Atmospheric Cloud Forecasting in Support of Space Based Applications

Randall J. Alliss, Duane Apling, Heather Kiley and Billy Felton
Northrop Grumman Corporation, 7555 Colshire Drive, McLean, VA 22102

ABSTRACT

Many space based applications from imaging to communications are impacted by the atmosphere. Atmospheric impacts such as optical turbulence and clouds are the main drivers for these types of systems. For example, in space based optical communications, clouds will produce channel fades on the order of many hundreds of decibels (dB) thereby breaking the communication link. Optical turbulence can also produce fades but can be compensated for by adaptive optics.

The ability to forecast the current and future location and optical thickness of clouds for spaced based to ground optical communications is therefore critical in order to achieve a highly reliable system. We have developed an innovative method for producing such forecasts. These forecasts are intended to provide lead times on the order of several hours so that communication links can be transferred from a current clear ground location to another more desirable ground site. This modeling system is referred to as the Cloud Propagator Forecast (CPF) and it operates on successive, satellite remotely sensed, cloud analyses. The CPF produces probability forecasts of future cloud cover conditions at each point location.

The forecasting algorithm is a combination of empirical Lagrangian and Eulerian regression over multiple spatial scales, but treats time auto-regressively. Input cloud masks are transformed into proxies first. A cloud cover proxy is a variable which has a more Gaussian distribution than literal cloud cover. For a given pixel, the cloud cover proxy is computed first by determining whether at the initialization time the pixel was clear or cloudy. Clear pixels will be assigned only positive proxies; cloudy pixels will be given only negative proxies. The degree the assigned proxy is different than zero depends on the fraction of pixels in a small neighboring space which have similar cloudy/clearness. The neighboring space is approximately the spatial scale of a skydome and has a temporal scale of one hour. Pixels which are unlike their neighbors will have proxies close to zero, those largely identical to their neighbors will has proxies close to plus or minus one. Final cloud proxies are computed using a non-linear transform to stretch out the extremes into a pseudo-Gaussian distribution. The model then decomposes the proxy fields into scale-filtered components. Longer spatial scale patterns are expected to be more predictable over time; shorted scales less so. Differentiating them allows the model to retain the maximum predictive skill through training.

The resulting forecasts have several desirable characteristics. First, they evidence substantial skill when compared to persistence. Additionally, these forecasts extrapolate movement of cloud features, and also allow for degradation of fine scale features without compromising more predictable larger scales. The forecasts are reliable, in that specific probability categories will assess at their stated probabilities, and also consequently unbiased. Details of the algorithm and results used for a realtime spaced based application will be shown at the conference.

Keywords: optical communications, lasercom, cloud forecasting, space imaging

1. INTRODUCTION

Strategies to support high-availability laser communications for future missions from space to Earth are increasingly receiving attention. Such missions will generate an ever increasing amount of data that must be transferred to ground locations on Earth. As an alternative to the current use of radio communications, deep-space to ground optical communications will provide a higher bandwidth to transfer these data with smaller power mass and power consumption subsystems. However, optical communications may be interrupted by the presence of cloud cover. Typical clouds have optical fades that far exceed three dB. Therefore, it may not be feasible to include enough link margin in the link budget to prevent a link outage. It should be noted that some cirrus clouds may have optical fades less than three dB when

averaged over many minutes. However, an optical communications link directed through the sky may encounter "knots" or areas within thin cirrus that may far exceed three dB. Therefore, a mitigation strategy ensuring a high likelihood of a cloud-free line of site (CFLOS) between a ground station and the spacecraft is needed to maximize the transfer of data and overall availability of the network.

One strategy to address this problem is predicting the cloud condition at each ground site. An accurate cloud forecast for each site will allow communication links to be transferred from one site to another in a preemptive mode thus ensuring *make before break*. Make before break is desired in order to minimize disruptions to the link and therefore maximizing system availability. The length of the desired cloud forecast is highly dependent on the mission conops but could range from a few minutes (Low Earth Orbit space link) to many hours for a deep space to earth link. Techniques for cloud forecasting range from a) persistence forecasting, b) advection forecasting and c) numerical weather prediction. Persistence forecasting is the most basic. It says what ever the sky condition is now will stay the same for the duration of the forecast. By default, the persistence forecast does not predict change. This may be very desirable for very short range forecasting but will almost certainly fail at forecast durations beyond an hour. Advection forecasting looks at the recent cloud motion typically from remote sensed images of clouds and applies a vector approach to ascertain recent motion. The CPF algorithm is an advection type forecasting scheme. Numerical Weather Prediction uses sophisticated physical models and parameterizations of the atmosphere to predict future atmospheric states. These type forecasts are used for every day weather forecasting. The advantage of this approach is that it allows for longer duration forecasts out to many hours and days. This paper focuses on the zero to two hour forecast time frame using the advection approach.

2. Technical Approach

2.1 SATELLITE DATA

Geostationary meteorological imagery is used to produce cloud analyses for scenes of interest. The NOAA Geostationary Operational Environmental Satellite (GOES) is utilized to develop the input cloud analyses necessary for the CPF algorithm. GOES imagers have five bands: visible (0.6 μm), shortwave infrared (SWIR; 3.9μm), water vapor (6.7 μm), longwave infrared (LWIR; 10.7 μm), and split window (11.2 μm). The water vapor channel, which is not used for cloud detection, is replaced with the reflectivity product during the day and the fog product at night (see below and section 2.3 for more detail on these products). The spatial resolution of the visible band is 1 km and that for the other bands is 4 km. For our purposes, the 4 km data is resampled to 1 km resolution so that it is comparable to the visible band. The following discusses at a high level the Cloud Mask Generator (CMG) algorithm.

2.2 CLEAR SKY BACKGROUND

Our cloud analysis techniques for the GOES data are described in detail by Alliss et al.[1]. All cloud tests consist of comparing satellite image values to dynamically computed clear sky background (CSB) values pixel by pixel in the regions of interest. The CSB is discussed below and main cloud test algorithms (albedo, LWIR, fog, and reflectivity) are discussed in section 2.3.

The CSB is defined as the amount of radiation emitted and/or reflected from a surface that reaches a satellite sensor when no clouds are present. The CSB varies spatially and temporally and is influenced by the radiative properties of the surface material, surface temperature, terrain height, soil moisture, and solar illumination angle. Because of these variations, the CSB must be calculated for each region separately, on a pixel by pixel basis, as a function of the above-mentioned factors to generate accurate cloud masks. For example, if the albedo test used a fixed threshold for typical differences between the observed and calculated CSB albedos for all locations, then false cloud detections would be likely over naturally highly reflective regions such as White Sands, NM or the salt flats of northern Chile.

Four CSBs are estimated in the CMG: albedo, reflectivity, LWIR, and fog[1]. The CSB is calculated for each pixel by using data from clear times over the previous 30 days at a given analysis time (e.g., 1400 GMT). This approach provides sufficient clear sky data and reduces the effect of diurnal and seasonal cycles of temperature and illumination, in particular, on the calculated CSB. The database from which clear times are determined includes not only the satellite imagery, but also ancillary surface and ship observations collected by the National Weather Service (NWS), World Meteorological Organization (WMO), and at several telescope observatories in South America.

The albedo CSB is calculated by identifying the darkest 10 % of albedo values from the previous 30 days of visible images. The selected albedo values are averaged to define the CSB for each pixel. The reflectivity CSB is determined only during the day and when snow cover is not likely present. Like the albedo, the darkest 10 % reflectivity product values from the previous 30 days are selected and averaged to generate the CSB.

To develop the fog product CSB, the warmest 10 % of LWIR values for the pixel over the previous 30 days are selected. The corresponding fog product values are then averaged to give the fog CSB. Note that the procedure used to generate the fog product CSB differs from that used to generate the albedo and reflectivity products in which clear pixels are chosen based on the albedo and reflectivity values themselves. Both fog product extremes indicate clouds and the selection of the 10 % warmest or coldest values will not provide the needed information; therefore, the two-step process is used for the fog product CSB.

The LWIR CSB is determined as the average of the difference between the LWIR temperature from the satellite for a given pixel and the LWIR CSB temperature estimated from a linear regression model. The regression model is developed with data from clear sky pixels that are used as prototypes. These prototype pixels are selected by a series of tests that find pixels with a high probability of being clear, even without the benefit of any of the cloud tests. The coefficients of the regression model for twelve predictors are fit with the data from the prototype pixels. The predictors include satellite data, time, terrain, and regional observations such as cloud cover and air temperature from the NWS and WMO. The LWIR regression model estimates the clear sky LWIR brightness temperature for each pixel. The LWIR residuals are the differences between the regression model temperatures and the measured imager LWIR temperatures. The warmest 10 % of the LWIR residuals are averaged to determine the LWIR residual CSB that is used in the LWIR cloud tests.

2.3 CLOUD TESTS

The CSB values and the satellite data are compared in four main cloud tests in the CMG: the LWIR test, the albedo product test, the fog product test, and the reflectivity product test. The LWIR test is applied at all times of the day, unlike the albedo, reflectivity, or fog product tests. A pixel is considered to be cloudy if the LWIR CSB for a given pixel exceeds the LWIR from the satellite by the threshold value or greater. This test cannot easily detect fog/low clouds at night because the cloud top temperatures are very similar to the surface temperatures. It is unlikely that clouds will radiate in the LWIR at temperatures greater than 300K. A pixel is deemed clear if the LWIR temperature is greater than 300K, even if the LWIR cloud test indicates that it is cloudy.

As mentioned above, the detection of fog and low stratus clouds at night is difficult with the LWIR. The fog product test is a multi-spectral test that compares values of the fog product calculated as the difference between the LWIR and the SWIR brightness temperatures[2,3,4]. The temperature differences result mainly because clouds observed in the SWIR have an emissivity that is 20%-40% lower than clouds observed in the LWIR[5]. Therefore, at night, liquid stratiform (low) clouds appear colder in the SWIR than they do in the LWIR. Typical T_{LWIR}-T_{SWIR} for fog and low stratus are approximately 2 K or larger[6]. The fog product can also detect ice clouds, which are highly transmissive and therefore appear warmer in the SWIR. Typical values for ice clouds are T_{LWIR}-T_{SWIR} are approximately -5 K or lower[6]. The daytime SWIR is dominated by reflected solar SWIR and therefore, the fog product is only useful at night.

The albedo test, which uses visible data, is applied when the solar zenith angle is below 89^{o}. This test will detect clouds if the pixel is more reflective than the albedo CSB and the difference is greater than a predefined threshold for that pixel. If the difference between the calculated albedo and the CSB is less than the threshold, the pixel is deemed clear. The albedo test may falsely detect snow as clouds.

The shortwave reflectivity product is implemented during the day to decide if a pixel is cloudy or if the surface is snow-covered. This product indicates the amount of reflected solar SWIR detected and is derived by removing the thermal component from the SWIR[5,6]. Water clouds are highly reflective in the SWIR while ice clouds are poorly reflective in the SWIR. As a result, water clouds appear as bright white and poorly reflective ice clouds and snow appear as dark gray or black in the resulting images. The reflectivity product, then, can easily distinguish between low clouds and snow cover. The reflectivity test is only applied when and where snow cover is likely and can override a false cloud detection for snow cover indicated by the albedo test. To ensure that high ice clouds (which also appear dark in the reflectivity test) are not present, the LWIR test must not indicate the presence of high clouds for a pixel to be considered clear.

With the CSB and satellite data, the CMG performs the necessary tests to determine the cloud masks. During the day, for example over southern Italy and Sicily (Figure 1), the LWIR and albedo products are used to detect clouds with the resulting mask accurately showing the location of clouds. At night when low clouds cannot be adequately detected by the LWIR, the fog product is vital to developing accurate cloud masks. In fact, in a cloud scene from Hawaii (Figure 3), the low clouds over the land would not have been detected without the fog product.

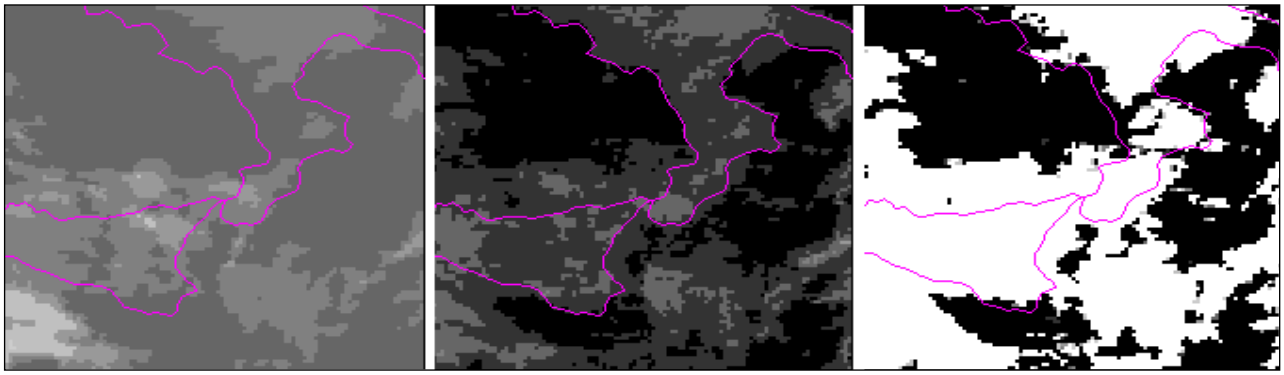

Figure 1: A sample cloud scene during the day for southern Italy and Sicily. The image on the left is the LWIR image from MSG during the day. The image in the center is the corresponding visible image from MSG. For these two images, the lighter grey images indicate clouds. The cloud mask on the right shows clouds as white and was generated with the CMG.

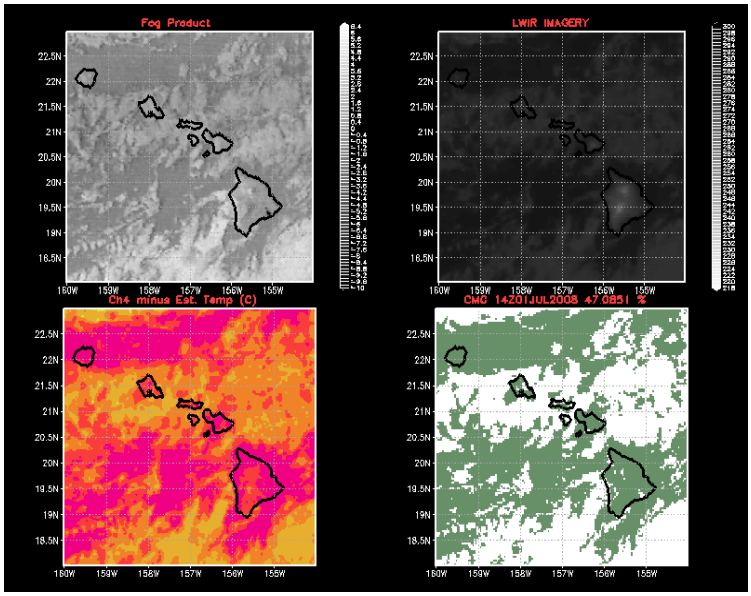

Figure 2. Fog Product (upper right), LWIR imagery (upper left), longwave infrared modeling (lower left), and composite cloud test (lower right) of a nighttime scene over Hawaii.

2.4 CPF

The CPF operates on successive cloud masks to produce reliable probability forecasts of future cloud cover conditions at each point location (single mask grid cell) or for the expectation of the amount of skycover in a local skydome about each point location. This discussion covers the point probability mode, with exceptions for skydome mode presented at the conclusion.

The CPF makes reads CMG gridded data, from either the historical or real time-modes, and produces a single forecast time slice valid a specified time in the future. The output from the CPF is placed into identically formatted output files, with forecast fields stored in the time positions equal to the valid times of the forecasts. The CPF retains trained model data from previous runs for the same tile, and is capable of boot-strapping itself rapidly if no prior run model data are available.

The forecasting algorithm is a combination of empirical Lagrangian and Eulerian regression over multiple spatial scales, but treats time auto-regressively. Input cloud masks are transformed into proxies first. A cloud cover proxy is a variable which has a more Gaussian distribution than literal cloud cover. For a given pixel, the cloud cover proxy is computed first by determining whether at the initialization time the pixel was clear or cloudy. Clear pixels will be assigned only positive proxies; cloudy pixels will be given only negative proxies. The degree the assigned proxy is different than zero depends on the fraction of pixels in a small neighboring space that have similar cloudy/clearness. The neighboring space is approximately the spatial scale of a skydome and has a temporal scale of one hour. Pixels which are unlike their neighbors will have proxies close to zero, those largely identical to their neighbors will has proxies close to plus or minus one. Final cloud proxies are computed using a non-linear transform to stretch out the extremes into a pseudo-Gaussian distribution.

The model then decomposes the proxy fields into scale-filtered components. Longer spatial scale patterns are expected to be more predictable over time; shorted scales less so. Differentiating them allows the model to retain the maximum predictive skill through training. The training is performed in pure-hindsight. A hindsight forecast is constructed from past data and is evaluated against current proxies to update motion vectors and regression weights. Motion vectors are computed on each scale using a fast local search which maximizes offset spatial pattern correlations. The motion vectors are an exponentially deweighted running mean. Regression predictors are extracted from the predictor field centered at the motion vector offset, with a 3x3 or 5x5 array (configurable) of samples taken spaced at the characteristic length of the spatial component. The prior regression models are deweighted exponentially, similarly to the motion vectors, so that evolving weather patterns can be followed rapidly by the modeling. The regression models produce minimum error estimates of the current time proxy, as a function of a proxy offset from the past. If multiple forecast lead times are configured for the run, then multiple motion/regression models are independently used.

Additionally, the hindsight proxy forecasts to the initialization time are tracked in concert with actual cloud/no-cloud statuses of the hindsight training data. These are accumulated in reliability tables which have separate slots for eight three-hourly times of day to allow forecast accuracy to be diurnally variable. These tables are also exponentially deweighted, but with a time-scale over days rather than hours. Finally, training is not executed for time steps where input data quality is thought to be low. This decision is made by examining the absolute change in regional cloud cover over a single time step and comparing it to a user configurable threshold. The motion/regression models are saved, and then the model is run from the current time, forward, to create a future proxy forecast. The proxies are translated into cloud cover by interpolated look-up into the reliability tables. The results are shown as probabilities of cloud at each pixel in the forecast image.

3. Results

We ran the CPF on several months of images centered on White Sands, New Mexico (WSC). The period of study was March – July 2013. Over 13,000 independent forecasts were generated during this time for each forecast length. Table 1 shows a summary of the results in terms of a series of contingency tables. The tables are shown by forecast length starting with 15 minutes through 120 minutes. The contingency tables indicate the frequency of which the CPF reports clear and cloudy versus truth as defined by the CMG for skydome cloud fraction centered on WSC. In this case clear is defined as a skydome less than 30% and cloudy greater than 30%. Results indicate that CPF forecasts clear and cloudy conditions correctly approximately 93% of the time at 15 minutes but by 120 minutes the agreement drops to approximately 83%. There appears to be very little bias (over or under forecasting of clouds) relative to truth. At 15 minutes 3.2% of the CPF clear forecasts were actually cloud while 3.4% of the CPF cloudy forecasts were actually clear. This ratio is nearly the same regardless of the forecast length. At 120 minutes the CPF clear forecasts were incorrect 8.3% of the time while the CPF cloudy forecasts were incorrect about 8.8% of the time.

Figure 3 shows how the CPF performs relative to a simple persistence forecast. A persistence forecast predicts that the cloud state at time 1 will remain the same through time – N. For example, if WSC is cloud free at time 1, persistence predicts it will remain cloud free through 120 minutes.

15 Min. Forecasts	CPF Clear	CPF Cloudy	30 Min. Forecasts	CPF Clear	CPF Cloudy
CMG Truth Clear	62.5%	3.4%	CMG Truth Clear	61.6%	4.3%
CMG Truth Cloudy	3.2%	30.9%	CMG Truth Cloudy	4.6%	29.5%
45 Min. Forecasts	**CPF Clear**	**CPF Cloudy**	**60 Min. Forecasts**	**CPF Clear**	**CPF Cloudy**
CMG Truth Clear	60.5%	5.3%	CMG Truth Clear	59.9%	5.8%
CMG Truth Cloudy	5.5%	28.7%	CMG Truth Cloudy	6.3%	28.0%
75 Min. Forecasts	**CPF Clear**	**CPF Cloudy**	**90 Min. Forecasts**	**CPF Clear**	**CPF Cloudy**
CMG Truth Clear	59.1%	6.5%	CMG Truth Clear	58.6%	7.3%
CMG Truth Cloudy	7.1%	27.3%	CMG Truth Cloudy	7.5%	26.6%
105 Min Forecasts	**CPF Clear**	**CPF Cloudy**	**120 Min Forecasts**	**CPF Clear**	**CPF Cloudy**
CMG Truth Clear	57.9%	8.0%	CMG Truth Clear	57.3%	8.8%
CMG Truth Cloudy	7.7%	26.4%	CMG Truth Cloudy	8.3%	25.6%

Table 1. CPF versus CMG contingency tables for eight forecast lengths ranging from 15 minutes to 120 minutes.

Figure 3 shows the Root Mean Square Error (RMSE) for both the CPF and persistence forecast as a function of forecast length and by each month in the study. In most cases the CPF (solid line) outperforms the persistence forecast (dash line). RMSE's range from 0.03 – 0.07 at 15 minutes to nearly 0.08 - 0.15 at 120 minutes. RMSE's of the persistence forecasts range from 0.05 – 0.07 at 15 minutes to 0.13-0.23 at 120 minutes. As expected the longer the forecast the more persistence underperforms relative to CPF because of the increased likelihood of a change in the cloud condition. By definition, persistence does not predict change.

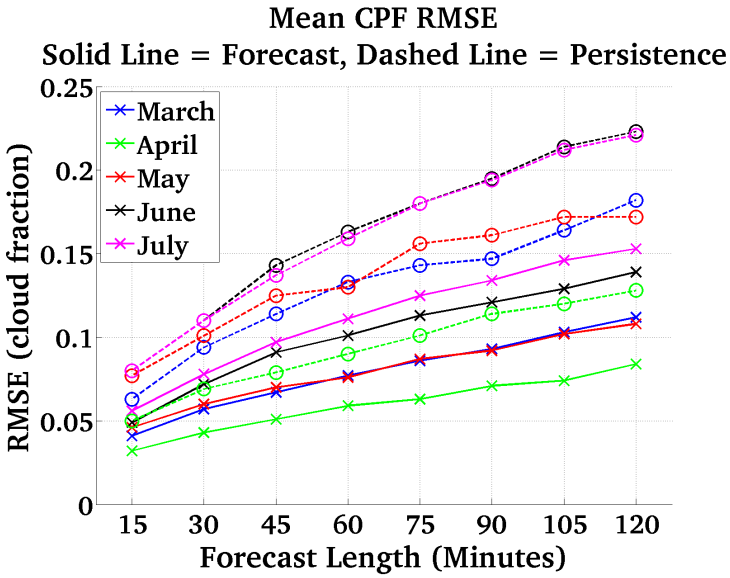

Figure 3. RMSE of the CPF (solid line) and persistence (dash line) forecast techniques as a function of forecast length and month of the study.

As reported earlier, the bias or the tendency for CPF to over or under predict clouds relative to truth is relatively low. Figure 4 shows the bias for each month in the study as a function of forecast length. A slight positive bias (over prediction of clouds) is noted in the CPF but generally ranges less than 0.04 at 120 minutes.

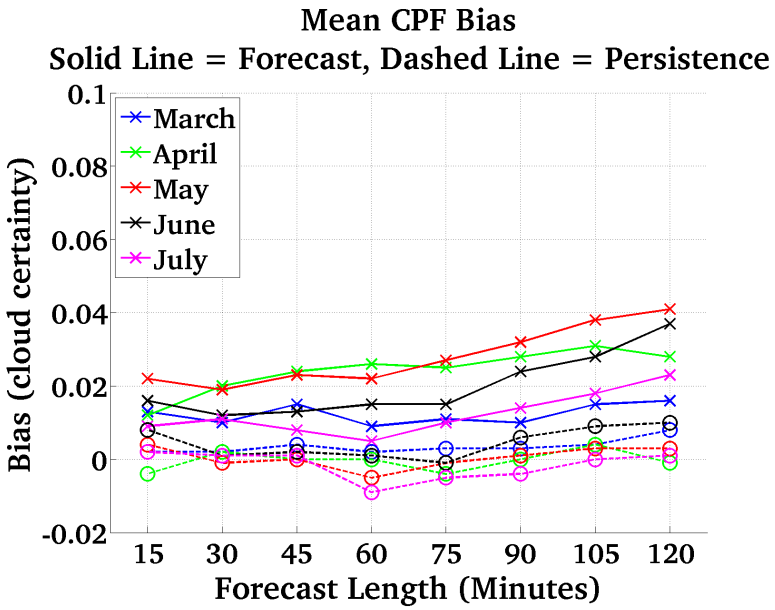

Figure 4. The mean bias in CPF (solid line) forecasts relative to persistence (dash line).

The CPF has demonstrated it works very well during times when WSC is either clear or cloudy. However, it is desired for many applications including spaced based imaging and optical communications, that CPF perform well during transition states. A transition state refers to times when the site is becoming cloudy or when the site is becoming clear over some short period of time. To investigate this isolated all the times when the atmosphere was transitioning from clear to cloudy (crossing the 30% skydome cloud fraction threshold) and cloudy to clear (crossing the 30% skydome cloud fraction threshold). Over the course of 15 minutes this occurred approximately three percent of the time while at 120 minutes it occurred nearly 16% of the time. Figure 5 indicates the frequency that the CPF algorithm successfully predicted the change in cloud condition during these times. The data is broken into two categories, 1) those times when the site went from being cloudy to clear and, 2) those times when the site went from being clear to cloudy. Results indicate the clear to cloudy transitions were only predicted by CPF between 10% and 20% of the time depending on the forecast length. CPF predicted the cloudy to clear transitions between 33% and 40% of the time. These results are not surprising given that the CPF in general is not trying to forecast the detailed edges of clouds (considered noise to the algorithm) rather it is looking for more bulk cloud features to track.

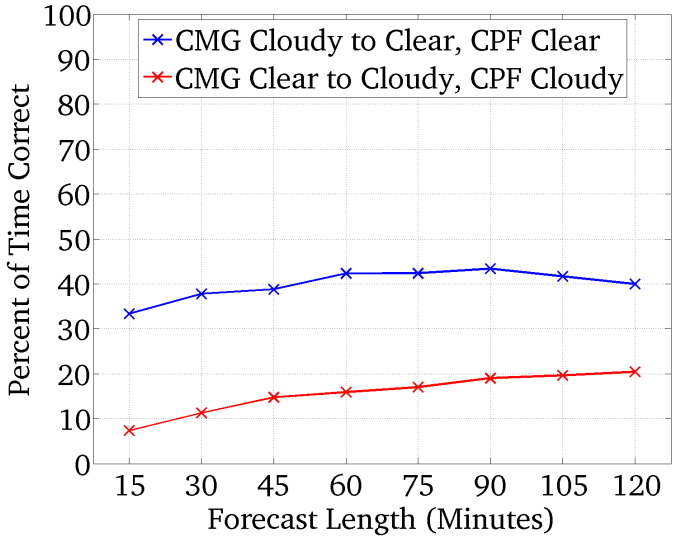

Figure 5. The frequency CPF forecasts the change in state of clouds at WSC.

4. Summary

Many space based applications from imaging to communications are impacted by the atmosphere. Atmospheric impacts such as optical turbulence and clouds are the main drivers for these types of systems. For example, in space based optical communications, clouds will produce channel fades on the order of many hundreds of decibels (dB) thereby breaking the communication link.

The ability to forecast the current and future location and optical thickness of clouds for spaced based to ground optical communications is therefore critical in order to achieve a highly reliable system. We have developed an innovative method for producing such forecasts. These forecasts are intended to provide lead times on the order of several hours so that communication links can be transferred from a current clear ground location to another more desirable ground site. This modeling system is referred to as the Cloud Propagator Forecast (CPF) and it operates on successive, satellite remotely sensed, cloud analyses. The CPF produces probability forecasts of future cloud cover conditions at each point location. Initial results of this algorithm are promising (agreement with truth between 83% - 93%) but more work is needed to raise its agreement with truth. The main weakness of the current algorithm is its inability to consistently predict changes in the cloud state. A revised algorithm is currently under development which will better track cloud features by introducing cloud height information.

REFERENCES

1. R. J. Alliss, M. E. Loftus, D. Apling, and J. Lefever, "The development of cloud retrieval algorithms applied to goes digital data," in *10th Conference on Satellite Meteorology and Oceanography*, pp. 330–333, American Meteorological Soc., January 2000.
2. G. P. Ellrod, "Advances in the detection and analysis of fog at night using goes multispectral infrared imagery," *Weather Forecasting,* **10**, pp. 606–619, 1995.
3. G. E. Hunt, Radiative properties of terrestrial clouds at visible and infrared thermal window wavelengths," *Quarterly Journal of the Royal Meteorological Society*, **99**, 346-359.
4. T. F. Lee, F. J. Turk, and K. Richardson, "Stratus and fog products using GOES-8 3.9 μm data," *Weather Forecasting,* **12**, pp. 664–677, 1997.
5. R. C. Allen, Jr., P. A. Durkee, and C. H. Wash, "Snow/cloud discrimination with multispectral satellite measurements," *Journal of Applied Meteorology,* **29**, pp. 994–1004, 1990.
6. M. Setvak and C. A. Doswell, III, "The AVHRR channel 3 cloud top reflectivity of convective storms," *Monthly Weather Review,* 119, pp. 841–847, 1991.