Altitude-dependent probability of vertical, cloud-free line-of-sight from European Centre for Medium Range Weather Forecasting Re-Analysis-interim cloud cover

Adam Willitsford,* Gregory Hicks, and Walt Bowen
Johns Hopkins University Applied Physics Lab, Laurel, Maryland, United States

Abstract. A method to calculate the altitude-dependent, vertical, cloud-free line-of-sight (CFLOS) using the fractional cloud cover from the European Centre for Medium Range Weather Forecasting Re-Analysis Interim (ERA-I) dataset has been developed. This method enables users of airborne and satellite collections of optical ground data to understand the statistical coverage limitations of these collection systems by informing them of global probabilities of CFLOS versus altitude as well as time of year. This method is accurate for regions between ± 60 deg of latitude; it should not be applied to polar regions due to limitations in the underlying ERA-I data. Our CFLOS calculations have been compared to the results of CloudSAT/Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation analysis and with Moderate Resolution Imaging Spectroradiometer (MODIS) total cloud cover data. It is shown that the ERA-I reports on average less cloud cover by about 7.5% (absolute) for regions within ± 60 deg of the equator relative to MODIS cloud cover retrievals. Our CFLOS calculation leverages the resolution and diversity of ERA-I that enables spatial coverage as well as frequency of occurrence CFLOS calculations for nearly all non-polar regions on earth. [DOI: 10.1117/1.JRS.16.028502]

Keywords: cloud-free line-of-sight; European Centre for Medium Range Weather Forecasting Re-Analysis Interim; cloud-free; Moderate Resolution Imaging Spectroradiometer; Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation.

Paper 210707G received Nov. 2, 2021; accepted for publication May 13, 2022; published online May 28, 2022.

1 Introduction

The availability of ground observations of both active and passive airborne and satellite electro-optic and infrared (EO/IR) sensors can be limited by the sensor's cloud-free line-of-sight (CFLOS) to the ground. Depending on the wavelength of interest, the clouds could be responsible for scattering and/or absorptive losses in addition to providing significant increases in thermal background radiation. To understand the availability and performance of airborne and satellite sensors, it is important to understand the spatial and temporal frequency of clouds and their structure from the ground to the top of the atmosphere. To assess the performance, we present a method to determine the altitude-dependent, vertical, cloud-free line-of-sight (CFLOS) using European Centre for Medium Range Weather Forecasting Re-Analysis Interim (ERA-I) cloud data. The uniqueness of this approach is that it enables spatial (lat/lon), as well as altitude dependence and temporal coverage of the variability of the global vertical, cloud-free line of sight. The Moderate Resolution Imaging Spectroradiometer (MODIS) satellite data enables global total cloud cover comparisons at high (lat/lon) resolution but provides no information with respect to the altitude dependence. Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO)/CloudSAT can be used to construct altitude-dependent CFLOS but is sparse in time with orbit repeats on the order of 16 days. The approach presented here utilizing ERA-I data enables global analytics on the native 0.7 deg field in addition to altitude dependence and temporal variation with model updates every 6 h for the entire globe. The following describes the methodology/approach, mathematical process model, and resultant global

*Address all correspondence to Adam Willitsford, adam.willitsford@jhuapl.edu

comparisons to MODIS and CALIPSO/CloudSAT. This method of utilizing the ERA-I cloud fields enables statistical views of cloud-free line of sight as a function of altitude as well as time of year for nearly anywhere on the globe.

Over the years, methods to determine the CFLOS have been developed that use statistical analysis of whole sky imaging (WSI) cameras[1–5] to determine CFLOS from the ground to space. The works of Lund and Shanklin[1,2] allowed for the creation of master probability matrices that describe the CFLOS from the ground to space for any elevation angle given total sky cover. As noted, this is for a surface observer to space and does not contain any information with respect to the altitude-dependent CFLOS. In addition, the WSI data have been processed to provide persistent statistics on CFLOS as well as calibrated whole sky irradiance values.[6] The resultant data provide much needed inputs to statistical propagation analytics for active systems such as high-energy laser weapons[7] as well as free-space optical communications[8,9] and passive imaging systems for EO/IR detection as well as astronomical viewing. The results of WSI are quite useful for air to space or space to ground links but have limited utility in determining CFLOS as a function of slant and vertical paths that are less than the entire atmosphere. Other researchers[10,11] utilized the results of Lund's master probability matrices and cumulative clouds to construct CFLOS at limited altitudes. However, their data are highly dependent on relative cloud types and frequency that were determined by Lund. To improve this data, one can look across the globe at other sources of cloud information, but this is complicated by the sparseness of the data and diversity of these reporting metrics.[12]

An ongoing challenge has been to determine a method of calculating CFLOS as a function of altitude. A detailed analysis of CFLOS versus altitude as well as look angle was performed by Reinke et al.[13] utilizing CloudSAT cloud profiling radar and coincident CALIPSO lidar data. The data have proven very useful for global trends and comparisons and is one of the best estimates of CFLOS as a function of both look angle as well as altitude. Unfortunately the CloudSAT data are sparse in time, as the repeat rate of the satellite is 16 days; thus, only long-term statistics can be evaluated via analysis of years of data. Here, we present a method to calculate CFLOS based on an isotonic interpolation of probabilities. We utilize the ERA-I dataset to construct altitude-dependent CFLOS and construct comparisons to CloudSAT CPR CFLOS[13] as well as total cloud cover comparisons to MODIS data.[14]

2 Overview of ERA-I Data Set

The probability of CFLOS has been analyzed using the European Centre for Medium Range Weather Forecasts Re-Analysis or ERA-Interim (aka ERA-I) dataset. The ERA-Interim dataset[15,16] supplies cloud fraction (coverage) as a function of pressure/altitude. The first level starts at 0.1 hPA ($\approx$ 65 km alt.) and increases in pressure to 1012 hPA (ground level) in a quasi-logarithmic fashion yielding higher spatial resolution near the ground. This results in vertical steps of hundreds of meters to 1 km for the first 10 km of altitude and 1- to 2-km steps up to 20 km of altitude. The altitude-dependent cloud fraction is the total cloud cover within each altitude bin when viewed from above (fractional areal coverage). When looking at the cloud fractional data on a global scale within the ERA-I dataset, no clouds were found above 20 km, which yields 39 useful vertical bins (ground to 20-km altitude). In addition to fractional cloud coverage versus altitude, ERA-I provides four other useful cloud data fields: the low (LO) cloud cover (clouds < 2 km); medium (MED) cloud cover (2 to 6.5 km); high (HI) clouds (> 6.5 km); and total cloud cover, which in our case is simply 1 minus the CFLOS from 20 km to the ground along a vertical path. The ERA-I data set that was utilized provided several cloud metrics on 0.7 deg $\times$0.7 deg spatial grid with data every 6 h. The ERA-I data from 2017 was analyzed for the globe comprising: $1460 \times 60 \times 256 \times 512$ (time, pressure/altitude, latitude, and longitude) data points.

2.1 *Calculation of CFLOS from ERA-I*

To determine the CFLOS, we need to take fractional cloud coverage as a function of altitude and construct a downward-looking probability of cloud obstruction (PCO), i.e., (1 − CFLOS),

Table 1 PIE calculation for 3 levels from the low, mid, and high cloud fields provided by the ERA-I dataset.

3 Level PIE—Probability of Cloud Obscuration (1-CFLOS)	Altitude range
$PCO(0) = 0$	0 to < 2 km
$PCO_{LO} = P_A$	2 to 6.5 km
$PCO_{MED} = P_A + P_B - (P_A \cap P_B)$	6.5 to 20 km
$PCO_{HI} = P_A + P_B + P_C - (P_A \cap P_B) + (P_A \cap P_C) + (P_B \cap P_C) + (P_A \cap P_B \cap P_C)$	Total cloud cover

as we go up in altitude. This is done in a two-part process, described in brief here, and in more detail below. First, we compute the probability of cloud obstruction for the course LO, MED, and HI altitude regimes using the values provided by ERA-I in Table 1. This uses the combinatorics principle of inclusion and exclusion (PIE) to determine the PCO at each coarse altitude when looking down. Next, we calculate PCO(1) through PCO(39) for the high-resolution altitude bins 1 to 39. The calculations of PCO for these high-resolution altitude bins 1 to 39 are normalized with values calculated in the first step so that they are consistent with PCO_{LO}, PCO_{MED}, and PCO_{HI}.

STEP 1: Calculation of LO, MED, and HI altitude PCO

We employ the combinatorics PIE to compute PCO_{LO}, PCO_{MED}, and PCO_{HI}. This is a method of combining probabilities—in our case, fractional cloud coverage as a function of altitude—to compute a singular probability. Our use of PIE is only valid for statistics where the individual events are not correlated, and so it cannot be applied directly to the higher spatial resolution cloud coverage ($\approx$ 39 useable altitude bins) because the cloud structure is correlated on the vertical scale of hundreds of meters to perhaps as large as 2 km. There are methods to calculate PIE accounting for correlation in the data; however, they become computationally intensive and perhaps more importantly, rely very heavily on knowing the correlation between every single layer given by the 39 altitude bins. This would imply the user has a priori knowledge of the vertical cloud structure correlation not only seasonally (temporally) but also globally (spatially). In fact, one can show that a naïve direct application of PIE to the high-resolution bins will yield results that contradict the total cloud cover provided by ERA-I.

However, the principle of inclusion/exclusion can be used to combine the probability of a cloud obstruction (1-CFLOS) given by the ERA-I lower spatial resolution LO, MED, and HI cloud coverage values. This is approximately valid because the LO, MED, and HI clouds are vertically separated by a large enough spacing to be sufficiently uncorrelated to use PIE combinatorics. A detailed look at global cloud vertical and horizontal structure was performed by Guillaume et. al.[17] utilizing CloudSAT data and a summary of the vertical cloud extent is shown in Table 2 of their work. The clouds present between each of the three coarse levels exhibit vertical structure on average that is less than the step size from LO to MED to HI in the ERA-I database. This provides for additional confidence in our assumption that the LO, MED, and HI clouds are sufficiently uncorrelated for combinatorics. Two notable exceptions are nimbostratus (NS) clouds with a mean thickness of 4.7 km[17] and deep convection (DS) with a mean thickness of 9.3 km.[17] In general these are LO probability cases as DS occurs <5% to 10% of the time[18] and NS is rare except in polar regions > 60 deg,[18] a region we are not considering for our CFLOS calculation.

Figure 1 shows how we construct a three-level PCO by assessing the Venn diagram for each level. Starting at Cloud A (ERA-I LO cloud coverage), we simply determine the PCO to be the spatial (areal) coverage provided by Cloud A. As we go up through the atmosphere to a point just above Cloud B, we then need to determine the combination of probabilities of seeing a cloud $P_A + P_B$ and subtract the region that has been double-counted $P_A \cap P_B$. This approach is continued for our third level, and we thus have the probability of cloud obscuration for three distinct altitudes and is outlined in Table 1.

An additional field, total cloud cover, reported in the ERA-I dataset was plotted against the PCO(HI) value, derived total cloud cover from combinatorics, and was found to be linear with

Table 2 ERA-I minus MODIS total cloud cover global statistics.

Total Cloud Cover Global Comparisons for Latitudes between ±60 deg		
Month for 2017	Mean (ERA-I minus MODIS)	Std. dev. (ERA-I minus MODIS)
January	−0.075	0.128
February	−0.084	0.115
March	−0.075	0.106
April	−0.074	0.108
May	−0.071	0.100
June	−0.079	0.115
July	−0.079	0.120
August	−0.078	0.118
September	−0.070	0.108
October	−0.073	0.105
November	−0.075	0.110
December	−0.083	0.121

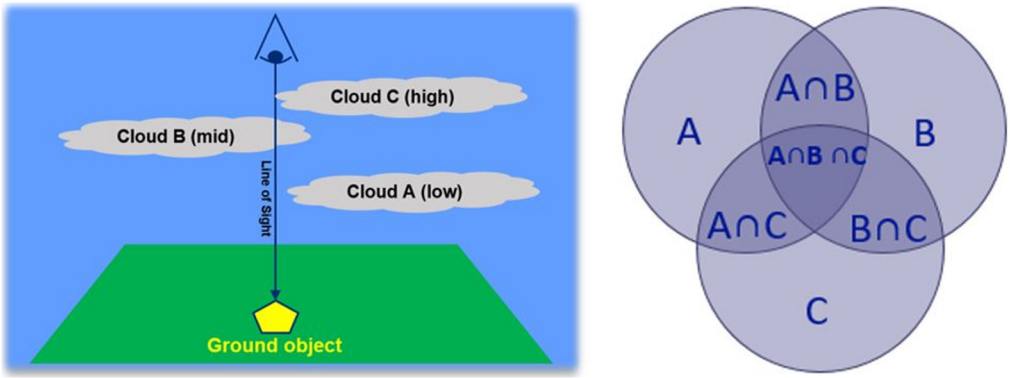

Fig. 1 The three-level PIE to determine the probability of seeing a cloud as a function of altitude.

a fit of $x = 0.999\ y$ ($R^2 > 0.999$), see Fig. 2. This further strengthens the argument that the clouds in the LO, MED, and HI cloud cover bins can be considered uncorrelated within the ERA-I data set.

STEP 2: Calculation of PCO for high-resolution altitude bins 1 to 39

In our application, as we increase in altitude the PCO either remains the same or increases as we go through cloud layers. That is, the probability of obstruction monotonically increases, and this is a hard constraint. For this reason, we develop an isotonic interpolation model to approximate the vertical CFLOS from any of the altitude bins 1 to 39 looking down to the ground.

The isotonic interpolation is performed between the three boundary conditions: PCO_{LO}, $PCLOS_{MED}$, and PCO_{HI} at altitudes of 2, 6.5, and 20 km, respectively. To perform this calculation, we take the cumulative sum of the fractional (area) cloud coverage altitude bins from with-in the ERA-I model from the ground to 2 km. Using the cumulative sum on the higher spatial resolution scale allows the data we actually do have to inform the rate at which our interpolator increases/accumulates probability between each of these coarse altitude bins. We then normalize the cumulative sum at the 2 km point to the 2-km PIE value—knowing that the cumulative sum of fractional cloud cover should equate to the cloud cover from the 2-km PIE point.

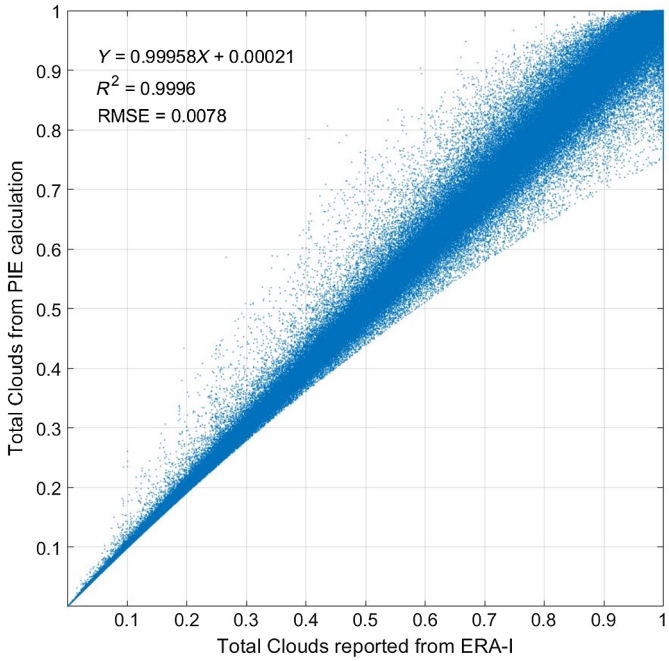

Fig. 2 The total probability of seeing a cloud as derived from the three-level PIE approach is plotted against the total cloud field reported by the ERA-I data. The data shown represent nearly five million points around the contiguous United States for the year 2017 (1460 times, 38 latitudes, and 89 longitudes).

That is, we enforce the boundary condition to ensure that the endpoints maintain correct cloud coverage values. In our case, the ground has no clouds and the 2-km point is equivalent to the LO cloud $\mathrm{PCO_{LO}}$ value. The points between the ground and 2 km are then interpolated by the normalized cumulative sum from the ground to 2 km—resulting in a cloud obstruction profile (1-CFLOS) from 0 to 2 km in altitude.

$$\mathrm{cumulativeCloudFraction(alt)} = \mathrm{ccf(alt)} = \sum_{\mathrm{altStart}}^{\mathrm{altEnd}} \mathrm{cloudFraction(alt)}. \tag{1}$$

This process is repeated from the 2- to 6.5-km mMED cloud level (normalizing the cumulative sum profile endpoints to the end point PIE values $\mathrm{PCO_{LO}}$ (2 km) and $\mathrm{PCO_{MED}}$ (6.5 km) and finally, for the 6.5- to 20-km cloud level ($\mathrm{PCO_{HI}}$). We concatenate the results and create a PCO (1-CFLOS) as a function of altitude with vertical resolutions equivalent to the original ERA-I data. These calculations are summarized by the piecewise defined interpolation function

$$\mathrm{PCO(alt)} = \begin{cases} \frac{\mathrm{ccf(alt)}}{\mathrm{ccf(2\ km)}} \times (\mathrm{PCO_{LO}} - \mathrm{PCO}_G) + \mathrm{PCO(0\ km)} & \mathrm{alt} \in [0,2] \\ \frac{\mathrm{ccf(alt)}}{\mathrm{ccf(6.5\ km)}} \times (\mathrm{PCO_{MED}} - \mathrm{PCO_{LO}}) + \mathrm{PCO(2\ km)} & \mathrm{alt} \in (2, 6.5] \\ \frac{\mathrm{ccf(alt)}}{\mathrm{ccf(20\ km)}} \times (\mathrm{PCO_{HI}} - \mathrm{PCO_{MED}}) + \mathrm{PCO(6.5\ km)} & \mathrm{alt} \in (6.5, 20] \end{cases}, \tag{2}$$

where ccf stands for the cumulativeCloudFraction as calculated using Eq. (1).

This calculation is shown graphically in Fig. 3, where the cloud fraction versus altitude from the ERA-I data on the left and the resultant CFLOS is shown on the right. In vertical regions where clouds exist the CFLOS reduces to account for the existence of clouds and when the fractional cloud coverage goes to zero, we see the CFLOS remain constant as a function of altitude until an additional cloud is found. There is a specific set of cases where this will result in inaccurate results: noteworthy is any region where we have two significant fractional cloud (>0.75) layers separated by a region of no clouds within either the LO, MED, or HI altitude zones. The cumulative summation and normalization will smooth over the first cloud, resulting

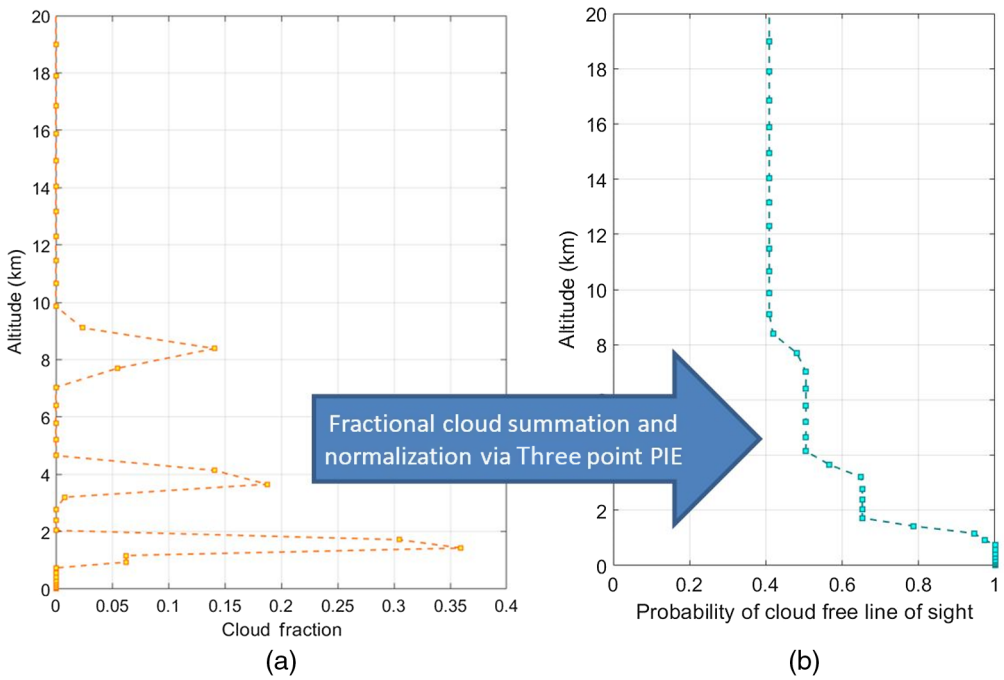

Fig. 3 (a) Fractional cloud coverage from ERA-I data. (b) Derived CFLOS using the three-level PIE and cumulative summation normalization.

in clearer conditions at the lower altitude bin than otherwise should be seen; fortunately, this is found to occur $< 1\%$ of the time.

3 Process Model Justification

In the preceding section, we described an isotonic interpolation of probability. We now describe this modeling in a more rigorous fashion and provide a reasonable justification for its use.

3.1 *Orthographic Projection and Geometric Probability*

The topic of geometric probability is multi-disciplined and covers several niche probability problem types.[19] However, for our purposes, here, we are using the term geometric probability to describe a specific ratio of measures that might be found in the subject. Consider a cell, or element of volume if one prefers, R_{ijk} in an over-approximating rectangular partition of a region R of the atmosphere. And let us then define the right rectangular solid

$$S_{\sqsubset\sqsupset}^{k_l,k_u} = \{(x,y,z) \in R | x_{i-1} \le x \le x_i,\ y_{j-1} \le y \le y_j, z_{k_l} < z \le z_{k_u}\}, \tag{3}$$

which is the union of $R_{ij(k_l+1)}, \ldots, R_{ijk_u}$. The intersection of the clouds C with $S_{\sqsubset\sqsupset}^{k_l,k_u}$ is simply $C_{k_l,k_u} = C \cap S_{\sqsubset\sqsupset}^{k_l,k_u}$. The upper face F_{k_u} of the cylinder will be a rectangle in the plane $z = z_{k_u}$. We orthographically project C_{k_l,k_u} onto F_{k_u} via the mapping $P_{k_l,k_u}: S_{\sqsubset\sqsupset}^{k_l,k_u} \to F_{k_u} : p = (x,y,z) \mapsto (x,y,z_{k_u})$. Let μ denote the Lebesgue measure on F_{k_u}. Then, the geometric probability of cloud coverage for the cylinder $S_{\sqsubset\sqsupset}^{k_l,k_u}$ is given as

$$q_{k_l,k_u} = \frac{\mu(P_{k_l,k_u}(C_{k_l,k_u}))}{\mu(F_{k_u})}, \tag{4}$$

so long as the projection $P_{k_l,k_u}(C_{k_l,k_u})$ is measurable. An interpretation of this probability is that it is the likelihood q_{k_l,k_u} that sighting along a vertical ray from altitude k_l to altitude k_u will be

blocked by the clouds, PCO. Let k_0 be the altitude index of ground and define $q_{k_0,k_0} = 0$. We will let p_{k_l,k_u} denote the complementary probability of CFLOS for $S_{\sqsubset\sqsupset}^{k_l,k_u}$.

3.2 *Bernoulli Processes*

Let $D = \{(x,y)|(x,y,z) \in S_{\sqsubset\sqsupset}^{k_l,k_u}\}$, a rectangle of side lengths Δx and Δy. Then, for any $(x,y) \in D$ chosen uniformly at random, we will assign to B_{k_l,k_u} the value of 1 if we have CFLOS through $S_{\sqsubset\sqsupset}^{k_l,k_u}$ and 0 else. $\{B_{k_l,k_u}\}$ is then a two-parameter family of Bernoulli random variables with probability of 'success' p_{k_l,k_u} and failure q_{k_l,k_u}.

Recalling that $S_{\sqsubset\sqsupset}^{k_l,k_u} = \bigcup_{k=k_l+1}^{k_u} R_{\text{ijk}}$ and keeping firmly in mind the definitions made heretofore, it should be clear that there are strong relationships between members of the family $\{B_{k_l,k_u}\}$ and the associated distribution parameters consistent with the underlying geometry. Indeed, we have in general the constructive relationship $p_{k_l,k_u} = P(B_{k_l,k_u} = 1) = P(B_{k_l,k} + B_{k,k_u} = 2)$. We will single out of the family the one step variables $B_{k-1,k}$ and denote them simply B_k. The recurrance probability relation is then calculated as

$$p_{k_l,k_u} = P(B_{k_l,k_u} = 1) = P(B_{k_l,k_u-1} + B_{k_u} = 2) = \ldots = P(B_{k_l+1} + \ldots + B_{k_u} = k_u - k_l). \quad (5)$$

Contextualizing within our application, we now observe the following. Our problem data is the distributional parameterization $\{p_k\}_{k=k_l+1}^{k_u}$ of the Bernoulli process $\{B_k\}_{k=k_l+1}^{k_u}$, the fractional (area) cloud coverage altitude bins from with-in the ERA-I model. We are aiming to determine the distributional parameterization $\{p_{k_l,k}\}_{k=k_l+1}^{k_u}$ of the Bernoulli process $\{B_{k_l,k}\}_{k=k_l+1}^{k_u}$, and Eq. (5) provides the connection between what we are given and what we desire. The goal of our isotonic interpolation is in fact to model $p_{k_l,k}$ in terms of the p_l, $l \in [k_l k]$, where $[k_l k] \doteq k_l + 1, \ldots, k$.

3.3 *Probability Models*

We will now introduce reasonable arithmetic and geometric recurrence models for $p_{k_l,k}$ in terms of the p_l, $l \in [k_l k]$, and then indicate why we proceed with the arithmetic model as the basis for our isotonic interpolation model and how this arithmetic model gets configured into an interpolant.

3.3.1 *Arithmetic model development*

To develop the arithmetic model, we examine the complementary probability, which can be expressed using De Morgan's law as

$$q_{k_l,k_u} = P((B_{k_l,k_u-1} = 0) \cup (B_{k_u} = 0)), \quad (6)$$

$$= P(B_{k_l,k_u-1} = 0) + P(B_{k_u} = 0) - P((B_{k_l,k_u-1} = 0) \cap (B_{k_u} = 0)), \quad (7)$$

$$= P(B_{k_l,k_u-1} = 0) + [P(B_{k_u} = 0) - P(B_{k_u} = 0|B_{k_l,k_u-1} = 0)P(B_{k_l,k_u-1} = 0)], \quad (8)$$

where the second equality follows from the PIE. Rewriting the last member of the equation in terms of distributional parameters and rearranging, we have the difference/recurrance relation $q_{k_l,k_u} - q_{k_l,k_u-1} = d$ where

$$d = q_{k_u} - P(B_{k_u} = 0|B_{k_l,k_u-1} = 0)q_{k_l,k_u-1}. \quad (9)$$

We focus on difference field d. Appealing to the geometry $q_{k_u} - q_{k_l,k_u-1}$ is a difference in the measures of the cloud projections of the k_u partition strata and the cloud projections from the strata within a preceding right rectangular cylinder (a column of atmosphere below). Apparently, this difference can be negative whereas the difference field cannot. The probability

$P(B_{k_u} = 0 | B_{k_l,k_u-1} = 0)$ is the premodifier on the prior that appropriately conditions the field so that this never occurs. In the case of positive associations between the random variables B_{k_u} and B_{k_l,k_u-1}, this probability is larger and reduces the difference field. In the case of negative associations between these random variables, the probability is slight and increases the difference field. In effect, which is the case amounts to whether we are dealing with the same clouds or not as we enter the k_u'th strata. Let us assume that we are dealing with a portion of atmosphere of a size that suggests positive association between cload masses. The $P(B_{k_u} = 0 | B_{k_l,k_u-1} = 0)$ then serves to diminish the positive attibution of the difference field to the accumulation captured by this difference equation.

The remaining issue is that we do not have access to any data in our problem to provide $P(B_{k_u} = 0 | B_{k_l,k_u-1} = 0)$. This difference field modifier cannot be sensibly computed from the q_l. For a given $0 \leq m_p \leq 1$, there are any given number of regions P_{m_p} in F_l that yield the geometric probability m_p. And while those regions cannot be dispersed, we know of no bases for an argument that they have centrality or tendency with respect to location in F_l. Generally speaking, we do not know what the blotches will look like or where they will be, so we cannot make inferences about the projection of a cloud distribution in a cell, or the cloud distribution in the cell itself, using q_l. We lose the geometry in the measure of the projection. Based on this observation, we can only say that the extent to which q_{k_u} may potentially contribute positively to the difference $q_{k_l,k_u} - q_{k_l,k_u-1}$ is directly proportional to q_{k_u}. Defining a_{k_u} as the proportionality constant, we posit

$$q_{k_l,k_u-1} = q_{k_l,k_u} - a_{k_u} q_{k_u}. \tag{10}$$

We assume consensus across strata with respect to the constant of proportionality, which we will denote by a. We have then the arithmetic recurrance model

$$q_{k_l,k_u-1} = q_{k_l,k_u} - a q_{k_u}. \tag{11}$$

3.3.2 *Geometric model development*

Our geometric model comes to us directly from the Conway–Maxwell–Poisson (CMP) binomial distribution.[20] This distribution generalizes the Poisson binomial distribution, which characterizes Bernoulli sums, to the case were the probabilities of success are associated.[21] We obtain from the CMP binomial distribution

$$p_{k_l,k_u} = P(B_{k_l,k_u-1} + B_{k_u} = 2) = \frac{1}{C^{\nu}_{k_l,k_u-1}} p_{k_l,k_u-1} p_u, \tag{12}$$

where

$$C^{\nu}_{k_l,k_u-1} = q_{k_l,k_u-1} q_u + 2^{\nu-1}(q_{k_l,k_u-1} p_u + p_{k_l,k_u-1} q_u) + p_{k_l,k_u-1} p_u. \tag{13}$$

If our CFLOS events over strata were independent, ν and $C^{\nu}_{k_l,k_u-1} = 1$, and we get a very clear-cut product relationship between p_{k_l,k_u} and the p_l. In the case of our assumed positively associated events, however, $-\infty \leq \nu < 1$, which yields $0 < C^{\nu}_{(k_u-k_l)} < 1$, scaling the product up appropriately. This scale factor is nonlinear in ν and dependent on k_l, k_u. We have no clear synthesis method through which to obtain $\nu(k_l, k_u)$ from our problem data. This is a beautiful, but complicated model. But treating with it as we did with the difference model, we have, upon defining b_{k_u} as the proportionality constant

$$p_{k_l,k_u} = b_{k_u} p_{k_l,k_u-1} p_{k_u}. \tag{14}$$

Assuming consensus across strata in the correction factor, call it b, we have the geometric recurrance model

$$p_{k_l,k_u} = b p_{k_l,k_u-1} p_{k_u}. \tag{15}$$

3.4 *Sensitivity Analysis and Justification of Arithmetic Model Use*

Solving the arithmetic recursion model yields us

$$q_{k_{n-1},k_{n+1}} = q_{k_{n-1},k_n} + \sum_{l=k_n+1}^{k_{n+1}} a_l q_l, \tag{16}$$

while solution of the geometric recursion model provides

$$p_{k_{n-1},k_{n+1}} = \left(\prod_{l=k_n+1}^{k_{n+1}} b_l\right)\left(\prod_{l=k_n+1}^{k_{n+1}} p_l\right) p_{k_{n-1},k_n}. \tag{17}$$

Let us now compute the relative sensitivities of these recursion solutions with respect to their respective proportionality constants. In the case of the arithmetic recursion direct computation provides

$$\frac{a_k}{q_{k_{n-1},k_{n+1}}}\frac{\partial q_{k_{n-1},k_{n+1}}}{\partial a_k} = \frac{a_k q_k}{q_{k_{n-1},k_n} + \sum_{l=k_n+1}^{k_{n+1}} a_l q_l} \leq 1, \tag{18}$$

when defined. Equality only holds in the case that q_{k_{n-1},k_n} and all q_l, $l \neq k$, in the summation are 0. In the case of the geometric recursion direct computation provides

$$\frac{b_k}{p_{k_{n-1},k_{n+1}}}\frac{\partial p_{k_{n-1},k_{n+1}}}{\partial b_k} = 1, \tag{19}$$

when defined.

The takeaway is the following. In assuming consensus in the proportionality constant, we are inducing a perturbation away from the more ideal strata dependent proportionality constants. The percent change in the correct probabilities due to a percent change in the parameters due to these pertubations is significantly less in the case of the arithmetic model than would be seen with the geometric model. Simply put, the geometric model is appreciably more sensitive to our gross practical simplifications. For this reason alone, we make use of the arithmetic model instead of the geometric model here, even though the latter has a more philosophically pleasing basis.

3.5 *Arithmetic Recurrence Model Use in Interpolation*

Suppose then that we have a two-point boundary value problem with given initial condition q_{k_{n-1},k_n} and final condition $q_{k_{n-1},k_{n+1}}$ and that we seek $q_{k_n,k}$ for $k \in [k_n k_{n+1}]$. Let us return to our Eq. (16) solution to the arithmetic recurrance relation with the consensus assumption back in force. We have

$$q_{k_{n-1},k_{n+1}} = q_{k_{n-1},k_n} + a \sum_{l=k_n+1}^{k_{n+1}} q_l, \tag{20}$$

from which we readily obtain

$$a = \frac{q_{k_{n-1},k_{n+1}} - q_{k_{n-1},k_n}}{\sum_{l=k_n+1}^{k_{n+1}} q_l}, \tag{21}$$

and for $k \in [k_n k_{n+1}]$, we have

$$q_{k_n,k} = q_{k_{n-1},k_n} + \left(\frac{q_{k_{n-1},k_{n+1}} - q_{k_{n-1},k_n}}{\sum_{l=k_n+1}^{k_{n+1}} q_l}\right) \sum_{l=k_n+1}^{k} q_l, \tag{22}$$

as a isotonic interpolant between the boundary data. The final piece of mathematical detail is realizing that this is the interpolant used repeatedly along with accurate PIE generated boundary data to obtain isotonic interpolation Eq. (2).

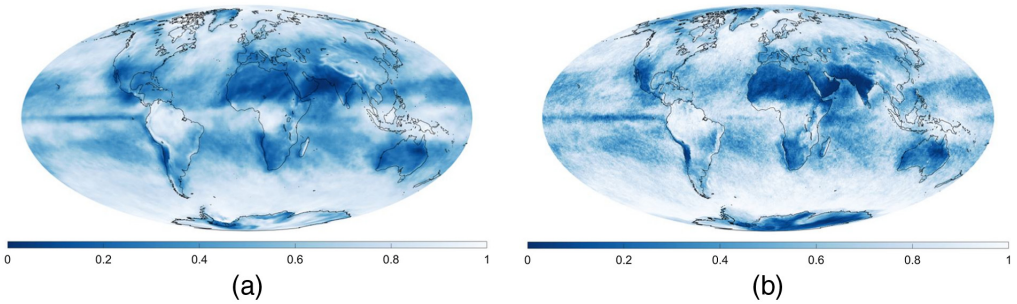

Fig. 4 Mean total cloud cover for April 2017. Left: ERA-I data analysis results from this study using PIE combinatorics and cumulative-summation normalization. Right: MODIS on NASA's Terra Satellite.[14]

4 Comparisons: MODIS Total Cloud Comparison

The global total cloud cover from the ERA-I data was averaged on a monthly basis for 2017 and compared to the MODIS data[14] for the same period by looking at the difference in cloud cover on a latitude and longitude grid. The $0.1-$ deg MODIS data were interpolated and downsampled to the $0.7-$ deg ERA-I data and directly compared by taking the difference between the ERA-I total cloud cover and MODIS total cloud cover. The mean difference was calculated for latitudes between ± 60 deg and was found to be -0.074 with a standard deviation of ≈ 0.11. This confirms other researchers' findings[22–24] that the ERA-I data are generally in good agreement with a bias toward clearer conditions and are relatively inaccurate at the poles. By utilizing the MODIS total cloud cover as a source of calibration data, we can feel confident in our PIE combinatorics approach to the ERA-I data set that uniquely enables global altitude-dependent CFLOS metrics with temporal variation as the cloud fields in ERA-I are updated every 6 h. Figure 4 shows the mean total cloud cover for the month of April 2017; on the left is the ERA-I data and on the right is the MODIS data. Figure 5 shows the difference in total cloud cover for December 2017—the color scale has been set using the mean as the center with the limits set to ± 2 standard deviations. A summary of the total global cloud cover and and statistical comparisons to MODIS broken down by month is shown in Table 2. The monthly breakdown shows little variability in the relative total cloud cover for ERA-I and MODIS with a consistent bias of ERA-I toward marginally clearer conditions. See Appendix A for ERA-I total cloud cover minus MODIS total cloud cover for each month of 2017.

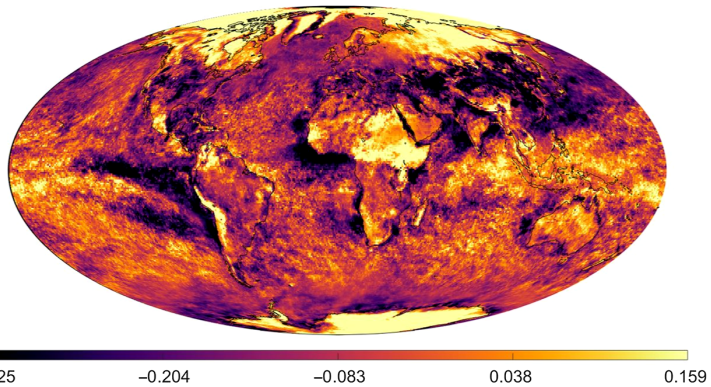

Fig. 5 Difference between ERA-I and MODIS total cloud cover for December 2017. The color scale has been set using the mean difference as the center and the limits set to $\pm 2\sigma$ as described in Table 2 for December. See Appendix A for a detail of each month.

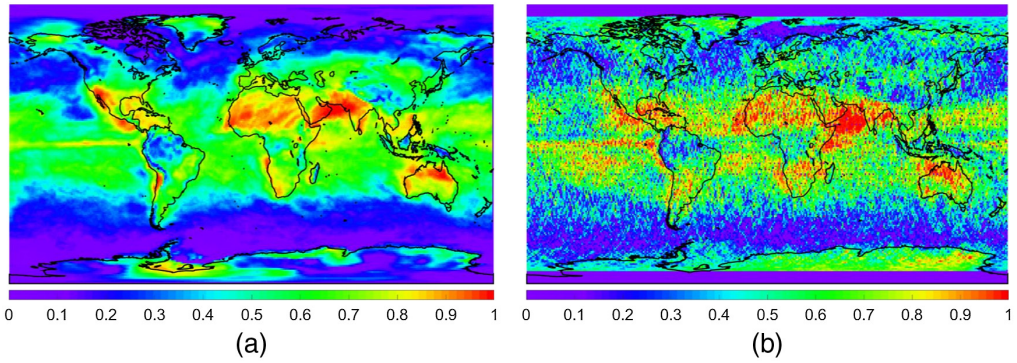

Fig. 6 Probability of vertical CFLOS from 9.6-km altitude in April. (a) ERA-I data analysis results from this study with three-point PIE and cumulative-summation normalization. (b) CloudSAT/CALIPSO (Reinke et al.).[25]

5 Comparisons: CloudSAT Comparisons

We also compared the ERA-I derived vertical CFLOS to the CFLOS determined by Reinke et al.[13] using CloudSAT data. Figure 6 shows the average over the month of April of the vertical CFLOS from 9.6-km altitude; on the left is the ERA-I-derived CFLOS and on the right is the CloudSAT-derived CFLOS. The data were then directly compared by taking the difference between the ERA-I and CloudSAT data. This was performed by interpolating the $0.7-$ deg gridded ERA-I data to the $1-$ deg gridded CloudSAT data. The mean difference was calculated for latitudes between ± 60 deg and was found to be <0.025 for most altitudes below 10 km with larger differences at higher altitudes. The standard deviation was larger with $\sigma \approx 0.2$ for each altitude. As noted earlier in the comparisons to MODIS, the total cloud cover in ERA-I has a bias toward clearer conditions; this is also directly seen in the altitude-dependent CFLOS calculated with ERA-I and compared to CloudSAT/CALIPSO. Comparative statistics of CFLOS for each altitude are shown in Table 3. Analyzing the results of this table, we can see that the difference in CFLOS is seen to grow mostly between ≈ 11 to ≈ 16 km. The higher-altitude cirrus clouds within the ERA-I dataset are underrepresented relative to CloudSAT/CALIPSO and lead to lower total cloud cover when compared to MODIS. Figure 7 shows the difference in CFLOS for April from 9.6-km altitude—the color scale has been set using the mean as the center with the limits set to ± 2 standard deviations. Upon careful inspection, one will notice that the

Table 3 Difference in vertical CFLOS (ERA-I minus CloudSAT) for the month of April as a function of altitude for latitudes between ± 60 deg.

Alt km	Mean (ERA-I minus CloudSAT)	Std. dev. (ERA-I minus CloudSAT)	Alt km	Mean (ERA-I minus CloudSAT)	Std. dev. (ERA-I minus CloudSAT)
0.96	0.07	0.155	10.56	−0.027	0.175
1.92	0.001	0.176	11.52	−0.032	0.177
2.88	−0.018	0.177	12.48	−0.039	0.177
3.84	−0.020	0.177	13.44	−0.048	0.175
4.80	−0.021	0.177	14.40	−0.060	0.173
5.76	−0.020	0.176	15.36	−0.072	0.174
6.72	−0.021	0.175	16.32	−0.082	0.177
7.68	−0.016	0.176	17.28	−0.088	0.179
8.64	−0.013	0.176	18.24	−0.088	0.179
9.60	−0.025	0.174	19.20	−0.088	0.179

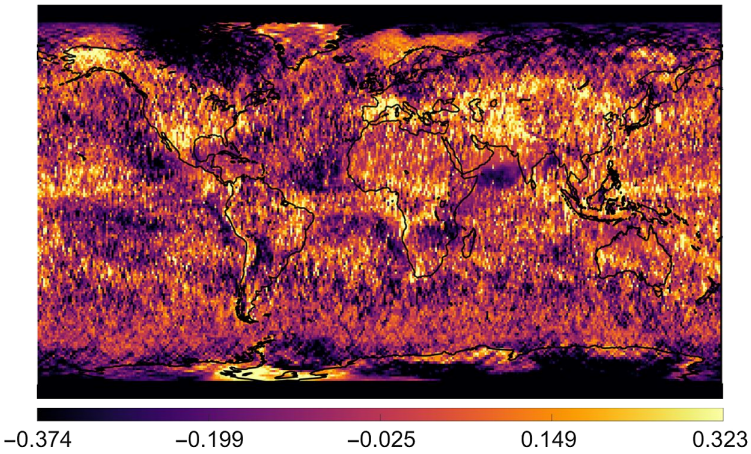

Fig. 7 Difference between ERA-I and CloudSAT CFLOS for April from 9.6-km above the ground. The color scale has been set using the mean difference as the center and the limits set to $\pm 2\sigma$ as described in Table 3 for altitude = 9.6 km. Appendix B contains global maps for additional altitude levels.

CloudSAT/CALIPSO data have a linear structure that is clearly the result of the satellite's orbit. Additional global plots of the altitude-dependent difference in ERA-I CFLOS and CloudSAT CFLOS can found in Appendix B.

In addition to global analysis, a detailed look at the United States was performed for the ERA-I data from 2017. Figure 8 shows the mean CFLOS from 9.6-km altitude for 2017 over the United States. Outside of the southwest United States, one would expect to have a CFLOS to 9.6-km altitude approximately half of the time.

6 Temporal Analysis

One of the major advantages of using the ERA-I data is the 6 h temporal resolution that is available. Utilizing the time domain data, we are able to assess the frequency of occurrence of clouds around the globe. We approach this by calculating a family of percentile based curves. Each family of color curves represent the percentage of time that a certain CFLOS will be observed. Thus, the cloud-free line-of-sight can be thought of as the amount of ground or sky in a vertical slice one will observe that is cloud-free, while the color curves tell you how often you would expect to see that much. Figure 9 shows the CFLOS derived herein from the ERA-I data as a function of altitude. The left side is for the Johns Hopkins Applied Physics Laboratory facility

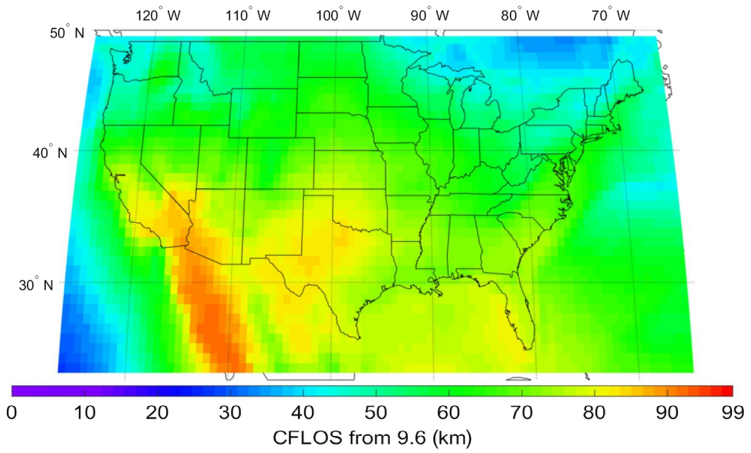

Fig. 8 The 2017 year-long mean vertical CFLOS from 9.6-km altitude—derived from the ERA-I cloud data as described herein.

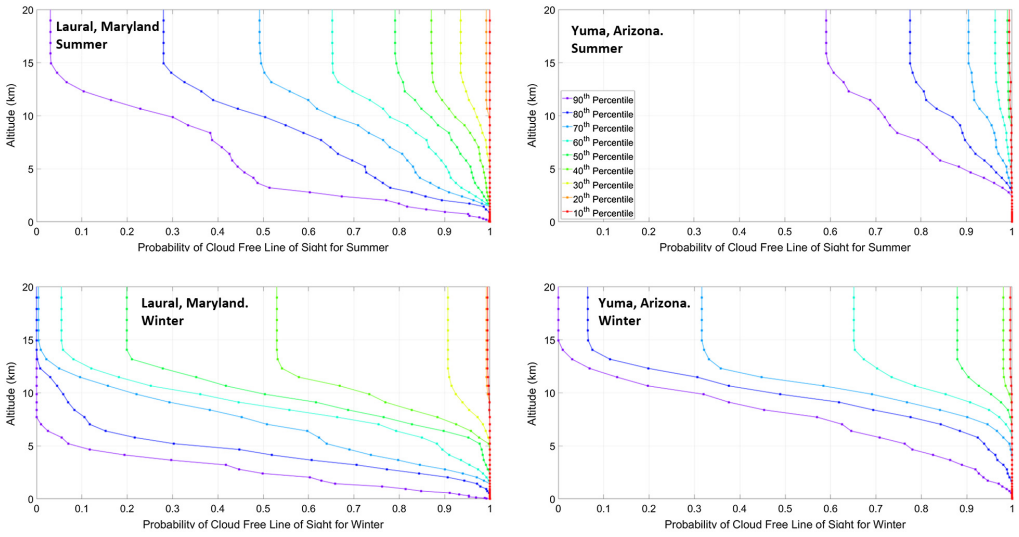

Fig. 9 Vertical CFLOS for summer and winter. Left: JHU/APL Laurel Maryland. Right: Yuma, Arizona. Winter in general tends to have the most clouds in the US. Large drops in CFLOS as a function of altitude are indicative of cloud layers. The families of colored curves represent the percentage of time one would expect to have a certain CFLOS to the ground.

located in Laurel, Maryland, and the right side of Fig. 9 shows Yuma, Arizona. The top plots are a result of aggregating the CFLOS curves for the summer (June, July, and August) and the bottom sets of curves are from aggregating the winter CFLOS (December, January, and February). The data show that in general, winter has more clouds in both Laurel, Maryland, and Yuma Arizona. If one focuses on the plot of the CFLOS for Yuma Arizona in the winter (bottom-right of Fig. 9), the lower altitudes are generally cloud-free. As we go up in altitude, the middle percentile curves start to turn toward lower CFLOS between 7 and 12 km, indicative of cloud layers. In any of these types of plots, the cloud layers become immediately obvious by looking for large drops in CFLOS as a function of altitude.

7 Conclusions

A method of calculating the CFLOS from the ERA-I data has been developed. The results of the ERA-I total cloud cover have been compared on a global scale to MODIS for latitudes between ± 60 deg and in general are shown to under-predict clouds relative to MODIS with total cloud cover $\sim$7.5% less than MODIS with a standard deviation of 10%. Additionally, the CFLOS derived from the ERA-I dataset has been compared to CloudSAT CPR data as a function of altitude on a global scale for latitudes between ± 60 deg and shows on average good agreement with a larger standard deviation of 0.2. In addition to spatial analysis of CFLOS (latitude, longitude, altitude), we have looked at the ERA-I data every 6 h over multiple years, thus enabling the end-user to look at the temporal distribution of CFLOS as well. The families of percentile-based curves were developed to represent the percentage of time one would expect to have a certain cloud-free-line-of-sight to the ground as a function of altitude. It is understood that the ERA-I cloud derivation is not perfect and in some rare ($< 1\%$ of the time) instances it will over-predict the CFLOS for lower altitudes. However, our method of retrieving altitude-dependent CFLOS from ERA-I data enables many engineering tasks not previously possible. The ERA-I technique provides for vertical, altitude-dependent CFLOS on a global scale with temporal density; e.g., profiles every 6 h. CloudSAT/CALIPSO retrievals provide for very accurate measures of the vertical structure but are temporally sparse. The ability to look at the vertical structure on a more dense temporal scale provides for operational decisions with respect to airborne sensors for optimum altitudes of operation utilizing the statistics of CFLOS for day versus night, winter versus summer etc. Future iterations would include additional knowledge of vertical cloud coherence to refine the probability models.

8 Appendix A: ERA-I Minus MODIS Total Cloud Cover

The global total cloud cover difference for ERA-I minus Modis has been analyzed to understand statistical variability over a years? worth of data. The monthly breakdown shows little variability in the relative total cloud cover for ERA-I and MODIS with a consistent bias of ERA-I toward marginally clearer conditions (Fig. 10).

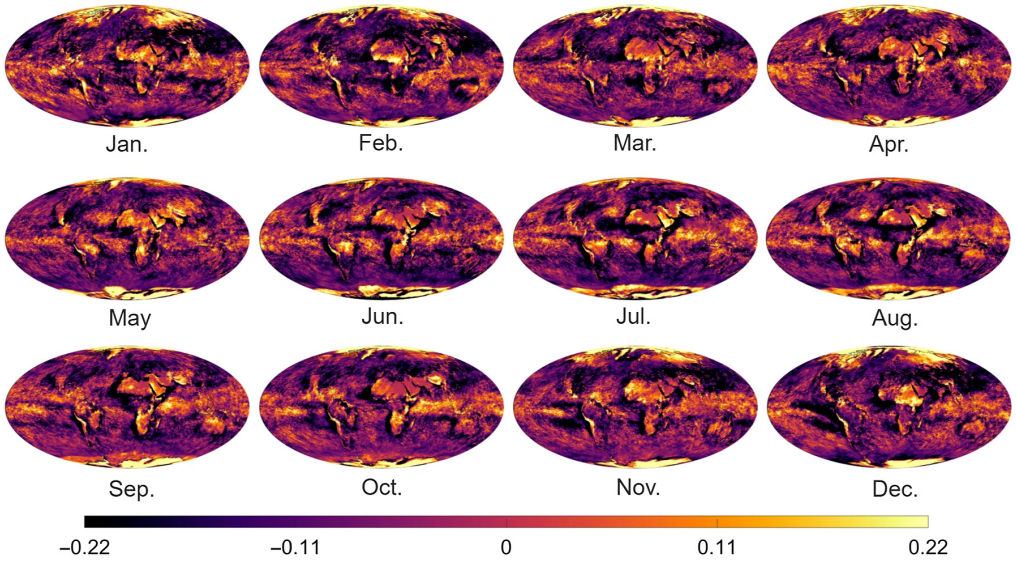

Fig. 10 The ERA-I minus MODIS total Cloud Cover, monthly average for 2017. Color scale has been set to a zero mean difference with $\approx \pm 2\sigma$ limit. Statistics for each month are described in Table 2.

9 Appendix B: ERA-I CFLOS Minus CloudSAT CFLOS

Global altitude dependent differential CFLOS for ERA-I and Cloudsat. Lower altitudes show very close agreement with altitudes above 10 km yielding marginally larger differences due to ERA-I under representing the higher altitude cirrus clouds (Fig. 11).

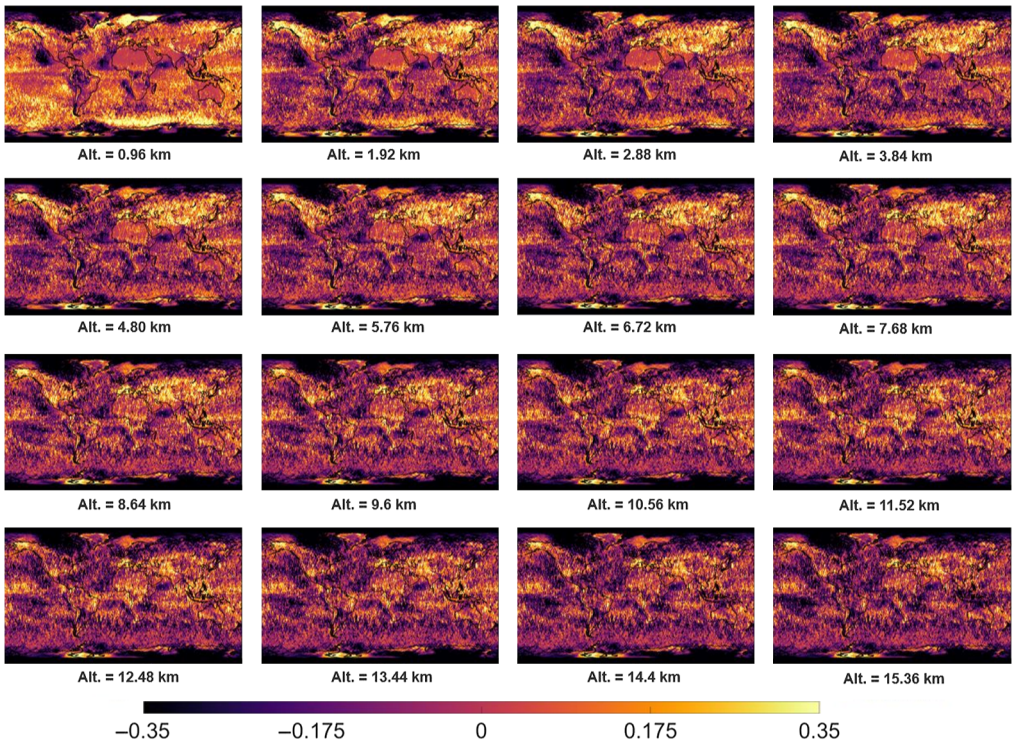

Fig. 11 The ERA-I CFLOS minus CloudSAT CFLOS as a function of altitude. Color scale has been set to a zero mean difference with $\approx \pm 2\sigma$ limit. Statistics for each altitude are described in Table 3.

Acknowledgments

The authors thank Don Reinke for supplying the CloudSAT/CPR CFLOS data for the month of April. The work presented here was funded by The Strategic Capabilities Office. The authors have no conflict of interests.

References

1. I. Lund and M. Shanklin, "Photogramettrically determined cloud-free lines-of-sight through the atmosphere," *J. Appl. Meteorol.* **11**, 773–782 (1972).
2. I. Lund and M. Shanklin, "Universal methods for estimating probabilities of cloud-free lines-of-sight through the atmosphere," *J. Appl. Meteorol.* **12**, 28–35 (1973).
3. I. Lund, "Persistence and recurrence probabilites of cloud-free and cloudy lines-of-sight through the atmosphere," *J. Appl. Meteorol.* **12**, 1222–1228 (1973).
4. J. Shields et al., "New cloud free line of sight statistics measured with digital whole sky imagers," *Proc. SPIE* **5891**, 58910M (2005).
5. J. Shields et al. "Measurement and evaluation of cloud free line of sight with digital whole sky imagers," in *The Battlepace Atmos. and Cloud Impacts on Mil. Oper. (BACIMO)*, NRL, Monterey, California (2005).

6. J. E. Shields et al., "Day-night whole sky imagers for 24-h cloud and sky assessment: history and overview," *Appl. Opt.* **52**(8), 1605–1616 (2013).
7. D. Nahrstedt, "Cloud modeling for laser weapon propagation analysis," *Proc. SPIE* **4034**, 1–13 (2000).
8. R. Venkat and D. W. Young, "Cloud-free line-of-sight estimation for free space optical communications," *Proc. SPIE* **8732**, 873205 (2013).
9. R. Alliss and B. Felton, "The mitigation of cloud impacts on free space optical communications," *Proc. SPIE* **8380**, 83800S (2012).
10. L. Yunying and S. Litan, "Calculation of probability of cloud-free lines-of-sight at given heights in foshan, china," *Adv. Atmos. Sci.* **16**, 297–303 (1999).
11. S. Yu, K. Case, and J. Chernick, "Methodologies and comparisons for lund's two methods for calculating probability of cloud-free line-of-sight," *J. Clim. Appl. Meteorol.* **25**, 389–397 (1986).
12. M. Free and B. Sun, "Time-varying biases in U.S. total cloud cover data," *J. Atmos. Ocean. Technol.* **30**, 2838–2849 (2013).
13. D. Reinke et al., "Probability of cloud-free line of sight (pcflos) derived from cloudsat cloud profiling radar and coincident lidar data," in *Preprints, 17th Conf. Satell. Meteorol.*, American Meteorological Society, Annapolis, Maryland (2010).
14. R. Stockli, "Nasa's Earth observatory - modis atmospheric imagery," https://neo.sci.gsfc.nasa.gov/view.php?datasetId=MODAL2_M_CLD_FR (accessed 2020/06/01).
15. P. Berrisford et al., "The era-interim archive version 2.0," 2011. https://www.ecmwf.int/node/8174 (accessed 2020/06/01).
16. D. Dee et al., "The era-interim reanalysis: configuration and performance of the data assimilation system," *Q. J. R. Meteorol. Soc.* **137**, 553–597 (2011).
17. A. Guillaume et al., "Horizontal and vertical scaling of cloud geometry inferred from cloudsat data," *J. Atmos. Sci.* **75**, 2187–2197 (2018).
18. K. Sassen and Z. Wang, "Classifying clouds around the globe with cloudsat radar: 1-year of results," *Geophys. Res. Lett.* **35**, 1–5 (2008).
19. H. Solomon, *Geometric Probability*, Society for Industrial and Applied Mathematics (1978).
20. G. Shmueli et al., "A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution," *J. R. Stat. Soc.: Ser. C (Appl. Stat.)* **54**(1), 127–142 (2005).
21. J. B. Kadane, "Sums of possibly associated Bernoulli variables: the Conway–Maxwell–binomial distribution," *Bayesian Anal.* **11**(2), 403–420 (2016).
22. C. Naud, J. Booth, and A. D. Genio, "Evaluation of the era-interim and merra cloudiness in the southern ocean," *J. Clim.* **27**, 2109–2124 (2014).
23. M. Stengel et al., "Comparing era-interim clouds with satellite observation using a simplified satellite simulator," *Atmos. Chem. Phys* **18**, 17601–17614 (2018).
24. M. Free, B. Sun, and H. Yoo, "Comparison between total cloud cover in four reanalysis products and cloud measured by visual observations at U. S. weather stations," *J. Clim.* **29**, 2015–2021 (2016).
25. D. Reinke, "Probability of cloud free line of sight," 2021, https://www.cloudsat.cira.colostate.edu (accessed 1 March 2022).

Adam Willitsford received his BS degree in electrical and computer engineering from Lafayette College (2002) and his PhD in electrical engineering from The Pennsylvania State University (2008). He is a member of the Senior Professional Staff at the John's Hopkins University Applied Physics Lab in the Optics and Photonics Group. In addition, he teaches fundamentals of lasers in the Engineering for Professionals program at John's Hopkins University. His research interests include active and passive remote sensing and imaging systems as well as statistical analysis of EO/IR atmospheric propagation.

Gregory Hicks was awarded his PhD in applied mathematics from North Carolina State University (2003). He is member of the senior professional staff the John's Hopkins University Applied Physics Lab and is the current supervisor of the Advanced Concepts Section of the Air and Missile Defense Sector's Guidance, Navigation, and Control Group. His current research is principally focused on geometric and algebraic control synthesis methods for multi-agent systems, formal methods, stochastic systems, and optimization.

Walt Bowen received his bachelor's and master's degrees in electrical engineering from Virginia Tech in 1986 and 1988, respectively. He is currently a member of the principle professional staff at the Johns Hopkins University Applied Physics Laboratory in the Guidance, Control, Navigation, and Seekers group in the Force Projection Sector. He is the chief engineer of this group. His work currently focuses on the integration of optical and RF sensors into flight guidance systems.