8A.8 PROBABILISTIC CLOUD FORECASTING USING LOGISTIC REGRESSION

Eric M. Kemp* and Randall J. Alliss
Northrop Grumman Information Technology/TASC, Chantilly, Virginia

1. INTRODUCTION

Over the past decade, TASC has pursued applied research in the areas of cloud detection, simulation, and forecasting (Alliss et al. 2000a,b, 2004; Kelly et al. 2001; Kemp and Alliss 2005; Link et al. 2005; Wojcik et al. 2005, 2006). This research has included investigations on how best to use numerical weather prediction (NWP) models in forecasting clouds at select locations out to 48 hours. Because of the computational cost of running a NWP model at cloud resolving resolutions ($\Delta x \leq 4$ km) – as well as the difficulty of correctly initializing such a model with cloud scale initial conditions – it has proved necessary to adopt a hybrid approach combining mesoscale NWP output with statistical post-processing. In this paper we discuss our approach, which uses the Weather Research and Forecasting (WRF) NWP model (Skamarock et al. 2007), the Cloud Mask Generator (CMG) developed by TASC (Alliss et al. 2000a), and a statistical technique called logistic regression (Hosmer and Lemeshow 2000).

2. FORECAST TECHNIQUE

Figure 1 shows a high level summary of the forecast technique. The procedure begins by collecting the previous 20 days of 0–48 hr NWP forecasts along with objective cloud analyses. For each NWP forecast and location, a spatial/temporal cloud fraction is constructed from the cloud analyses over a 400 km $\times$ 400 km area centered on the location of interest, with a time window of $\pm$2 hours from the NWP forecast valid time. These cloud fractions are then compared to selected NWP variables to fit a logistic regression equation for each forecast hour and location. Once all the equations are fit, data from a new NWP forecast run is fetched and passed to the equations. The result is a time series of mesoscale cloud forecasts that can be interpreted either as probabilities or as cloud fraction.

Corresponding author address: Eric M. Kemp, Northrop Grumman Information Technology/TASC, 4801 Stonecroft Blvd, Chantilly, VA 20151. E-mail: eric.kemp@ngc.com

The training period length, skydome area size, and time window can easily be modified, and indeed these settings have evolved as our technique has been developed and tested. The 20 day training period was originally selected to maximize the number of cloud forecasts available for validation, with the constraint that TASC-produced NWP simulations only go back to mid-January 2006. In time, it will be practical to increase the training period and still have enough validation data to draw conclusions from. As for the skydome area and time window, the current values were selected after early tests with smaller time windows and areal coverages showed strong sensitivity to NWP propagation errors.

Figure 2 shows a sample forecast time series for Washington-Dulles International Airport (KIAD) as well as a verification using independent cloud analyses. In this example, the forecasts exhibit trend errors during the first 12 hours, but otherwise are in reasonable agreement with the observed cloud fraction. Further discussion of forecast results is postponed to section 3; below we describe the components of the technique in more detail.

a. WRF

The technique uses Version 2.1 of the Advanced Research WRF model (WRF-ARW) maintained by the National Center for Atmospheric Research (Skamarock et al. 2007). This is a 3-D, fully compressible non-hydrostatic NWP model that uses a hydrostatic-pressure based vertical coordinate (Laprise 2002) on an Arakawa-C staggered grid. The model is initialized twice a day at 0000 UTC and 1200 UTC directly from the 40-km (Grid 212) North American Mesoscale (NAM) analysis produced by the National Weather Service. Lateral boundary conditions are provided out to 48 hours by three-hourly NAM forecasts. WRF-ARW is configured to use a 36 km horizontal resolution domain covering all of the continental United States (CONUS) from the ground up to 50 mb ($\sim$20 km above mean sea level). Time integration is performed using a third-order Runga-Kutta split-time scheme (Wicker and Skamarock 2002)

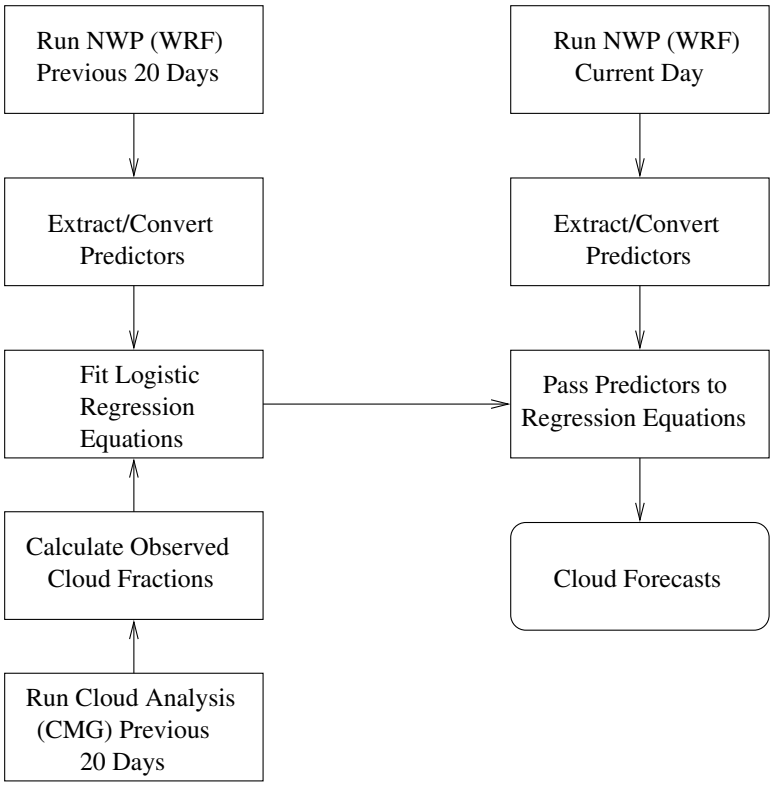

Figure 1: Cloud Forecast Algorithm

with a time step of 120 s. The physics selected are: the WSM5 ice microphysics scheme (Hong et al. 2004), the updated Kain-Fritsch cumulus parameterization (Kain 2004), the Noah land surface model (Ek et al. 2003), the YSU planetary boundary layer scheme (Hong et al. 2006), the MM5 shortwave radiation scheme (Dudhia 1989), and the RRTM longwave radiation package (Mlawer et al. 1997).

b. CMG

The CMG (Alliss et al. 2000a) is designed to objectively classify the presence or absence of cloud along various lines of site between the Earth's surface and a weather satellite. This is accomplished by comparing multispectral satellite imagery and derived products with clear sky background fields created from the previous thirty days of satellite data. Typically the CMG is run using four channels and two derived products from the Geostationary Operational Environment Satellites (GOES): visible (0.6 μm), shortwave infrared (3.9 μm), longwave infrared (10.7 μm), the split window channel (11.2 μm), the nighttime multispectral fog product and the daytime shortwave reflectivity product (Lee et al. 1997). Each channel and product has its own strengths and weaknesses in discriminating between clouds and other phenomena (e.g., snowpack, extreme cold surface temperatures, etc.); therefore, a series of single and multispectral tests are performed and intercompared before making a final cloud determination. See Alliss et al. (2000a) for more details. GOES-derived cloud masks have been generated for the entire CONUS from 1995 to present, with a temporal (spatial) resolution of 15 min (4 km). More limited cloud masks for other parts of the globe have also been created using GOES and Meteosat data (Alliss et al. 2004; Link et al. 2005; Wojcik et al. 2005, 2006).

c. Logistic Regression

Logistic regression is a procedure for fitting one or more predictors to a binary outcome (in this case, "cloud" or "no cloud"). The function that is fit is an "S-shaped" or sigmoid function constrained to lie in the range (0,1). The logistic equation is defined as:

$$\ln\left[\frac{\pi}{1-\pi}\right] = \alpha + \beta_1 x_1 + \cdots + \beta_p x_p$$

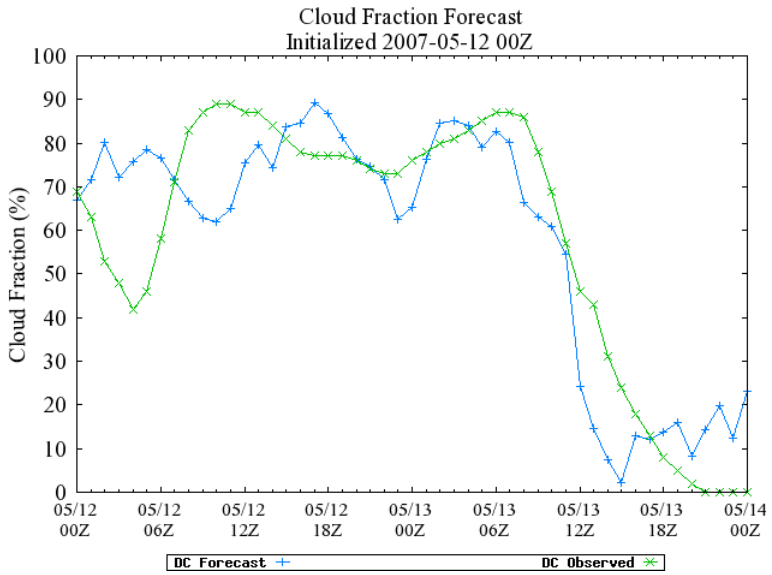

Figure 2: 00–48 hr cloud fraction forecasts (blue) and analyses (green) for KIAD, beginning at 0000 UTC 12 May 2007.

or equivalently:

$$\pi = \frac{\exp(\alpha + \beta_1 x_1 + \cdots + \beta_p x_p)}{1 + \exp(\alpha + \beta_1 x_1 + \cdots + \beta_p x_p)}$$

Here π is the probability of a cloud occurring, x_i is the value of the ith predictor x, β_i is the associated weight given to the predictor, and p is the total number of predictors. α is a constant term that accounts for "persistence" , i.e., the average amount of cloud coverage that occurred throughout the training period. An example logistic equation, using a single predictor and giving no weight to persistence, is plotted in Figure 3.

The α and β terms are fit to maximize the log-likelihood function L, which is defined as:

$$L = \sum_{i=1}^{n} \{y_i \ln[\pi_i] + (1 - y_i) \ln[1 - \pi_i]\} .$$

Here y_i represents the presence (1) or absence (0) of cloud at pixel i, π_i is the corresponding forecast from the logistic regression model, and n is the total sample size of pixel forecasts. An exact analytic solution for this problem does not exist; however, the α and β terms can be fit using the Miller (1992) iterative weighted least squares algorithm (see http://users.bigpond.net.au/amiller/ for Fortran 90 source code).

There are two major advantages to using logistic regression instead of the more traditional linear regression (Hosmer and Lemeshow 2000). First,

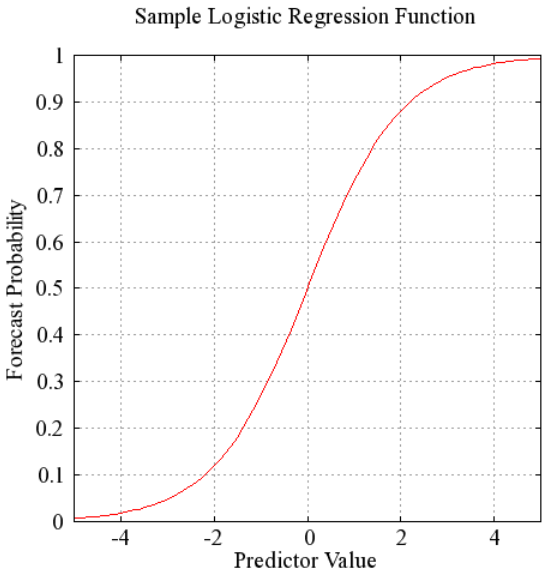

Figure 3: Example logistic function with $p = 1$, $\alpha = 0$, and $\beta_1 = 1$.

linear regression assumes that errors are normally distributed with a constant variance. However, binary outcomes require binomial error distributions with a $\pi[1 - \pi]$ variance, and this is assumed with logistic regression. Second, output from a linear regression equation is not guaranteed to remain between 0 and 1, violating the definitions of probability and cloud fraction. In contrast, logistic regression constrains the output to lie between 0 and 1

(although outcomes of exactly 0 or 1 are not permitted, as this would cause $L = \infty$).

The number of regression predictors (excluding so-called "interaction terms") is determined from the length m of the training period through the $m/10$ rule (Harrell 2001). Thus the current training set length of 20 days is used to fit two independent regression predictors. The first predictor x_1 is always the maximum relative humidity w.r.t. ice (RHI), and relates cloudiness to model resolved moisture. The second predictor x_2 is a function of the north-south wind component v, and apparently relates cloudiness to the presence of a trough west of the location of interest. The specific function used for x_2 (e.g., v at the level of maximum RHI, cosine of 700 mb wind, cosine of 300 mb wind, etc.) varies by forecast hour, and is based on early test results for KIAD. For some forecast hours an "interaction term" of $x_3 = x_1 \times x_2$ is also included, again based on early testing.

3. EVALUATION

In this section we validate 00-hr, 06-hr, 12-hr, 18-hr, and 24-hr forecasts for KIAD from early February 2006 to mid May 2007. (Forecasts beyond 24 hours have only been recently attempted and will be validated in future work.) The relevant validation metrics depend on whether we interpret the output as forecast probabilities or as forecast cloud fractions. Both approaches will be summarized below.

a. Probabilistic Validation

If we interpret the forecasts as probabilities, an appropriate validation metric is the Relative Operating Characteristic (ROC) curve (Mason 1982). The ROC curve shows the relationship between Probability of Detection (POD; ratio of correct cloudy forecasts to all cloudy events) and Probability of False Detection (POFD; ratio of incorrect cloudy forecasts to all clear events) as a function of forecast threshold. Figure 4 shows the ROC curves for the KIAD forecasts. All the forecast hours exhibit the desired behavior of POD $>$ POFD for nearly all thresholds, evidence that the forecasts can discriminate between cloudy and clear events.

More information can be found by calculating the area under the ROC curve (AUC). The AUC is interpreted as the likelihood that, for a randomly selected "yes" (cloudy) event and "no" (clear) event, the forecast probability for the "yes" event will be higher than that for the "no" event (Mason and Graham 2002). Perfect discrimination is represented by an AUC of 1.0, while an AUC of 0.5 indicates no discrimination. The KIAD forecast AUCs – calculated analytically following Mason and Graham (2002) – are displayed in Figure 5. These plots indicate a fair amount of discrimination in the forecasts, with AUCs close to 0.8.

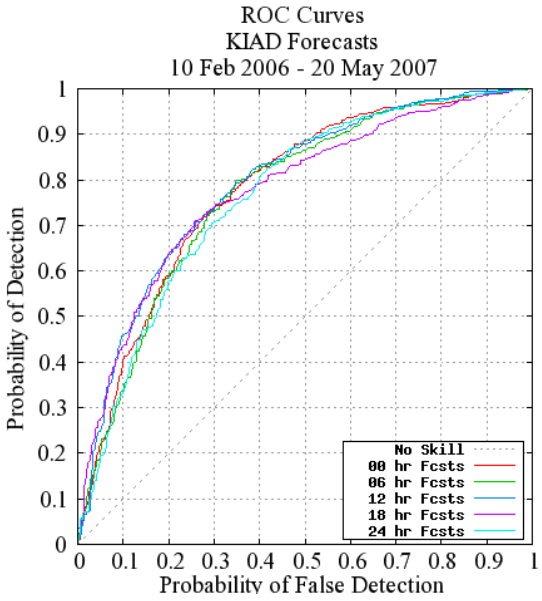

Figure 4: ROC curves for KIAD 00, 06, 12, 18, and 24 hour forecasts.

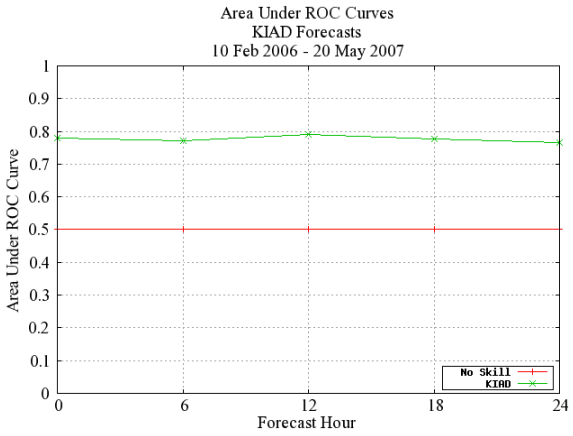

Figure 5: Area Under ROC Curves for KIAD 00, 06, 12, 18, and 24 hour forecasts.

In addition to the ROC curves, reliability diagrams can be used to evaluate probability forecasts (Wilks 1995). A reliability diagram illustrates the relationship between forecast probability (π) and conditional observed probability (also called the observed relative frequency); the latter is the probability of cloud being observed given a particular value (or range of values) of forecast probability.

As its name implies, this diagram summarizes the reliability of a forecast system, by showing how closely particular forecast values correspond to actual events. Perfectly reliable forecasts lie on the $y = x$ line, while points above (below) that line indicate a clear (cloudy) conditional bias.

Figure 6 shows the reliability diagram for the KIAD forecasts. Most of the points are rather close to the $y = x$ line, but the 6 hr forecasts exhibit a clear bias of about 15% when forecasting probabilities of 0.25–0.35. Also, the 18 hr forecasts for the range 0.45–0.55 are too cloudy by a little over 10%, while all forecasts above 0.85 are likewise too cloudy.

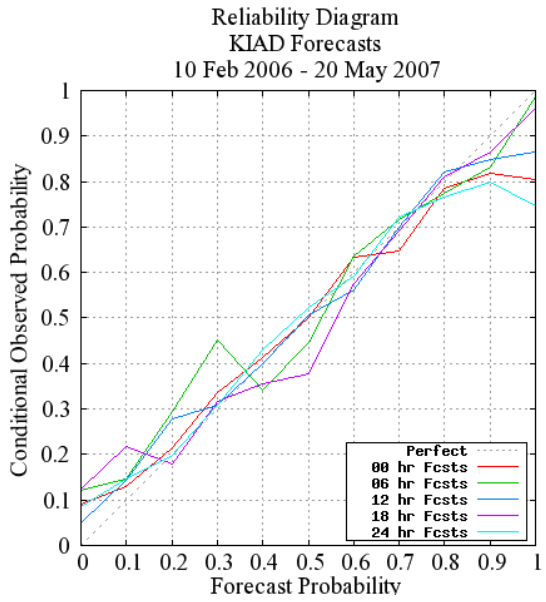

Figure 6: Reliability diagram for KIAD 00, 06, 12, 18, and 24 hour forecasts. Forecast probability bins are 0.0–0.05, 0.05–0.15, 0.15–0.25, ..., 0.85–0.95, 0.95–1.0.

b. Cloud Fraction Validation

If we interpret the forecasts as cloud fraction, the accuracy can be summarized using the absolute skydome error, defined here as:

$$A = \left| \pi - (1/n) \sum_{i=1}^{n} y_i \right|$$

where a perfect forecast yields $A = 0$. The errors for each forecast hour can then be summarized by plotting the empirical cumulative distribution functions (CDF), allowing for determination of any quantile (e.g., the median) from visual inspection.

Figure 7 summarizes the CDF of absolute skydome errors, and depicts considerable accuracy from the technique: 50% of the forecasts have absolute skydome errors of 15% or less; 75% of the forecasts have A of 25% or less; and extreme errors of 50% are restricted to less than 5% of the forecasts.

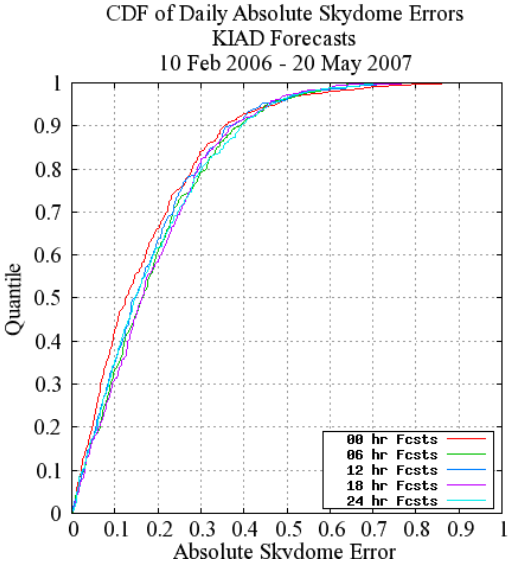

Figure 7: Plots of the empirical cumulative density functions of absolute skydome errors with KIAD 00, 06, 12, 18, and 24 hour forecasts.

4. SUMMARY

A forecast technique combining NWP data, objective cloud analyses, and logistic regression has been developed to produce cloud forecasts for select locations. The forecasts can be interpreted as cloud probabilities or alternatively as cloud fractions. Forecasts for Washington-Dulles International Airport have been compared with independent cloud analyses, and show considerable discrimination, reliability, and accuracy. Future work will include forecast production and validation for other regions in CONUS, increasing the training period and number of predictors, and validating forecasts past 24 hours.

Acknowledgements. We thank Alan Miller for providing his logistic regression source code. Danny Felton, Heather Kiley, and Robert Link provided constructive criticism and feedback that aided the development of this forecast technique. Martha Tonkin and Joe Greenseid provided computer support.

References

Alliss, R. J., R. Link, and M. E. Craddock, 2004: Mitigating the impact of clouds on optical communications. *13th Conference on Satellite Meteorology and Oceanography*, American Meteorological Society, Norfolk, VA. [Available on-line at http://ams.confex.com/ams/htsearch.cgi].

Alliss, R. J., M. E. Loftus, D. Apling, and J. Lefever, 2000a: The development of cloud retrieval algorithms applied to GOES digital data. *Preprints, 10th Conference on Satellite Meteorology and Oceanography*, American Meteorological Society, Long Beach, CA, 330-333. [Available on-line at http://ams.confex.com/ams/htsearch.cgi].

— 2000b: The use of satellite data in an optimal interpolation assimilation system. *Preprints, 10th Conference on Satellite Meteorology and Oceanography*, American Meteorological Society, Long Beach, CA. [Available on-line at http://ams.confex.com/ams/htsearch.cgi].

Dudhia, J., 1989: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 2077–3107.

Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarplay, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.*, **108**, doi:10.1029/2002JD003296.

Harrell, F. E., 2001: *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer Series in Statistics, Springer, New York, 568 pp.

Hong, S.-Y., J. Dudhia, and S.-H. Chen, 2004: A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Mon. Wea. Rev.*, **132**, 103–120.

Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341.

Hosmer, D. W. and S. Lemeshow, 2000: *Applied Logistic Regression*. Wiley Series in Probability and Statistics, John Wiley and Sons, Inc., New York, Second edition, 375 pp.

Kain, J. S., 2004: The Kain-Fritsch convective parameterization: An update. *J. Appl. Meteor.*, **43**, 170–181.

Kelly, M. A., R. J. Alliss, M. E. Loftus, and J. C. Lefever, 2001: A quantitative comparison of MM5 cloud forecasts and GOES cloud analyses. *Preprints, 18th Conference on Weather Analysis and Forecasting/14th Conference on Numerical Weather Prediction*, American Meteorological Society, Ft. Lauderdale, FL. [Available on-line at http://ams.confex.com/ams/htsearch.cgi].

Kemp, E. M. and R. J. Alliss, 2005: Enhancements to a three-dimensional cloud analysis scheme. *21st Conference on Weather Analysis and Forecasting/17th Conference on Numerical Weather Prediction*, American Meteorological Society, Washington, DC. [Available on-line at http://ams.confex.com/ams/htsearch.cgi].

Laprise, R., 2002: The Euler equations of motion with hydrostatic pressure as an independent variable. *Mon. Wea. Rev.*, **120**, 197–207.

Lee, T. F., F. J. Turk, and K. Richardson, 1997: Stratus and fog products using GOES–8 3.9 μm data. *Wea. Forecasting*, **12**, 664–677.

Link, R. L., M. E. Craddock, and R. J. Alliss, 2005: Mitigating the impact of clouds on optical communications. *2005 IEEE Aerospace Conference*, IEEE, Big Sky, MT.

Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303.

Mason, S. J. and N. E. Graham, 2002: Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quart. J. Roy. Meteor. Soc.*, **128**, 2145–2166.

Miller, A. J., 1992: Algorithm AS 274: Least squares routines to supplement those of Gentleman. *Appl. Statist.*, **41**, 458–478.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16663–16682.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and

J. G. Powers, 2007: A description of the Advanced Research WRF version 2. Technical report, National Center for Atmospheric Research, NCAR/TN-468+STR. 88pp. [Available on-line at http://www.mmm.ucar.edu/wrf/users/].

Wicker, L. J. and W. C. Skamarock, 2002: Time splitting methods for elastic models using forward time schemes. *Mon. Wea. Rev.*, **130**, 2088–2097.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. International Geophysics Series, Academic Press, San Diego, 464 pp.

Wojcik, G. S., H. L. Szymczak, R. J. Alliss, R. P. Link, M. E. Craddock, and M. L. Mason, 2005: Deep-space to ground laser communications in a cloudy world. *Proceedings of the SPIE Free-Space Laser Communications V*, SPIE, volume 5892, 17–27, San Diego, CA, doi:10.1117/12.615435.

Wojcik, G. S., H. L. Szymczak, R. J. Alliss, and M. L. Mason, 2006: The role of satellite-derived climatologies in deep-space to ground laser communications. *14th Conference on Satellite Meteorology and Oceanography*, American Meteorological Society, Sante Fe, NM. [Available on-line at http://ams.confex.com/ams/htsearch.cgi].