Development of Methodology and Tools for Determining the Impact of Cloud-Cover on Satellite Sensors

Luke A. McGuire
National Weather Center Research Experience for Undergraduates, Norman, Oklahoma
Bucknell University, Lewisburg, Pennsylvania

Hilary E. Snell and T. Scott Zaccheo
Atmospheric and Environmental Research, Inc.
Lexington, Massachusetts

ABSTRACT

Several studies have addressed the problem of optimizing field of view (FOV) size and sampling area of infrared sensors with the goal of achieving a higher percentage of cloud-free measurements. This study focuses on developing a tool to use global cloud analysis data in order to better understand the effects that different FOV sizes and satellite tracks have on the percentage of cloud-free measurements and the expected altitude of clouds that distort the signal of interest. This paper specifically discusses the situation of a satellite taking nadir measurements with a square FOV. The probability of a cloud contaminated measurement is estimated within 12-km grid boxes, making up a domain centered over the continental United States, using cloud fraction, cloud top altitude, and cloud base altitude values. The data confirms that the probability of a cloud contaminated FOV increases with an increase in FOV size. Compared to seasonal and diurnal variations, data suggests that FOV size has a relatively small effect on the expected value of cloud top and base altitudes. Increased understanding of factors effecting cloud contamination can improve scanning strategies and future satellite-based sensor designs.

1. Introduction

The effects of cloud contamination on measurements from satellite sensors, including “active” sensors such as LIDAR (Light Detection and Ranging), have been recognized as a problem for some time. Studies have shown that different scanning techniques and fields of view have an impact on the probability of obtaining a cloud free measurement (Smith 1996). The inability to obtain accurate sounding data through clouds leads to questions about the effect that satellite tracks, including the timing of measurements, and field-of-view size have on the percentage of cloud-free measurements. In this paper, techniques are presented to use cloud analysis data to estimate the probability of a cloud-free FOV, and the expected value of cloud top and base altitude given the event of cloud contamination, from a satellite taking nadir measurements with a square FOV. Several satellite tracks are analyzed to show the importance that FOV size and satellite path have on the percentage of cloud-free measurements.

The presence of clouds can significantly degrade the performance of satellite-based remote sensing devices by directly blocking the signal or by adding a component to the measured radiance that distorts the signal of interest. Studies have shown that optimizing the sampling area (SA) can reduce processing demands, simplify data assimilation, and improve overall performance of the sensor (Huang et all,

1996). Therefore, when simulating the performance of a satellite-based sensor it is important to consider the effects of clouds. In the past, to simplify processing, data from cloud contaminated footprints have not been used. However, recent efforts have been made to develop techniques which are able to accurately and efficiently utilize information, even if it is not obtained from a cloud-free measurement. (McCarty et al, 2007). In these cases, the altitude of the clouds is meaningful, as a lower cloud enables the majority of the atmospheric column to be sensed without cloud interference.

In general a smaller FOV should increase the percentage of cloud-free measurements, but it also has drawbacks, namely causing an increase in the signal-to-noise ratio. This trade-off of FOV size and signal-to-noise ratio has lead to studies with the focus of finding an optimal value for the FOV which will maximize the number of clear measurements (Smith 1996). Even though global cloud data is available from a variety of sources (e.g., MODIS and ISCCP data), it would be helpful, especially for sensor design studies, to have a tool that allowed one to change the FOV size of a potential sensor and try to optimize the FOV relative to how often the sensor would be taking cloud contaminated measurements. This paper discusses one approach which may be used to address this problem.

In section 2, the dataset is described and the methodology behind developing a tool to analyze the effects of cloud contamination on sensors with various FOV sizes is presented. Section 3 consists of the analysis of sample satellite tracks and general statistics relating the percentage of clear measurements to FOV size and time of day. This is followed by a discussion section, section 4.

2. Data/Methodology

We used data from geosynchronous and polar-orbiting satellites to provide knowledge of the effects of cloud cover with respect to potential designs of future satellite-based remote sensing instruments. These data are collected and processed by the Air Force Weather Agency and the final product includes a cloud analysis (cloud/no-cloud determination for a particular location), cloud fraction, and cloud top and base altitude, on an hourly basis. Global data are available, but the region chosen for this study is centered over the Continental United States and is broken up into 12-km grid boxes, or cells. January and July of 2003 are picked as representative months. The terms afternoon and morning hours refer, in general, to the time periods 3PM-5PM and 7AM-9AM local time. However, time zones within the domain were not taken into account with great detail. For simplicity, every 15° longitude was assumed to correspond to the time zone border.

A cloud fraction for three different layers, high, middle, and low, is associated with each cell as well as cloud top and base altitudes. Given restrictions on time of day and month, data from every hour and day within the desired interval were used to calculate an average cloud fraction for each cell layer. In regard to cloud top and base altitude, the average of all nonzero values is taken and later used to compute the expected value of a cloud top and base given a partially cloudy (a partially cloudy or completely overcast situation) FOV in a particular cell.

The probability of a clear FOV in one layer depends on two values. The first is the area in the cell that a FOV could be placed into while remaining cloud-free, which will be referred to as clear area. The second is the total area that a FOV could be placed into in that particular cell layer, which will be called total area. The total area is $144 km^2$ for the high layer, but for the middle and lower layers it is dependent upon the vertical and horizontal cloud structure of the higher layer(s) and will be discussed in more detail later. First, the amount of clear area in each layer is found as a function of FOV and the average cloud fraction in that given cell layer. Clearly a uniform distribution of cloud in small amounts throughout the cell would make the probability of a cloud-free FOV impossible. A completely random distribution would also be unrealistic. Therefore, the horizontal cloud distribution in each layer is estimated as an average of three different methods that place the

appropriate area of cloud or clear sky into the cell in varying formations. If the cloud fraction is less than fifty percent, then the cloud amount is distributed throughout the cell. (For numerical purposes, if the cloud fraction is greater than fifty percent the distribution is done with the proper amount of clear area). In the following examples it is assumed that the cloud fraction is less than fifty percent and the cloud amount is placed into the cell in different arrangements.

In this study, anything within one FOV length of the border between a cloud filled area and a clear area is added to the cloud amount for the cell because placing a FOV inside of that area would result in a contaminated measurement. Consequently, anything within one FOV of the estimated horizontal cloud formations is added to the cloud amount for that layer. The area added to the cloud fraction due to this horizontal distribution of cloud will be referred to as the variable H.

Method (1) involves computing the cloud amount within the grid cell from the cloud fraction and then distributing that area into the cell in the shape of a square. The perimeter of the square, representing the length of the border between clear sky and cloudy sky, is then calculated and multiplied by one FOV length, to obtain an approximate value for the area H. The same is done with a cloud formation consisting of two squares, whose total area adds up to the correct cloud amount for that cell, then one circle, and finally two circles.

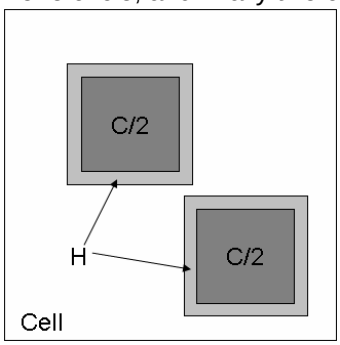

Figure 1: Illustrates method (1) with two square cloud formations inside the cell. The region denoted H is the area added to the cloud amount due to the horizontal distribution. Similarly, C denotes cloud amount.

In the case of circles, with radius R, the region H, is not calculated directly from circumference. Instead, it is determined by subtracting the area of a circle with radius R from the area of a circle with radius R+F, where F is the length of one FOV. The cloud amount is never broken up into more than two regions due to the size of each cell.

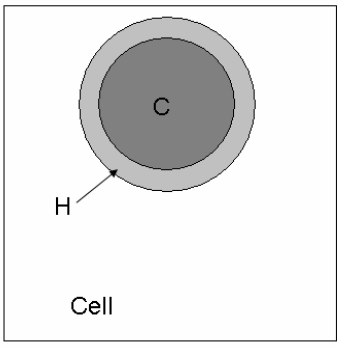

Figure 2: Demonstrates method (1) with one circular cloud formation inside the cell. The region denoted H is the area added to the cloud amount due to the horizontal distribution. Similarly, C denotes cloud amount.

The four values obtained from these calculations are averaged, and the result is recorded. In all cases, these cloud filled circles and squares are assumed to be at least one FOV length away from the cell's edges and any other cloud formations within the cell.

In method (2), the appropriate area of cloud is placed into the cell in the shape of a triangle with two sides of equal length positioned along the edge of the cell, leaving the hypotenuse as the only border between clear sky and cloud.

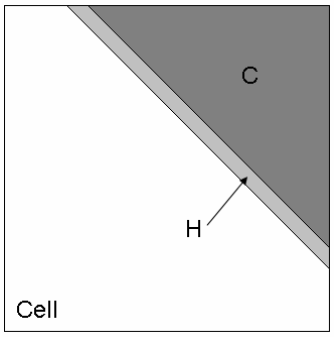

Figure 3: Illustrates method (2) with one triangular cloud formation inside the cell. The region denoted H is the area added to the cloud amount due to the horizontal distribution. Similarly, C denotes cloud amount.

Again, this value is multiplied by the FOV length and a measurement in the resulting area will be considered, from a probability standpoint, to be partially cloudy. A similar technique could have been used to determine H when considering square cloud formations, but is not used for simplicity purposes.

In method (3), the final result obtained from method (1) is divided by two to account for the possibility of border conditions with other cloud formations inside the cell or the cell's edges. It is a simple way to approximate the area H if the cloud formations discussed in method (1) were within two FOV lengths of each other or within one FOV length of the cell's border in such a way that it decreased. Depending on the size of the cell, the value that method (1) is divided by (in this study method (1) is divided by two) should vary. In general, the possibilities for horizontal distribution of the cloud amount within the cell should depend on the type of cloud formations that one might expect to see.

Finally, values obtained from the three different methods are averaged to provide an estimate for the amount of area inside the cell that should be added to the cloud amount. As noted above, if the cloud fraction for a given layer is greater than fifty percent, clear sky is distributed into the cell in various forms instead of doing so with clouds. In some instances, the result of adding the area H to the cloud amount is greater than $144km^2$. In these cases, the result is simply changed to $144km^2$ since it is clearly not possible to have a cloud fraction greater than 100%.

Vertical cloud distribution is approximated with four different methods similar to techniques discussed by Mace and Benson-Troth (Mace et al., 2002). Although their findings were geared toward estimates in atmospheric General Circulation Models, they are used in this study to accomplish a similar goal, achieving a better estimate of vertical cloud structure. The amount of vertical overlap is dependant upon several factors including the distance between two partially cloudy layers, and whether or not there is a clear layer separating two partially cloudy layers. However, in this dataset a layer is not defined as having a cloud fraction of zero unless there are no remaining partially cloudy layers. Here, the only variable used for determining vertical overlap is the distance between cloud layers: If the distance between layers, defined by subtracting the average cloud base altitude of the higher layer from the average cloud top altitude of the lower layer, is less than 2km then maximum overlap is assumed. If the distance between layers is between 2km and 4km, a 75% overlap is assumed. For a distance of greater than 4km between consecutive cloud layers, a random percentage of overlap is generated. This is based off of the statistics compiled by Mace and Benson-Troth that find the vertical overlap between cloud layers is best modeled by a random value as the distance between cloud layers increases (Mace et al., 2002). For simplicity, when determining the percentage of vertical overlap for the lowest layer, the distance between the high and low layers is used and the third layer is assumed to overlap as appropriate with both layers above it. If, in any case, the estimates of cloud overlap cause the probability of a cloud contaminated measurement to be greater than 1, then the probability is simply decreased to 1. This is not of great concern since the probability will only be greater than 1 if an assumption is made that sets the vertical overlap to a value that is less than the minimum.

Once estimates are made for cloud distribution within the cell, the overall probability of a cloud contaminated measurement can be calculated. Letting C_1 and C_2 be the cloud amounts in the high and middle layers respectively, V be the percentage of vertical overlap between the high and middle cloud layers, and A be the total area of the cell, the conditional probability of a partially cloudy FOV in the middle layer, p_2, given the event, L_1, that a clear FOV is maintained through the high layer, is broken up into two cases. If C_2 is greater than C_1 then,

$$P(p_2 \mid L_1) = \frac{C_2 - (C_1 * V)}{A - C_1}.$$

If C_2 is less than C_1 then,

$$P(p_2 \mid L_1) = \frac{C_2 - (C_2 * V)}{A - C_1}.$$

The conditional probability of obtaining a partially cloudy FOV in the lowest layer, p_3, given the event, L_2, that a clear FOV is maintained through the high and middle layers, is similarly determined with two cases. The percentage of the cell that is covered by clouds from the high and middle layers is:

$$R = C_1 + [C_2 * (1 - V)].$$

If C_3 is greater than R then,

$$P(p_3 \mid L_2) = \frac{C_3 - (R * V_2)}{A - R}.$$

If C_3 is less than R then,

$$P(p_3 \mid L_2) = \frac{C_3 - (C_3 * V_2)}{A - R}.$$

Satellite tracks are then given in terms of latitude and longitude points, from which the number of FOVs needed to progress through each cell in its path is determined. The probability of a clear FOV inside of a given cell is treated as an independent event in all cases since the area of a single FOV is always at least two orders of magnitude smaller than the area of the cell. A cloud-free FOV is the result with probability, P_{clr}, which is defined as

$$P_{clr} = (1 - \frac{C_1}{A}) * (1 - P(p_2 \mid L_1)) * (1 - P(p_3 \mid L_2)).$$

Since independence is assumed, the expected number of cloudy FOVs inside each cell is calculated using the expected value of the binomial distribution. Letting N be the total number of measurements taking in a given cell, the number expected to be cloudy is $N * (1 - P_{clr})$. The number of cloud contaminated FOVs across all cells are added together and compared to the total number of FOVs in the satellite track to determine the percentage of partially cloudy and clear FOVs. The expected value of cloud top and base altitude given the event of a partially cloudy FOV in a particular cell is calculated as a weighted average, depending on the probability of a partially cloudy FOV resulting from clouds in the higher, middle, and lower layers.

In order to begin to understand the impact of field-of-view size and time of day on the probability of achieving a cloud-free measurement, a set of sample satellite tracks. These are created by beginning at the following latitude and longitude points:

Track 1: 8.0° N, 150.0° W
Track 2: 15.0° N, 150.0° W
Track 3: 22.0° N, 150.0° W
Track 4: 29.0° N, 150.0° W
Track 5: 36.0° N, 150.0° W
Track 6: 8.0° N, 137.0° W
Track 7: 10.0° N, 120.0° W

The subsequent latitude and longitude points along the paths are then determined by incrementing the latitude value by 0.10° and decreasing the longitude value by 0.15°. For instance, the second point in satellite track 2 would be 8.1° N, 149.35° W. Due to the generic nature of the satellite tracks and for comparison purposes, only one measurement is assumed to be taken in each cell along the satellite paths.

3. Results

In all cases regarding the probability of a partially cloudy measurement, the difference between a 120m, 600m, and 1.2km FOV is clear. Varying FOV size during the afternoon hours in July resulted in the largest difference in the percentage of clear measurements, assuming the entire domain is sampled (Table 1). The smallest change occurs when each cell in the domain is sampled once during the morning or afternoon hours in January (Table 1).

A relatively linear increase in probability due to an enlarged FOV is reflected in Figures 4, 5, and 6, which show the probability of obtaining a cloud contaminated measurement along every cell in seven different sample satellite tracks. The graphical data averaged over the July afternoon hours indicates high probabilities of cloud contamination with a 1.2km FOV, especially in tracks 6 and 7. The data averaged over the same time period when a 120m FOV is used shows areas embedded

inside regions of high probability (of cloud contamination) that are noticeably more favorable for obtaining cloud free measurement than surrounding regions (Figure 7). With a 1.2km field of view, the same areas are not as well defined (Figure 8).

In general, FOV size did not appear to have a large impact on the expected value of cloud top and base altitude given the event of a partially cloudy measurement, although there are small differences suggested in all cases. In comparing figures 10 and 11, showing expected cloud top altitudes for 120m and 1.2km FOVs in January, small changes are evident in track 1 and the northeast portion of track 7. Several interesting seasonal trends included a sharp increase in the expected value of cloud top and base altitude from January to July, especially during the afternoon hours (Figures 9 and 10).

FOV	January (AM)	January (PM)	July (AM)	July (PM)
120m	.577	.556	.547	.544
600m	.499	.476	.461	.455
1.2km	.399	.378	.355	.347

Table 1: Percentage of cloud-free FOVs expected to be achieved if the entire domain is sampled

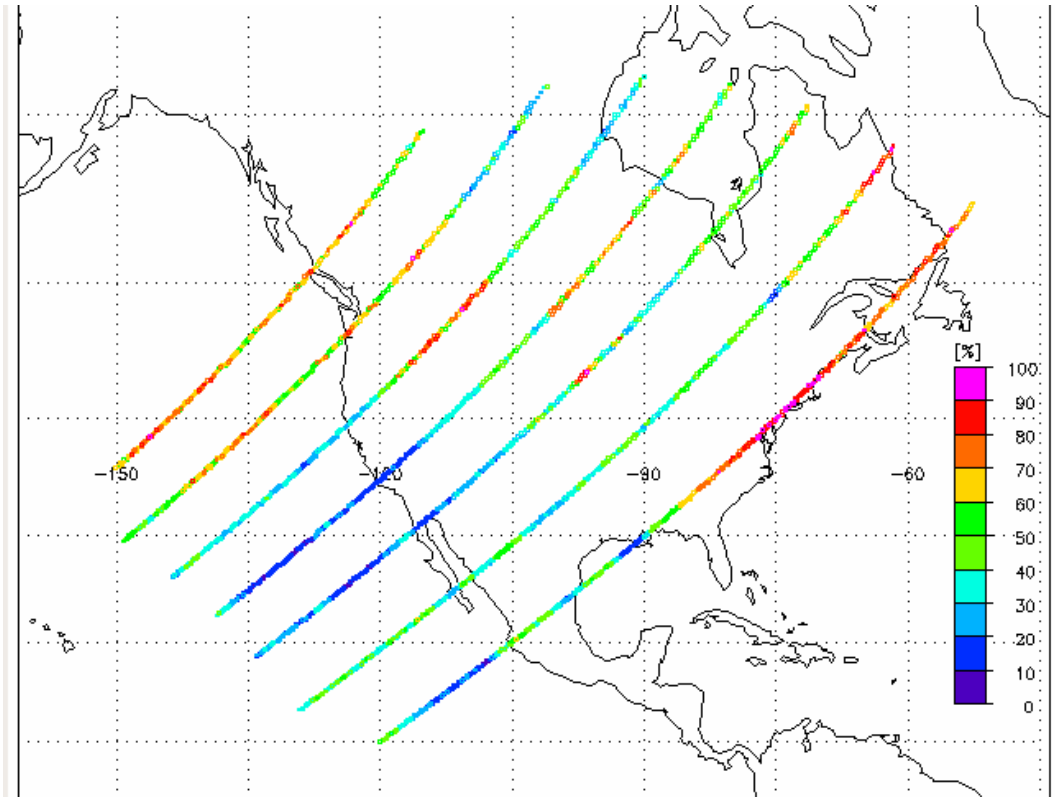

Figure 4: Shows the estimated probability of obtaining a partially cloudy 120m FOV. Data averaged over January morning (7AM-9AM) hours to determine the probability of a cloud contaminated FOV in every cell along the sample satellite paths shown above.

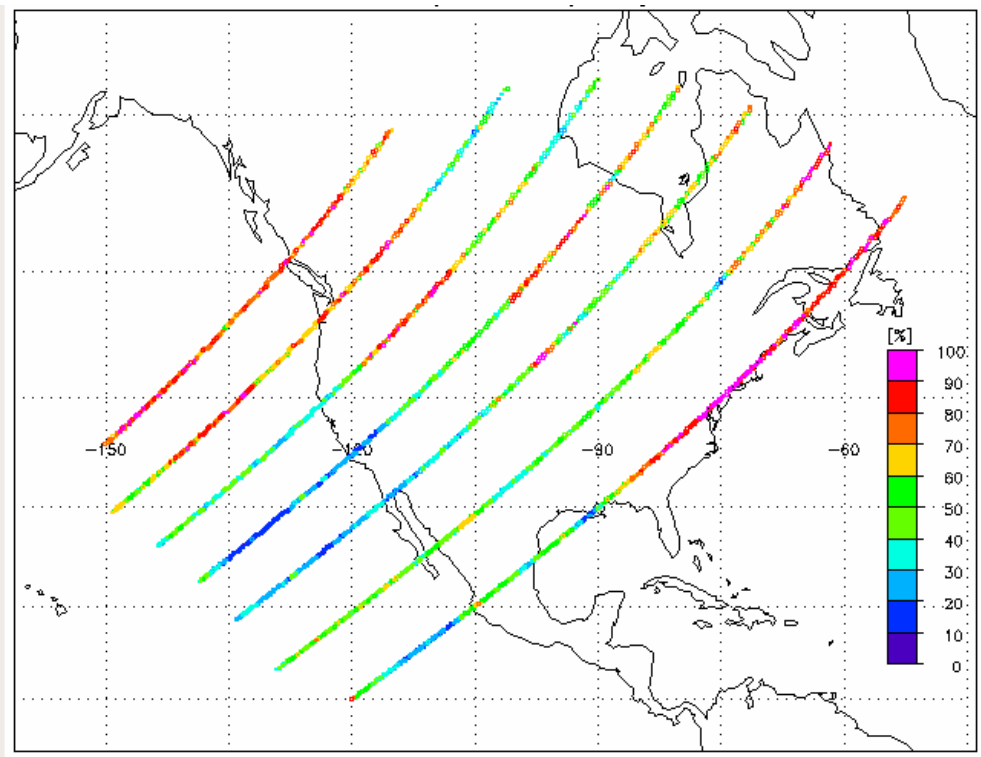

Figure 5: Data averaged over January morning (7AM-9AM) hours to determine the probability of obtaining a partially cloudy 120m FOV in every cell along the sample satellite paths shown above.

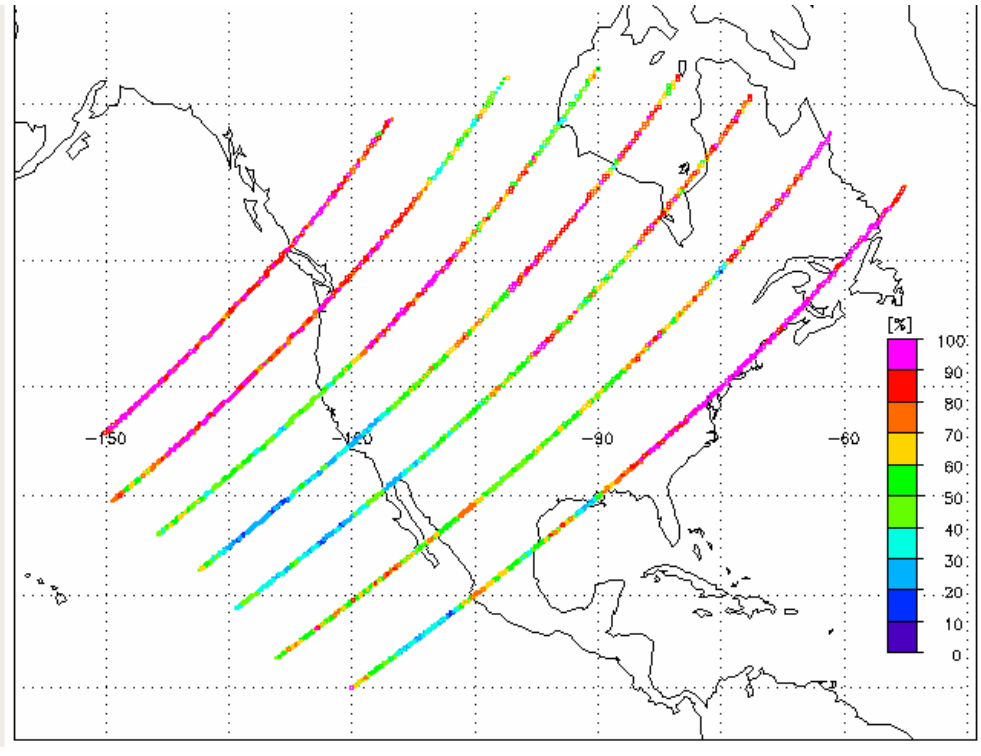

Figure 6: Data averaged over January morning (7AM-9AM) hours to determine the probability of a partially cloudy 120m FOV in every cell along the sample satellite paths shown above.

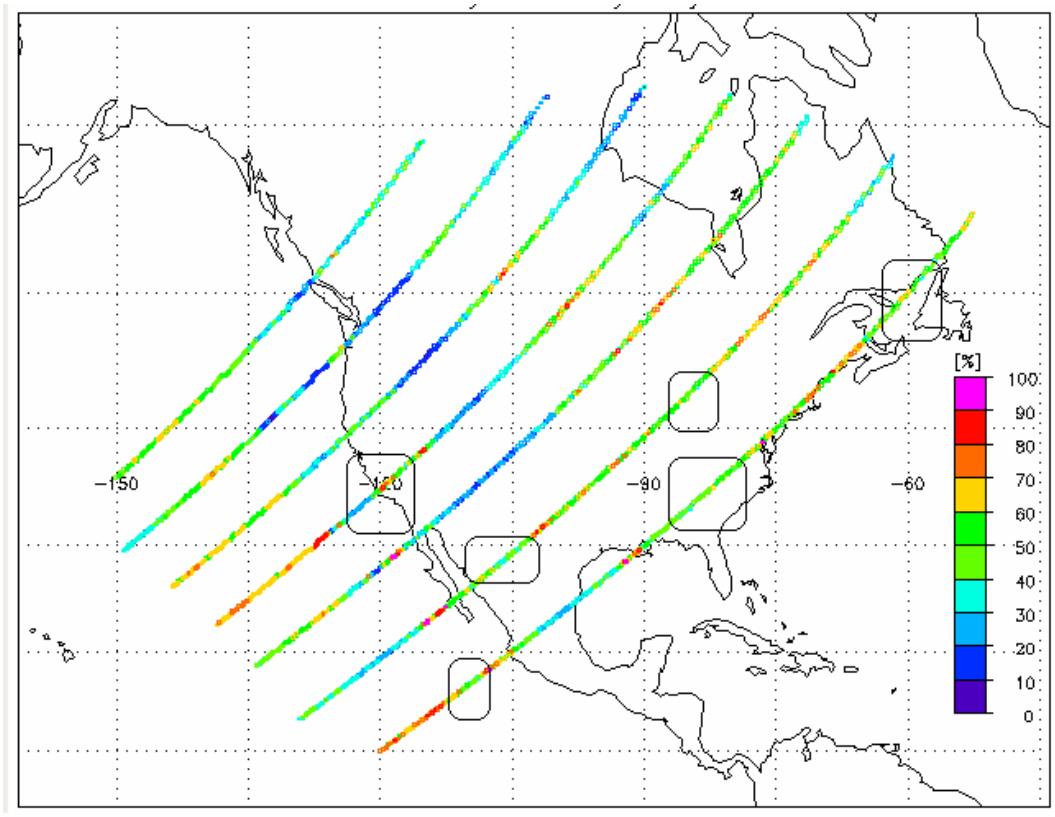

Figure 7: Tracks 1-7 are pictured above with along with the probability of obtaining a partially cloudy 120m FOV provided for every cell along each satellite track. Track 1 is the farthest to the northwest and track 7 is the farthest to the southeast. Several areas are pointed out where probabilities for clear measurements are relatively good compared to neighboring regions.

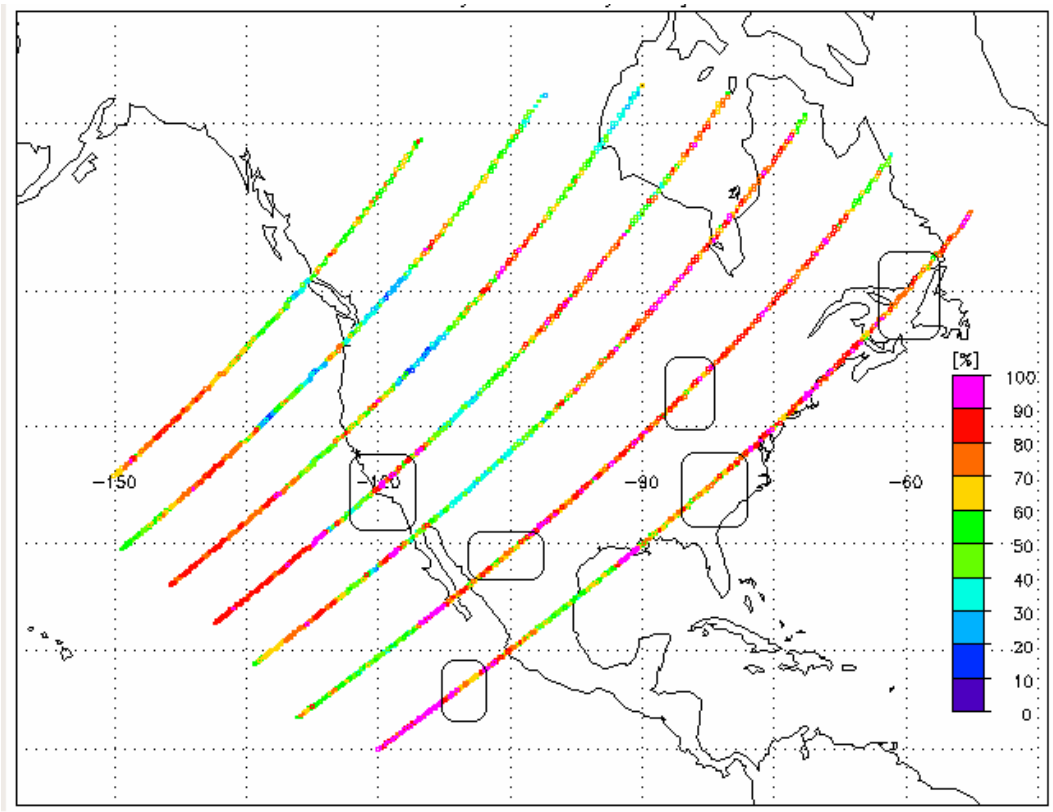

Figure 8: Shows tracks 1-7 and indicates the probability of obtaining a partially cloudy 1.2km FOV provided for every cell along each satellite track. Track 1 is the farthest to the northwest and track 7 is the farthest to the southeast. The same areas are pointed out as in Figure 4 for comparison purposes.

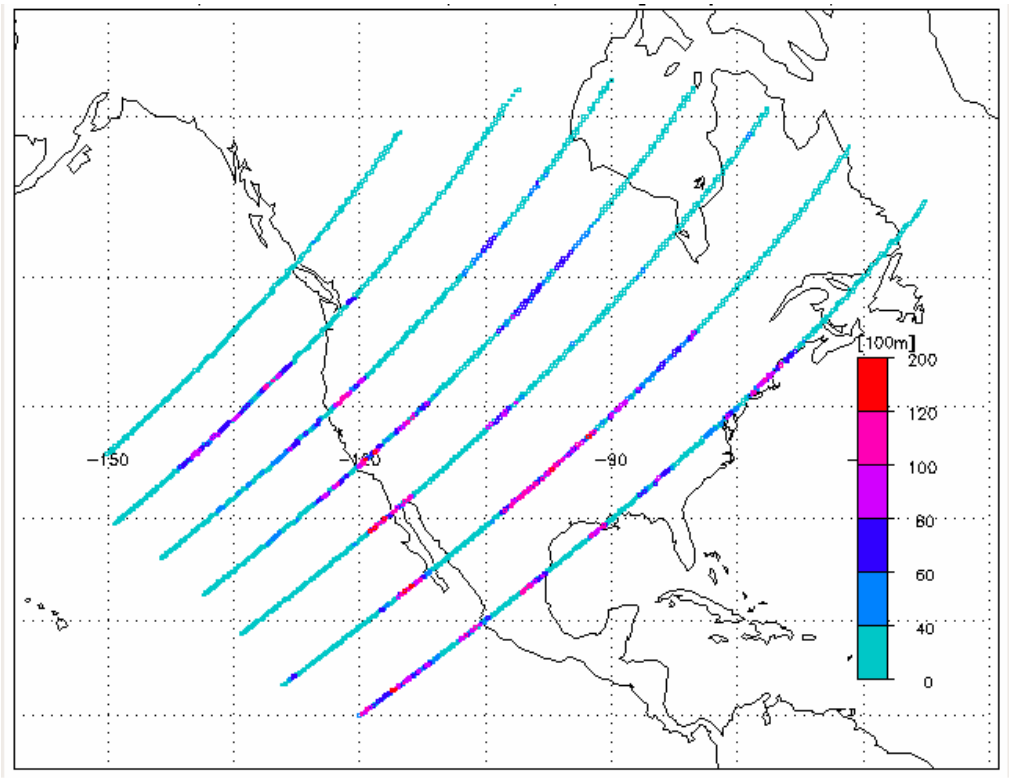

Figure 9: Expected value of cloud top altitude using a 1.2km FOV. Calculated from data averaged over the afternoon hours (3PM-5PM) of July.

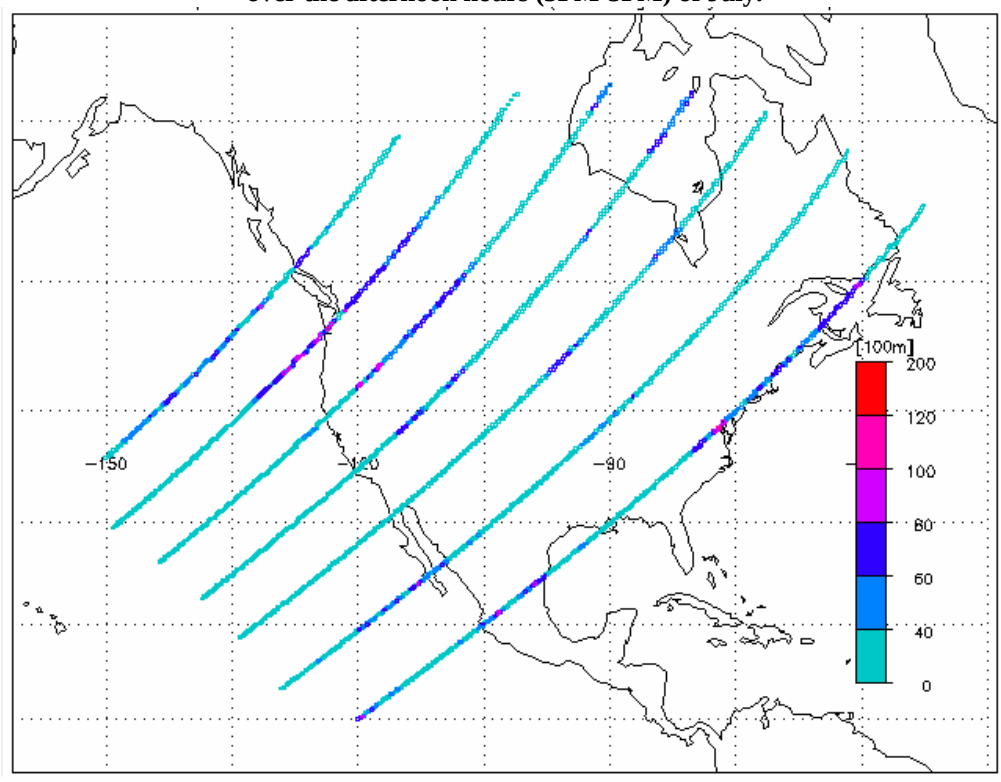

Figure 10: Expected value of cloud top altitude using a 1.2km FOV. Calculated from data averaged over the afternoon hours (3PM-5PM) of January.

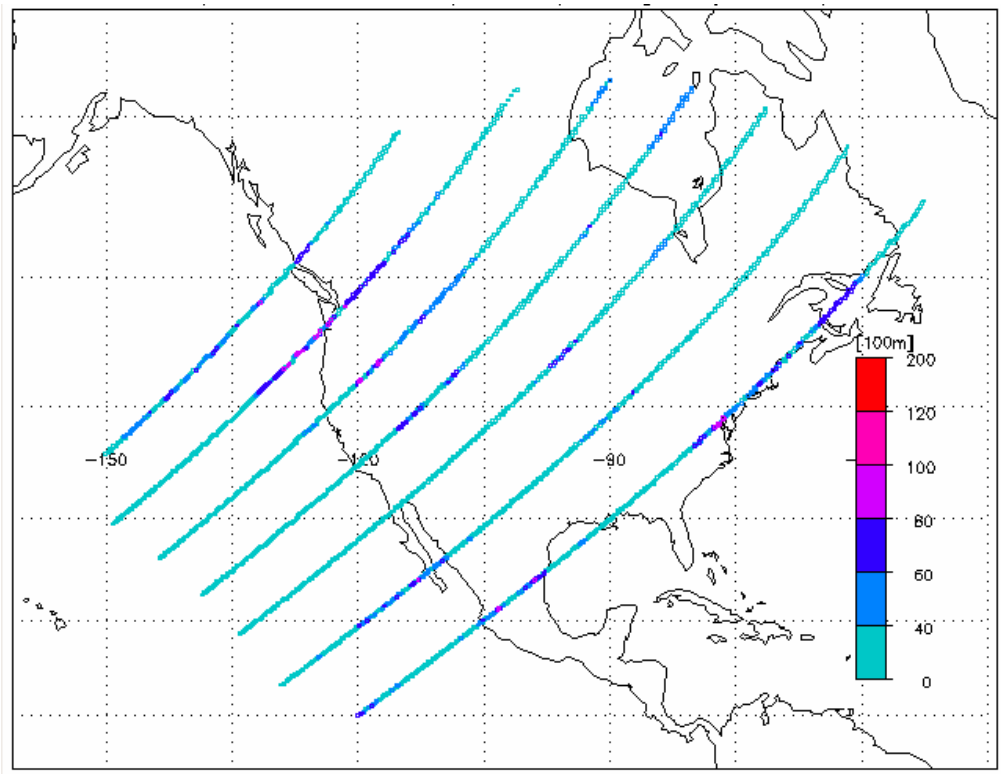

Figure 11: Expected value of cloud top altitude assuming a partially cloudy 120m FOV is obtained. Calculated from data averaged over the afternoon hours (3PM-5PM) in January.

4. Discussion

Although data (Table 1) suggests the relationship between FOV size and the probability of a cloud-free measurement is approximately linear, there are several issues to consider including the methods used to estimate horizontal cloud distribution. Recall that (assuming the cloud fraction is less than 50 percent) the cloud amount for a particular layer is distributed within the cell in the shape of either a triangle or one or two circles or squares. Anything within one FOV of these cloud formations is referred to as H and is added to the cloud amount for that layer. In the case of a square, H is approximated by multiplying the perimeter of the square by the FOV length, thus setting up a linear relationship between FOV size and the amount of clear area in that cell layer. The same is true in the case of a triangle (method 2). With a circle, however, the relationship between FOV size and H is not linear due to the fact that H is determined by subtracting the area of a circle with radius R from that of a circle with radius R+F, where F is the FOV length. Since the relationship between the radius and area is not linear, neither is the relationship between FOV size and H.

In reality, it is expected that the perimeter of a cloud formation is not linearly related to its area and this idea suggests the relationship between FOV size and the probability of a cloud contaminated measurement is also not linear. When a non-linear relationship between FOV size and H is accentuated throughout multiple layers, especially if there is minimal overlap, it is expected that a more non-linear trend would develop.

Although the data did not suggest, on a large scale, dramatic differences in the expected value of cloud top and base altitudes caused by enlarged FOV sizes, there were definite contrasts between the July and January data. It is important to note that these differences exist and could indicate that the amount of the atmosphere that is able to be accurately measured by a

sensor is dependent upon the time of year, among other variables. Also, there are areas, including those pointed out in figures 6 and 7, where it appears inefficient to 'hole-hunt' for clear measurements unless a small (120m) FOV is used. It appears that the model is able to resolve important trends in probability of achieving clear measurements and expected values for cloud top and base altitude. This is necessary if it is to be used in the future to aid in the evaluation of satellite-based sensor designs.

This tool uses hourly, global cloud cover information to provide an assessment of the impact of clouds associated with different sensor fields-of-view and orbit parameters, such as time of day and measurement repeat time (how often the sensor will view the same location). The current software and processing algorithm is limited to nadir-viewing sensors. Future improvements include the addition of an off-nadir capability to simulate scanning sensors and to account for FOVs that do not fall completely within any one cell. Being able to identify situations where a FOV is located in multiple cells is essential if accurate evaluations of larger FOV sizes are to be made without increasing the size of the grid boxes in the domain. Near-term research efforts will focus on evaluating cloud cover probabilities over the course of a year for a realistic satellite track.

Acknowledgements: This material is based upon work supported by the National Science Foundation under Grant No. ATM-0648566. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The first author would like to thank mentor Ned Snell for all of his help and advice and Scott Zaccheo for his work in making this project possible.

References

Huang, H., R. Frey, W. L. Smith, D. K. Zhou, H. Bloom, 2003: Defining Optimal Spatial Resolution for High-Spectral Resolution Infrared Sensors

Mace, G. G., and Benson-Troth, S. , 2002: Cloud-Layer Overlap Characteristics Derived from Long-Term Cloud Radar Data: *Journal of Climate*, **15**, 2505-2515.

McCarty, W, Jedlovec, G. J., Molthan, A. L., LeMarshall, J. F., 2007: An Investigation of the Characteristics of Cloud Contamination in Hyperspectral Radiances. Preprints, *The 87th AMS Annual Meeting*, San Antonio, TX, American Meteorological Society.

Smith, W.L., Huang, H. L., and Jenney, J. A., 1996: An Advanced Sounder Cloud Contamination Study: *Journal of Applied Meteorology,* **35**, 1249-1255.