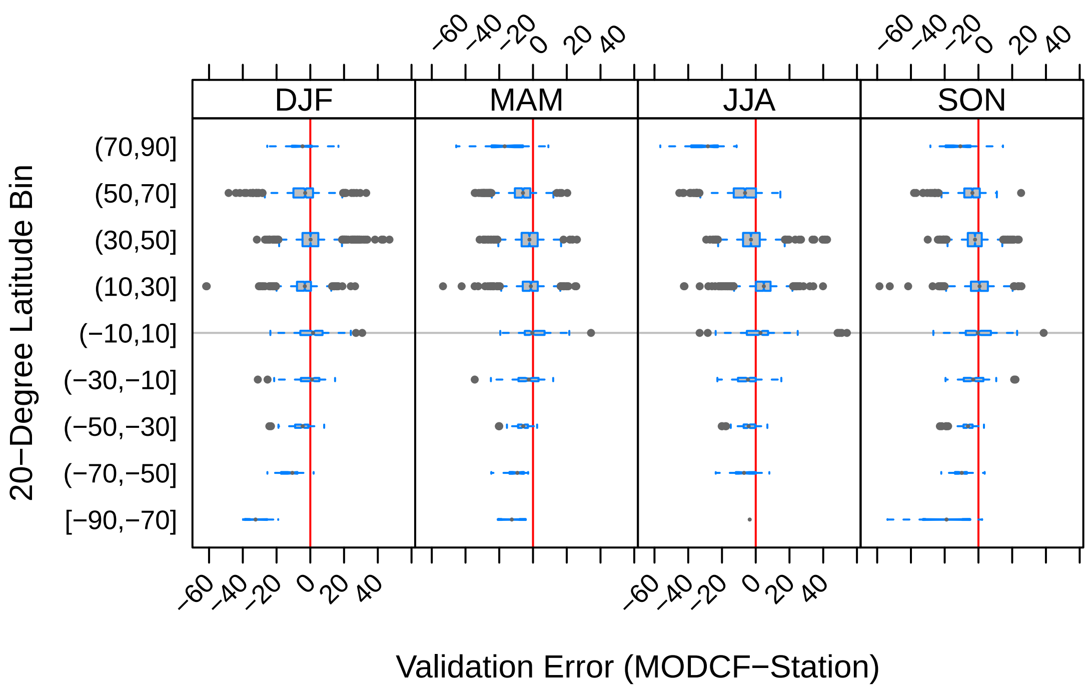


Figure SM5: Boxplots of MODCF-Station anomolies by season and 20-degree latitudinal bin. Boxplot width is proportional to the number of available validation data. Boxplot notches indicate approximate confidence intervals around the mean value in each group.

5.2 Land-Use Land-Cover Effects

Land Use - Land Cover	DJF	MAM	JJA	SON
Barren or sparsely vegetated	9.9 (440)	10.7 (294)	10.2 (576)	9.9 (441)
Cropland/Natural vegetation mosaic	11.2 (1264)	8.8 (847)	9.1 (1701)	9.2 (1298)
Croplands	7.6 (2633)	5.4 (1817)	8.3 (3582)	6.8 (2659)
Deciduous Broadleaf forest	8.5 (60)	6.8 (43)	6.3 (81)	6.3 (61)
Deciduous Needleleaf forest	20 (166)	14.4 (108)	9.8 (221)	10.4 (169)
Evergreen Broadleaf forest	10.2 (306)	9.5 (208)	9.9 (412)	10.1 (306)
Evergreen Needleleaf forest	9.8 (158)	5.6 (111)	7 (216)	4 (167)
Grasslands	12.1 (1582)	8.7 (1074)	9.8 (2113)	8.9 (1633)
Mixed forest	10.3 (1312)	6.6 (873)	7.4 (1769)	6.4 (1362)
Open shrublands	10.6 (898)	9.5 (624)	13.2 (1262)	8.1 (950)
Permanent wetlands	7.8 (32)	6.1 (22)	12.4 (44)	4.3 (31)
Savannas	11.1 (255)	8.3 (172)	7.6 (348)	10.1 (259)
Snow and ice	20.6 (18)	13.3 (17)	27 (21)	14.3 (24)
Urban and built-up	8.2 (420)	7.5 (282)	9.6 (570)	7.9 (428)
Water	8.4 (2896)	8.2 (2006)	11.5 (4032)	8.1 (3042)
Woody savannas	9.2 (724)	7.7 (496)	11.1 (992)	7.1 (750)

6 Regional Comparisons

References

Douglas, M. (2013). A high spatial resolution satellite-based cloud climatology for biogeographical applications. In *6th International Conference*, Miami, FL. International Biogeography Society.

Eastman, R. and Warren, S. G. (2012). Land Cloud Update, 1997-2009, Appended to Cloud Climatology for land stations worldwide, 1971-1996. Technical Report NDP-026D, Climate Change Research Division Office of Biological and Environmental Research U.S. Department of Energy, Oak Ridge, Tennessee.

Fehrenbach, J. and Weiss, P. (2013). Processing stationary noise: model and parameter selection in variational methods. *arXiv preprint arXiv:1307.4592*.

Fehrenbach, J., Weiss, P., and Lorenzo, C. (2012). Variational Algorithms to Remove Stationary Noise: Applications to Microscopy Imaging. *IEEE Transactions on Image Processing*, 21(10):4420–4430.

Foster, M. J. and Heidinger, A. (2012). PATMOS-x: Results from a Diurnally Corrected 30-yr Satellite Cloud Climatology. *Journal of Climate*, 26(2):414–425.

Gregg, W. W. and Casey, N. W. (2007). Sampling biases in MODIS and SeaWiFS ocean chlorophyll data. *Remote Sensing of Environment*, 111(1):25–35.

Knapp, K., Ansari, S., Bain, C., Bourassa, M., Dickinson, M., Funk, C., Helms, C., Hennon, C., Holmes, C., Huffman, G., and others (2011). Globally Gridded Satellite Observations for Climate Studies. *Bulletin of the American Meteorological Society*, 92(7):893.

Liu, R. and Liu, Y. (2013). Generation of new cloud masks from MODIS land surface reflectance products. *Remote Sensing of Environment*, 133:21–37.

Mulligan, M. (2006). MODIS MOD35 pan-tropical cloud climatology.

Petitcolin, F. and Vermote, E. (2002). Land surface reflectance, emissivity and temperature from MODIS middle and thermal infrared data. *Remote Sensing of Environment*, 83(1–2):112–134.

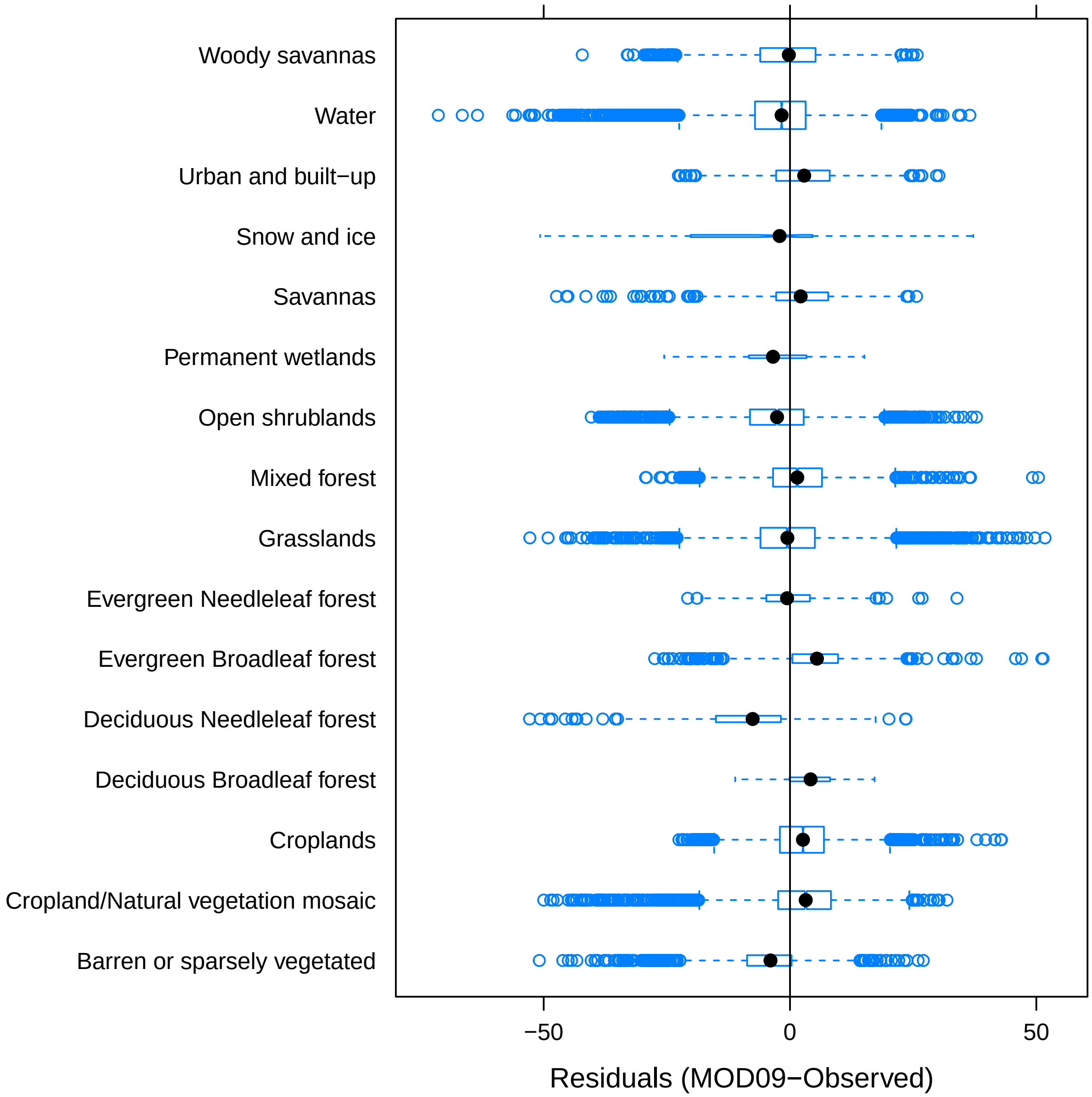


Figure SM6: Boxplot showing residuals (MOD09-Station) by land cover type.

Roger, J. and Vermote, E. (1998). A Method to Retrieve the Reflectivity Signature at 3.75 μm from AVHRR Data. *Remote Sensing of Environment*, 64(1):103–114.

Stubenrauch, C. J., Rossow, W. B., Kinne, S., Ackerman, S., Cesana, G., Chepfer, H., Di Girolamo, L., Getzewich, B., Guignard, A., Heidinger, A., Maddux, B. C., Menzel, W. P., Minnis, P., Pearl, C., Platnick, S., Poulsen, C., Riedi, J., Sun-Mack, S., Walther, A., Winker, D., Zeng, S., and Zhao, G. (2013). Assessment of global cloud datasets from satellites: Project and database initiated by the GEWEX Radiation Panel. *Bulletin of the American Meteorological Society*, page 130117123745009.

Vermote, E. F., O'Bannon, J., Davis, R., PetitColin, F., Ray, J., and El Saleous, N. (2001). Land surface reflectance code development, testing and delivery. Technical Report Contract Number NAS5-96062, University of Maryland, University of Maryland.

Wilson, A. M., Parmentier, B., and Jetz, W. (2014). Systematic land cover bias in Collection 5 MODIS cloud mask and derived products — A global overview. *Remote Sensing of Environment*, 141:149–154.

Wylie, D., Jackson, D. L., Menzel, W. P., and Bates, J. J. (2005). Trends in global cloud cover in two decades of HIRS observations. *Journal of climate*, 18(15):3021–3031.