10.1 AUTOMATED ANALYSIS AND FORECAST TECHNIQUES FOR CEILING AND VISIBILITY ON THE NATIONAL SCALE

Paul H. Herzegh*[1], Gerry Wiener[2], Rich Bankert[1], Richard Bateman[1], Beth Chorbajian[1] and Matthew Tryhane[1]

[1]National Center for Atmospheric Research, Boulder, CO 80307
[2]Naval Research Laboratory, Monterey, CA 93943

1. INTRODUCTION

Recent development of an exploratory automated ceiling and visibility (C&V) analysis and forecast system by the FAA's National Ceiling and Visibility (NCV) product development team utilizes expert system methodology to merge numerical and observational inputs in the synthesis of current analyses and forecasts out to twelve hours. Trial products covering the continental U.S. have yielded encouraging early results and useful insight into directions for future development.

This paper provides a brief overview of the methodologies used for analysis and forecast product generation in the NCV system. We also describe early development steps associated with two new techniques: (*i*) use of Knowledge Discovery in Databases (KDD) to improve real-time analysis of ceiling in data-poor areas, and (*ii*) an observations-based forecast technique that utilizes forecast rulesets derived from data mining of long-term observational records at selected sites.

2. OVERVIEW OF THE NCV SYSTEM

The NCV product development team is researching and developing products targeted for operational use directly by the flight service station briefer, pilot, dispatcher, controller, and other end user. Since automation is key to enabling frequent product updates and around the clock operation, our work relies upon unattended, computer-aided techniques. These include *(i)* expert system methods to conditionally manipulate data inputs and manage functional interactions among them, and *(ii)* fuzzy logic techniques to formulate a consensus product (*e.g.*, analysis, or forecast), generally based upon the selective merging of individual data and product sub-elements.

Gridded Analyses

The NCV system produces gridded analyses of current ceiling, visibility and flight category conditions. These are provided at the RUC forecast model native grid resolution (currently 20 km) and provide ready access to supporting interactive data overlays. The concept here is to provide rapid (15 min) updates of current C&V conditions in graphical form while incorporating tools that allow concurrent examination of METARs, TAFs, AIRMETs, and satellite and NEXRAD imagery. The analysis of present conditions given by the current NCV prototype product covering the continental U.S. (CONUS)

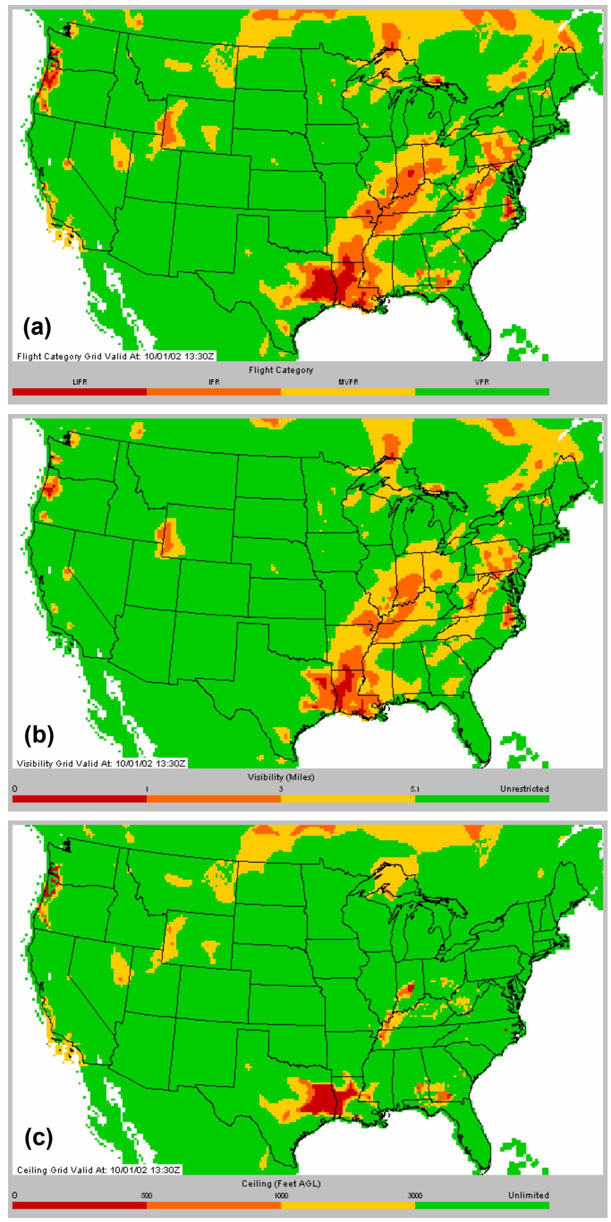


Figure 1. Current conditions as given by METAR reports analyzed by the NCV automated system. (a) Flight Category (Low IFR, IFR, Marginal VFR and VFR) as determined from the values in (b) and (c). (b) Visibility in statute miles. (c) Ceiling in feet AGL. Prototype display is continuously accessible at www.rap.ucar.edu /projects/cvis.

* *Corresponding author address*: Paul H. Herzegh, NCAR, P.O. Box 3000, Boulder, CO 80307; Email: herzegh@ucar.edu

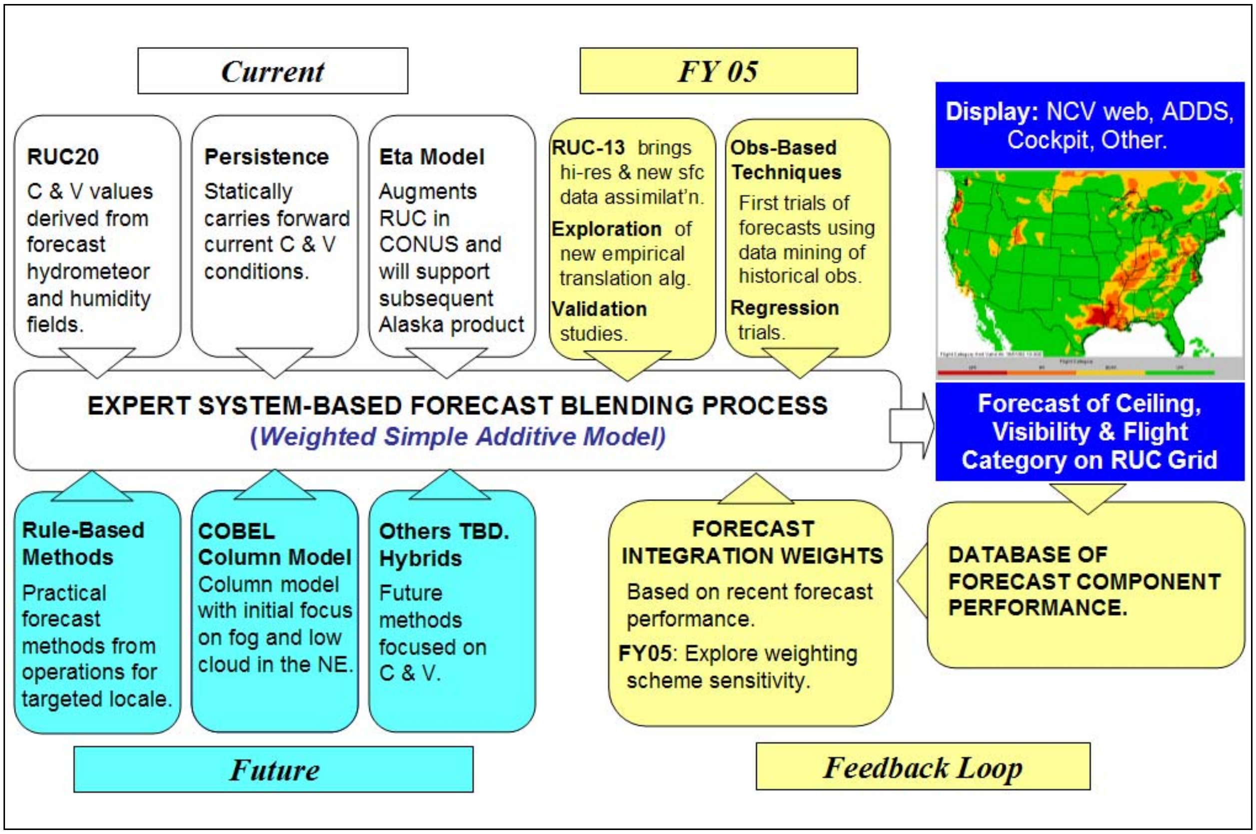


Figure 2. Schematic representation of NCV forecast system components for CONUS C&V in use today and those planned for future implementation. Forecast components flow to the expert system-based automated merging process shown at center. Real-time scoring of component forecast skill feeds back to optimize the weighting of forecast components in the automated merging process.

is illustrated in Fig. 1.

1-12h Forecasts

The NCV forecast product is formulated as a consensus among a variety of parallel forecast techniques comprised of numerical modeling and observations-based approaches. The forecast product is updated hourly. The NCV plan makes use of several forecast components and techniques, with future additions planned for implementation. The key elements of the conceived forecast system are illustrated schematically in Fig. 2, which also shows the fuzzy logic-based system currently in place to weight and merge forecast information through an additive model.

3. IMPROVING CEILING ANALYSES USING KNOWLEDGE DISCOVERY IN DATABASES

Current NCV Practice

Characterizing ceiling behavior in the regions *between* routine METAR observations is one of the key challenges to be met in improving regional analyses of ceiling height needed for flight planning and in-flight guidance. The NCV system uses a natural neighbor interpolation scheme to estimate ceiling height between METAR sites. This interpolation scheme is based on Voronoi polygons and takes into account the geometry of METAR site location with respect to the grid point in question. The ceiling value at a grid point is formed as a weighted sum of neighboring METAR reports. The choice of neighboring METAR sites affecting a grid point is determined by an analysis of the Voronoi polygons, and the weights are determined by an area weighting scheme.

Where GOES data indicate a cloud-free region, the interpolated ceiling height is raised to a value corresponding (effectively) to unlimited ceiling conditions. This is a simple, first-order approach to gap-filling. The approach can successfully represent clear areas within the analysis, but takes no step toward improvement of ceiling values in cloudy regions between METAR sites. Any information in the GOES data beyond that indicating cloudy vs. clear conditions is unused.

A second area under development is the accurate derivation of ceiling values (which are not a direct product of model predictions) from model-predicted meteorological fields. This translation is a critical step in extracting predicted ceiling fields from numerical model results. Common practice today is to apply the Stoelinga and Warner (1999) translation algorithm (or a related adaptation of that approach) to model output to derive ceiling and visibility fields. The Stoelinga-Warner (hereafter SW) technique utilizes theoretical and empirical relationships between light extinction and hydrometeor characteristics to translate model-predicted hydrometeor fields to useable ceiling and visibility values.

Overview of a KDD Approach

The problem of estimating cloud ceiling heights at locations where no ground observations exist is a difficult one for both satellite and numerical model applications. While the modeling of weather phenomena has been driven by physical laws verified by data, some parameters (e.g, cloud ceiling height) the conceptual modeling can be data-driven. Through proper analysis, data relationships representing the physics implicit in the data are empirically discovered.

Data mining methods, used in a Knowledge Discovery from Databases (KDD) procedure, are applied to the cloud ceiling height assessment problem (Bankert, et al, 2004). The KDD process involves collection and processing of data and the application of data mining tools to stored data records to uncover the relationships that represent physical laws implicit in the data (Fayyad, et al, 1996).

In the previous research at the Naval Research Laboratory (NRL), satellite (including GOES-10) and numerical model (COAMPS^{TM}) data were collected within the KDD procedure to determine those relationships that could provide estimates of cloud ceiling heights at California METAR stations. While there is no *direct* indication of ceiling height in GOES data, it is justified to reason that specific factors such as the existence of cloud and a variety of its detectable characteristics and patterns of occurrence should reflect significantly on the probability that certain ceiling characteristics are associated with the cloud. These detectable characteristics might include, for example, its type (*i.e*. stratus, cirrus, etc.), optical thickness, the height of cloud top, observed ceiling values at neighboring METAR sites, the ceiling values associated with similar GOES cloud signatures in the same region in the past, and many others. KDD techniques provide systematic means to find and categorize the patterns and relationships among factors that are found to affect a targeted characteristic – in this case the existence of a cloud ceiling and its height. In addition, KDD techniques provide means to develop a simple model from existing data and apply that model to retrieve estimates of the targeted characteristic. Such a data-derived model for ceiling height is outlined below.

The NRL KDD approach to cloud ceiling estimation begins with selection of data sources and collection of hourly data over a multi-year period for the domain of interest. In the test example described here, a U.S. west coast domain was selected. A 2.5 year record of data was compiled to relate hourly cloud ceiling observations made at 18 California METAR sites with corresponding data (coincident in time and location) from various satellite platforms and the COAMPS^{TM} numerical model. The data can be described as follows:

- METAR ceiling height is the ground truth parameter (dependent variable) used in the KDD process for this study. Other METAR observation parameters (e.g, visibility) are collected and could be used similarly in future studies.
- Satellite data covering each METAR site at each hour were extracted from GOES-10, NOAA's polar-orbiting Advanced Very High Resolution Radiometer (AVHRR), and the Defense Meteorological Satellite Program's Special Sensor Microwave Imager (SSM/I). Satellite data selected included all sensor fields, plus results from a microwave sensor satellite rain rate algorithm, a cloud optical depth algorithm, a low cloud product, and a cloud top height algorithm. Only GOES-10 parameters will be used in the data mining described below.
- Model data covering each METAR site were taken hourly from COAMPS^{TM} (Hodur, 1997; Hodur *et al*., 2002) triply-nested (81, 27, 9 km) mesoscale model runs (12 hour). Forty-two model parameters were selected based on a priori assumptions about which parameters might have most influence on cloud ceiling height.

The data outlined above reside in a database whose structure is optimized to support efficient data mining. The database is continuously updated (daily) as new observations and model results are acquired.

Data mining was performed on the database to uncover relationships among the variables that balanced predictive skill with model generality for an algorithm focused on the estimation of cloud ceiling. Classification models (represented as a decision tree) were produced through use of the Rulequest Research C5.0 data mining tool (Quinlan, 1993; Rulequest Research, 1997-2004). Rule-based predictive models (numerical output) were produced through use of the Rulequest's Cubist algorithm. Applying both C5.0 and Cubist, and through repeated testing and combination, a three-step system for estimating cloud ceiling conditions was developed. After establishing the algorithms at each step, the decision/estimation process for a given data record presented to the three-step system can be summarized as follows:

Step 1: Classification Algorithm (C5.0) – Ceiling vs. No Ceiling?

If classified as ceiling, proceed to Step 2; otherwise, no ceiling is output.

Step 2: Classification Algorithm (C5.0) – Is Ceiling Below 1000 m or Above 1000 m?

If ceiling is below 1000 m, proceed to Step 3; otherwise "high ceiling" is output.

Step 3: Rule-based Predictive Algorithm (Cubist) to Estimate Ceiling Height.

Cloud ceiling height estimate is output.

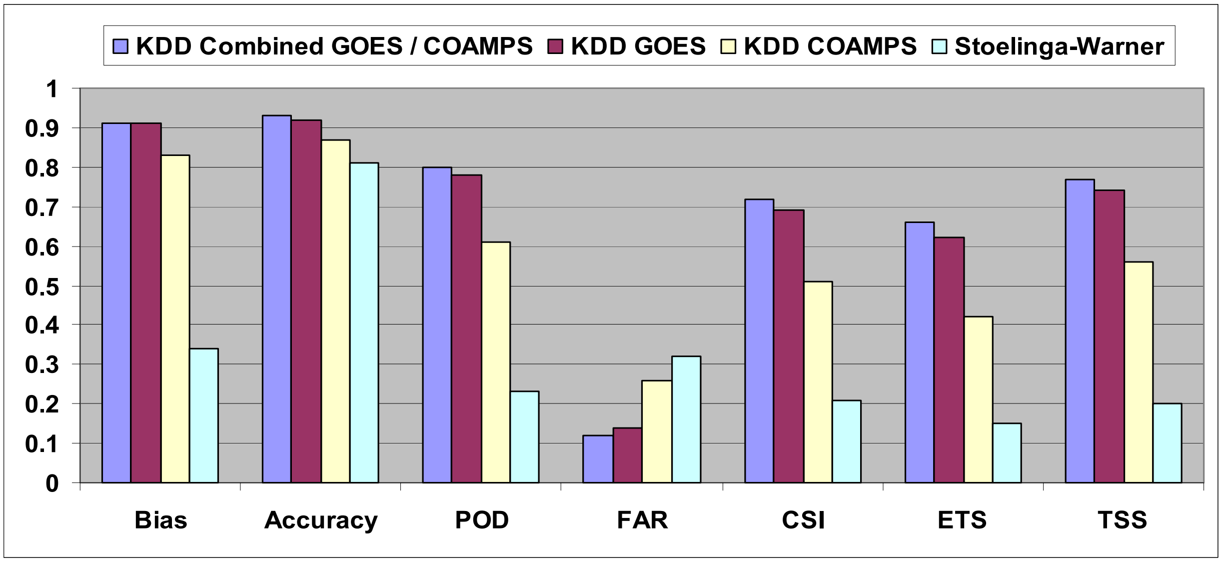


Fig. 3: Comparison of results of four methods for ceiling vs no ceiling classification. The three KDD-based methods use GOES-10 input data (only), COAMPS model input (only), and both GOES-10 and COAMPS input, respectively. The SW method applies the Stoelinga-Warner translation algorithm to COAMPS model results. POD – Probability of Detection; FAR- False Alarm Ratio; CSI – Critical Success Index; ETS – Equitable Threat Score; TSS – True Skill Score

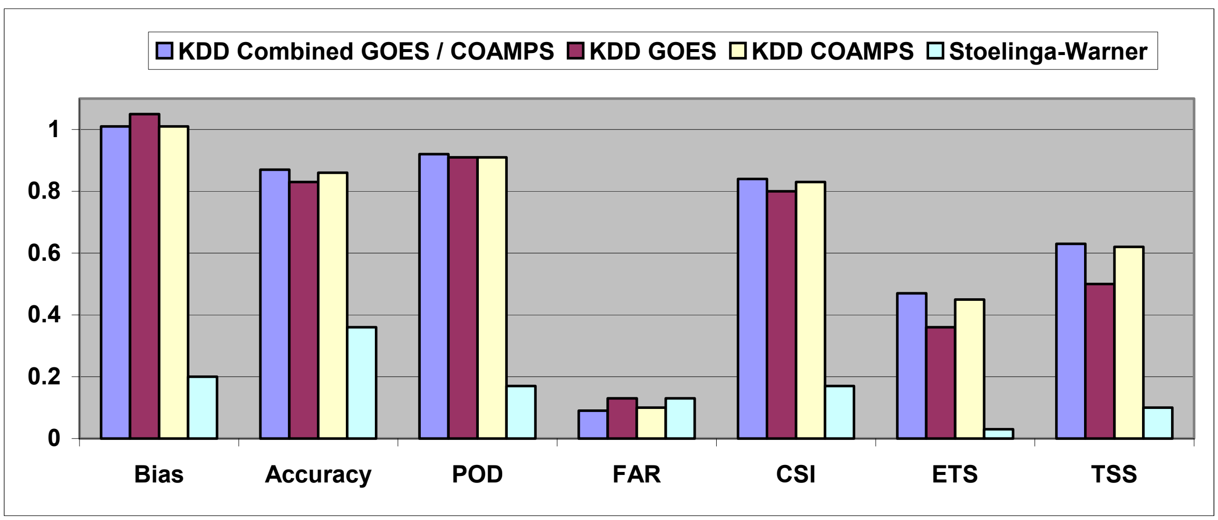


Fig. 4: Comparison of results of four methods for low ceiling vs high ceiling classification. The three KDD-based methods use GOES-10 input data (only), COAMPS model input (only), and both GOES-10 and COAMPS input, respectively. The SW method applies the Stoelinga-Warner translation algorithm to COAMPS model results. POD – Probability of Detection; FAR- False Alarm Ratio; CSI – Critical Success Index; ETS – Equitable Threat Score; TSS – True Skill Score

KDD Ceiling Experimental Results

To obtain an estimate of the performance capabilities of the 3-step system, the hourly data records were divided into training and testing sets for each step. Test results for each step under daytime conditions over the 18 California METAR stations have been excerpted from Bankert *et al.* (2004). Results from three KDD-derived models for ceiling height estimation (using GOES-10 inputs only, COAMPS™ inputs only, and GOES-10 plus COAMPS™ inputs) are compared with corresponding results of the SW translation algorithm (Stoelinga and Warner, 1999) as applied to COAMPS model results. Fig. 3 is a bar chart of various performance measures for ceiling vs no ceiling classification (Step 1 of the cloud ceiling algorithm). The KDD algorithm using both data types produced the best results with all KDD algorithms outperforming SW.

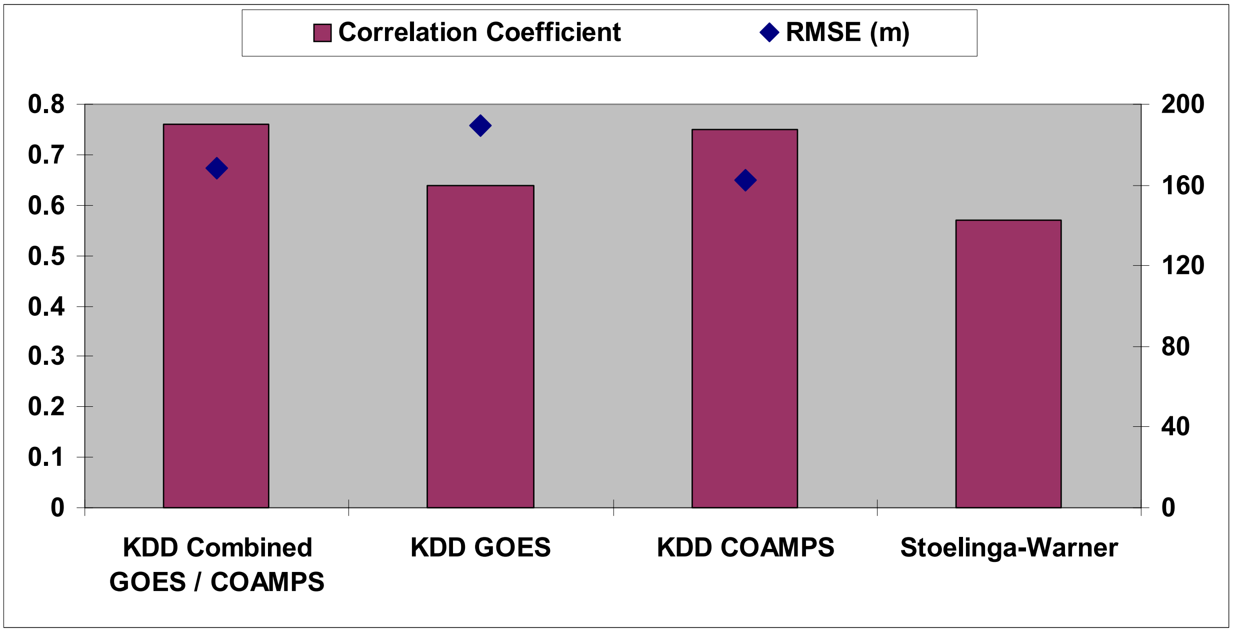
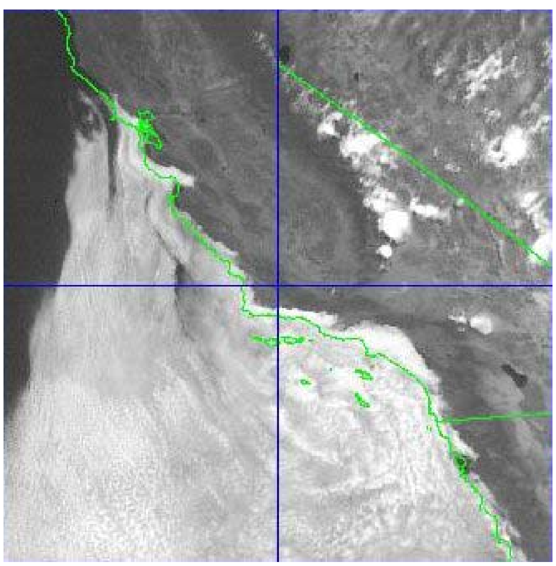


Fig 5: Comparison of results of four methods for low cloud ceiling height estimation. The three KDD-based methods use GOES-10 input data (only), COAMPS model input (only), and both GOES-10 and COAMPS input, respectively. The SW method applies the Stoelinga-Warner translation algorithm to COAMPS model results. The RMSE for SW (not shown on the graph) is 715 m.



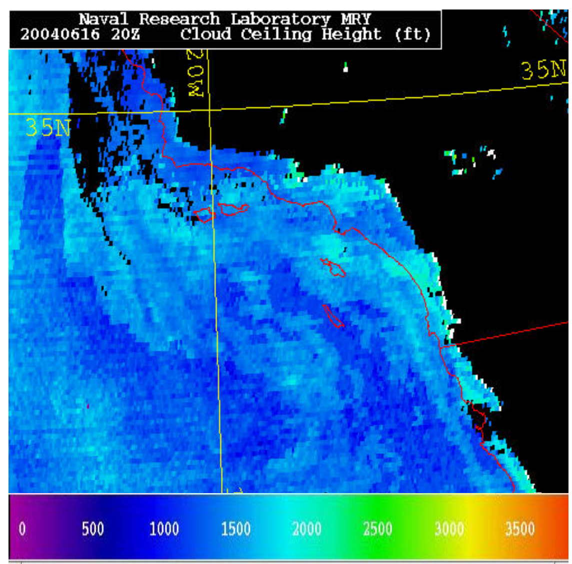


Fig 6: 16 June 2004 20 UTC visible image (left) and corresponding KDD GOES-10 cloud ceiling image (right) of Southern California and adjacent water with heights estimated in feet.

Similar results were found for Step 2 (low ceiling vs high ceiling classification) as seen in Figure 4. For step 3 – estimation of low cloud ceiling height – the correlation coefficient (CC) and root mean square error (RMSE) were computed (Figure 5). Again, the KDD algorithms outperform SW with the combined GOES/COAMPSTM and COAMPS^{TM}–only algorithms producing similar results.

Along with performance analysis, the relationships discovered through the data mining process and used in the derived system can also be analyzed. For example, the COAMPS^{TM} parameter representing the difference in temperature between 10 m and 1500 m is a dominant variable in the low/high ceiling classifications. This variable is required to be relatively small for low ceilings and may be representing the cool, less stable environment in low cloud ceiling situations. Further discussion

METAR LOCATION	*METAR CEILING HEIGHT (FT)*	*KDD GOES-10 CEILING HEIGHT (FT)*
KSBA (Santa Barbara)	*1300*	*1204*
KVNY (Van Nuys)	*1800*	*1915*
KLAX (Los Angeles)	*1800*	*1525*
KLGB (Long Beach)	*1400*	*1315*
KSNA (Santa Ana)	*1600*	*1355*
KOKB (Oceanside)	*1900*	*2014*
KSAN (San Diego)	*1600*	*1630*

Table 1: Comparison of observed cloud ceiling heights (METAR) and KDD GOES-10 cloud ceiling heights at various locations on 16 June 2004 at 2000 UTC.

of the results and analysis can be found in Bankert, *et al.* (2004).

These results provide strong encouragement for continued development of the KDD technique and corresponding establishment of data collection and database functionality as part of the NCV data handling infrastructure. Application of GOES-10 and GOES-12 KDD models to perform gap-filling between METAR sites should significantly benefit the ceiling analysis function within the NCV system. Since the NCV system makes use of the operational RUC20 model rather than COAMPS™, development of KDD techniques for NCV use is currently addressing a change of model inputs. Use of a GOES plus RUC20 KDD model will enable comparison with the current RUC20 plus SW ceiling height prediction. Based upon the results shown in Fig. 3-5, it is expected that the KDD model will achieve significantly improved results, and thus improved skill in NCV ceiling height predictions.

The four geographic regions planned for data collection and subsequent data mining are Iowa, Northeast Texas, Gulf Coast, and Mid-Atlantic region east of Appalachians. These regions were chosen due to the quantity and density of METAR stations and the homogeneous nature (both geographically and climatologically) of each region. Half of the stations in each region will be used for training both ceiling and visibility algorithms and the other half will be the testing set. GOES-12, RUC, and METAR data will be retrieved from archived data sets comprising one to two years of hourly data. Additional research will involve development of KDD-derived ceiling and visibility translation algorithms for the RUC model at individual station locations.

Real-Time Application

As a follow-on to the NRL research, real-time application of the developed algorithm is underway. The NRL GOES-10 cloud ceiling algorithm is currently producing hourly output of cloud ceiling heights for the Southern California region. Figure 6 is an example of these products. Comparisons to ground observations at various METAR locations for this example (16 June 2004, 2000 UTC) indicate a fairly reliable representation of the cloud ceiling heights (Table 1).

4. AN EMERGING DATA MINING-BASED FORECAST TECHNIQUE

Wiener et al. (2004, this volume) describe the methodology employed in a newly-developed technique for 1-12h forecasting at a given site utilizing rulesets derived from data mining of the long-term record of observations at that site. In this section we further describe preliminary results from use of this data mining (DM) forecast technique for ceiling and visibility at six cities in the U.S.

Single-Station Trials

Individual forecast rulesets are established for each forecast site, for each of the forecast lead times used in this trial (1, 3, 5 and 7 h), and for each target forecast parameter (in this case, ceiling or visibility). Each ruleset is the result of a separate data mining exercise over the long-term archive of observed and derived parameters at the forecast site.

We begin with a simple implementation of the technique in which data mining is performed over a site archive containing the target forecast parameter, associated operational meteorological observations (*e.g.*, temperature, humidity, winds, precipitation, pressure, etc.), and derived parameters such as dewpoint depression and tendencies. No additional forecast information is used in this simple trial approach. For five of the six sites reported here, forecasts were initiated at each hour through the two-year test period 2003 through 2004. The rulesets used were derived from data mining of the training data set at each site, which covered the period 1980 to present, but excluded the 2003-2004 test period. Due to the move of ASOS facilities upon the opening of Denver International Airport, our Denver studies are centered at the former Stapleton Airport site, and the forecast test period used was 1993-1994.

The DM forecasts specified either IFR or VFR conditions for ceiling and visibility. Specification of IFR conditions was considered an IFR event, while specification of VFR conditions was considered a null event. We chose to compare DM forecast skill with that of persistence, which is itself a particularly skillful forecast method over short lead times such as those examined here.

A selection of skill metrics for DM and persistence forecasts of IFR events at the six sites are shown in Figs. 7 and 8. The metrics include bias, probability of detection (POD), false alarm ratio (FAR) and critical success index (CSI). Given that these trials come at an

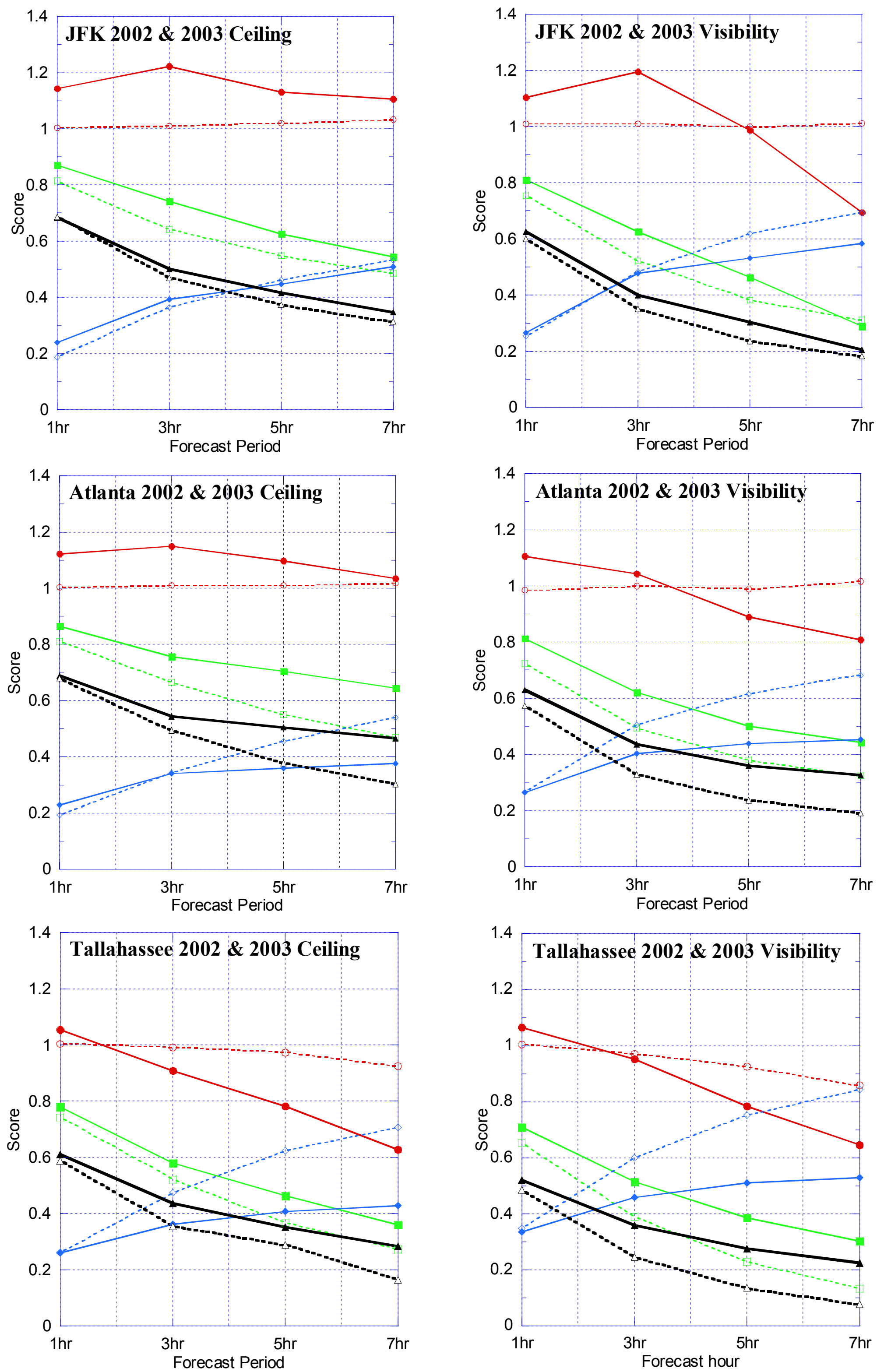


Figure 7. Plots of ***bias*** *(red),* ***probability of detection*** *(green),* ***false alarm ratio*** *(blue) and* ***critical success index*** *(black) for single-site DM forecast lead times of 1, 3, 5 and 7 hours for the occurrence of IFR conditions in ceiling and visibility at the sites shown. Solid lines show the DM forecast scores. Dashed lines show the corresponding persistence forecast scores.*

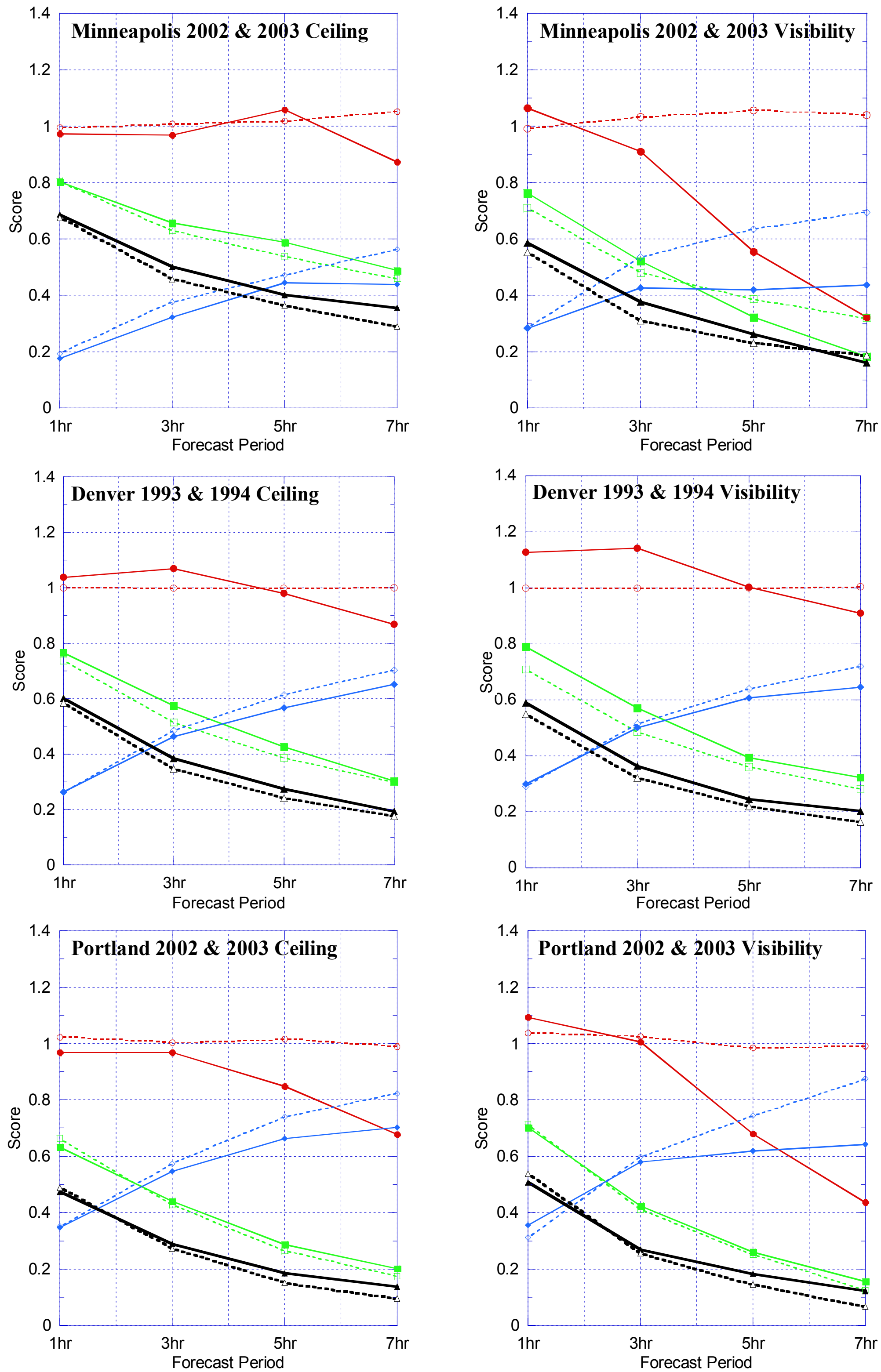


*Figure 8. As in Fig. 7. Plots of **bias** (red), **probability of detection** (green), **false alarm ratio** (blue) and **critical success index** (black).*

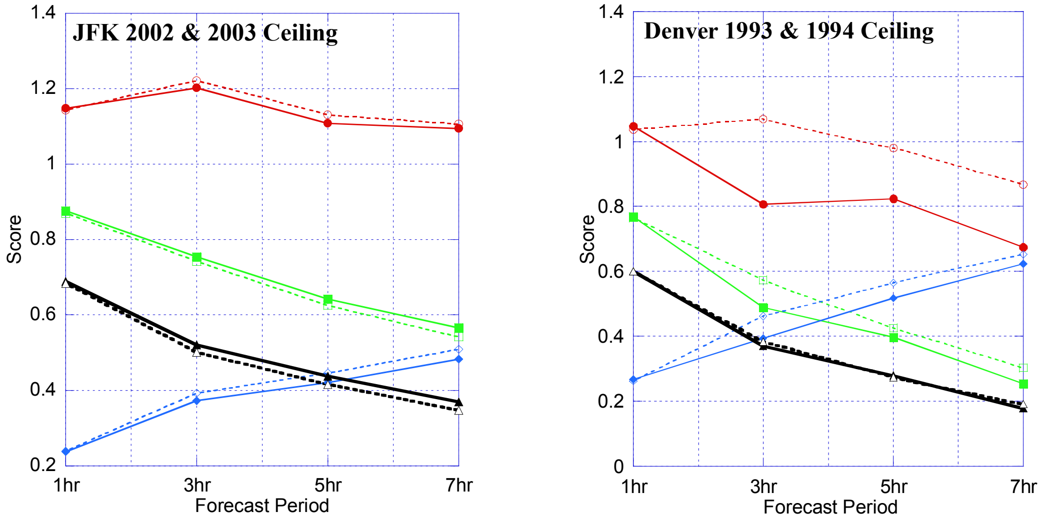


*Figure 9. Plots of **bias** (red), **probability of detection** (green), **false alarm ratio** (blue) and **critical success index** (black) comparing single-site and multi-site DM forecasts at JFK and Denver. Solid lines show forecast scores for multi-site method. Dashed lines show forecast scores for single-site method.*

early stage with little effort yet to explore further adjustments or techniques to improve forecast performance, we find the results in Figs. 7 and 8 to be very encouraging. The principal points drawn from examination of these results to date are as follows.

- DM forecast performance as indicated by CSI scores meets or exceeds persistence at 1h at all sites except Portland. At all sites except Minneapolis (7h visibility), DM CSI scores exceed persistence at 3, 5 and 7 h.
- DM scores for probability of detection and false alarm ratio at 1h are equal or superior to those of persistence at all sites except Portland and JFK. These scores frequently further improve over persistence at 3, 5 and 7h. Minneapolis (visibility) is a clear exception, where POD for the DM method falls below that for persistence at 7h.
- Whereas persistence yields bias scores generally close to 1.0 (indicating little tendency to over or under-forecast IFR conditions), the DM technique produces biases that vary far more from site to site. JFK ceiling and Minneapolis visibility, for example, show strong tendencies to over-forecast and under-forecast, respectively. Ongoing work is examining the roles of weighting parameters (set within the data mining process), additional derived forecast parameters, and the use of neighboring sites in reducing the extremes in DM bias values and improving overall forecast skill.

<u>Multi-Station Trials</u>

Observing the areal behavior of ceiling and visibility provides evidence that use of observations across a set of neighboring sites may yield improvement in forecasts for a target site. For example, Leyton and Fritsch (2003) found improvement in statistical forecasts of ceiling and visibility when using neighboring sites surrounding target sites in the upper midwest.

We believe that the multi-station approach is an important path for examination of DM forecast improvement. The simple preliminary tests made to date are centered on forecasting for JFK and Denver. At JFK we utilized the long-term (1980-present) data archives for neighboring LaGuardia and Newark airports. At Denver's former Stapleton Airport site, we used Buckley Field and Arapahoe Airport as the neighboring sites. In the Denver case, the training data ran from 1984 through 1992.

Our initial results demonstrate that there seems to be potential in this multi-site approach, but also, as expected, that non-optimal use of neighboring sites can degrade overall forecast performance as well. The forecast results shown in Fig. 9 for JFK compare performance *with* use of neighboring sites (solid line) and without use of neighboring sites (dashed line). Slight improvement of a few per cent is shown in each of the scores shown for 3, 5 and 7h. Improvement at 1h was essentially nil. In contrast, the results for Denver in Fig. 9 show that use of the two neighboring sites degraded bias scores substantially below 1.0. As might be expected, the bias change was accompanied by a significant decreases in POD and FAR. The CSI score declined very slightly at 3, 5 and 7h.

A key aspect of these preliminary tests is that the choice of the neighboring stations was quite arbitrary, based more on the availability of data than on any conceptual model of ceiling or visibility behavior at the target sites. The shorter period of archived data available for the Denver case is also a disadvantage. Further tests will explore the impact of this shorter training on the representativeness of the forecast results.

Further development of the DM forecast approach, in both single and multi-site forms, requires systematic evaluation of the contributing roles of observed and derived meteorological parameters, potential use of forecasted parameters within the DM forecast method, weights assigned within the DM ruleset formation process, and optimal use of neighboring sites. Rather than 'one size fits all', our experience indicates that the differing circumstances and regimes from one site to another will demand that the DM process be optimized on a site-by-site basis. As a result, it is also necessary to balance optimization research/effort against anticipated performance gains.

5. SUMMARY

This paper outlines the plans and early results of a long-term R&D program directed toward improved automated analysis and forecast tools to aid avoidance of in-flight C&V hazards. We principally target the needs for C&V information within the general aviation community, where improved access and utilization of briefing and in-flight guidance information can lead to a significant improvement in flight safety.

A three-step KDD (Knowledge Discovery from Databases) process has been developed to estimate ceiling height in regions where only satellite and model data are available. The process utilizes decision steps based upon data mining of satellite and model data archives. Initial trials of the method at 18 sites in California have yielded very encouraging results, showing ceiling height estimation performance that is consistently better than that derived from application of the Stoelinga-Warner translation algorithm to the model data available for the sites. Further work is directed toward trials to evaluate the performance of the method at a broader range of sites and meteorological regimes across the U.S.

A second new technique utilizes data mining of long-term data archives to produce rulesets for 1-12h forecasting of ceiling and visibility at selected sites. Though the first trials reported here include little effort to optimize the method to improve performance, the results are encouraging in that the data mining forecasts frequently exceed the performance of persistence at 1-7h. A preliminary trial utilizing data from neighboring sites to contribute to the forecast at a target site showed slight improvement over single-site forecasts at JFK airport, and lowered performance for forecasts at a Denver site. The data mining forecast results as a whole show excellent promise, but also demonstrate the importance of systematic diagnostic trials in developing predictors and compatible neighboring sites that most improve performance, rather than degrade it.

Acknowledgements

The authors gratefully acknowledge contributions to this work by Stan Benjamin (NOAA/FSL) and Ted Tsui (NRL). This research is in response to requirements and funding by the Federal Aviation Administration (FAA). The views expressed are those of the authors and do not necessarily represent the official policy or position of the FAA.

REFERENCES

Bankert, R.L., M. Hadjimichael, A.P. Kuciauskas, W.T. Thompson and K. Richardson, 2003: Remote cloud ceiling assessment using data mining methods. J. Appl. Meteor., accepted for publication.

Fayyad, U.M., Piatetsky-Shapiro, G., and Smyth, P. 1996: From data mining to knowledge discovery: an overview. In Advances In Knowledge Discovery And Data Mining , eds. U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, AAAI Press/The MIT Press, Menlo Park, CA., 1-34.

Hadjimichael, M., R.L. Bankert, A.P. Kuciauskas, K.L. Richardson and G.N. Vogel, 2003: Application of knowledge discovery from databases to remote weather assessment. Proc. 19th Conf. on Interact. Info. Proc. Systems., AMS, Boston.

Hodur, R.M., 1997: The Naval Research Laboratory's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS). Mon. Wea. Rev., 125, 1414-1430.

*Hodur, R.M., J. Pullen, J. Cummings, X. Hong, J.D. Doyle, P. Martin and M.A. Rennick, 2002: The coupled ocean/atmosphere mesoscale prediction system (COAMPS^{TM}). Oceanography, **15**, 88-89.*

*Leyton, S.M. and J.M. Fritsch, 2003: Short-term probabilistic forecasts of ceiling and visibility utilizing high-density surface weather observations. Weather and Forecasting, **18**, 891-902.*

Quinlan, J.R, 1993: C4.5: Programs for machine learning. Morgan Kaufmann Pub., San Mateo, 302 pp.

Rulequest Research , 1997-2004: C5.0 and Cubist. http://www.rulequest.com

*Stoelinga, M.T. and T.T. Warner, 1999: Nonhydrostatic mesobeta-scale model simulations of cloud ceiling and visibility for an east coast winter precipitation event. J. Appl. Meteor., **38**, 385-404*

Wiener, G., P. Herzegh, R. Bateman and B. Chorbajian, 2004: Utilizing site-based data mining in national ceiling and visibility forecasting. Preprints, Aviation and Range Meteor. Conf., Hyannis, MA, AMS.