

Randall Alliss*, Robert Link, and Mary Ellen Craddock
Northrop Grumman Information Technology TASC, Chantilly, Virginia

ABSTRACT

NASA Jet Propulsion Laboratory (JPL) is interested in adding optical communications to its deep space communications network. Clouds adversely affect the transmission of optical communications; in order to mitigate the effects of clouds and achieve reliable communications, a geographically diverse set of ground receiver stations is needed. To study cloud effects on optical communications we have developed a high-resolution cloud climatology based on NOAA Geostationary Environmental Operational Satellite (GOES) imager data. The GOES imager includes multi-spectral channels, one visible and four infrared, at 4-km spatial resolution and 15-minute time resolution. Cloud detection is accomplished by modeling the radiance of the ground in the absence of clouds and comparing the actual radiance values from the imagery. A composite cloud decision is formed by objectively combining the results of the tests from the individual channels. Ground site selection studies are accomplished using the Lasercom Network Optimization Tool (LNOT). LNOT applies a discrete optimization algorithm to the cloud climatology dataset to find the optimal number and locations of ground stations for a given concept of operations. Applying LNOT to the JPL problem we find that 90% availability could be achieved with 4–5 ground stations in the continental US and Hawaii. We also present the results of a pilot study that includes 6 months of cloud data over South America.

Keywords: Optical communications, cloud detection, site selection studies

1. INTRODUCTION

Future deep space probes are expected to generate ever larger volumes of data, making high-bandwidth data links necessary to return the data to analysts on earth. Free-space laser communications provide an attractive option for achieving the necessary bandwidth without imposing unfeasible power requirements on the

spacecraft. However, unlike radio communications, optical communications are interrupted by clouds. Therefore, any high availability laser communications system must include a strategy for ensuring a Cloud-free line of sight (CFLOS).

Even the most cloud-free locations are cloudy about 30% of the time, so achieving system availabilities higher than 70% requires a mitigation strategy.** The most effective strategy is “site diversity”, which is having redundant sites so that if one is clouded out another can be used as a backup.

The availability gained by site diversity depends greatly on the number and location of the sites in the network. For example, large-scale meteorological patterns ensure that sites within a few hundred kilometers of one another will have highly correlated occurrence of clouds. Therefore, stations should be placed far enough apart to avoid correlations, even if that means foregoing sites that are individually better than the sites that are less correlated. At the same time, stations need to be close enough to give significant overlap in their probe coverage. The purpose of the calculations presented here is to find networks of ground stations that meet the JPL performance benchmark of 90% availability with the fewest possible stations.

There are further constraints that the JPL concept of operations imposes on the system. At low elevation angles the increased air mass in the line of sight makes a lasercom link untenable; consequently, we consider a station to be available for communication only if the elevation angle of the probe exceeds a minimum elevation angle, taken to be 20° in this study. During the daytime, background sky brightness also interferes with establishing a lasercom link. The background of scattered light can be reduced by restricting eligible ground station sites to locations at high altitudes. The reduced pool of eligible sites exacts a toll in availability; we compare the availabilities for minimum altitudes of 0 km (i.e., no restriction), 1.5 km, and 2 km. Finally, JPL has a number of sites that are preferred as ground station locations because of preexisting infrastructure. We examine the availability of networks constrained to include one or more of the preferred sites.

In the remainder of this paper we discuss the data used to perform this study and the methods and results of the study itself. Section 2 describes the satellite data and the automatic cloud analysis used to process the data into a database of cloud decisions. Section 3 explains the optimization process by which high-availability networks of ground stations are identified. Section 4 gives the results of applying these methods to the JPL problem and discusses their implications.

*Corresponding author address: Randall Alliss, Northrop Grumman Information Technology TASC, 4801 Stonecroft Blvd., Chantilly, VA 20151; e-mail: randall.alliss@ngc.com.

**For our purposes “availability” refers to the fraction of time that a CFLOS is available to the system. Other sources of system outage (e.g., mechanical failures) could further reduce the fraction of time the system is usable for communication.

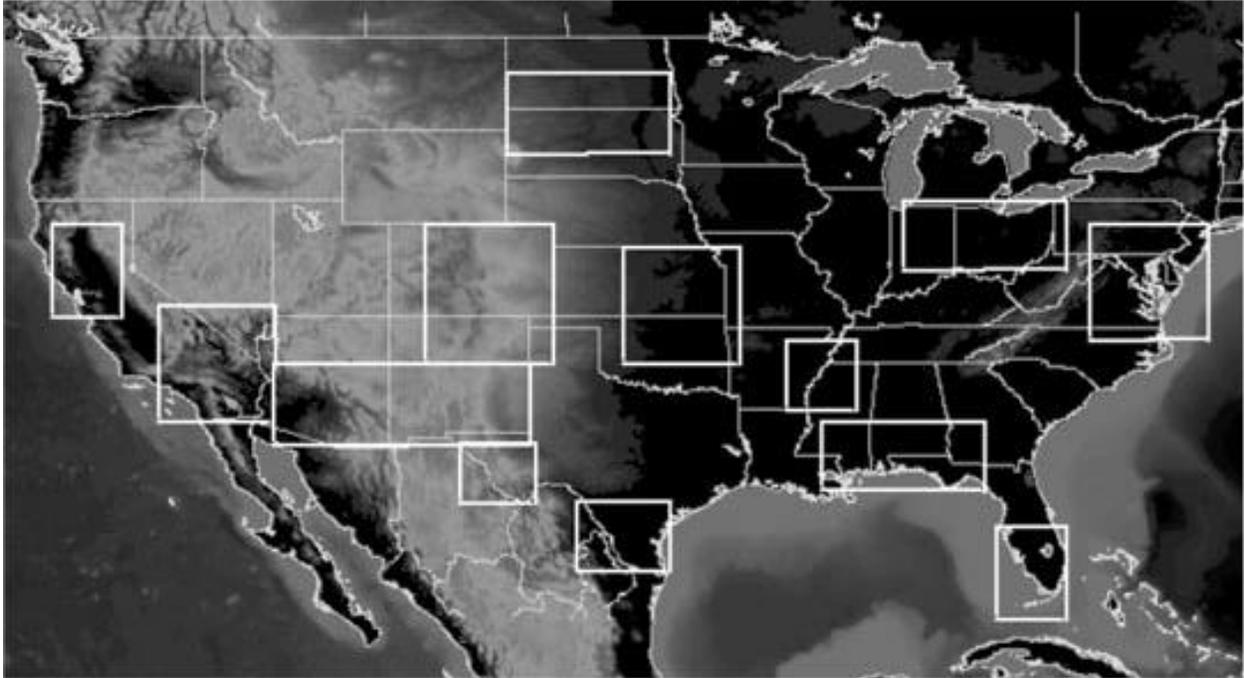


Figure 1. The continental US regions in which cloud analysis was performed for this study. Cloud analyses were also performed over Hawaii.

2. CLOUD ANALYSIS

The cloud analysis is performed in selected regions (fig. 1) across the continental United States and Hawaii. The data are available at 30-minute time resolution from 1997–1999 and at 15-minute resolution from 2000–present. In order to evaluate the benefit of including southern hemisphere ground stations in the network we began collecting and archiving data over South America in March 2003. Cloud analysis over South America was performed in the regions depicted in figure 2.

2.1 Cloud Tests

We derive cloud analyses from the NOAA GOES imager data using the algorithms described by Alliss et al. (2000). The GOES imager has 5 bands: visible ($0.6\ \mu\text{m}$), shortwave infrared ($3.9\ \mu\text{m}$) (SWIR), water vapor ($6.7\ \mu\text{m}$), longwave infrared ($10.7\ \mu\text{m}$) (LWIR), and split window ($11.2\ \mu\text{m}$). The water vapor channel, is not used for cloud detection and is replaced by a multispectral fog product at night, and a shortwave reflectivity product during the day. The resolution of the visible band is 1 km, and the other bands are at 4 km. In the cloud detection algorithms the 1 km data is resampled to 4 km so that it may be readily combined with the data from the other bands. All of the cloud tests are made by comparison to a dynamically computed clear sky background (CSB), which we describe in section 2.2.

The visible channel is used when the solar zenith angle is less than 89° ; however, for solar zenith angles between 89° and 81° cloud detections in this band are deweighted, due to the low signal-to-noise ratio when

the scene is illuminated at low solar elevations. If the calculated albedo exceeds the CSB by a predefined threshold the pixel is deemed cloudy. Conversely, if the albedo is less than the CSB by more than the threshold, the pixel is deemed clear (i.e., cloud detections from other tests may be negated).

The LWIR is used directly in a cloud detection test, in addition to being used in the multispectral tests. A pixel is considered cloudy if the LWIR CSB for the pixel exceeds the LWIR temperature by a predefined threshold. Unlike the visible and multispectral tests, the LWIR test is usable at any time of day.

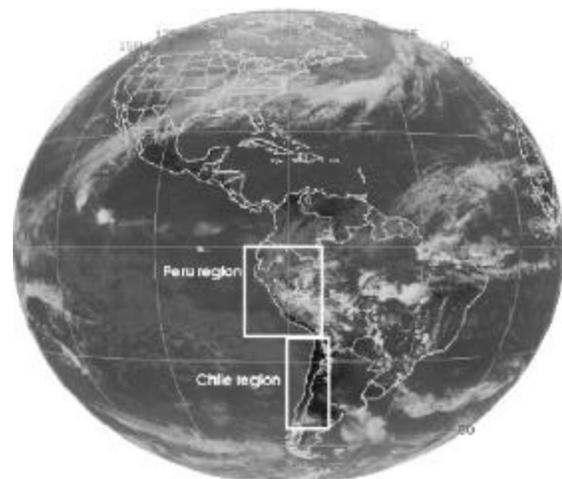


Figure 2. The South America regions in which cloud analysis was performed for this study.

The fog product is calculated as the difference between the LWIR and SWIR brightness temperatures (Ellrod, 1995). The emissivity of water clouds in the SWIR is lower than in the LWIR; therefore, low clouds produce colder SWIR temperatures, resulting in $T_{LW} - T_{SW} \gtrsim 2K$ (Lee et al., 1997). (The exact threshold is determined by the clear-sky background model described below.) This product can detect clouds that have LWIR temperatures too similar to the ground temperature to be detected by the LWIR alone. The fog product is also useful for detecting high ice clouds. These clouds are transmissive and therefore appear warmer in the SWIR, resulting in $T_{LW} - T_{SW} \lesssim 5K$ (Lee et al., 1997). Because the SWIR is dominated by reflected radiation during the day, the fog product is usable only at night.

The shortwave reflectivity product is calculated by subtracting the thermal component from the SWIR, leaving only the reflected solar component (Allen et al., 1990; Setvak and Doswell, 1991). Because water clouds are highly reflective in the SWIR, while ice is poorly reflective in the SWIR, the reflectivity product can readily distinguish between low clouds and snow cover. Absent the reflectivity product, the visible channel could misidentify the latter as cloud.

2.2 Clear Sky Background

The threshold tests described above require knowledge of the CSB, the radiation received by the GOES sensor in the absence of clouds. This background can be reflected, emitted, or a combination of both. The reflective and emissive properties of the ground vary from place to place; therefore, using fixed thresholds in the cloud tests will produce faulty cloud decisions in some places. For example, an albedo threshold tuned to detect clouds over "typical" terrain will consistently produce spurious clouds over the highly reflective surface of White Sands, NM. Similarly, seasonal variations in ground temperature will affect the LWIR background. Terrain height, soil moisture, and illumination angle also affect the CSB. In order to account for these differences the CSB must be modeled separately for each pixel at each time (Alliss et al., 2000).

In order to minimize the effects of diurnal cycles, the CSB is processed using data from the previous 30 days at a single analysis time (e.g., 1200 UTC). This scheme isolates most of the diurnal variation in temperature and illumination. A separate CSB is calculated for each band or multispectral product in use at the particular analysis time: LWIR, visible, reflectivity product, and fog product.

The albedo CSB is the average of the darkest ten percent of albedo values from the previous 30 days for the pixel being analyzed. The 30-day data window represents a compromise between making the sample large enough to be likely to include several clear observations and making the sample small enough to be sensitive to seasonal variations.

The reflectivity CSB is calculated using the darkest ten percent of reflectivity product values from the previous 30 days. The calculation is in other respects similar to the calculation for the albedo CSB.

The fog product CSB is calculated by identifying the warmest 10 percent of LWIR values for the pixel over the previous 30 days. The fog product values for the selected times are averaged to form the fog product CSB. This procedure differs from the albedo and reflectivity versions (which choose clear pixels based on the albedo and reflectivity themselves) because both extremes of the fog product values indicate clouds.

The LWIR CSB is determined with the aid of the LWIR regression model, in which each pixel's LWIR temperature is estimated using a linear regression model. The regression model is populated with prototypical clear sky pixels from the entire analysis region. These prototypes are chosen using a series of tests that detect only pixels that have a high probability of being clear (i.e., even without the benefit of thresholds from the regression modeling they are clearly cloud-free.) We use the prototype pixels to fit coefficients of a linear regression model with twelve predictors, including pixel level data from the GOES imager, regional data from the NWS surface reports, time, and terrain.

The LWIR regression model is used to estimate the clear sky LWIR brightness temperature in each pixel. The differences between the regression model temperature and the measured GOES LWIR temperature are the LWIR residuals. The warmest ten percent of the LWIR residuals are averaged to obtain the LWIR residual CSB that is used in the LWIR cloud test.

Figure 3 shows an example of a cloud analysis from 11 Dec 2003 at 17:15Z. The scene is in the daytime, so visible, LWIR, and reflectivity products are in use. The algorithms detect the widespread cloud cover over Delaware, Maryland, and Pennsylvania, as well as the small-scale clouds associated with lee waves observed over the Appalachian Mountains.

3. NETWORK OPTIMIZATION AND EVALUATION

Achieving a high-availability communications link will require a network of ground stations, both to provide longitude coverage and to mitigate against clouds. It is not sufficient simply to choose networks composed of locations that have individually high cloud-free fractions. Incidence of clouds is typically correlated between locations, even over relatively large areas. Networks including such locations will achieve much less availability than might be expected from the cloud-free fractions of the stations alone (fig. 4). Widely separated ("geographically diverse") locations will tend to be less correlated, and microclimate effects can produce locations that are slightly anti-correlated. The task of network optimization is to identify such combinations of

locations, in order to produce networks with the highest possible availability.

Finding an optimal ground station network is a discrete optimization problem; there are more than 250 thousand pixels in the cloud database, admitting over 4×10^{21} four-site networks. The JPL constraints (§ 1) reduce this number significantly, but the number of networks is still too large to search exhaustively for the network with maximum availability. Therefore, the optimization algorithm we adopt must be able to find the desired networks after searching only a small fraction of the network configuration space.

There are two properties that are particularly desirable in a discrete optimization algorithm. First, the configuration space should be defined such that good configurations (as measured by the figure of merit function adopted for the problem) lie close together in the space. This property allows the algorithm to make progress; without it the one-thousandth guess is no better than the first. Second, the search algorithm should resist getting trapped in local extrema in the configuration space. We will refer to the first property as “locality” and the second as “robustness”.

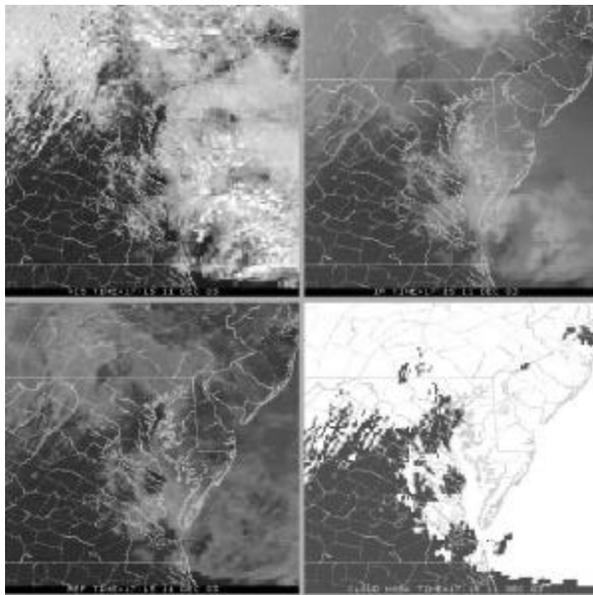


Figure 3. An example cloud scene centered over northern Virginia, showing the visible (top left), LWIR (top right), reflectivity product (bottom left), and the composite cloud decision mask. The algorithms detect the widespread cloud cover over Delaware, Maryland, and Pennsylvania, as well as the small-scale clouds associated with lee waves observed over the Appalachian Mountains.

In general, and for this problem in particular, there is a trade-off between locality and robustness. Consequently, we divide our optimizations into two stages. The first stage is performed using an algorithm that sacrifices some locality in favor of robustness, while the second

stage is performed using an algorithm that sacrifices robustness in favor of locality. The strategy is to use the first algorithm to search widely over the configuration space. Once we believe we are in the vicinity of the solution we switch over to the second algorithm to find the best configuration in the neighborhood of the last configuration found by the first algorithm. In the second stage we are not concerned about the algorithm’s lack of robustness, since we expect the optimal configuration to be close by.

A typical optimization run evaluates the availability of over 40 million networks. These calculations make several simplifying assumptions in order to speed up the individual availability calculations. Once the optimization algorithm has identified a small number (10–20) of candidate networks, a more comprehensive evaluation is performed for each network. This evaluation includes a detailed line-of-sight calculation including the effects of parallax between the GOES imager and the probe, complete network status at every time period, statistical calculations of uncertainties, intra-network correlations, serial correlations, and distributions of outages. The results of these calculations are what is quoted as “the availability” of the networks in section 4.

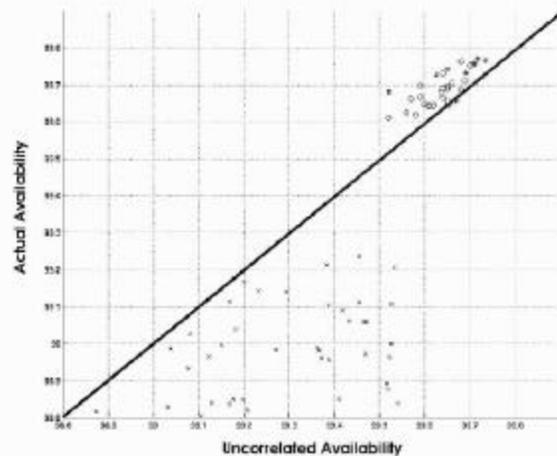


Figure 4. Comparison between the availability calculated by assuming no correlation between the stations in the network, and the actual availability achieved by the networks. The \times symbol marks randomly chosen networks; the $^{\circ}$ and $*$ symbols mark networks returned by the optimization algorithm. Most networks significantly under-perform the availability expected absent correlations, but a few outperform the no-correlation case.

The optimization calculations require the position of the probe as a function of time in order to determine whether the probe is at high enough elevation angle, as seen from a ground station, to establish a communications link. The evaluation calculations also require the position of the probe, both for this reason and in order to account for the parallax between the GOES imager and the probe. The line of sight to the probe is calculated by assuming the probe to be in a circular orbit at 0° inclination to the ecliptic, with a radius of 1.5237 AU.

This orbit is close enough to the orbit of Mars for our purposes and is much faster to calculate than an elliptical, inclined orbit.

A further consideration is that our cloud analyses are derived solely from the GOES East and GOES West satellites. Notionally, the JPL ground station network would consist of stations distributed around the full 360 degrees of longitude; however, without cloud analyses in areas not covered by the GOES satellites, we cannot determine the disposition of stations in those areas. It is not sufficient to restrict the eligible station locations to areas covered by GOES imagery because such networks will have their availability penalized for all the times when the probe is below the horizon for the entire network. This is not an accurate measure of availability because probe coverage at these times would be provided by other stations in the regions without cloud analyses. Therefore, we restrict our calculations to those times of day when North and South America and Hawaii have exclusive coverage of the probe; i.e., times when the probe is below the minimum elevation angle for notional ground stations in Europe and Australia. Times when North and South America have nonexclusive coverage or no coverage at all are counted neither for nor against the availability of the system. This results in an average of 6 hours of data used in the availability calculation per day in the period of record.

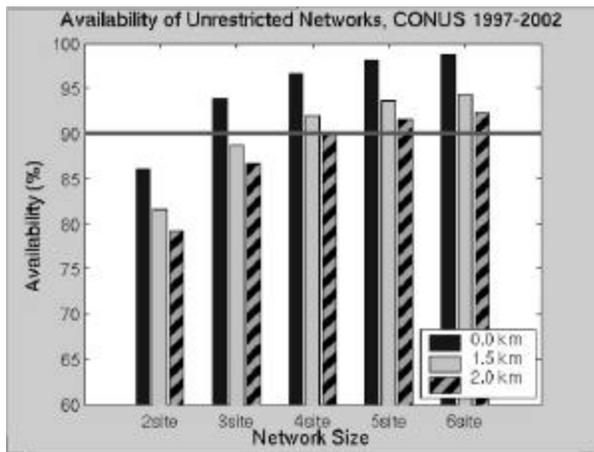


Figure 5. Availabilities for networks with sizes ranging from two to six sites. In each case the availability is shown for the best networks with 0 km, 1.5 km, and 2 km altitude constraints. The heavy horizontal line highlights the 90% availability requirement.

4. RESULTS

4.1 6-Year Continental US and Hawaii Data

The LNOT algorithms were applied to the 6-year cloud decision database covering the continental United States and Hawaii. The minimum elevation angle for communication was 20° , and sites were permitted to be located anywhere within the cloud analysis regions (fig. 1), subject to the altitude constraint. Figure 5 shows a comparison of the availability results for minimum site

altitudes of 0 km, 1.5 km, and 2 km, for network sizes ranging from two to six sites. We see that for the 2 km case five sites are required to achieve the 90% availability requirement, while the 1.5 km and 0 km case require four and three sites, respectively.

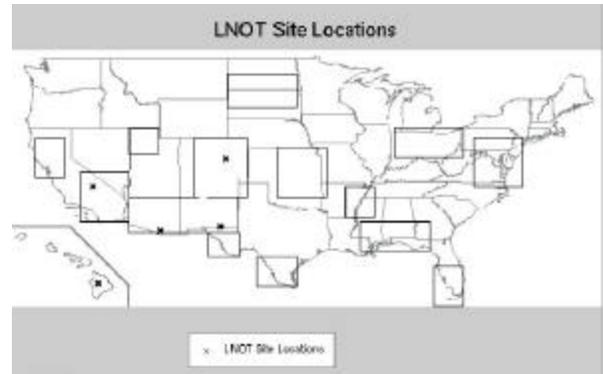


Figure 6. Locations of stations for the best 5-site network meeting the 2 km altitude constraint. The eastern analysis regions have no sites meeting the altitude constraint, and are therefore effectively excluded from the calculation.

Figure 6 shows the station locations for the best 5-site network with a 2 km altitude constraint, and figure 7 shows monthly average availabilities for the same network. Although the network meets the 90% requirement on average, the network experiences three periods of significantly reduced availability during mid-1998, early 2000, and late 2001–early 2002. These periods correspond to the times when the probe is at its southernmost declination. During these times the daily period of coverage for northern hemisphere stations is shortest; consequently, there is less overlap in coverage between stations, and therefore less opportunity for site diversity to increase availability. Sites in the southern hemisphere would have better probe coverage at these times, which would boost the availability considerably.

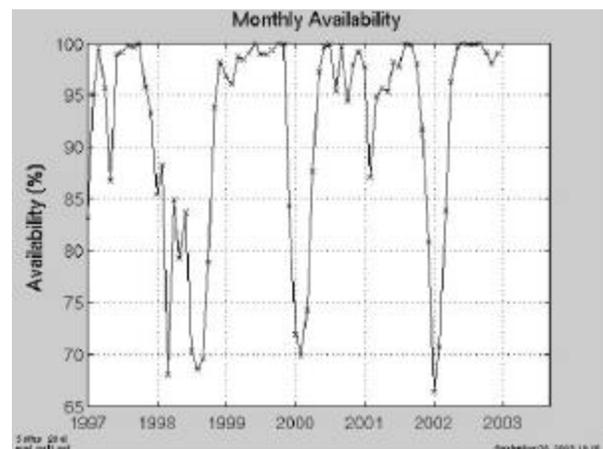


Figure 7. Monthly average availabilities for the best 5-site network meeting the 2 km altitude constraint. Note the periodic dips in performance when the probe is at southern declination.

4.2 6-Month Continental US, Hawaii, and South America Data

Having southern hemisphere stations in the network is expected to improve the probe coverage when the probe is at southern declination. Moreover, South America is further east than North America, resulting in better coverage overlap with notional stations in Europe or Africa. To evaluate the availability added by South American stations we have begun collecting and analyzing cloud data over Peru and Chile.

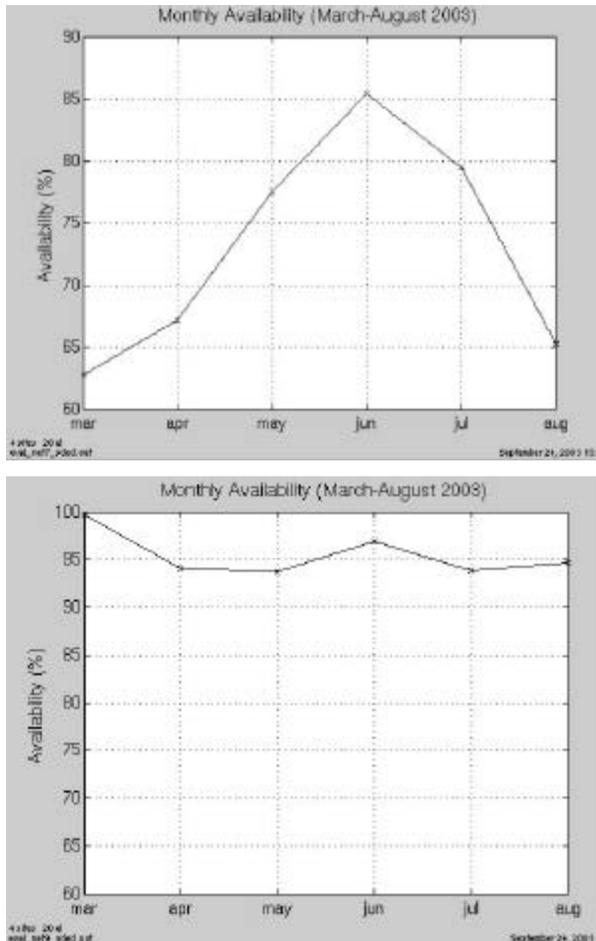


Figure 8. Top: monthly average availabilities for the 6-month period of record for the optimal 4-site network made up of sites in the continental US and Hawaii only. Bottom: monthly average availabilities for the optimal 4-site network made up of sites in the continental US, Hawaii, and South America.

Since there are only six months of data in the period of record, no conclusions can be drawn from this analysis; the results are, however, suggestive. Figure 8 shows a comparison between the monthly average availability for the optimal 4-site network made up entirely of North American (including Hawaii) stations and the monthly averages for the optimal 4-site network made up of North and South American stations. During the period

covered by the calculation the probe was between 15 and 23 degrees south declination. At least in this case, including a South American Station resulted in a substantial improvement in availability.

It remains to be seen whether the gain in availability when the probe is in the southern sky outweighs the loss of availability when the probe is in the northern sky. However, the six-year calculations provide some guidance in this area, since we expect that the loss in availability from replacing a northern station with a southern station should be less than the loss from removing a northern station entirely (i.e., replacing it with nothing). With this in mind, and examining figure 5, it seems likely that including southern hemisphere stations will be necessary if the system is to have high availability at all phases of the probe's orbit.

4.3. Summary

To achieve high availability in a lasercom system it is necessary to mitigate against clouds interrupting the line of sight. The best way to create a ground network of redundant sites so that if one site is cloudy another can be used as a backup. For this to work the ground stations must be placed so as to minimize correlations in the incidence of clouds between stations. At the same time, stations must be placed with regard to the amount of overlap in coverage of the probe as the probe rises and sets throughout the day. Finally, there are constraints on the probe's elevation angle and on the altitude of the ground stations.

We have used the LNOT code to solve the discrete optimization problem of finding the optimal ground network. We find that over the 6-year period of record 90% availability can be achieved with five ground stations, for a minimum station altitude of 2 km. For lower minimum station altitudes 90% availability can be achieved with as few as three ground stations.

Networks made up of stations in the continental United States and Hawaii suffer periods of greatly reduced availability when the probe is at southern declination. This reduced availability is caused by the shortened periods of daily coverage at each individual northern hemisphere station. This explanation suggests that the loss of availability could be rectified by including a station in the southern hemisphere. A pilot study using six months of satellite imagery collected over South America suggests that this is indeed the case; however, the 6-month period of record is too short to draw any definite conclusion. Collection efforts are continuing over South America, in order to produce a data set that can show conclusively the value of southern hemisphere stations.

5. ACKNOWLEDGMENTS

The authors would like to thank Keith Wilson of NASA-JPL for his feedback on these results. This research was funded by NASA-JPL.

6. REFERENCES

- Allen, Jr., R. C., Durkee, P. A., and Wash, C. H., 1990: Snow/cloud discrimination with multispectral satellite measurements. *J. Appl. Meteor.*, **29**, 994–1004.
- Alliss, R. J., Loftus, M. E., Apling, D., and Lefever, J., 2000: The development of cloud retrieval algorithms applied to goes digital data, in *10th Conference on Satellite Meteorology and Oceanography*, 330–333, American Meteorological Soc.
- Doswell, III, C. A., and Setvak, M., 1991: The AVHRR channel 3 cloud top reflectivity of convective storms. *Monthly Weather Review*, **119**, 841–847.
- Ellrod, G. P., 1995: Advances in the detection and analysis of fog at night using goes multispectral infrared imagery. *Weather Forecasting*, **10**, 606–619.
- Lee, T. F., Turk, F. J., and Richardson, K., 1997: Stratus and fog products using GOES-8 3.9 μm data. *Weather Forecasting*, **12**, 664–677.