



NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

**NOWCASTING CLOUD FIELDS FOR U.S. AIR FORCE
SPECIAL OPERATIONS**

by

Sean L. Zoufaly

March 2017

Thesis Advisor:
Second Reader:

Wendell A. Nuss
Eric A. Hendricks

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			<i>Form Approved OMB No. 0704-0188</i>	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE March 2017		3. REPORT TYPE AND DATES COVERED Master's thesis
4. TITLE AND SUBTITLE NOWCASTING CLOUD FIELDS FOR U.S. AIR FORCE SPECIAL OPERATIONS			5. FUNDING NUMBERS	
6. AUTHOR(S) Sean L. Zoufaly				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB number ____N/A____.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (maximum 200 words) Nowcasting is a trending subset of numerical weather prediction that aims to produce a highly accurate analysis of current conditions along with a short-term forecast. One of the greatest challenges to a nowcast system operating in data-sparse regions is that of accurately forecasting clouds. Clouds significantly impact a variety of operations, particularly intelligence, surveillance and reconnaissance. A prototype nowcast system is developed and tested on a case of summertime stratus clouds over the Monterey Bay in California. This system ingests high-resolution geostationary satellite data and mesoscale model fields to produce gridded 06-h forecasts of cloud reflectance and probability of cloud. A statistical post-processing technique is applied using Bayesian estimation to train the system from a set of past predictor variables and observed imagery. This approach demonstrates skill over a climatology-based approach and shows an ability to accurately forecast non-typical cloud patterns. It proves to be very computationally feasible for nowcasting. This study lays down the initial framework for a highly accurate nowcast system that can operate anywhere in the world to enable mission success while reducing costs.				
14. SUBJECT TERMS operational nowcasting, low cloud forecasting, cloud reflectance, ISR, Bayesian estimation, statistical post-processing, machine learning			15. NUMBER OF PAGES 111	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UU	

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

**NOWCASTING CLOUD FIELDS FOR U.S. AIR FORCE SPECIAL
OPERATIONS**

Sean L. Zoufaly
1st Lieutenant, United States Air Force
B.S., United States Air Force Academy, 2013

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN METEOROLOGY

from the

**NAVAL POSTGRADUATE SCHOOL
March 2017**

Approved by: Wendell A. Nuss
Thesis Advisor

Eric A. Hendricks
Second Reader

Wendell A. Nuss
Chair, Department of Meteorology

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Nowcasting is a trending subset of numerical weather prediction that aims to produce a highly accurate analysis of current conditions along with a short-term forecast. One of the greatest challenges to a nowcast system operating in data-sparse regions is that of accurately forecasting clouds. Clouds significantly impact a variety of operations, particularly intelligence, surveillance and reconnaissance.

A prototype nowcast system is developed and tested on a case of summertime stratus clouds over the Monterey Bay in California. This system ingests high-resolution geostationary satellite data and mesoscale model fields to produce gridded 06-h forecasts of cloud reflectance and probability of cloud. A statistical post-processing technique is applied using Bayesian estimation to train the system from a set of past predictor variables and observed imagery.

This approach demonstrates skill over a climatology-based approach and shows an ability to accurately forecast non-typical cloud patterns. It proves to be very computationally feasible for nowcasting. This study lays down the initial framework for a highly accurate nowcast system that can operate anywhere in the world to enable mission success while reducing costs.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
A.	IMPORTANCE OF NOWCASTING CLOUD FIELDS	1
B.	CUSTOMER NEEDS FOR NOWCASTING	1
C.	MOTIVATION AND SCOPE OF RESEARCH.....	3
D.	BENEFITS OF STUDY.....	5
II.	BACKGROUND	7
A.	NOWCASTING	7
B.	MACHINE LEARNING AND BAYESIAN INFERENCE IN NUMERICAL WEATHER PREDICTION.....	9
C.	PREVIOUS CLOUD FORECASTING RESEARCH.....	13
III.	DATA/METHODOLOGY	17
A.	OVERVIEW	17
B.	DATA SET.....	18
1.	NOAA GOES 15 (Geostationary Operational Environmental Satellite) Imagery	18
2.	13 km Resolution NCEP Rapid Refresh (RAP)	20
C.	ANALYSIS PROGRAM	20
1.	VISUAL	20
2.	GARP	21
3.	Grid and Multi-quadric Interpolation	21
D.	SATELLITE PROCESSING AND INGEST METHODS.....	23
E.	FORECAST METHODS	26
1.	Overview	26
2.	Bayesian Estimation Method	26
3.	Application of Bayesian Estimation	29
4.	Predictor Variable List.....	31
5.	Three Data Grouping Schemes.....	33
F.	DATA ANALYSIS AND SCORING.....	38
IV.	RESULTS	41
A.	OVERVIEW	41
B.	GROUPING SCHEME CROSS VALIDATION.....	41
1.	Pre-processing	41
2.	Generalized Linear Model	43
3.	Performance	50

C.	PERIOD LENGTH CROSS VALIDATION.....	54
1.	Pre-processing	54
2.	Generalized Linear Model	58
3.	Performance	61
D.	GENERAL PERFORMANCE	70
V.	CONCLUSION	81
A.	SUMMARY AND CONCLUSIONS	81
B.	RECOMMENDATIONS FOR FURTHER RESEARCH	83
	APPENDIX: NRL GOES-15 IMAGERY DIMENSIONS	85
	LIST OF REFERENCES	87
	INITIAL DISTRIBUTION LIST	91

LIST OF FIGURES

Figure 1.	Schematic depicts the PEMDAS NOWcasting system bridging the gap between conventional models and actual conditions. Source: PEMDAS (2015).....	2
Figure 2.	PEMDAS 00-h analyses from August 26 2015 2100Z displayed using the VISUAL graphics program	4
Figure 3.	Example of PPD. Source: Warner, M. (n.d.).	10
Figure 4.	TASC forecast process. Source: Kemp and Alliss (2007).....	15
Figure 5.	NRL 0.6- μm GOES 15 Monterey Bay Vis sector (JPEG)	19
Figure 6.	NRL (a) 10.20–11.20- μm (FIR) and (b) 3.80–4.00- μm (NIR) TeraScan images	20
Figure 7.	NPS nowcast 100 x 100 km grid.....	22
Figure 8.	Comparison of original imagery (a) and (c) to imagery after the reflectance conversion (steps 1 and 2) was applied (b) and (d).....	24
Figure 9.	Mapped coastline error shown after processing steps 1 and 2 were applied.....	25
Figure 10.	Summary of NPS nowcast Bayesian estimation prediction process.....	27
Figure 11.	NPS nowcast sub-sampled 144 point grid	30
Figure 12.	Comparison of June 01 2016 1400Z reflectance imagery to RAP 06-h forecast fields.....	32
Figure 13.	Data set 6 normalized LCLD versus reflectivity with no grouping.....	34
Figure 14.	Data set 6 normalized LCLD versus reflectivity with land-sea grouping	35
Figure 15.	Data set 6 normalized LCLD versus reflectivity with <i>k</i> -means clustering.....	37
Figure 16.	Data set 7 silhouette analysis for $k = 6$	38
Figure 17.	Pre-processed normalized model predictor versus observed reflectance	42

Figure 18.	Pre-processed (land-sea) normalized model predictor versus observed reflectance.....	43
Figure 19.	Pre-processed (<i>k</i> -means cluster) normalized model predictor versus observed reflectance.....	43
Figure 20.	No grouping scheme data set 6 a) jumping kernel variance and b) MCMC acceptance rate.....	44
Figure 21.	No grouping θ_2 (LCLD) a) sampling convergence process and b) PDF	46
Figure 22.	No grouping θ_3 (OMEGA) PDF	47
Figure 23.	Land-sea θ_3 (OMEGA) PDFs	48
Figure 24.	No grouping θ_4 (RH) PDF	49
Figure 25.	No grouping θ_5 (variance) PDF	49
Figure 26.	Forecast absolute error distributions of a) no grouping b) land-sea grouping and c) <i>k</i> -means cluster	51
Figure 27.	June 6 2016 2100Z forecast for each data grouping scheme versus observed reflectance imagery	53
Figure 28.	Data set 6 pre-processed training and verification data for a) LCLD, b) OMEGA, and c) RH versus observed reflectance	56
Figure 29.	Data set 9 pre-processed training and verification data for a) LCLD, b) OMEGA, and c) RH versus observed reflectance	57
Figure 30.	Relationship between normalized RH and LCLD in a) data set 6, and b) data set 10	61
Figure 31.	Absolute error of mean forecast value of a) set 6 b) set 7 c) set 8 d) set 9, and e) set 10.....	63
Figure 32.	Data set 9 scatter plots of a) mean forecast value versus observed reflectance, and b) absolute error grouped by cluster	65
Figure 33.	Data set 10 scatter plot of mean forecast value versus observed reflectance	66
Figure 34.	Data set 9 June 20 2016 1700Z a) forecast image, b) observed image, c) probability of cloud product, d) Brier score map, and e) reference map.....	67

Figure 35.	Data set 8 scatter plot of mean forecast value versus observed reflectance.....	69
Figure 36.	Data set 8 June 13 2016 2100Z observed image with reference map included.....	69
Figure 37.	Data set 7 June 06 2200Z grey-scale image a) forecast and b) observed	70
Figure 38.	Data set 8 land-sea “climocast”	71
Figure 39.	Land-sea climatology error distribution for each data set	73
Figure 40.	Probability of cloud heat map for a) data set 9 and b) data set 10 with c) reference map included (continued on next page)	74
Figure 41.	Brier score heat map for a) data set 9 and b) data set 10 with c) reference map included	76
Figure 42.	MAE heat map for a) data set 9 and b) data set 10	78
Figure 43.	Violin plots of absolute error for each valid time in a) data set 9 and b) data set 10	80

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

Table 1.	Data set 6 MAE and Brier scores for each scheme.....	50
Table 2.	Total time of “training step”	58
Table 3.	Model parameter (θ) means	60
Table 4.	MAE and Brier scores for each data set.....	64
Table 5.	Skill scores	72

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ASAPS	Atmospheric Sensing and Prediction System
BCT	bispectral composite threshold method
BE	Bayesian estimation
BMA	Bayesian model averaging
CMG	Cloud Mask Generator
COAMPS	Coupled Ocean Atmosphere Mesoscale Prediction System
ESRL	NOAA Earth System Research Laboratory
FIR	far infrared channel
FITL	Forecaster-in-the-loop
GARP	GEMPAK Analysis and Rendering Program
GLM	generalized linear model
GOES	Geostationary Operational Environmental Satellite
HDI	highest density interval
HRRR	High-Resolution Rapid Refresh model
HWD	Horizontal Weather Depiction
IC	Intelligence community
IR	infrared channel
ISR	intelligence, surveillance and reconnaissance
KAF	k-nearest neighbors analogue forecast algorithm
KDE	kernel density estimation
k-nn	<i>k</i> -nearest neighbors
LCLD	low cloud cover percentage
MAE	mean absolute error
mb	millibar
MCMC	Markov Chain Monte Carlo sampling
MM5	Mesoscale Model 5
MOS	model output statistics
MSE	mean squared error
NCAR	National Center for Atmospheric Research
NCEP	NOAA National Centers for Environmental Prediction

NEMEFO	Neural Meteorological Forecast model
NIR	near infrared channel
NOAA	National Oceanic and Atmospheric Administration
NPS	Naval Postgraduate School
NRL	Naval Research Laboratory
obs	observations
PDF	probability density function
PPD	posterior predictive distribution
RAP	Rapid Refresh model
RH	relative humidity
RUC	Rapid Update Cycle model
TASC	The Analytic Sciences Corporation owned by Northrop Grumman
UAV	unmanned aerial vehicle
VIS	visible channel
WRF	Weather Research and Forecasting model
Z	ZULU, military indicator for Universal Time

ACKNOWLEDGMENTS

I would like to thank Dr. Wendell Nuss. Thank you for helping me to come to NPS as an intern five years ago and to be able to return as a graduate student. Thank you for always making time for me out of your busy schedule; your dedication to your students is extraordinary. It has been a pleasure and a privilege to work with you.

To LCDR Robert “Travis” Wendt, thank you for introducing me to the world of Bayesian estimation, as well as for the countless hours of modifying and running your program. Without your brilliance and dedication, the Results section would be empty. I have enjoyed learning from you, working with you, and getting to know you. Thank you to your family for letting me borrow your time.

Thank you to Dr. Eric Hendricks for your support, patience, hard work and honest feedback. Thank you to Mr. Kurt Nielsen for all of your hard work with the satellite imagery processing, as well as for your friendship and support.

Thank you to Mr. Kim Richardson and the Naval Research Laboratory, for getting us the specific satellite imagery we needed. This research would not have gotten off the ground without it. You were very friendly and helpful. Thank you to Dr. Michael Gauthier of PEMDAS and to Dr. Jeffrey Cetola for your feedback and support. Thank you to Mr. Bob Creasey and Dr. Tom Murphree for your contributions and support.

Thank you to all my professors for being very helpful and putting up with my many questions. Thank you to my fellow students for your friendship and support.

Lastly, thank you Mom, Dad, Meagan, Crystal and Alex for all of your love and support. Thank you, God, for all that I have, and for all of the amazing people mentioned above.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. IMPORTANCE OF NOWCASTING CLOUD FIELDS

Nowcasting is a trending subset of numerical weather prediction that aims to produce a highly accurate analysis of current weather conditions along with a short-term forecast. The goal of nowcast systems is to achieve higher accuracy by focusing efforts on a limited forecast area and time period, as opposed to conventional forecast models that are designed for larger time periods and forecast areas. Producing highly accurate analysis and short-term forecasts is one of the most difficult forecast challenges. This challenge increases the value of human forecasters for their ability to look at and react quickly to real-time data such as in-situ observations and satellite imagery.

Clouds significantly impact a variety of military and civilian activities. The intelligence community (IC) is particularly interested in clouds because they can be obstacles to optical and thermal sensing systems that perform intelligence, surveillance and reconnaissance (ISR). For instance, unmanned aerial vehicle (UAV) sorties can be rendered unsuccessful if the UAV cannot observe its target because clouds obscure its view. UAV operators must carefully consider cloud cover before deciding to sortie, so that time and fuel are not wasted. Related problems to cloud forecasting include turbulence and icing, which also represent significant threats to UAV missions and a variety of other flying operations. Therefore, UAV missions and many other military operations require highly accurate cloud field nowcasting.

B. CUSTOMER NEEDS FOR NOWCASTING

A civilian company called PEMDAS Technologies and Innovations has developed a nowcast system for the U.S. Air Force. The goal of their “NOWcasting” system is to produce accurate analysis and short-term forecasts of current weather conditions to be used primarily by UAV operators in data-sparse regions (Lockhart 2015). The PEMDAS NOWcasting system employs observational nudging techniques to adjust high-resolution regional model forecasts with real-time data from various in-situ observations including sensors attached to UAVs. The NOWcasting system also has the

advantage of updating hourly, whereas conventional models update every six hours. Figure 1 visually depicts how the NOWcasting system bridges the gap between conventional models and reality. Note that conventional model runs occur every 6 or 12 hours and do not reach the forecaster until 3 to 6 hours after the analysis (00-h) time. PEMDAS conducted a major test of its NOWcasting system in Barrow, Alaska, in the summer of 2015.

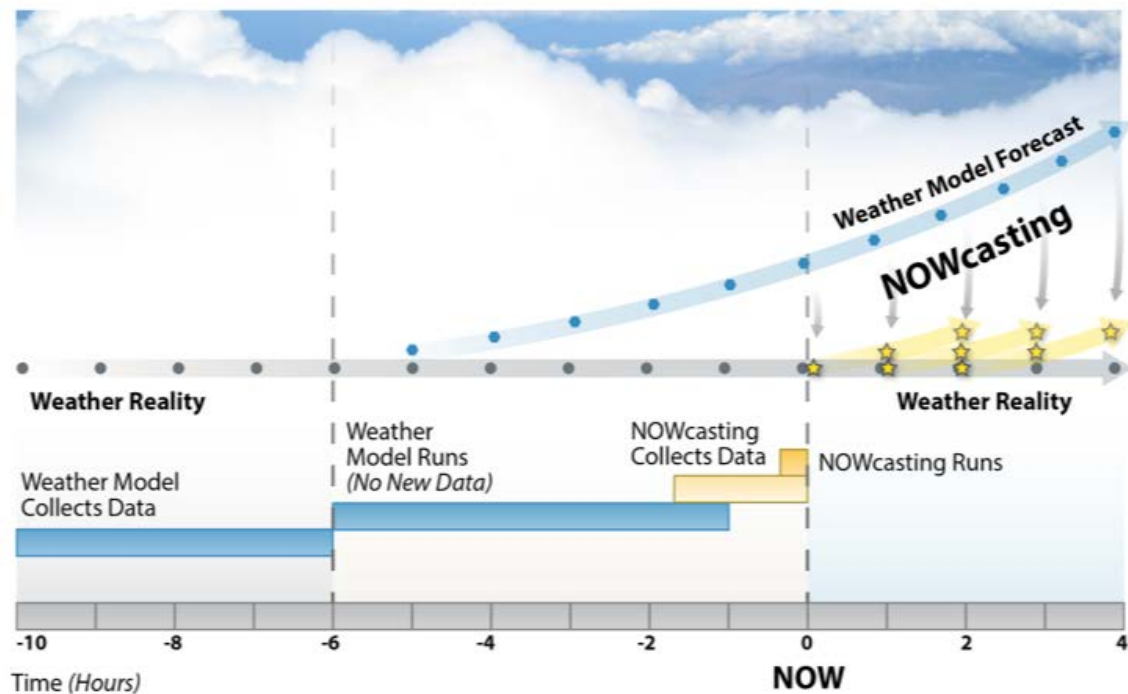


Figure 1. Schematic depicts the PEMDAS NOWcasting system bridging the gap between conventional models and actual conditions. Source: PEMDAS (2015).

Any nowcasting system designed for use by UAV operators would ideally predict cloud fields at high resolutions of 4 km or higher. In addition to horizontal extent, ISR forecasters need to have some information about the vertical extent of the cloud field. This is so that forecasters can determine whether the clouds are thick enough to have mission impacts, which ISR missions are impacted, and what flight level is ideal. At a minimum, this system should produce a forecast output for every hour in the 0–6 hour time frame, and be updated hourly as new observations come in. Clearly, it would need to

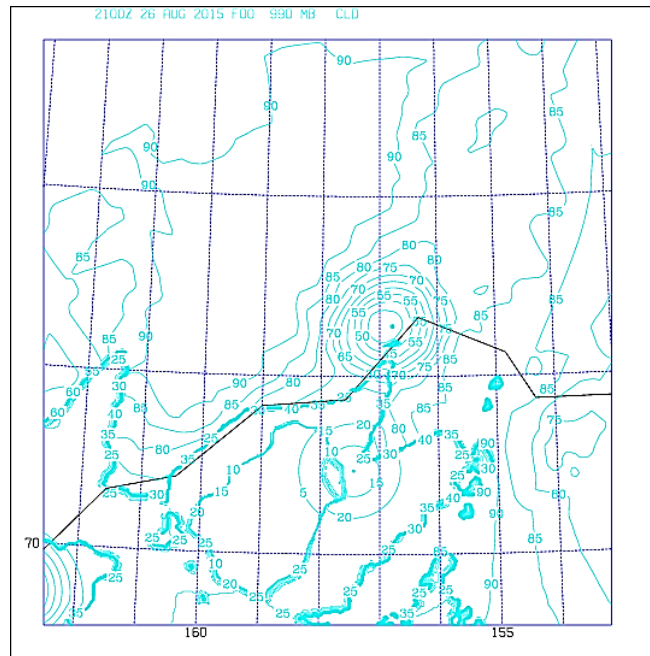
forecast more than just clouds; basic fields such as pressure, temperature and moisture should be included as well as accurate derived fields like icing, turbulence, visibility and precipitation.

It would require an easy-to-use visual output, ideally in a format that ISR weather forecasters are accustomed to. Model derived forecast “pseudo-satellite imagery” that looks just like a satellite image would be one such solution, as ISR forecasters are very adept at using satellite imagery. Another possible output format would be one similar to current AF horizontal weather depiction (HWD) charts, which depict cloud layers by sky coverage category and average layer heights. Forecaster-in-the-loop (FITL) format is also ideal; this is any automated format that can be adjusted by a forecaster before dissemination.

C. MOTIVATION AND SCOPE OF RESEARCH

A crucial challenge for any nowcasting system used to support ISR operations is the generation and depiction of accurate cloud fields despite the scarcity of available observations. As mentioned above, the success of UAV operations and other ISR missions depends upon accurate high-resolution cloud forecasts. These operations frequently occur over regions where ground-based weather observations are scarce, such as Afghanistan. An excellent example of the potential impact of incorporating only coarsely spaced surface observations with sub-optimal model fields (i.e., without leveraging satellite data) to produce cloud field analyses is shown in Figures 2.a and 2.b (derived from PEMDAS provided data generated during the Barrow, Alaska, exercise). Here, concentric circular contours of cloud fraction at given pressure levels highlight the adverse impact that the fusion of surface-based observational data has when satellite data is withheld from the fusion process. More accurate depictions are attainable when satellite data is leveraged in conjunction with surface based observations (not shown).

a) 990 mb (%) cloud cover (light blue) (northern Alaskan coastline shown in black)



b) 980 mb (%) cloud cover (light blue) (northern Alaskan coastline shown in black)

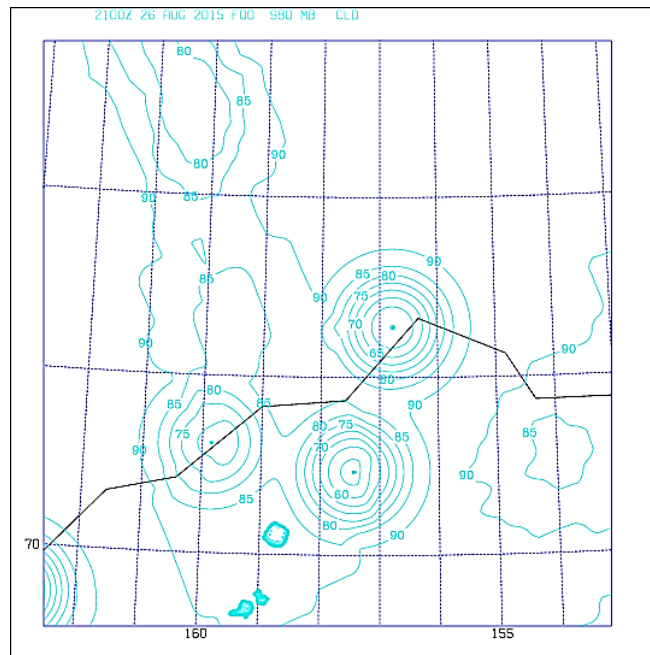


Figure 2. PEMDAS 00-h analyses from August 26 2015 2100Z displayed using the VISUAL graphics program

Another way to increase nowcast accuracy is using statistical methods. This is a popular approach in the field of numerical weather prediction due to limitations in conventional model forecasts. LCDR Robert Wendt, USN, is currently conducting innovative research in predictive modeling with Bayesian estimation (BE) to link predictor-variable fields to an outcome-variable field using a series of weights determined by past occurrences. This method can be applied to increase the accuracy of cloud field forecasts, given climatological data and cloud predictor variables derived from the forecast model.

If the two processes above can be successfully integrated into a nowcasting system, it may help to establish the framework for a highly accurate nowcast system that can be utilized by AF UAV operators in the data-sparse regions in which they operate. Such a system has the potential to greatly increase forecast accuracy and maximize sortie effectiveness while reducing the number of forecasters needed, thus reducing costs. It also has potential for use by various Department of Defense, Intelligence Community, and civilian agencies in forecasting for ISR and other missions.

The primary focus of the study is cloud field forecasting because that is a significant obstacle to a highly accurate nowcast system. The secondary focus of the study is to demonstrate the use of Bayesian estimation to produce more accurate cloud fields. If this can be shown to increase accuracy of cloud fields, it can be used to increase accuracy of other less complicated forecast fields as well.

D. BENEFITS OF STUDY

This study will determine the usefulness of such corrections in nowcasting for UAV operators. The improvements mentioned above could greatly increase forecast accuracy and maximize sortie effectiveness while reducing the number of forecasters needed, thus reducing cost. Because these forecasts are so vital, a human forecaster will still be desired as a sanity check to the model; however, the model may allow one forecaster to do with greater accuracy the work that four do now.

Relative to the operations community, this research will be of interest to agencies in the Department of Defense and intelligence communities that conduct ISR missions.

These techniques have the potential to be useful in other types of forecast models for a variety of civilian and military activities including land operations and maritime operations.

Relative to the meteorological community, this research will be of interest to the fields of numerical weather prediction and cloud forecasting. Bayesian estimation has shown great promise as a post-processing technique that can potentially improve model performance for a variety of forecast uses. It has yet to be applied to maritime cloud forecasting over a two-dimensional grid. Cloud forecasting utilizing high-resolution satellite ingest has not been studied extensively, and cloud forecasting that makes use of the visible channel is rare. Therefore, methods used in this study will lay down a framework for further BE and satellite ingest research that could lead to significant advances in numerical modeling practices.

II. BACKGROUND

A. NOWCASTING

Nowcasting is not a new concept but has developed over the years and has recently been trending due to technological advances that allow for increased short-term accuracy. The term “nowcast” is referenced in literature as far back as the mid-1970s. In its infancy, nowcasting focused on tracking convection and relied heavily on temporal extrapolation of meteorological radar imagery. Over time, more advanced algorithms were applied to track individual storm cells and eventually to account for storm growth and decay. Modern nowcast systems often combine mesoscale models with various real-time observations. Nowcasting capabilities have grown rapidly in recent years due to the development of higher resolution numerical models and denser observation networks (Mass 2012).

For example, the National Oceanic and Atmospheric Administration (NOAA) Rapid Update Cycle (RUC) originated in 1994 as an 80 km resolution model with updates every three hours (Mass 2012). It has since been replaced by the Rapid Refresh (RAP) model with 13 km resolution and hourly updates. In 2016, the High-Resolution Rapid Refresh (HRRR) was introduced operationally as a 3 km resolution version of the RAP that can also assimilate radar data from every 15 minutes (NOAA ESRL 2016). Five km and 1.67 km resolution versions of the Weather Research and Forecasting (WRF) mesoscale model have been utilized operationally by Air Force Weather. The Penn State University/National Center for Atmospheric Research (NCAR) Mesoscale Model (MM5) also boasts resolution as high as 4 km. The United States and other developing countries have seen a huge increase in the density of their weather observation networks in the past several decades (Mass 2012). Although other regions around the world remain data-sparse, there have still been advances that allow for increased data collection in these regions, much of which is not utilized by conventional models. This includes satellite data, lightning, Portable Doppler Radar, mobile observations, clandestine sensors, and data collected from UAVs, in addition to

conventional soundings and surface observations. All of these data sources can and should be exploited in a nowcast system designed for use by the ISR community.

Arguably the most advanced radar-based nowcast system is the National Center for Atmospheric Research (NCAR) Auto-Nowcast System. This system combines satellite, radar, lightning, numerical modeling, upper-air data and a variety of surface observations to output 0–1 hour forecasts of convective storm location and intensity. The satellite imagery ingested is from Geostationary Operational Environmental Satellite (GOES); the system identifies cumulus and cumulus congestus using cloud-type algorithms. Infrared (IR) channel brightness temperatures are also used to identify cloud top warming and cooling. All of the ingested data are processed into predictor fields. A statistical approach called fuzzy logic is employed to produce an overall likelihood from the observed predictor variables (Mueller et al. 2003). This nowcast system is similar to what is required by AF ISR, but more emphasis on non-convective clouds and additional output variable fields is needed.

A simple and commonly applied technique in recent nowcast systems is observational nudging—also known as Four Dimensional Data Assimilation (Mass 2012). Observational nudging takes a numerical model forecast as a first guess and utilizes additional observations to relax the model forecast fields in order to fit the observations. The nudged fields then become the nowcast analysis. The subsequent forecast hours (which are calculated using the nudged fields) become the nowcast forecast hours. The simplicity and low computational costs of this approach allow for frequent new nowcast runs (Schroeder et al. 2006). One system that employs this technique is the Rapidly Relocatable Nowcast Prediction System (RRNPS) produced for the U.S. Army; it uses the Penn State/NCAR MM5 as its first guess and can be run for any location on the globe. The NCAR Four Dimensional Weather System (4DWX) is a similar model created for use at Army test ranges; it updates every three hours.

The PEMDAS NOWcasting system is perhaps the latest system to use observational nudging. It was tested using both the NOAA RAP and the Navy's Coupled Ocean Atmosphere Mesoscale Prediction System (COAMPS) as a first guess over Barrow, Alaska. The RAP was run on an 11.75 km resolution grid. The COAMPS model

was run at higher resolutions. The observed data used to nudge the first guess models included hourly surface observations, upper air soundings (twice per day), buoy data, and data observed by sensors mounted on small UAVs (Raven and Scan Eagle). PEMDAS calls this UAV data collection system ASAPS (Atmospheric Sensing and Prediction System). It can provide the model with continuous real-time data, which includes icing and cloud detection; it outputs a variety of easy to use displays for operators. Although a full evaluation of ASAPS and its usefulness is outside the scope of this study, it is worth noting that such a system could be very valuable to any nowcast system working in data sparse areas, specifically one operated by the ISR community (Housel et al. 2016). Satellite data and additional in-situ observations were collected by PEMDAS during the Barrow test but only for verification purposes; they were not utilized in NOWcasting predictions (PEMDAS 2015). However, the PEMDAS NOWcasting system is more than capable of ingesting and exploiting satellite data. According to PEMDAS Senior Scientist and expert on Remote Sensing technologies Mike Gauthier Ph.D., the NOWcasting system has the capability of ingesting “multiple channels” of satellite data at multiple spatial resolutions (M. Gauthier, personal communications, Mar. 29, 2017).

B. MACHINE LEARNING AND BAYESIAN INFERENCE IN NUMERICAL WEATHER PREDICTION

In 1963 mathematician Edward Lorenz famously showed that for any dynamical system (such as the atmosphere), “two [initially] indistinguishable states could eventually evolve into entirely different states” given time (Lorenz 1963). It follows that deterministic numerical weather prediction is limited due to the chaotic nature of the atmosphere. Stochastic modelers attempt to account for this problem using statistical methods to forecast probabilities rather than deterministic outcomes. Statistical post-processing refers to the family of techniques in which assumptions are made about future outcomes (the predictand) based on past observations (training data). These methods identify biases in a model and correct them accordingly to produce a more accurate and concentrated posterior predictive distribution (PPD), the distribution of predicted outcomes. Figure 3 shows an example of a PPD of predicted automobile fuel economy. Statistical post-processing methods include non-homogeneous Gaussian regression,

analogues, and kernel density estimation (KDE) such as ensemble dressing, ensemble regression and Bayesian model averaging (BMA). These methods are common in nowcasting research because they offer a computationally inexpensive way to increase the accuracy of mesoscale models and other predictors.

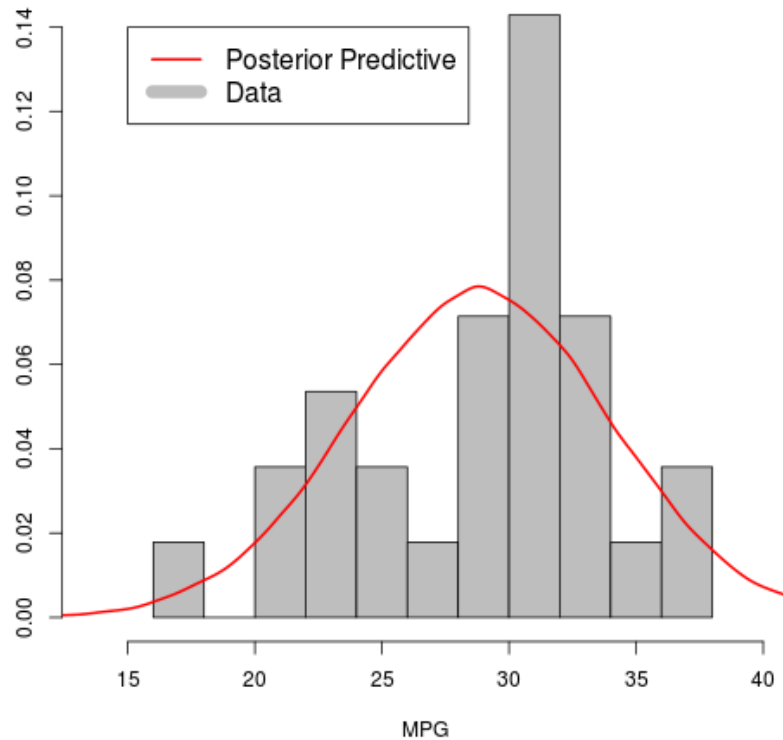


Figure 3. Example of PPD. Source: Warner (n.d.).

Although KDE and BMA have been popular and seemingly intuitive post-processing techniques, they have several deficiencies that have recently been highlighted. These deficiencies include: overfitting (Hamill 2007), problems with extreme forecasts (Bishop and Shanley 2008) and problems at long lead times (Wilks 2006), as observed by Hodyss et al. (2016). Hodyss et al. (2016) has also identified a tendency of BMA to overweight climatology when adjustments are applied in an incorrect order. It proposes a direct application of Bayes' rule, such as Bayesian estimation (BE), as an alternative that minimizes error variance (Hodyss et al. 2016). BE is also a less computationally expensive method and so is well suited for nowcasting. As Kruschke (2010) points out,

another advantage is that BE can handle redundancy in predictor variables; if there are correlations between two or more predictors, BE will still produce a valid probability distribution, whereas KDE and other methods would “explode” and not produce meaningful results. BE also makes it easy for the modeler to interpret results and identify relationships among predictor variables, so that the predictor variable list can be fine-tuned in later tests (Kruschke 2010).

BE is used to predict an observable outcome based on an inference about the model parameters. These model parameters are derived from past performance of a set of predictor variables. Bayes’ rule can be stated that “the *posterior probability* of a model parameter (θ) is the product of a *likelihood function* and a *prior probability* divided by the *evidence*”:

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)} \quad (1)$$

$p(\theta|Y)$ is the *posterior probability* and can be stated as “the probability of model parameter θ being observed given observation Y .” $p(Y|\theta)$ is the *likelihood function* and can be stated as “the likelihood of having observed Y given parameter θ .” $p(\theta)$ is the *prior probability* and can be stated as “the estimate of the probability of observing θ (before observing Y).” $p(Y)$ is the *evidence* or “the likelihood that Y will be observed at all.” The goal of the inference is to derive the posterior, $p(\theta|Y)$, from past data (referred to as “training period” data), so that we can then use that posterior to make a prediction about future observations given a set of current predictor variables (Kruschke 2010).

LCDR Robert Wendt (2017) is a Ph.D. candidate at the Naval Postgraduate School who has considered the findings of Hodyss et al. (2016) and others, and has been conducting research using Bayesian estimation. His application of predictive modeling utilizes BE to stochastically frame the forecast problem and a Markov Chain Monte Carlo (MCMC) sampling method to complete the inference (Wendt 2017). This method has been tested successfully in the North American Collegiate Weather Forecasting Competition (also known as The Weather Challenge) using predictor variables derived

from the NCEP Short Range Ensemble Forecast (SREF). Very similar methods will be applied in this study.

A trending practice in meteorology is the use of machine learning in numerical models. This involves employing statistical post-processing methods to “train” a model based on past (observed) results. Artificial neural networks embrace this concept and use feedback loops to produce continuous machine learning in the model. Eros Pasero and Walter Moniaci (2004) are two meteorological experts in the focus of neural network forecasting. They use the following analogy to describe the usefulness of neural networks in weather prediction:

The idea is that a native old fisherman usually is able to predict the weather of the harbour according to his experience during last years. It’s typical to ask to the old fisherman whether it’s a good idea to sail or to stay inside the harbour. He bases his “weather report” on his experience. His “neural network” knows which events influence the evolution of the weather, on a local basis, better than the national weather system.

The NEMEFO (Neural Meteorological Forecast) system is a neural network nowcast system based in Italy. It samples weather data every 15 minutes from a particular location and stores it in a historical database. Statistical methods must be used to analyze the historical data so that only the most relevant information is exploited to determine the probability of a particular phenomenon. Each variable in the data set is a potential predictor variable for the phenomenon in question. NEMEFO estimates the probability density function (PDF) of each predictor variable by “making a sum of Gauss kernels, each one centered on a record of the database” (Pasero and Moniaci 2004). This is a type of KDE known as the Parzen method. NEMEFO then compares two predictor variables at a time and selects the one whose PDF gives rise to the smallest entropy difference, eventually narrowing down the field of predictor variables to only the most significant (Pasero and Moniaci 2004).

There have been several studies in related fields performed using Bayesian statistics. Uddstrom et al. (1998) and English et al. (1999) applied it to cloud masks in order to retrieve surface irradiance for numerical models. Roquelaure and Bergot (2008, 2009), Roquelaure et al. (2009) and Chmielecki and Raftery (2011) studied Bayesian

model averaging techniques for visibility forecasting. Several studies including Pasini et al. (2001), Bremnes and Michaelides (2007), and Marzban et al. (2007) have applied neural networks to forecast visibility (Chmielecki and Raftery 2011).

C. PREVIOUS CLOUD FORECASTING RESEARCH

One of the first comprehensive studies of short-term (1–6-h) ceiling forecasting was performed at Penn State by Vislocky and Fritsch (1996) and employed statistical post-processing. A real-time observations (obs)–based statistical forecast method was tested against a traditional model output statistics (MOS)–based statistical approach and a persistence climatology statistical approach for 1, 3 and 6-h forecasts. The MOS-based approach takes model prediction, the latest observation and climatic tendency for the specific forecast site as predictor variables. The persistence climatology approach takes only the latest observation and climatic tendency as predictor variables. The obs-based approach takes a network of surface observations and climatic tendencies as predictor variables. All three sets of predictor variables are input into a least squares linear regression model. The relationship of the predictand, Y , to the predictor variables, X_p , is

$$Y = B_o + B_1X_1 + B_2X_2 + \cdots + B_nX_n \quad (2)$$

Coefficients, B_p , are determined as the values that minimize the least square error between the observed and predicted values (Vislocky and Fritsch 1996).

This study demonstrated the importance of real-time data in short-term statistical cloud forecasting. At all forecast hours, the obs-based method demonstrated considerable skill over the persistence climatology method, which had previously been considered a benchmark for accurate short-term cloud forecasting. It also beat the MOS-based method’s skill score by 4% on 1 and 3-h forecasts. The 06-h forecast hour was determined to be the “crossover” point at which the obs-based method no longer adds skill to the model based approach (Vislocky and Fritsch 1996). The obs-based statistical method was tested again on marine stratus at San Francisco International Airport and produced similar results (Hilliker and Fritsch 1999).

In 2008, the Northrup Grumman Analytic Sciences Corporation (TASC) took this a step further. Their forecast system combined the NCAR WRF 36 km resolution mesoscale model, NOAA Geostationary Operational Environmental Satellite (GOES) satellite data and statistical post-processing using logistic regression. The GOES data was used to derive cloud/no-cloud determinations made by the Cloud Mask Generator (CMG) developed by TASC. CMG utilizes six GOES products: visible 0.6- μm channel (VIS), near IR 3.9- μm channel (NIR), 10.7- μm and 11.2- μm far IR channels (FIR), derived nighttime fog, and derived daytime shortwave reflectivity. It essentially compares the current imagery to a “clear sky background” calculated from thirty days of previous GOES data (Kemp and Alliss 2007).

Logistic regression was utilized to determine cloud/no-cloud based on the two predictor variables, WRF and GOES data.

$$\ln \left[\frac{\pi}{1-\pi} \right] = \alpha + \beta_1 X_1 + \cdots + \beta_p X_p \quad (3)$$

Logistic regression offers advantages over linear regression in predicting a binary (yes/no) outcome. It assumes the $\pi[1 - \pi]$ variance required for binary outcomes. Also, it will always output values between 0 and 1. Figure 4 summarizes the TASC forecast system.

The TASC forecast system performed well in a test performed over Washington-Dulles International Airport. 75% of the forecasts displayed error 25% or less. Fifty percent of forecasts displayed error 15% or less. This study demonstrated that model data, GOES satellite imagery and statistical post-processing could be successfully combined to produce cloud/no-cloud forecasts (Kemp and Alliss 2007).

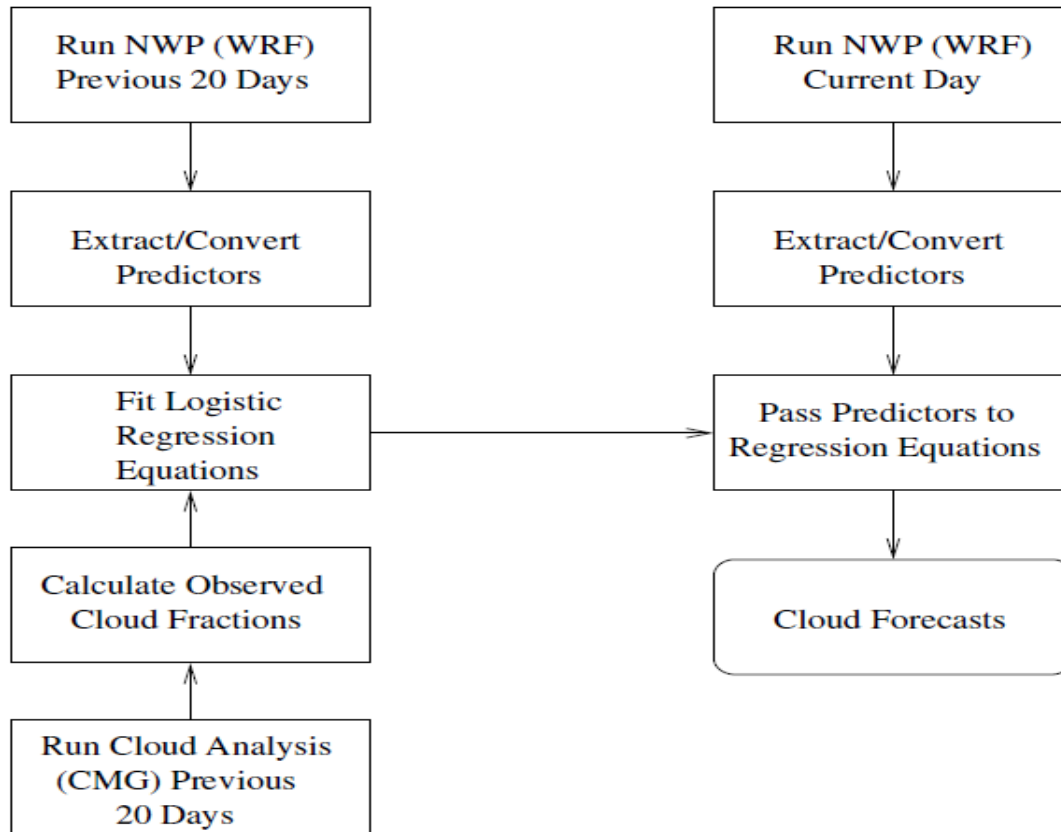


Figure 4. TASC forecast process. Source: Kemp and Alliss (2007).

Timothy Hall et al. (2010) of the Virginia based Aerospace Corporation Engineering and Technology Group conducted a study of short-range sky condition forecasting using statistical post-processing of model and GOES IR data. The end goal was a 1–5 hour probabilistic forecast of clear sky condition. A cloud mask algorithm was applied to 10.7- μm and 3.9- μm channel 4 km resolution satellite data to determine cloud or no-cloud for each pixel. This cloud mask technique is called the bispectral composite threshold (BCT) method established by Jedlovec et al. 2008 (Hall et al. 2010). This technique computes spatially and temporally varying thresholds of brightness temperature difference between the two channels. BCT was verified manually and shown to determine the true sky condition 87.6% of the time. It is worth noting that 82% of misses and 77% of false alarms came on low cloud predictions. The higher low cloud

false-alarm rate occurred over land, while the higher low-cloud miss rate occurred over ocean (Jedlovec et al. 2008).

The statistical post-processing technique applied by Hall et al. (2010) was the *k*-nearest neighbor (*k*-nn) analogue forecast algorithm (KAF); this scheme essentially identifies the *k* most analogous cases within a given set of training data and yields a probability of the observable (clear sky condition) based on the fraction of the analogue cases when the observable occurred (i.e., if clear sky condition occurred in 75 out of 100 analogues, the model would predict 75% chance of clear sky condition). The data set was made up of 105 predictor variables derived from satellite and the 40 km resolution National Centers for Environmental Prediction (NCEP) Eta Model Data Assimilation System (EDAS), then pruned down to 10–16 variables using data mining, heuristic elimination and random forest method. In every test, the most important predictor variables were determined to be satellite derived percent cloud coverage over the target. Other significant variables included percent cloud cover in the upwind direction, recent historical sky conditions, 1000–500mb thickness gradient and 6 hour change, mean sea level pressure gradient, static stability, time of day and maximum solar angle (Hall et al. 2010).

The study confirms that satellite imagery (GOES in particular) and statistical methods can add significant skill to model and observation based cloud forecasting. KAF outperformed persistence, conditional expectancy of persistence and satellite cloud climatology techniques on all five performance metrics. The results showed that this scheme can produce quality short-term sky condition forecasts for local and regional targets in two different geographic areas (Hall et al. 2010).

III. DATA/METHODOLOGY

A. OVERVIEW

The nowcast system created for this study combines 1 km resolution visible imagery brightness temperature values and 13 km resolution NCEP Rapid Refresh (RAP) data on a 1 km spacing lambert conformal grid using multi-quadric interpolation and Bayesian estimation (BE). The multi-quadric interpolation is used to map the RAP data to the nowcast grid at both analysis and forecast times. BE is used to predict out to 6 hours the presence of low cloud by statistically relating prior forecasts to cloud cover, then using this relationship to predict future cloud cover. This Naval Postgraduate School (NPS) nowcast system is tested on a case study of 2016 summer low-level clouds over the Monterey Bay in Central California using a 100 x 100 km grid.

This grid size was chosen to simulate a target region for a UAV mission. The small size also makes it easier to achieve meaningful results from BE, in order to establish proof of concept for this approach. The location and time period was chosen to maximize the presence of low clouds while including variation throughout the day as well as day to day. Low clouds represent significant obstacles to ISR and tend to be the most difficult to analyze and predict, as found by Jedlovec et al. (2008). Central California is also a mid-latitude location, representing latitudes where the military often operates and where geostationary imagery is the only reliable real-time satellite data.

Given the statistical nature of this approach, cross validation testing of various data grouping schemes and learning period lengths is performed in order to determine the ideal process for the NPS nowcast. Statistical analysis and side-by-side comparison are used to score NPS nowcast results and compare them to truth and climatology. Adjustments are made as necessary. All Bayesian estimation, model outputs and scoring were generated with Anaconda Python and the Seaborn library.

B. DATA SET

1. NOAA GOES 15 (Geostationary Operational Environmental Satellite) Imagery

NOAA GOES 15 (also known as GOES-WEST) nominal 1 km spatial resolution visible channel (0.55–0.75- μm) imagery and 4 km spatial resolution imagery from all other GOES 15 channels was provided by the Naval Research Laboratory (NRL) in Monterey, CA. The GOES 15 visible channel is accurate to within 5% of the maximum irradiance value (NOAA Satellite Information System 2013). Imagery is available at 30-minute intervals; several images from each day were not available due to information gaps and NRL processing. The satellite data were pulled from an archive and processed into TDF files by a TeraScan master machine. SeaSpace TeraScan is a software suite used to receive, process, and archive satellite data. Irradiance values were scaled 0–239 for TeraScan imaging (K. Richardson, NRL, personal correspondence, Dec. 12, 2016). The data was mapped to a lambert conformal map projection on an 800 x 800 km box over the Monterey Bay; NRL calls this the Monterey Bay Vis sector. The processed data was also provided in JPEG and netCDF file formats for reference purposes. A sun zenith angle correction was applied in producing the JPEG files, but no such correction was performed on the TDF data. Figure 5 shows a JPEG image of the Monterey Bay Vis sector.

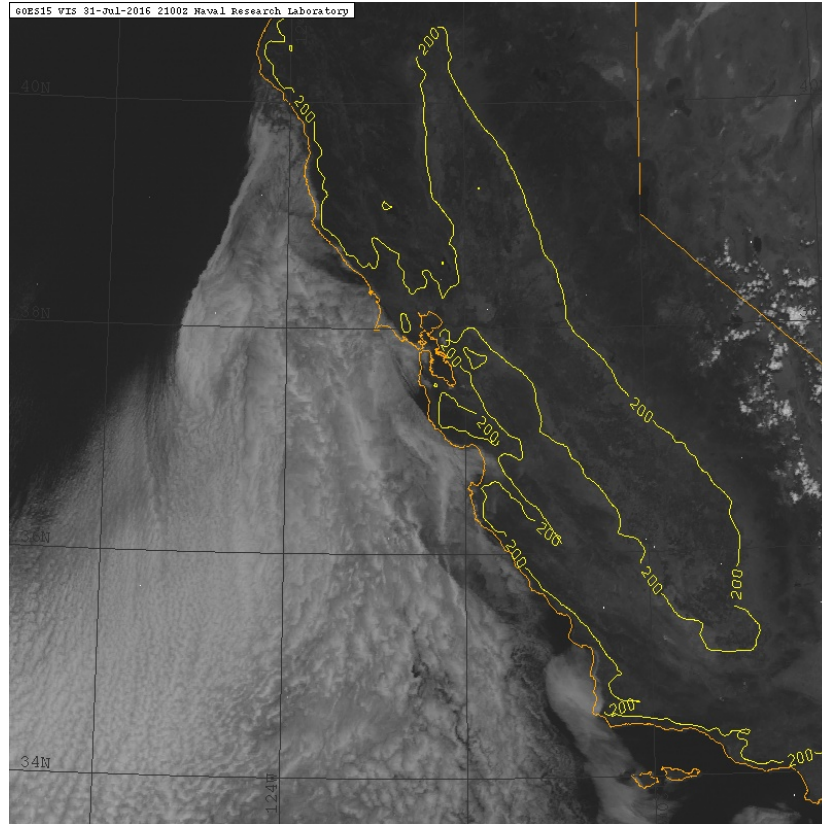


Figure 5. NRL 0.6- μm GOES 15 Monterey Bay Vis sector (JPEG)

Only the visible channel imagery is ingested and utilized in the NPS nowcast; however, other channels were received and were used to identify hours when reflectance values are contaminated by mid and high level clouds. These hours were identified through a subjective assessment of all hours used in this study. The 4 km resolution data from the 3.80–4.00- μm , 6.50–7.00- μm , 10.20–11.20- μm , and 11.50–12.50- μm channels is interpolated onto the same 800 x 800 km grid with 1 km grid spacing via TeraScan’s standard procedure. According to NRL satellite specialist Kim Richardson, it performs a “piecewise polynomial interpolation where the polygram size for the biquadratic polynomials is 9 x 9, which is the default: poly_size = 100. They say that the interpolation was found to be less [than] 0.15 kilometers [in error] with the default” (K. Richardson, NRL, personal correspondence, Dec. 12, 2016). Figure 6 displays (a) a 10.20–11.20- μm image and (b) a 3.80–4.00- μm image.

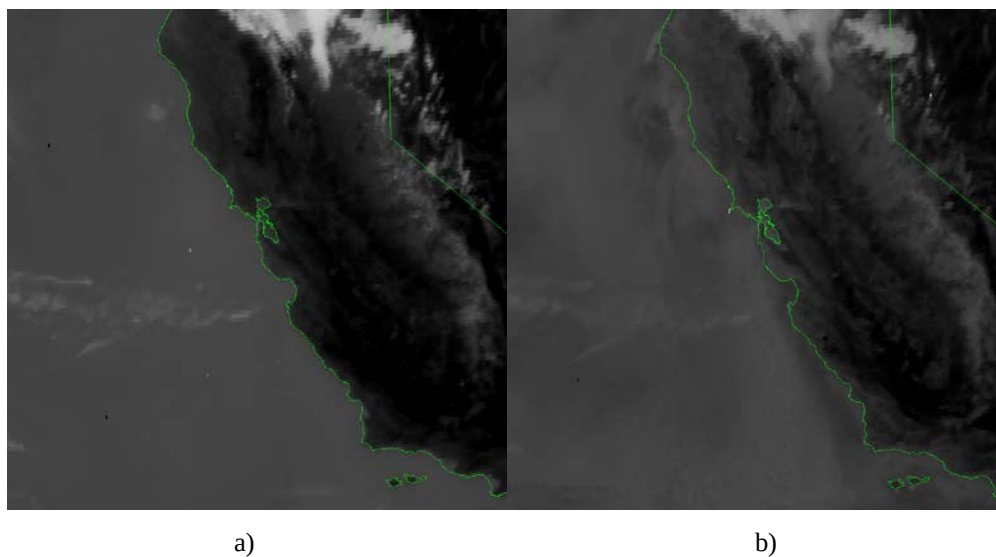


Figure 6. NRL (a) 10.20–11.20- μm (FIR) and (b) 3.80–4.00- μm (NIR) TeraScan images

2. 13 km Resolution NCEP Rapid Refresh (RAP)

13 km resolution RAP Version 2 data was downloaded and saved from NOAA NCEP's File Transfer Protocol (FTP) site. The data was saved in GRIB2 file formats for the period of 01 June–31 July 2016. RAP Version 2 is a non-hydrostatic grid point model that employs Hybrid Ensemble–3DVar (three-dimensional variational data assimilation) and Thompson v.3.4.1 microphysics. RAP is run hourly and outputs hourly forecasts out to 24-h each run (Benjamin et al. 2016). Data was available at every 25mb in the vertical. Only 00-h analysis and 06-h forecasts were saved for every available (hourly) model run. Several derived fields were not available, including liquid water content and cloud ice content.

C. ANALYSIS PROGRAM

1. VISUAL

VISUAL is a meteorological display program created by Wendell Nuss in 1986 and further developed over the years. It utilizes Graphical Kernel System (GKS) output primitives and NCAR Graphics routines (Nuss and Drake 1990). It was used to visually display and evaluate PEMDAS' NOWcasting output fields. (Figure 2, shown previously,

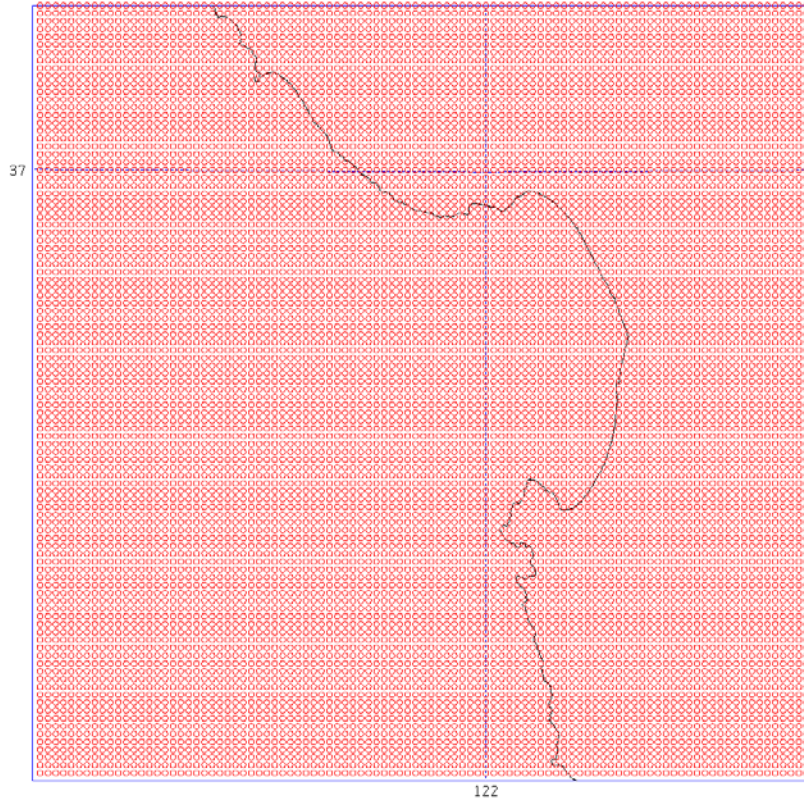
is an example of a VISUAL graphic used in this study.) It was also used to verify that data ingest techniques were functioning properly throughout the generation process of the NPS nowcast system. All displays used a uniform distance, non-staggered grid on a Lambert Conformal map projection.

2. GARP

GARP (GEMPAK Analysis and Rendering Program) is another display program that inputs GEMPAK file formats for visualization. It is designed for easy user interface. GARP was used to display RAP 13 km fields and compare to GOES 15 imagery in order to identify possible model predictor variables. Select valid time periods of potential model predictor variables were overlaid on the GOES visible imagery and subjectively assessed for their “fit” to the low cloud field. Although many potential variables could serve as predictors, this analysis resulted in three predictor variables, described in a later section.

3. Grid and Multi-quadric Interpolation

The grid employed in the NPS nowcast system is 100 x 100 km centered on the Monterey Bay with 1 km uniform spacing on a Lambert Conformal map projection, true at 30N and 60N latitude. The grid is referenced to latitude and longitude (36.8N; 122.0W) at grid point (x = 59, y = 57). This gives approximate corner points of (36.28N; 122.66W) in the lower left and (37.19N; 121.53W) in the upper right. Figure 7 is a VISUAL graphic of the NPS nowcast grid over the Monterey Bay coastline. Each red circle corresponds to a grid point.



This plot was created using VISUAL. Note that the first row (bottom) and column (far left) are omitted.

Figure 7. NPS nowcast 100 x 100 km grid

Multi-quadric interpolation is the technique used to interpolate the 13 km resolution RAP data onto the 1 km resolution grid. This method has been shown to produce accurate analyses in a variety of cases. It employs hyperboloid radial basis functions of the difference vector between the observation point and any other point. Due to it being much less computationally expensive and quick to run, it is much more applicable to nowcast models than methods commonly used at operational numerical weather prediction centers, such as optimum interpolation or 3D and 4D variational analysis. It has been shown to outperform other methods such as the Barnes and Cressman methods, which have been used for mesoscale analysis. It has also been demonstrated to perform reasonably well in data-sparse regions (Nuss and Titley 1994). Although this technique can be used for observational nudging, in this study the coarse resolution model fields were simply interpolated to the fine grid.

D. SATELLITE PROCESSING AND INGEST METHODS

After receipt from NRL, satellite data TDF files were prepped for data ingest via a four-step process:

- 1) A sun zenith angle correction was performed to standardize irradiance (*visalbedo*) values across all hours of the day:

$$\text{if } \text{sunzenith} < 85, \text{ then } \text{visalbedo} = \frac{\text{visalbedo}}{\cos(\text{sunzenith})},$$
$$\text{else } \text{visalbedo} = \text{badvalue}$$

- 2) Irradiance values (scaled 0–239) were converted to reflectance values (0-100): $\text{ref} = \frac{\text{visalbedo}}{155} * 100$.

Figure 8 compares satellite imagery before and after steps (1) and (2) are applied. Notice that before the correction, the 1400Z (0700 local time) and 2000Z (1300 local time) TDF images display very different irradiance values over the stratus clouds. After the correction, the reflectance values look similar over the stratus.

- 3) A 100 x 100 pixel box was cropped from the original to fit the 100 x 100 km NPS nowcast grid.

- 4) A uniform shift of 1 km north and 3 km east was applied to all images. All images with significant erroneous values or mapped coastline error were flagged and discarded.

A problem was encountered where the Terascan mapped coastline did not line up with the physical coastline apparent from visual appraisal of the imagery. A common issue that the analyst encounters when working with geostationary imagery is a slight “wobble” in the optics as the satellite performs its daily orbit of the earth. This wobble is not accounted for in the navigation parameters used to map the imagery and results in an imperfect alignment between the mapped satellite imagery and the physical coastline. This error was observed in the Terascan imagery and varied with no apparent pattern of predictability. Upon further manual inspection, most images displayed a mapped coastline offset of 1–5 km to the northeast of the physical coastline. The solution was to reduce this potential source of error by flagging and discarding image files with a

mapping error that was outside of this range. Figure 9 shows examples of (a) an acceptable mapped coastline error 1–5 km to the northeast and (b) an unacceptable coastline error 4 km to the west. After unacceptable image files were discarded, the uniform shift of 1 km north, 3 km east was applied to all images in order to minimize the coastline displacement error. Thus, each NPS nowcast forecast hour carries approximately a 0–2 km error due to this effect.

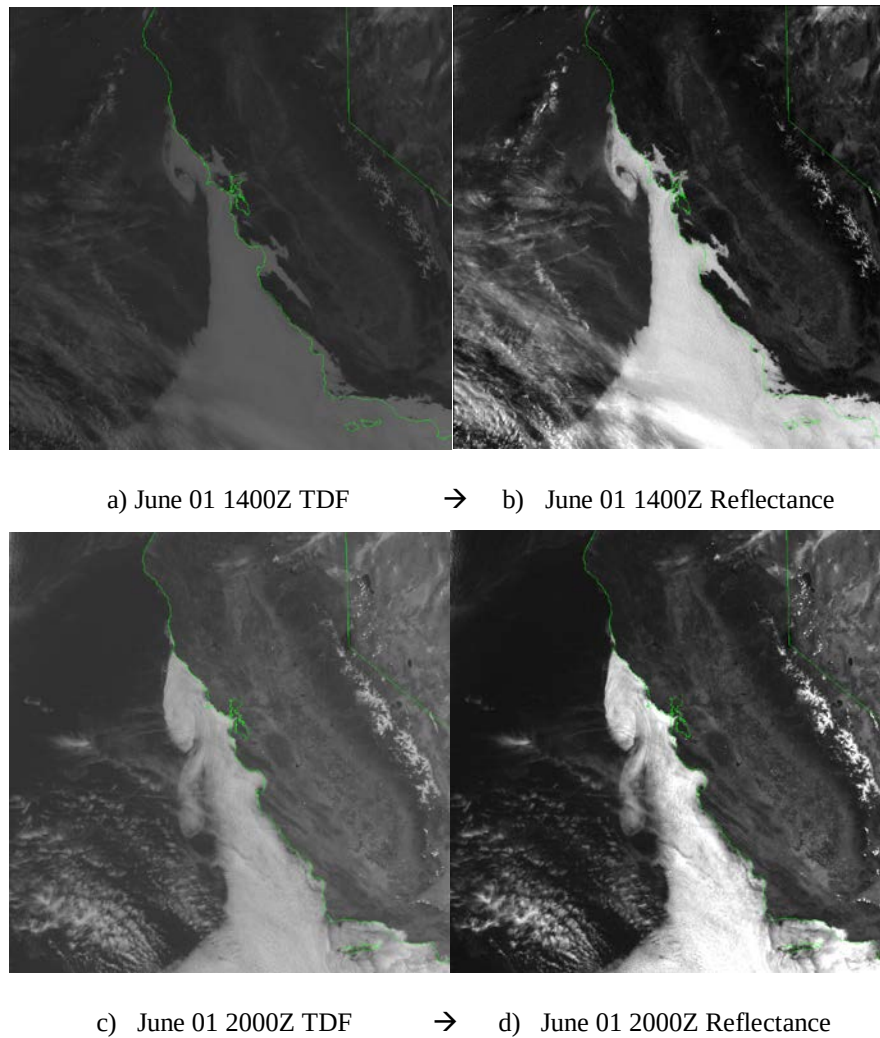
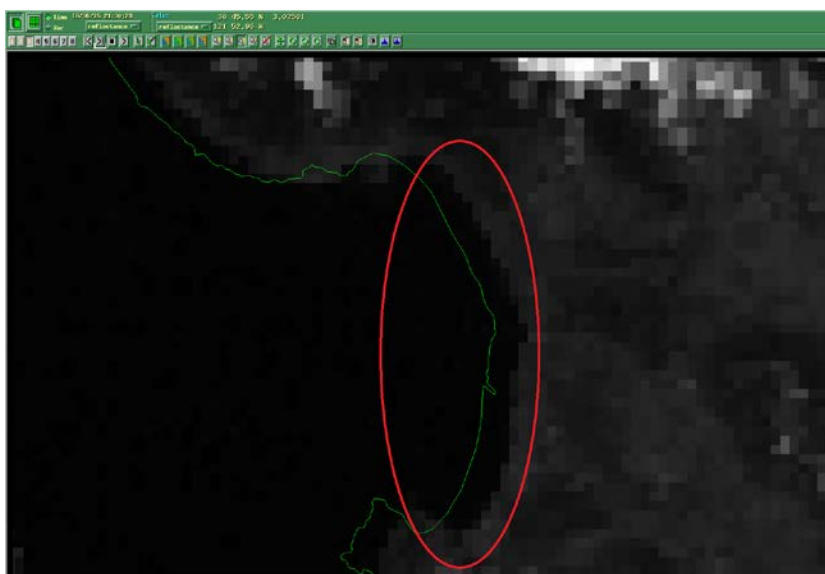


Figure 8. Comparison of original imagery (a) and (c) to imagery after the reflectance conversion (steps 1 and 2) was applied (b) and (d).

a) Example of acceptable mapped coastline error 1–5 km to the northeast



b) Example of unacceptable coastline error 4 km to the west



Note that both images are zoomed-in on the Monterey Bay coastline. The mapped coastline is in green.

Figure 9. Mapped coastline error shown after processing steps 1 and 2 were applied.

E. FORECAST METHODS

1. Overview

The NPS nowcast approach is characterized as a type of statistical post-processing of short-term numerical model forecasts. The short-term model forecasts used in this application are from the NOAA RAP model, but other models could be used. This approach does not generate new dynamical analyses and forecasts. Instead, statistical correction is applied based on past performance and current model data. The PEMDAS NOWcasting system that motivated this study employed a similar approach by using current observations to adjust the analyses and short-term forecasts. These approaches are common in nowcasting but are distinctly different than dynamical short-term analysis and forecast models such as the RAP, HRRR and Navy COAMPS systems. The purposes of statistical nowcast approaches are to reduce short-term error inherent in the dynamical analysis-forecast systems without greatly increasing computational costs.

2. Bayesian Estimation Method

The statistical post-processing method utilized by NPS nowcast is Bayesian estimation (BE) using a Markov Chain Monte Carlo (MCMC) sampling scheme adapted by Wendt (2017). This direct application of Bayes' Rule offers many advantages over Kernel Density Estimation (KDE) and other commonly used statistical post-processing methods. Predictor variables (from the NOAA RAP 06-h forecast fields) are compared to observed reflectance values (from GOES imagery) over the "training period." This training data is then used to infer a mathematical generalized linear model (GLM) that predicts future reflectance values given the corresponding RAP predictor variables. This prediction takes the form of a posterior predictive distribution (PPD), which is a probability distribution of possible observed reflectance values. This process is summarized visually by Figure 10.

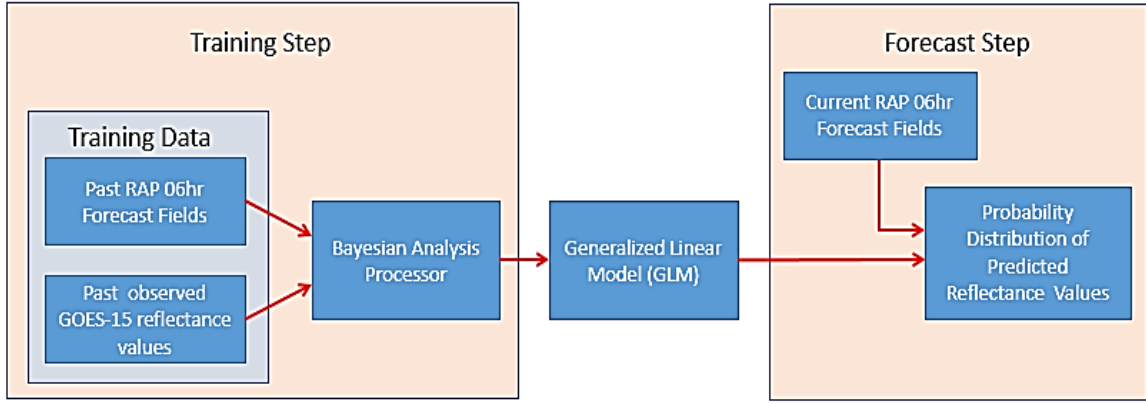


Figure 10. Summary of NPS nowcast Bayesian estimation prediction process

As stated above in equation (1), Bayes' Rule is given by:

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)}$$

In NPS nowcast, Y is the reflectance value. $\underline{\theta}$ is the set of five mathematical model parameters. θ_2 , θ_3 , and θ_4 correspond to the weighting coefficients of the predictor variables. In NPS nowcast, three predictor variables derived from RAP 06-h forecast fields are used. (The three predictor variables and the process used to choose them are described later in this chapter.) The two additional parameters, θ_1 and θ_5 , correspond to the intercept and the variance, σ^2 , of all the data points in the training period, respectively. Thus, this application of BE seeks to predict Y , reflectance values, by inferring information about the model parameters, $\underline{\theta}$, from the training data.

In the “training” step, Y is the observed reflectance value. $p(Y)$ can also be described: $p(Y) = \int p(Y|\theta)p(\theta)d\theta$. Thus, Bayes' Rule (1) can be re-written:

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{\int p(Y|\theta)p(\theta)d\theta} \quad (4)$$

Equation (4) is the inference used to obtain the probability of the parameters given the observed reflectance values, $p(\theta|Y)$. In order to complete the inference, MCMC discrete sampling methods are used to estimate the target distribution:

$$p(\theta|Y) \propto p(Y|\theta)p(\theta) \quad (5)$$

MCMC is commonly required to complete statistical inferences, especially complex BE with multiple predictor variables and multivariate structures (Gelman et al. 2013). It is a robust, sophisticated sampling method. MCMC methods have emerged in the last thirty years and have helped fuel a resurgence of BE techniques because MCMC is so well suited to BE (Robert and Casella 2011).

The specific MCMC method applied in this study is the Metropolis algorithm. It evaluates $p(\theta|Y)$ at a current state, θ_t , and a randomly selected proposed state, θ_{t+1} . If $r = \frac{p(\theta_{t+1}|Y)}{p(\theta_t|Y)}$ is greater than a randomly drawn number in the interval (0,1), then Metropolis accepts the jump to θ_{t+1} . If not, Metropolis rejects the jump, and θ_t is repeated as the next element of the chain. Each θ state has a corresponding T, transition kernel, describing the random “jump” from θ_{t-1} to θ_t . This chain of θ states and corresponding T’s is known as the Markov chain. It represents a “random walk” through the sample space. The chain is said to converge when the Metropolis algorithm reaches a stationary distribution (Wendt 2017). In the NPS nowcast, the Metropolis algorithm is continued more than 500,000 samples after convergence is deemed to have been reached.

Ten thousand Monte Carlo samples are then taken from the portion of the Markov chain succeeding the point of convergence. These samples yield a discrete probability distribution (i.e., histogram) of each parameter. KDE is used simply to estimate a continuous probability density function (PDF) (i.e., the curve over the top of the histogram) from the discrete distribution for visualization purposes. The PDFs of parameters θ_1 , θ_2 , θ_3 and θ_4 each correspond to the Betas (β)—also known as slope values or weighting coefficients. The PDF of θ_5 corresponds to the variance, σ^2 . The Betas and variance are parameters in the following Gaussian generalized linear model (GLM):

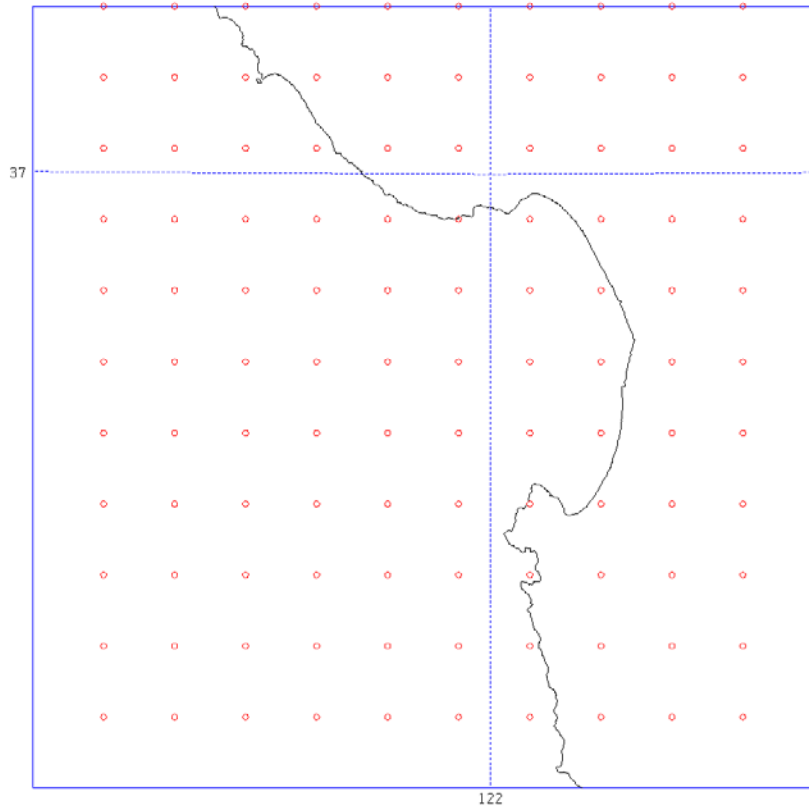
$$Y(X|\beta_o, \beta_1, \beta_2, \beta_3, \sigma^2) \sim N(\beta_o + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3, \sigma^2) \quad (6)$$

In the forecast step, this GLM outputs the prediction about the reflectance value that will be observed 6 hours in the future (i.e., the 06-h forecast value). Here, $N()$ is a Gaussian probability distribution function, $N(\mu, \sigma^2)$, describing the probability density of

observing reflectance value Y . The mean of the distribution, μ , is approximated by: $\mu \sim \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$. X_1 , X_2 , and X_3 are the 3 current predictor variable values (from RAP 06-h forecast fields). Notice that this GLM takes into account variance, σ^2 , whereas simple linear regression would not. The modeler also has the option of including prior assumptions about the observable, $p(Y)$; however, prior assumptions were not included in this application of BE. The final step of the forecast process is that the normal (Gaussian) distribution is transformed back into a log-normal distribution—since a log transformation was performed on the training data. In recap, the mathematical model (the GLM) parameters are inferred from training period predictor variables and observed reflectance values in the training step. Then in the forecast step, current predictor variables are input into the GLM, which outputs a predictive distribution of 06-h forecast reflectance values.

3. Application of Bayesian Estimation

In order to establish proof of concept, the BE procedure was simplified to decrease computational cost by using only every 9th grid point in the latitudinal and longitudinal direction as a sub-sample from the original 10,000. This creates a uniformly spaced grid of 144 representative points with 9 km spacing and the same external dimensions as the original 100 x 100 km box. Figure 11 is a VISUAL graphic of the 144 point sub-sampled grid. Note that the first row and column of the grid are omitted due to limitations of the VISUAL program. The data extracted at each of the 144 points consisted of 06-h RAP forecast fields and GOES observed visible channel (0.6- μm) reflectance values for the corresponding valid times. In addition, each of the 144 grid points was assigned a value of either 0 or 1 denoting either ocean or land surface, respectively. Lastly, each data point contains a valid time.



This plot was created using VISUAL. Note that the first row (bottom) and column (far left) are omitted.

Figure 11. NPS nowcast sub-sampled 144 point grid

The BE methods described above are applied using three different data grouping schemes described in further detail in a later section. For each scheme, multiple data sets (within the two month case study period) representing various training period lengths are tested. The purpose of this cross validation is to identify the ideal grouping scheme, training period and predictor variables. In every test, a PPD of reflectance values is produced for each of the 144 grid points, for each forecast valid time.

Two different forecast products are produced from these PPDs. Firstly, the mean reflectance value of each PPD is extracted and plotted to create a “pseudo-satellite image” for the forecast hour (in both grey-scale and color enhanced scales for added detail). The goal of this approach is to generate a false image that, as closely as possible, matches a reduced resolution version of the observed satellite image. This format would be easy for an intelligence, reconnaissance and surveillance (ISR) forecaster to use.

Secondly, each PPD is used to produce a ‘probability of cloud’ field. This product is primarily for the purposes of scoring and comparing to other forecasts. The following thresholds were determined through visual examination of the processed satellite imagery in TeraScan: for all sea grid points, *reflectance* < 5.0% = ‘no cloud’; for all land grid points, *reflectance* < 11.0% = ‘no cloud’. The 5% cloud threshold over sea is “safe” from a meteorological standpoint; the sun angle never produces a sun glint pattern over the Monterey Bay in June, so the maximum reflectance should remain under 5%. The surrounding land varies in albedo but is mostly vegetated, so 11% is also a “safe” land threshold based on meteorological reasoning.

4. Predictor Variable List

Three RAP 06-h forecast fields were used as predictor variables based on forecaster knowledge and side-by-side comparisons of model fields to satellite imagery:

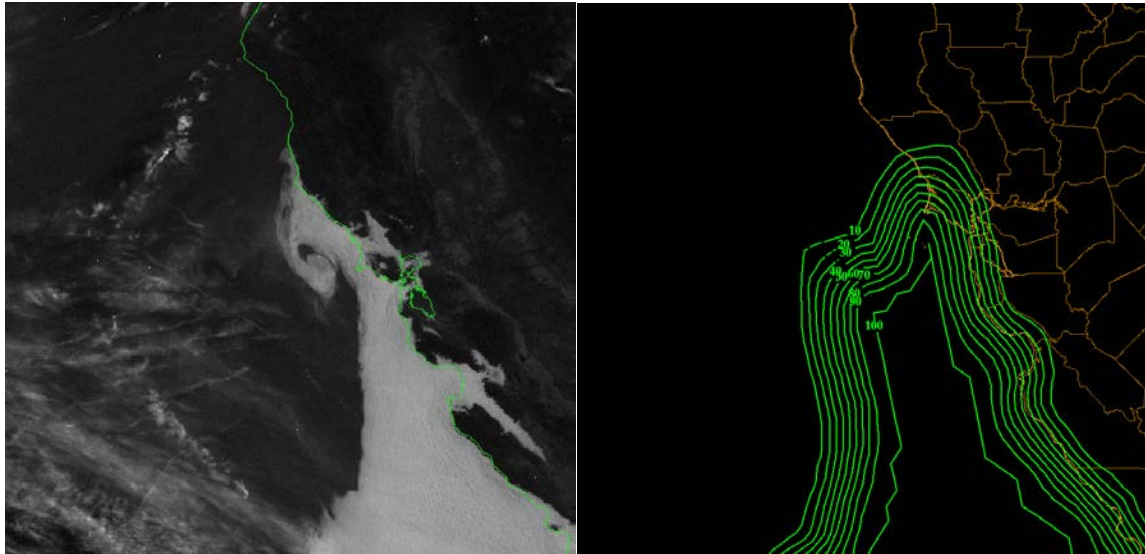
Low cloud fraction (LCLD) – this is a RAP derived field describing cloud cover below 642mb. The field is highly smoothed using a 40 km radius smoother. It is designed to match National Weather Service cloudiness forecasts. Values range from 0–100% (NOAA ESRL 2016).

Low level OMEGA (Ω) – the highest magnitude negative Ω value out of all isobaric levels between 1000mb and 850mb. OMEGA is a RAP derived field converted from vertical velocity (in m/s) using the formula: $\Omega = -\rho * g * w$, where ρ is air density, and $g = 9.80665 \text{ m/s}^2$. Negative values denote upward vertical velocity (NOAA ESRL 2016).

Low level Relative Humidity (RH) – the highest RH value out of all isobaric levels between 975mb and 700mb. Because the majority of grid points are located over ocean, the 1000mb height level was not used as it would have considerably decreased the variability of this predictor. At a majority of valid times, the highest RH value was observed in the 975mb field. RH is a RAP derived field defined with respect to saturation over water (NOAA ESRL 2016).

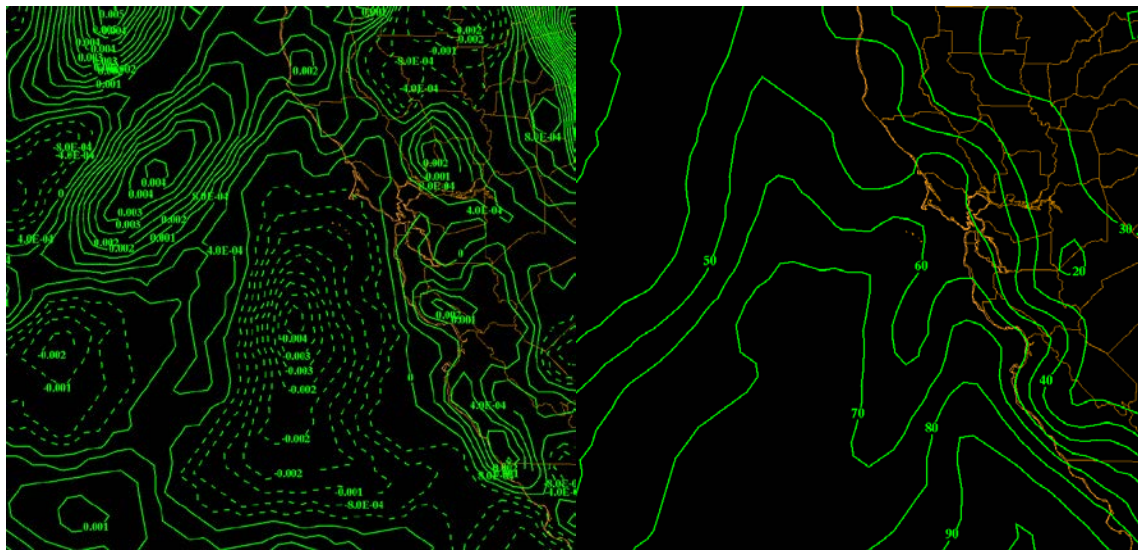
Figure 12 is one example of many visual comparisons that were made. From this example, it appears that the LCLD field (Figure 12.a) is correlated to the observed cloud but loses fine detail due to smoothing. The OMEGA field (12.b) looks like it may be loosely correlated as there are negative values (upward velocity) over most of the cloud region and positive values over land. 850mb is displayed here because this pressure level

contains the highest magnitude negative values at 1400Z. The RH field (12.c) also appears correlated. Notice it is similar to the LCLD field but contains more detail. 975mb is displayed here because it contains the highest RH values (not including 1000mb) at 1400Z.



a) June 01 1400Z reflectance image

b) June 01 06-h LCLD valid 1400Z



c) June 01 06-h 850mb OMEGA valid 1400Z

d) June 01 06-h 975mb RH valid 1400Z

Figure 12. Comparison of June 01 2016 1400Z reflectance imagery to RAP 06-h forecast fields

Other meaningful predictor variables were examined but not applied in order to maintain simplicity and establish proof of concept. Model derived variables that should be tested in a scheme predicting clouds at all height levels include:

- Mid level cloud fraction (MCLD)
- High level cloud fraction (HCLD)
- Mid level OMEGA
- High level OMEGA
- Mid level RH
- High level RH
- Surface based Lifted Index (LI)
- Best Lifted Index (BLI)

Satellite derived fields should also be tested, including reflectance value at grid point from the most recent visible image, brightness temperature value from the most recent FIR image, and a local average of reflectance (and brightness temperature) values. Cloud mask algorithms would need to be applied to identify what regime (low, mid, high or convective) each cloud falls into. A multivariate BE scheme could be used to produce a forecast for each regime so that height information is also provided to the forecaster.

5. Three Data Grouping Schemes

Without any grouping of the data points (i.e., the RAP predictor variable fields and corresponding observed reflectance values), each of the 144 grid points at each valid time in the verification period—the time period that the predictions are made for—would receive the same bias correction. That is to say that the same Betas would be used in the GLM to predict reflectance values for all grid points at all valid times. The only part of the GLM that would change with grid point or valid time is the current predictor variable values X_1 , X_2 , and X_3 , themselves. The problem with this approach occurs when the relationship between predictor variables and observed reflectance values varies across data points (i.e., if predictor variable bias changes with geographical location, time of day, time of month, or any other factor). These variations are highly likely from a meteorological standpoint; meteorological phenomena evolve differently depending on

geography, time of day, synoptic pattern, etc. Without grouping, the model will not identify these variations, and will calculate a less accurate bias correction for each point. Grouping similar data points together and generating a different set of Betas for each group can capture these variations and produce more accurate results. The disadvantages of grouping are that it can be more computationally expensive and can negatively impact forecast results if applied incorrectly (i.e., if the wrong groups are chosen). Three different grouping schemes are experimented with in this study.

1) No grouping. In the initial scheme, no grouping was applied. The linear regression was performed using all the data points (144 grid points multiplied by the number of hours in the training period). Thus, the same Betas are used in the GLM to predict reflectance values for all 144 grid points. The advantages of this approach are simplicity and low computational cost. The disadvantages are that it does not account for variation in predictor variable bias across location, time or other factors that impact forecast accuracy. As an example of this scheme, Figure 13 shows a scatter plot of normalized LCLD versus reflectivity for all points in data set 6. (Data set 6 is the smallest of 5 data sets which will be described in a later section.) Notice that one best fit line is drawn through the entire group.

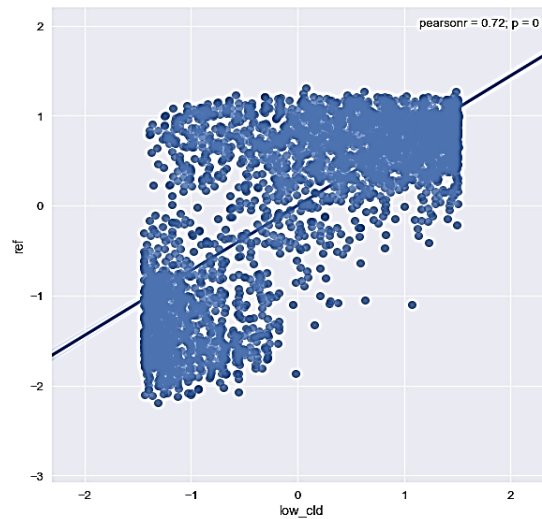


Figure 13. Data set 6 normalized LCLD versus reflectivity with no grouping

2) **Grouping by land and sea.** Data points were divided into two groups based on whether they occurred over a land or ocean surface. Each data point, from training and verification periods, was assigned either a 1 (land) or 0 (sea) and divided accordingly before Bayesian estimation was applied. Thus, two sets of Beta coefficients were calculated and used in the GLM, one for all land grid points and one for all sea grid point. The land surface is drier, less uniform and generally more reflective (of the $0.55\text{--}0.75\text{-}\mu\text{m}$ channel) than the ocean surface. Low clouds in the land regime are typically advected from the bay and so are much more dependent on winds and orographic effects. Thus, it was reasonable to expect different predictor variable biases from these two regimes. As an example of this scheme, Figure 14 shows a scatter plot of LCLD versus reflectivity, grouped by land (green) and sea (blue) for all points in data set 6. Notice the two best fit lines, one drawn through the land data points and the other through the sea points. The sea points tend to exhibit higher LCLD and reflectance values while the land points tend to exhibit lower values.

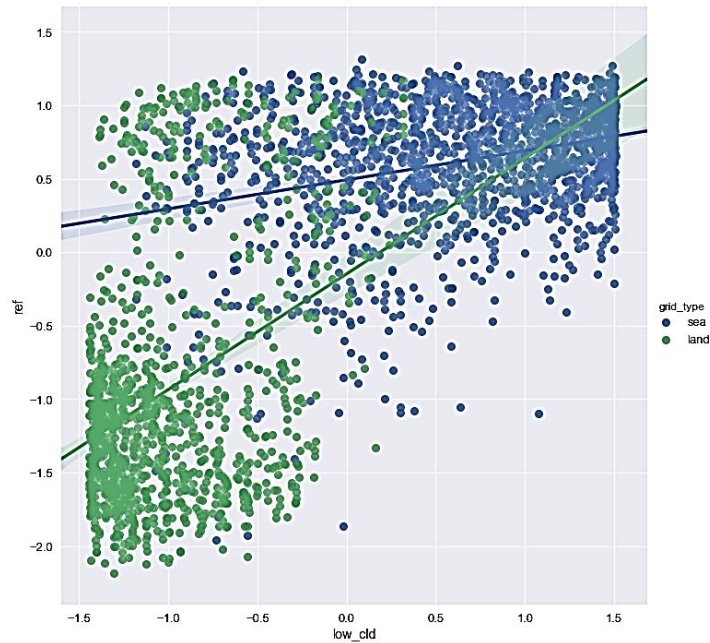


Figure 14. Data set 6 normalized LCLD versus reflectivity with land-sea grouping

3) ***k*-means clustering.** This approach uses data mining to group all training data points into k clusters. An algorithm determines clusters that yield the minimum sum of the distances (in data space) between all data points and the centroid of the cluster. This sum of the distances is known as the distortion. That data points are not grouped using observed reflectance values, because in reality the observed values in the verification data would not be known until after the forecast is verified. Instead, the four dimensions used in this algorithm are the three predictor variables as well as the time difference between the valid time and 2100Z. This variable is meant to account for model bias variations between midday and morning/nighttime hours, since 2100Z corresponds to 1400 local time. The k -means clustering scheme is similar to the land/sea approach in that it seeks to capture patterns in predictor bias variability. The difference is that it uses actual data to determine several groups rather than relying on a meteorological assumption to create two groups. The only foreseen disadvantage is that it requires a slight increase to computational cost. As an example of this scheme, Figure 15 shows a scatter plot of LCLD versus reflectivity, grouped in $k = 6$ clusters, for all points in data set 6. Notice the six color coded groups with a unique best fit line drawn through each.

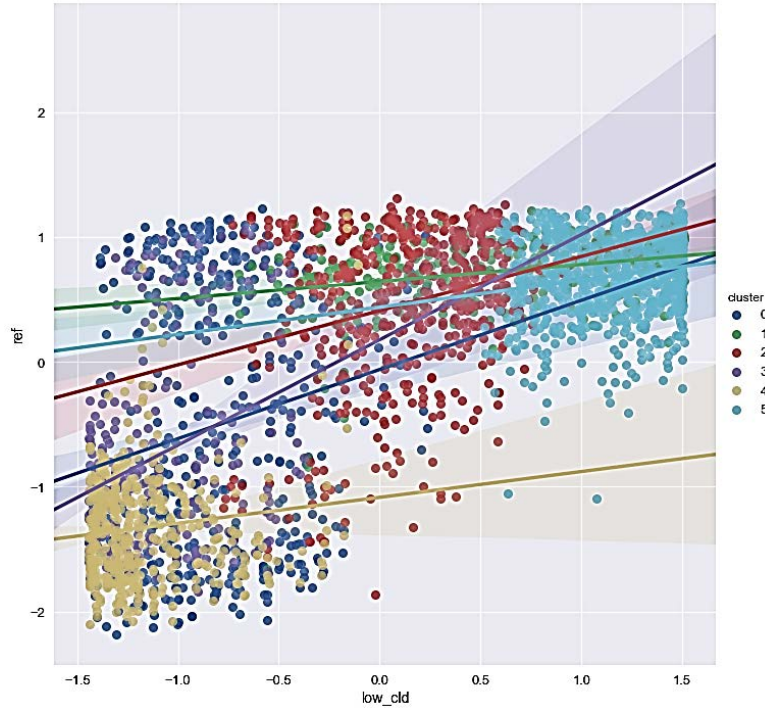


Figure 15. Data set 6 normalized LCLD versus reflectivity with k -means clustering

The number of clusters, k , was chosen by actually performing the clustering on each data set for $k = 2:8$ and conducting a silhouette analysis for each value of k . Silhouette analysis is a technique commonly used to determine the appropriate k value. Each data point is assigned a silhouette coefficient (on a -1.0 to 1.0 scale) based on how similar it is to the other points in its cluster. The ideal choice of k yields silhouettes with all positive or minimal negative coefficient values and high peak values. $k = 6$ was chosen as the ideal number of clusters for all data sets. Figure 16 shows the silhouette analysis for data set 7 when $k = 6$. (Data set 7 is the second smallest of 5 data sets described in the next section.) Small slivers of negative values in clusters 1 (indigo), 2 (light blue), 4 (yellow) and 5 (red) indicate minimal amounts of data points that were not similar—relatively speaking—to the other points in their cluster. The peaks of each silhouette range from coefficient values of 0.5 to 0.7, which was high compared to the silhouette analyses when k was less than 6.

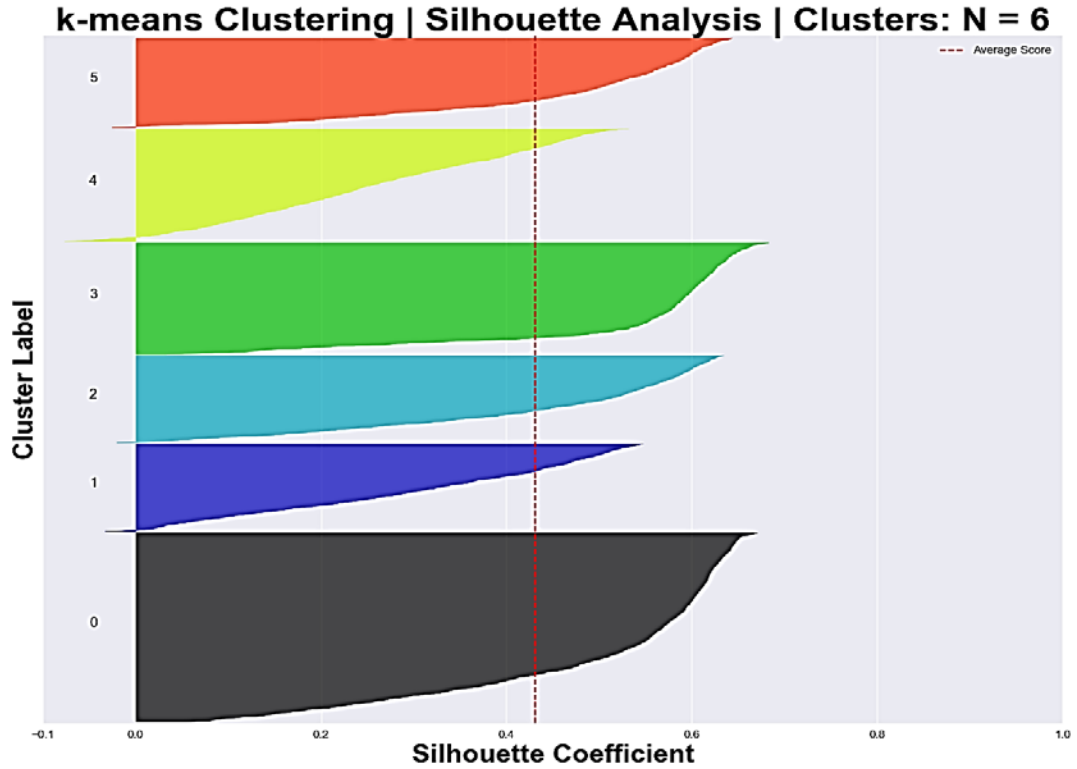


Figure 16. Data set 7 silhouette analysis for $k = 6$

F. DATA ANALYSIS AND SCORING

The data used to test NPS nowcast is grouped into 5 sets numbered 6–10. Each set contains training period data and verification (or forecast) period data. The “no grouping” and “land/sea grouping” schemes are tested on set 6 only. The k -means cluster scheme is tested on data sets 6–10 for cross validation purposes. The time periods contained in sets 6–10 are described here:

- Data set 6: 20 hours (10 training | 10 forecast), June 05–06
- Data set 7: 47 hours (37 training | 10 forecast), June 01–03, 05–06
- Data set 8: 98 hours (78 training | 20 forecast), June 01–13,
- Data set 9: 137 hours (112 training | 25 forecast), June 01–16, 18–20
- Data set 10: 201 hours (163 training | 38 forecast), June 01–16, 18–24, 26–28

Three data analysis steps were performed for each test. Firstly, the extracted training data is preprocessed to get an idea of the relationships between predictor

variables and observed reflectance. This is performed for each scheme. The scatter plots shown previously in Figures 13–15 are examples of this. For the *k*-means cluster scheme, additional analysis is performed to determine the most cost efficient value of *k*.

Secondly, after the “training step” is performed, the PDF of each model parameter, θ , is produced and analyzed visually. These PDFs determine the Beta, β , value for each predictor variable. Thus, the actual relationship between each predictor variable and the observed reflectance is determined. The relative “meaningfulness” of each predictor variable is assessed. Also, the Monte Carlo Markov Chain convergence process is graphed, and the amount of time that it takes for NPS nowcast to compute each PDF is recorded. This is performed for each cluster.

Lastly, after the “prediction step,” the forecast products are produced and analyzed. The results are visually compared to the observed imagery. The three primary metrics used to score NPS nowcast are mean absolute error (MAE), mean squared error (MSE) and Brier score. These metrics are commonly used to measure forecast results. MAE is the mean of the absolute values of the error at each data point (in space and time); it does not distinguish between over-forecasting and under-forecasting:

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| = \frac{1}{n} \sum_{i=1}^n |e_i| \text{ where } n \text{ is the total number of forecast data points}$$

(across space and time), f_i is the forecasted reflectance value (the mean of the PPD), y_i is the observed reflectance value, and e_i is the error (or the difference between them). MSE is the mean of the squared error values at each data point; MSE assigns more weight to larger individual errors than MAE does:

$$MSE = \frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2 = \frac{1}{n} \sum_{i=1}^n (e_i)^2 .$$

The Brier score is used only on the ‘probability of cloud’ output. It is applicable to probabilistic predictions of mutually exclusive outcomes. Thus, it does very well with yes/no

predictions like cloud/no-cloud: $BS = \frac{1}{n} \sum_{t=1}^n (f_t - o_t)^2$, where f_t is the probability of cloud

and o_t is the binary outcome (cloud = 1, no cloud = 0). Zero is a perfect Brier score, while 1.0 is the worst possible score. NPS nowcast results are also compared to that of “land-sea climatology.” This reference forecast simply outputs the mean value of all land (sea)

data points in the training period over land (sea). Lastly, further analysis is performed to determine what factors (geographic location, time, meteorological conditions, etc.) impact NPS nowcast results and how the results are impacted.

IV. RESULTS

A. OVERVIEW

While the overall goal of this study is to accurately predict the short-term low cloud field, the impacts of the grouping scheme and learning period length on forecast skill must be assessed in order to optimize performance. Thus, the results of this study are broken into three sections. Firstly, the impacts of each grouping scheme are examined to determine the ideal method of data grouping. Secondly, the role of learning period length is examined to determine its impact on forecast accuracy. These two sections are broken into three sub-sections: (1) pre-processing of the data, (2) the generalized linear model (GLM) that was produced from training period data, and (3) forecast output. Finally, overall performance is examined to demonstrate the viability of this nowcast approach and to identify its strengths and weaknesses.

B. GROUPING SCHEME CROSS VALIDATION

1. Pre-processing

“Pre-processing” refers to any techniques used to assess the data points before they enter the Bayesian processor. As part of the pre-processing step, several scatter plots were created to assess the relationships between each predictor variable and the observed outcome. Best fit lines were calculated to see if a linear relationship could be found. Figure 17 shows the relationship between each predictor variable and the observed reflectance values for the entire data set, the training and verification periods combined. All predictor variables and observed values are normalized. The scatter distribution (in Figure 17.b) of OMEGA versus reflectance does not lend itself well to a linear fit because its orientation is nearly vertical. The LCLD (17.a) and RH (17.c) scatter distributions lend themselves well to a linear fit, and so higher Beta values (indicating higher weighting) are expected for these two variables.

Figure 18 shows the same scatter plots with sea grid points in blue and land grid points in green. Very distinct groupings of sea and land points can be seen in each plot; the land points tend to cluster towards lower reflectance values as well as lower LCLD

(Figure 18.a), OMEGA (18.b), and RH (18.c) values. This indicates that the land-sea grouping approach may add skill to the model, particularly for θ_2 (corresponding to the LCLD β -coefficient) and θ_3 (corresponding to OMEGA β -coefficient) because the orientations of the sea and land groups look more linear than the original. Figure 19 shows the data points once again, this time grouped into 6 clusters by the k -means cluster approach. The clusters appear to be fairly distinct from each other as they should be by design. The clusters are not as distinct in terms of the range of reflectance values that they contain; for example, the LCLD plot (Figure 19.a) exhibits four clusters—shown in green, dark blue, red and light blue—that contain very similar ranges of reflectance values. Given that these clusters appear tightly grouped, they should lend themselves to a better linear fit and therefore increase forecast skill.

a) LCLD (X) v. reflectance (Y) b) OMEGA (X) v. reflectance (Y) c) RH (X) v. reflectance (Y)

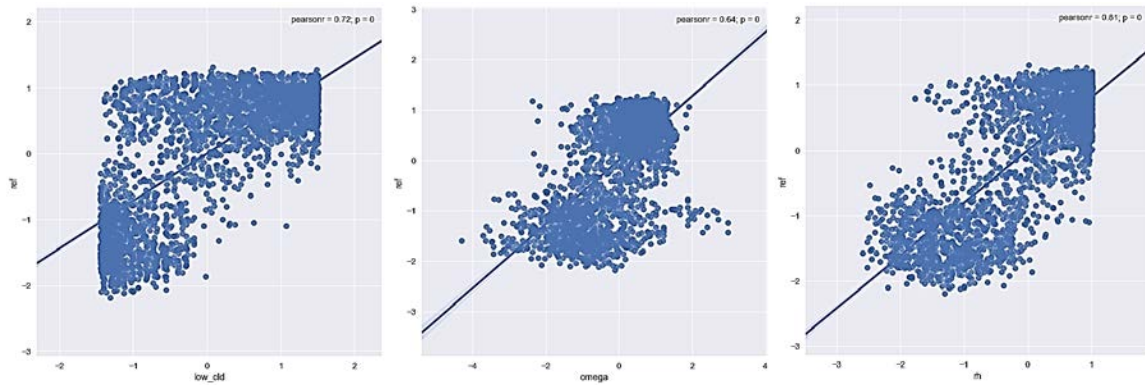


Figure 17. Pre-processed normalized model predictor versus observed reflectance

a) LCLD (X) v. reflectance (Y) b) OMEGA (X) v. reflectance (Y) c) RH (X) v. reflectance (Y)

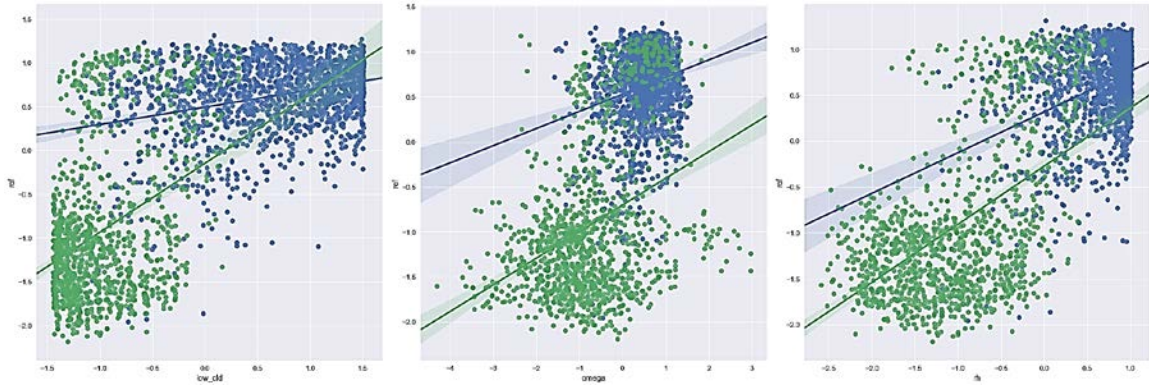


Figure 18. Pre-processed (land-sea) normalized model predictor versus observed reflectance

a) LCLD (X) v. reflectance (Y) b) OMEGA (X) v. reflectance (Y) c) RH (X) v. reflectance (Y)

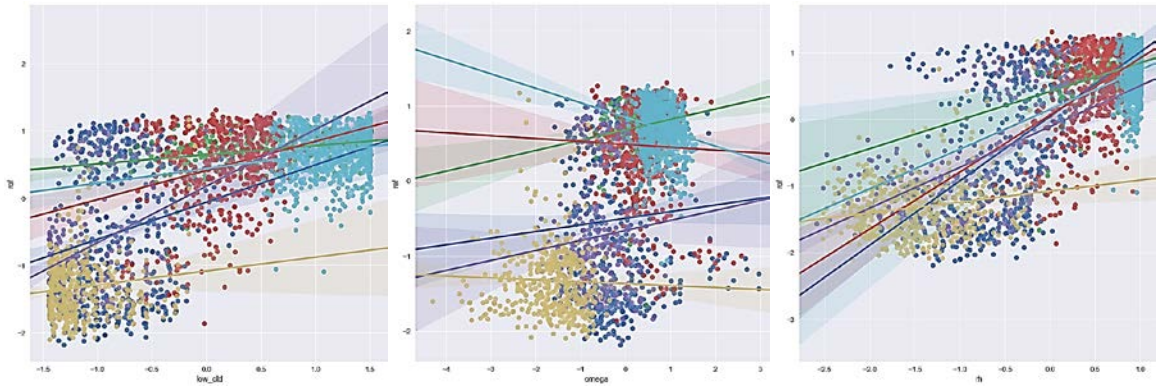


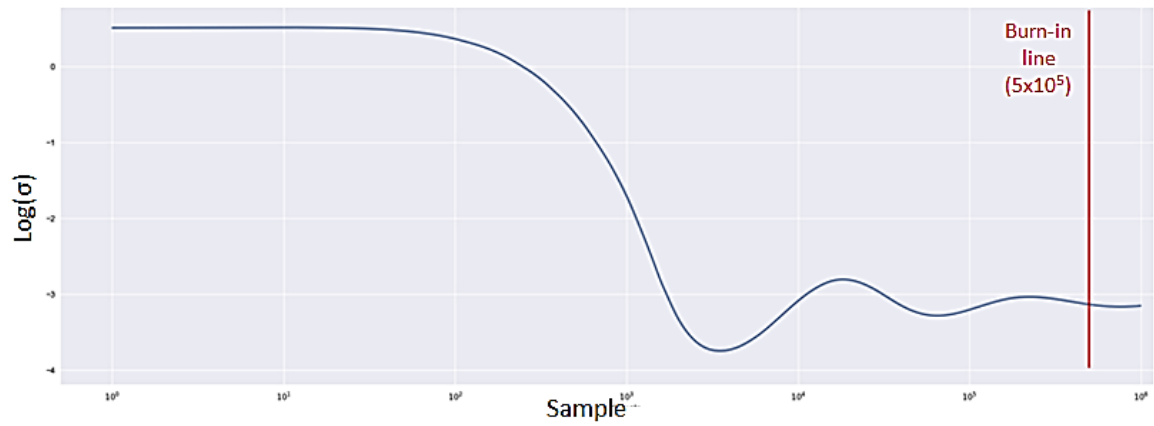
Figure 19. Pre-processed (*k*-means cluster) normalized model predictor versus observed reflectance

2. Generalized Linear Model

During the “training step,” training data is run through the Bayesian processor to generate the GLM. For each test, 1,000,000 MCMC samples were performed. The *posterior probability*, $p(\theta|Y)$, was then generated using 10,000 samples drawn from the last 500,000 samples. Graphical analysis of early trial runs indicated that convergence

occurred well-within the first 500,000 samples, and so 1,000,000 was chosen as a “safe” number of samples. Figure 20 shows the MCMC sampling process reaching convergence before the red vertical “burn-in” line (at 500,000 samples) for the no grouping scheme. The training step took a total time of approximate 3 minutes 20 seconds for the no grouping scheme, 6 minutes 20 seconds for the land-sea scheme, and 18 minutes for the k -means cluster scheme.

a) Jumping Kernel Variance



b) MCMC Acceptance Rate

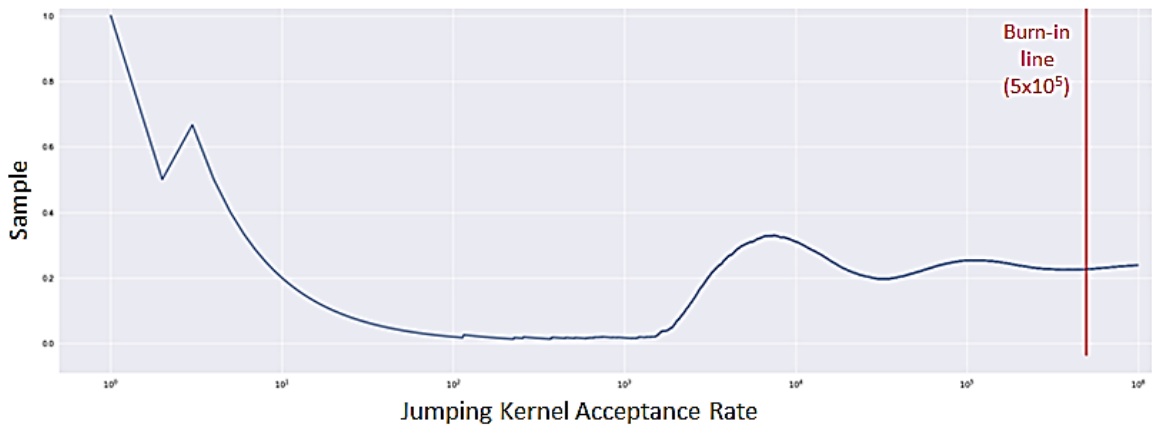
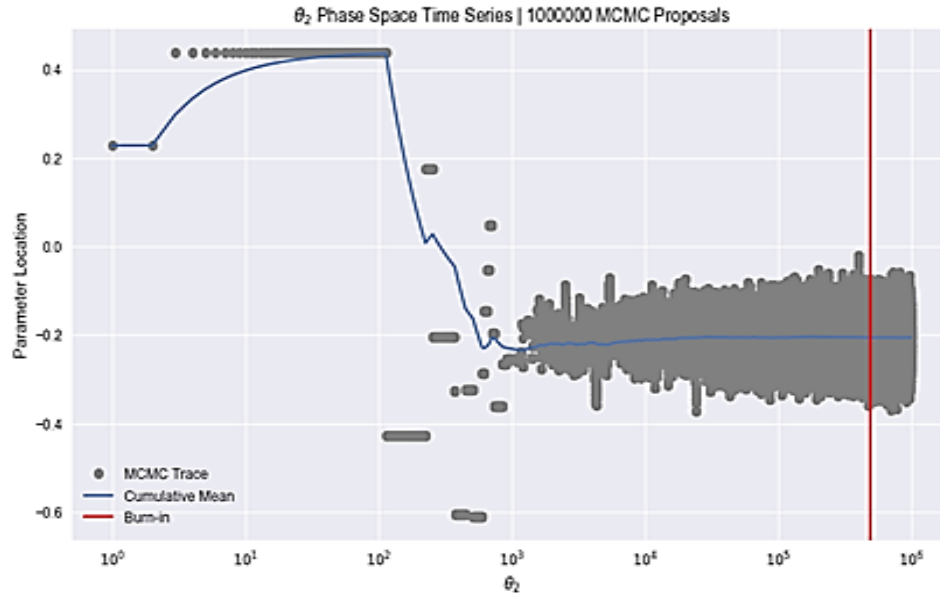


Figure 20. No grouping scheme data set 6 a) jumping kernel variance and b) MCMC acceptance rate

After the training data is run through the Bayesian processor, a posterior probability density function (PDF) is output for each model parameter (θ). θ_1 corresponds to the intercept and is always zero by design. θ_2 , θ_3 , and θ_4 correspond to the Betas of the three predictor variables; the Betas are essentially weighting coefficients assigned to each variable. θ_5 is the natural logarithm of variance, $\ln[\sigma^2]$, and indicates the squared standard deviation or the width of the posterior predictive distribution (PPD) that will be created; thus a small value of θ_5 is desired for predicting a singular reflectance value.

Figure 21 shows the PDF of θ_2 (the coefficient for LCLD) for the no grouping scheme and a graph of the sample convergence process used to generate this specific PDF. In Figure 21.a, notice that the cumulative mean (blue line) of the distribution stagnates on a single value as the distribution converges before the burn-in point (vertical red line) is reached. In Figure 21.b, the 95% highest density interval (HDI) is shown by the solid horizontal red line on the PDF and indicates the interval containing 95% of the distribution. The distribution mean is indicated by the dashed vertical red line on the PDF. The mean and HDI indicate the “meaningfulness” of LCLD relative to the other two predictor variables. The 95% HDI for θ_2 in the no grouping scheme ranges from -0.13 to -0.28 and the mean value is -0.21. In the land-sea scheme, these values were similar. In the *k*-means cluster scheme, they vary greatly with cluster; the mean ranges from -0.56 to +0.19, with several of the HDIs containing zero which indicates relatively no meaningfulness of LCLD for that cluster.

a) MCMC Sampling Convergence Process



b) PDF

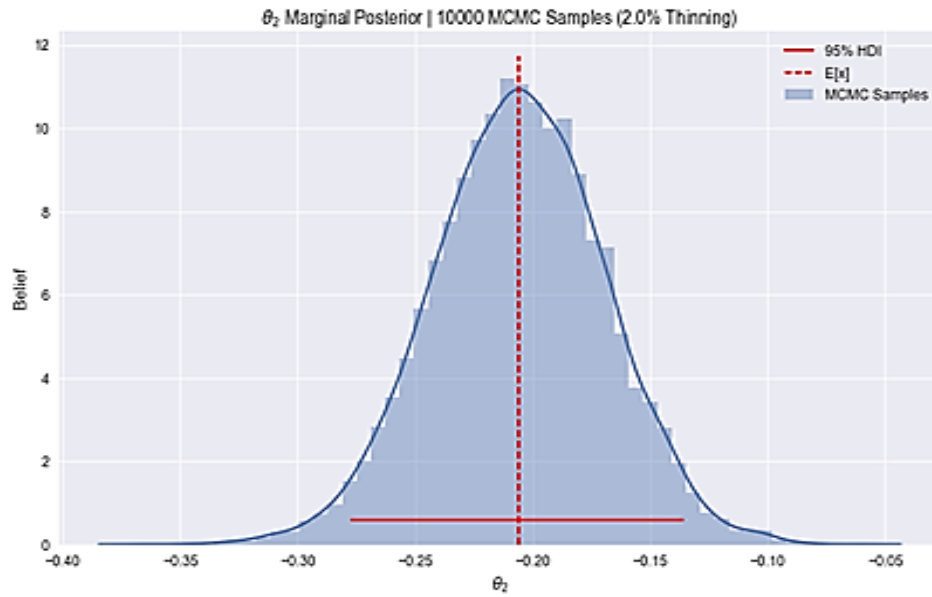


Figure 21. No grouping θ_2 (LCLD) a) sampling convergence process and b) PDF

Shown in Figure 22, the 95% HDI for θ_3 (the coefficient for OMEGA) in the no grouping scheme also contains zero. This indicates that OMEGA is not predictive of low cloud reflectance. However, in the land-sea scheme, θ_3 is predictive, and it is positive

over land and negative over sea as shown in Figure 23 for both the land (23.a) and sea (23.b) regimes, separately. This means that—for this location and training period on June 5—maximum upward vertical motion in the 1000mb–850mb layer is negatively correlated to low cloud over land, and positively correlated to low cloud over the ocean. Thus, the land-sea grouping scheme added significant skill to OMEGA. The *k*-means scheme values also varied with each cluster and added skill to OMEGA for several clusters.

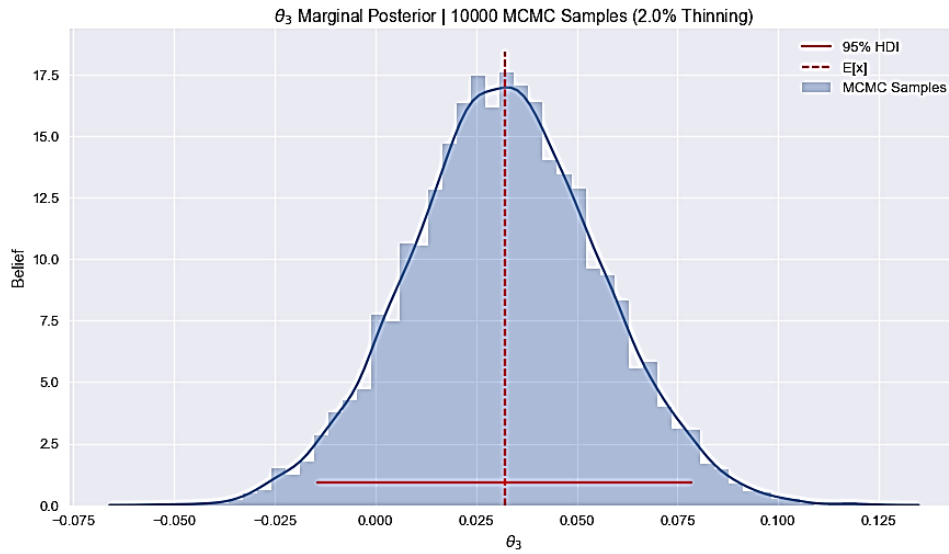
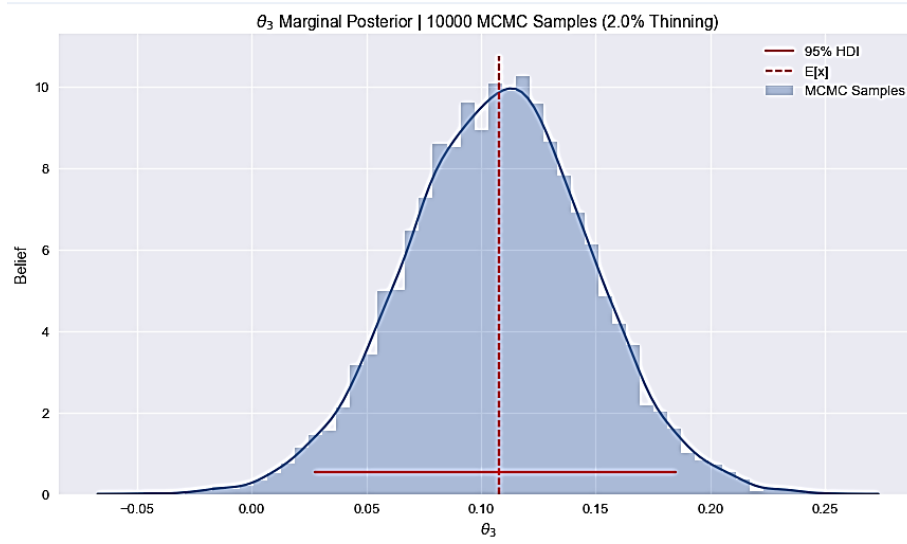


Figure 22. No grouping θ_3 (OMEGA) PDF

a) Land group



b) Sea group

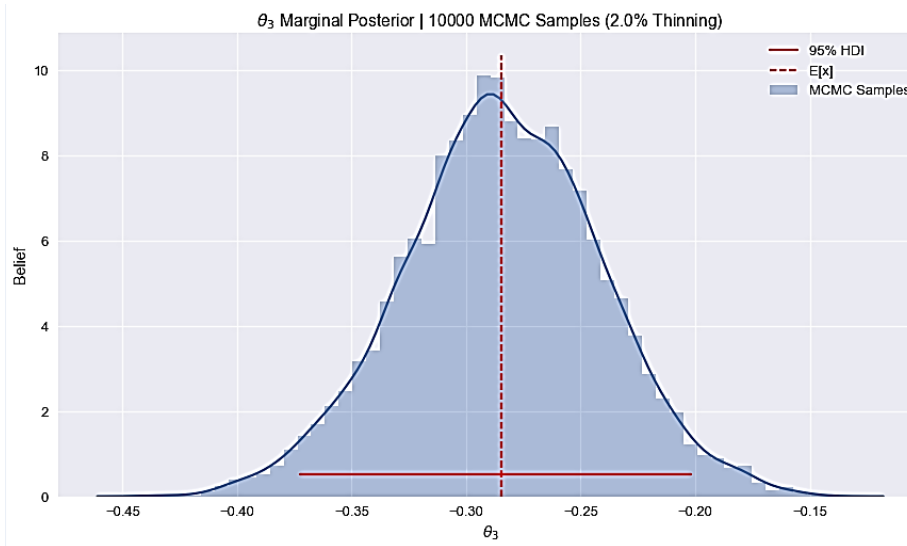


Figure 23. Land-sea θ_3 (OMEGA) PDFs

The 95% HDI for θ_4 (the coefficient for RH) in the no grouping scheme ranges from 0.90 to 1.06 with a mean of 0.98. This is shown in Figure 24. These values were lower in the land-sea scheme (means near 0.70 for both land and sea); however, this simply indicates that RH is relatively less predictive because LCLD and OMEGA are now more predictive in these schemes. In the k-mean scheme, values were all positive

and varied from near zero (not predictive) to 1.08 (very predictive) with cluster. Based on data set 6 only, RH is by far the most useful predictor variable for each data grouping scheme. Figure 24 shows the PDF of θ_5 (the coefficient for variance) for the no grouping scheme. The mean and HDI were similar for all three schemes. The magnitude of θ_5 is very high in data set 6 due to the very short training period (i.e., less data is used to train the model).

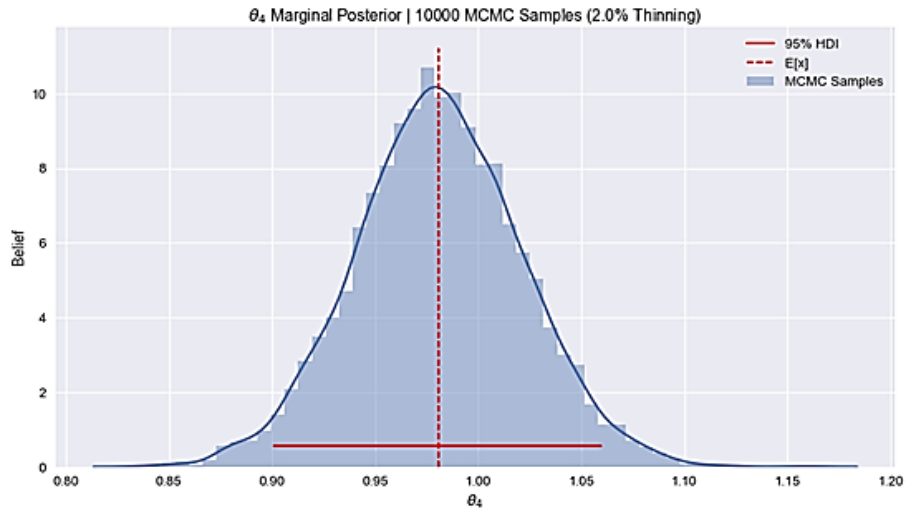


Figure 24. No grouping θ_4 (RH) PDF

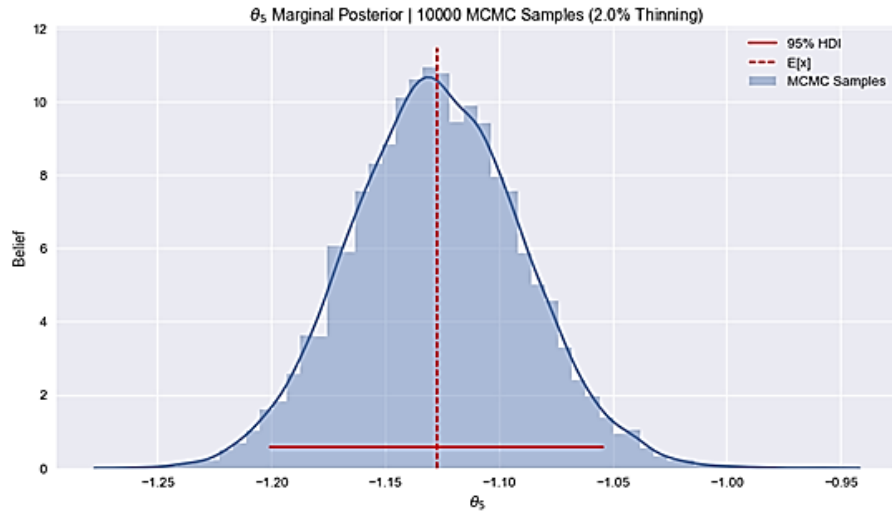


Figure 25. No grouping θ_5 (variance) PDF

3. Performance

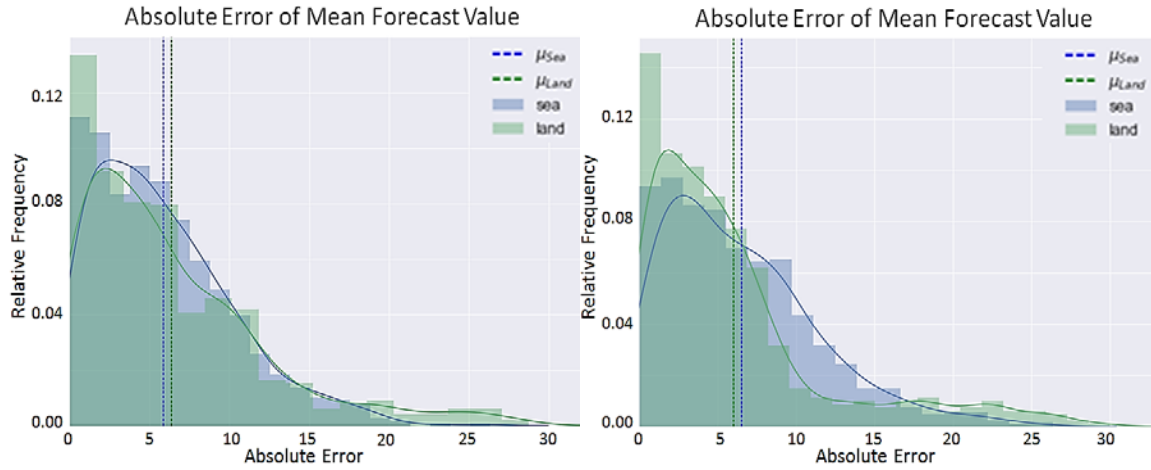
Data set 6 was a very small data set consisting of 10 hours of training data and 10 hours of verification data; it was easy to forecast for because there was little variation in the observed cloud field, which was present over the ocean for most of the period. All three grouping schemes performed similarly in terms of mean absolute error (MAE) and total Brier score on this very short and easy-to-forecast data set. Figure 26 shows the absolute error distribution for each scheme, broken into land (green) and sea (blue) regimes. The means of the land and sea distributions are denoted by the green and blue dashed vertical lines, respectively. The mode of each distribution is close to zero, and most of the mass of each distribution is shifted to the left (toward low error values), which is desirable. Table 1 shows total MAE and total Brier score as well as MAE and Brier score over land and over sea for each scheme. All MAE scores were within 0.7 of each other and all Brier scores were within 0.02 of each other. The no grouping scheme performed the best over sea grid points, which makes sense because there are more sea grid points than land. The land-sea scheme performed best over land, likely because the land group contains fewer data points and experienced little cloudiness throughout the period.

Table 1. Data set 6 MAE and Brier scores for each scheme

	Total MAE	Sea MAE	Land MAE	Total Brier Score	Sea Brier Score	Land Brier Score
No grouping	6.0716	5.8706	6.3697	0.0765	0.0023	0.1864
Land-sea	6.3285	6.5298	6.0298	0.0632	0.0023	0.1534
<i>k</i> -means	6.7561	6.8833	6.5674	0.0832	0.0032	0.2018

a) No grouping scheme absolute error

b) Land-sea grouping scheme absolute error



c) *k*-means cluster scheme absolute error

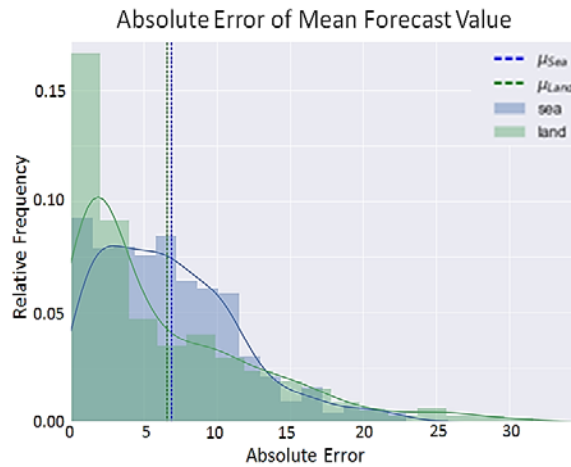


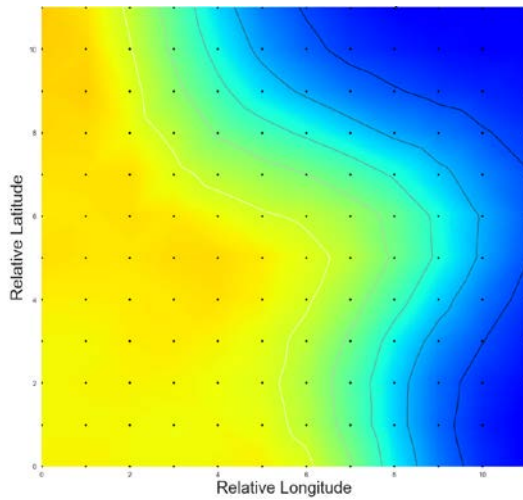
Figure 26. Forecast absolute error distributions of a) no grouping b) land-sea grouping and c) *k*-means cluster

All three schemes are able to capture the general shape and size of the cloud field, but do not necessarily capture the small variations in the extent of cloud cover within the bay and along the coastline. All three also struggled with pinpointing exact reflectance values within the cloud field (5 to 10 percentage points of absolute error were common). The *k*-means cluster scheme appeared to do best with resolving these finer details and some of the more extreme values. Figure 27 shows the NPS nowcast 06-h “pseudo-

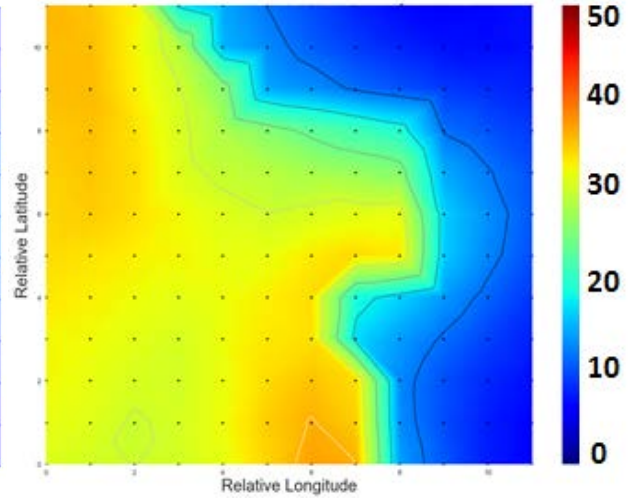
satellite” forecast output for each grouping scheme compared to a “false color” image of the observed cloud field (at the same resolution) for June 6 2100Z. Notice that the *k*-means cluster forecast better resolves the edge of the cloud field along the coastline and also captures some of the higher values within the cloud field. Also, it is worth noting that anything above 35% reflectance looks like thick low cloud to an ISR forecaster; therefore, there is some room for error with extreme values in this particular forecast application.

Although all three grouping schemes performed similarly on this easy “softball” data set, the *k*-means cluster scheme was selected for further testing because it has the greatest potential. In addition to resolving finer structure over water, more cloud edge detail and better handling of extreme values, it is most likely to yield accurate results for longer and more complex data sets. The no grouping and land-sea grouping schemes will likely falter in cases where the verification data points do not match with the training data. This will be explained further in the next section.

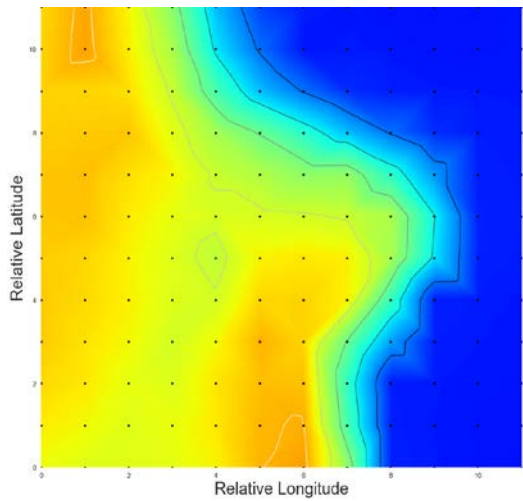
a) No grouping forecast



b) Land-sea grouping forecast



c) *k*-means cluster forecast



d) Observed image

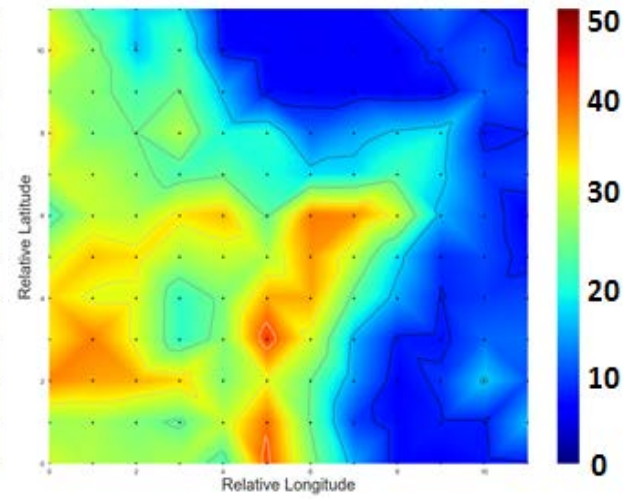


Figure 27. June 6 2016 2100Z forecast for each data grouping scheme versus observed reflectance imagery (continued on next page)

e) Map included for reference

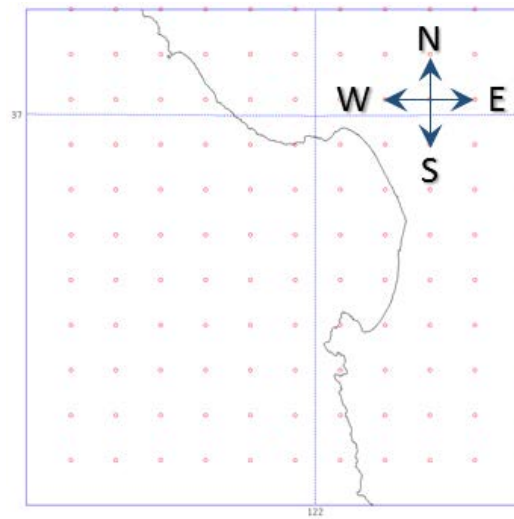


Figure 27. (Continued from previous page)

C. PERIOD LENGTH CROSS VALIDATION

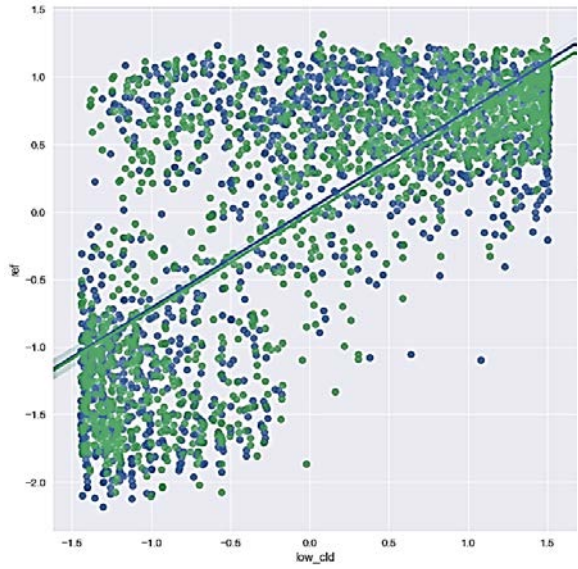
1. Pre-processing

Data set 6 consists of 10 hours of training data from June 5 and 10 hours of verification data from June 6. Clouds were present over much of the bay throughout the period. Both days were similar, and so good results were expected. Figure 28 shows a comparison of training data (blue) to verification data (green) for each predictor. As expected, the training and verification data match up very nicely because there was little variation in the observed cloud field throughout the period. If the verification data points were distributed very differently from the training data, this would be a reason to expect weaker results.

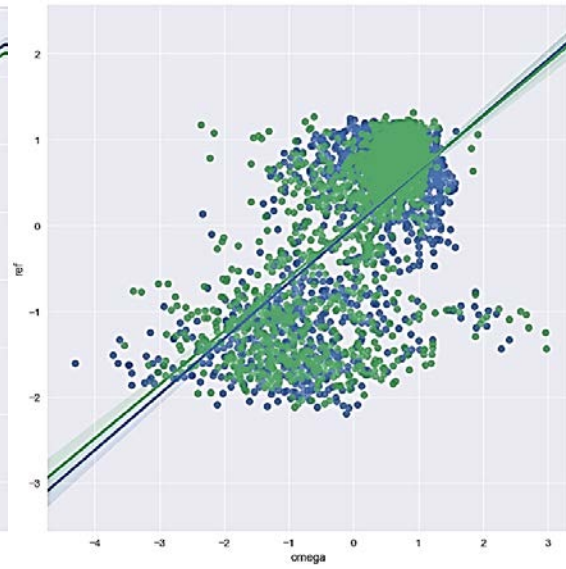
Data sets 7 appears similar in that the verification data matches up fairly well with the training data and a similar best fit line can be drawn through both for each predictor. The larger data sets contain much more variation in the observed cloud field and the corresponding predictor variables. Data set 9 represents a set in which the group of verification data points is notably different from the group of training data points. Figure 29 shows the comparison of training and verification data for set 9. It is apparent that the

observed reflectance values were anomalously low during the training period due to very little cloud cover present during this time. The best fit lines for the verification data are very different than that of the training data for each predictor. This is a cause for concern. In the no grouping scheme, this would mean an inappropriate bias correction and poor results. However, in the k-means cluster scheme, there is hope that the verification data will fall into clusters with the appropriate bias correction. For data sets 8 and 10, the training and verification data does not match up as nicely as in sets 6 and 7; however, the training and verification best-fit lines are not as dissimilar as in set 9.

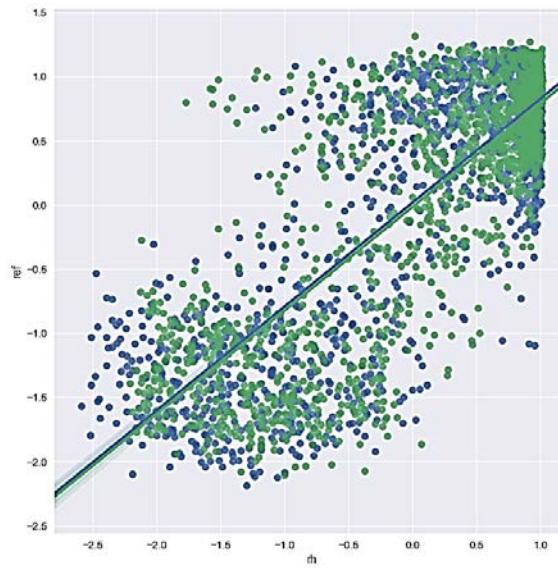
a) LCLD (X) v. reflectance (Y)



b) OMEGA (X) v. reflectance (Y)



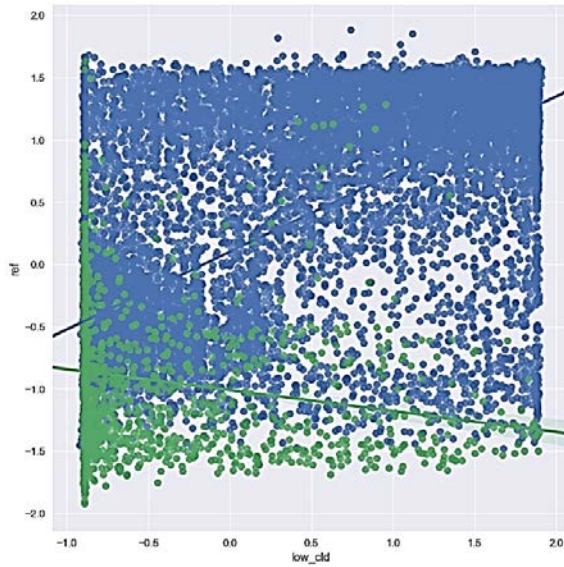
c) RH (X) v. reflectance (Y)



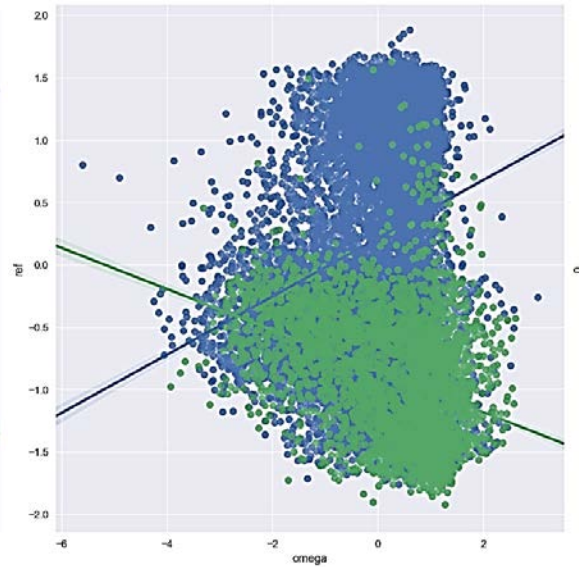
Note that data points are broken up into training (blue) and verification (green) points—not land and sea.

Figure 28. Data set 6 pre-processed training and verification data for a) LCLD, b) OMEGA, and c) RH versus observed reflectance

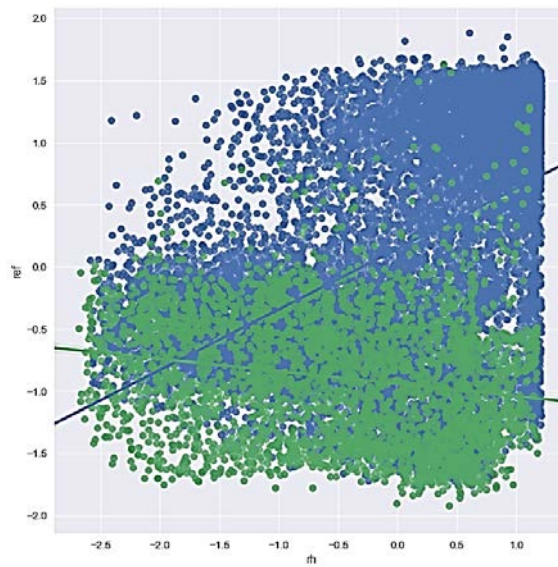
a) LCLD (X) v. reflectance (Y)



b) OMEGA (X) v. reflectance (Y)



c) RH (X) v. reflectance (Y)



Note that data points are broken up into training (blue) and verification (green) points—not land and sea.

Figure 29. Data set 9 pre-processed training and verification data for a) LCLD, b) OMEGA, and c) RH versus observed reflectance

2. Generalized Linear Model

Table 2 shows the total time it took to run each data set through the Bayesian processor in order to generate the GLM. Data set 10 (163 hours of training data) took the longest to run at 25.5 minutes. The time for each cluster varies with the amount of data in each and ranged from 2 minutes 38 seconds to 5 minutes 8 seconds. These times are very reasonable because the “training step” will not need to be run hourly for operational use. For instance, it could be run automatically once per day, and the updated GLM can be used for the next 24 hours until the next update. The results suggest that daily updating would be more than sufficient, as data sets 7–10 represent cases where the same GLM is used for multiple days without sacrificing performance as the period continues.

Table 2. Total time of “training step”

Data Set 6	Data set 7	Data set 8	Data set 9	Data set 10
~18 min	~17 min	~20 min	~23 min	~25 min 30 sec

In the *k*-means cluster scheme, the predictor variable Betas (corresponding to θ_2 , θ_3 and θ_4) varied greatly. Table 3 shows the mean parameter (θ) values (not including θ_1 , which equals zero) for each cluster within each data set. The mean value was not included if the 95% HDI contained zero. The variation of Betas which occurred between clusters was expected; the *k*-means clustering scheme is essentially designed to identify groups of data points with a distinct set of Betas. However, several of the clusters also yielded negative values of either θ_2 or θ_4 , but never both. This indicates that an increase in the variable (LCLD or RH) is correlated to a decrease in observed reflectance value in these clusters. This relationship is not readily meteorologically intuitive, but that does not mean that there is not statistical reasoning for it based on patterns of predictor variable bias in each cluster.

There were also large variations of the Betas across data sets, which was unexpected. In data set 6, θ_4 (the coefficient for RH) is relatively large and positive for

most of the clusters, while θ_2 (the coefficient for LCLD) is negative for all but one cluster. RH was clearly the dominant predictor variable for data set 6. As the size of the data set is increased, RH (corresponding to θ_4) becomes less dominant and LCLD (corresponding to θ_2) becomes more dominant. In data set 10, θ_2 is positive in every cluster, and LCLD is the dominant variable.

A possible explanation of this data set variation is that RH may perform better when more low cloud cover is present, and LCLD may perform better when less low cloud is present. Extensive low cloud cover was present over the bay on June 5–6. This is the time period of data set 6, but this time period is also contained in each data set. Therefore, if anomalously high amounts of cloud cover during these days caused RH to be very predictive and LCLD to not be very predictive, then adding more data (which included more clear sky conditions) would gradually decrease the influence of June 5–6 until RH was no longer the dominant variable. This could very well have something to do with the relationship between RH and LCLD, shown in Figure 30. This relationship is shown for (a) data set 6 and (b) data set 10 to show that it was fairly consistent across all data sets. Notice that at high RH values (at the far right of the plots) that correspond to high amounts of cloud, there are large variations in possible LCLD values; these large variations may inhibit the NPS nowcast’s ability to use LCLD as a meaningful predictor variable. On the other hand, when RH values are low (left side of both plots in Figure 30), corresponding to clear sky conditions, there is very little variation in LCLD, which may make it a better predictor variable. This relationship makes sense from a meteorological perspective; most sky cover conditions (0 to 100%) occur within about a 30% range of RH values.

For the most part, θ_3 (the coefficient for OMEGA) was negative and close to zero, indicating a relatively weak correlation between upward vertical motion and low cloud. As expected, the size of the HDI decreases with more data. θ_5 (corresponding to $\ln[\sigma^2]$) tended to decrease in magnitude with more data.

Table 3. Model parameter (θ) means

		Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Set 6	$\theta_2(\text{LCLD})$	-0.25	contains 0	contains 0	-0.56	contains 0	0.19
	$\theta_3(\text{OMEG})$	0.18	contains 0	-0.19	contains 0	contains 0	-0.34
	$\theta_4(\text{RH})$	0.82	0.41	0.47	1.08	contains 0	contains 0
	$\theta_5(\sigma^2)$	-0.61	-0.4	-0.23	-0.79	0.01	-0.10
Set 7	$\theta_2(\text{LCLD})$	-0.23	-0.31	-0.14	-0.19	-0.18	-0.20
	$\theta_3(\text{OMEG})$	-0.15	contains 0	0.21	contains 0	0.12	contains 0
	$\theta_4(\text{RH})$	0.52	0.19	0.45	0.34	0.60	0.65
	$\theta_5(\sigma^2)$	-0.34	-0.05	-0.17	-0.04	-0.30	-0.44
Set 8	$\theta_2(\text{LCLD})$	0.12	0.31	0.15	contains 0	-0.08	-0.22
	$\theta_3(\text{OMEG})$	contains 0	-0.18	-0.08	contains 0	-0.09	-0.10
	$\theta_4(\text{RH})$	0.35	-0.05	0.37	0.17	0.23	0.53
	$\theta_5(\sigma^2)$	-0.20	-0.15	-0.20	-0.03	-0.04	-0.31
Set 9	$\theta_2(\text{LCLD})$	0.05	0.38	0.55	0.19	-0.09	contains 0
	$\theta_3(\text{OMEG})$	contains 0	-0.20	-0.25	contains 0	-0.13	-0.05
	$\theta_4(\text{RH})$	0.31	-0.10	-0.14	-0.18	0.32	0.14
	$\theta_5(\sigma^2)$	-0.11	-0.24	-0.34	-0.02	-0.12	-0.03
Set 10	$\theta_2(\text{LCLD})$	0.13	0.52	0.08	0.47	0.11	0.45
	$\theta_3(\text{OMEG})$	-0.16	-0.27	-0.25	-0.29	-0.09	-0.35
	$\theta_4(\text{RH})$	0.05	-0.29	0.27	-0.19	0.13	contains 0
	$\theta_5(\sigma^2)$	-0.04	-0.25	-0.18	-0.43	-0.04	-0.50

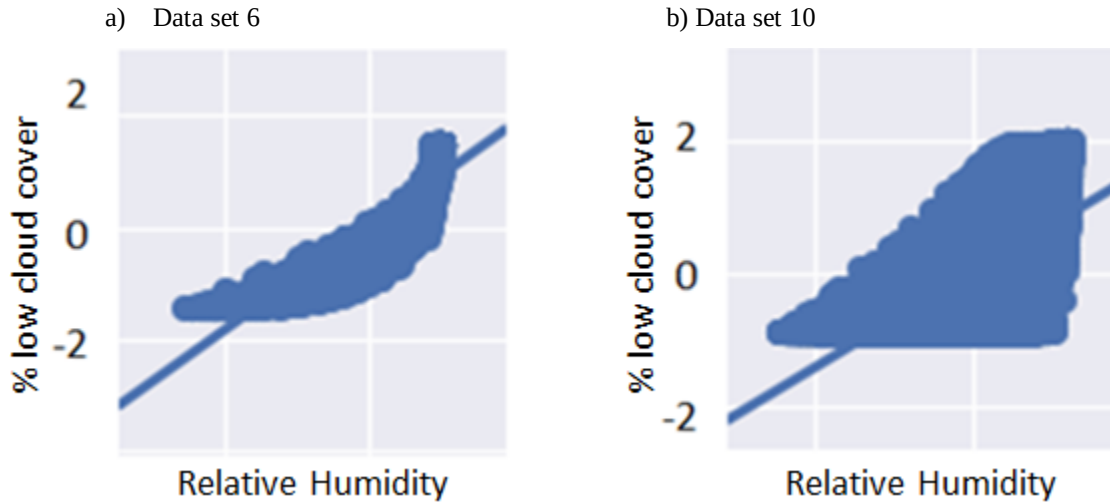


Figure 30. Relationship between normalized RH and LCLD in a) data set 6, and b) data set 10

3. Performance

Although the *k*-means clustering scheme was used to generate all of the following forecast results, many of the scores and graphics contain information about land and sea data points. It is useful to continue to split results into these two regimes in order to gain additional detail about performance over each. This allows the modeler to see how well the NPS nowcast accounts for the differences in these two regimes when *k*-means clustering is used.

As shown in Table 4 and Figure 31, the NPS nowcast achieved its lowest MAE score (4.38 total) on data set 9, even though pre-processing scatter plots looked concerning because the verification (or forecast) period data points did not match up well to training period data points. As shown in Figure 32.a, the forecast period for data set 9 exhibited very low observed reflectance values corresponding to very little observed cloud—most of the data points are grouped toward the lower left corner. The *k*-means clustering technique was able to account for this by grouping the majority of the forecast period data points into three clusters that yielded very low absolute error; this is shown in Figure 32.b. Each scatter point corresponds to a mean forecast value. Points are broken into land (green) and sea (blue). Although this success may also be attributed to the ability of the predictor variables to make accurate predictions when no cloud is present,

the fact that the majority of the data points were grouped into 3 clusters indicates that the *k*-means cluster scheme likely added skill over a no grouping or land-sea grouping scheme.

Data set 10 consisted of an even longer training period and also yielded impressive results. The MAE score was a bit higher (5.53 total), but the forecast period for this set contained much greater variability of observed values, particularly over the sea grid points, where set 10 did not perform as well. This increased variability is seen in Figure 33, the plot of mean forecast value versus observed for data set 10. Notice that unlike Figure 32.a, the sea data points in Figure 33 are spread throughout the plot.

Sets 9 and 10 did not do quite as well as sets 6 and 7 with Brier scores, particularly over the sea grid points. This likely has to do with the fact that there were more clear sky conditions in sets 9 and 10; it is more difficult to score a perfect Brier score with clear sky conditions because the window of reflectance values below the threshold ($< 5.0\%$ = 'no cloud' for sea grid points) is very small. This effect is apparent in Figure 34. Notice that the forecast image (Figure 34.a), which takes the mean value of the PPD, looks very similar to the observed (34.b); however, because these mean values are very close to the threshold (5% reflectance over sea), this translates to a non-zero probability of cloud (34.c) and therefore a non-zero Brier score in the same geographic location (34.d). Figure 34.d shows a map of individual Brier scores calculated for each grid point at 2100Z. It may be appropriate to choose a higher 'no cloud' threshold for the sake of fair scoring, even though 5.0% reflectance was the threshold observed from the GOES imagery.

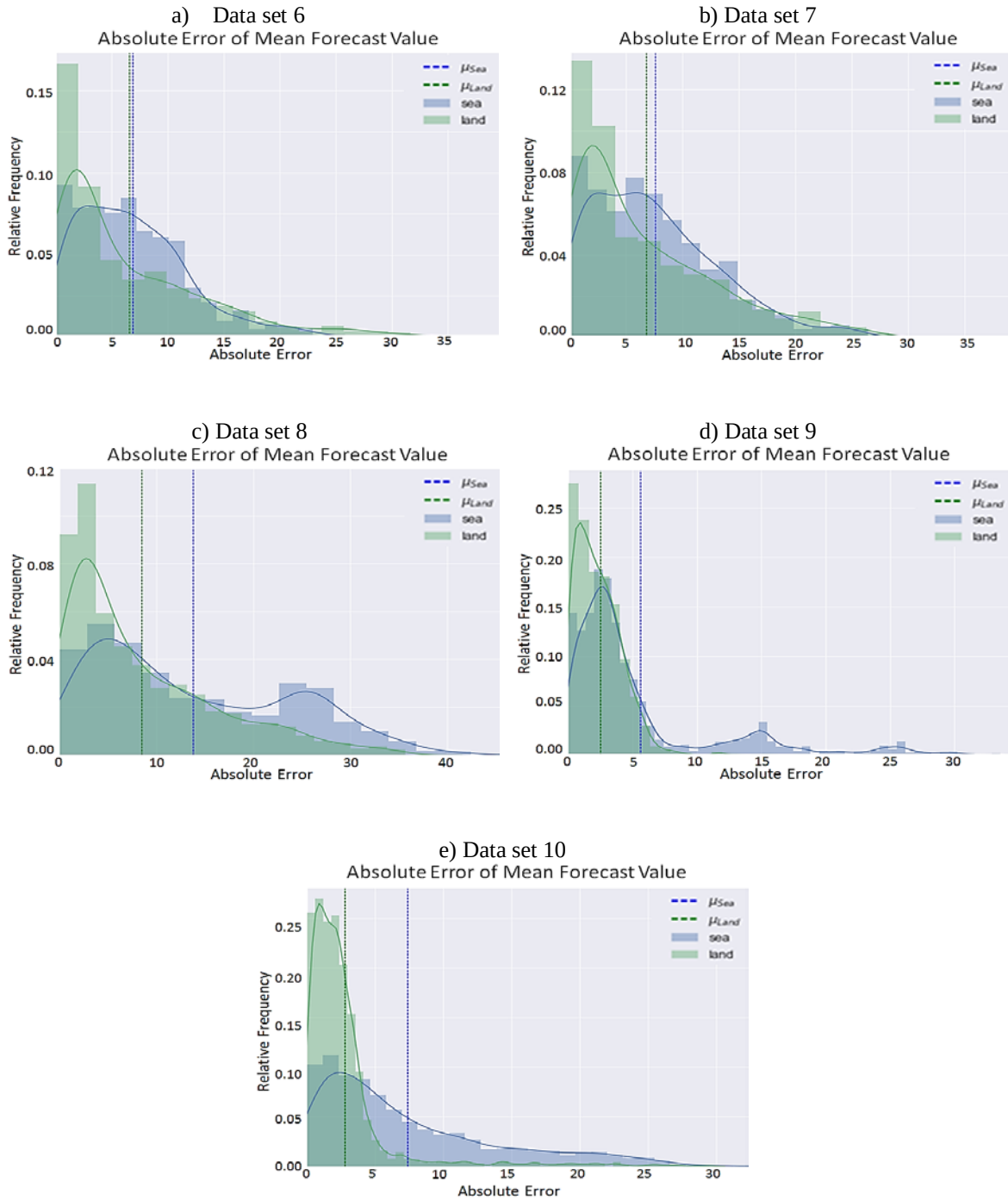


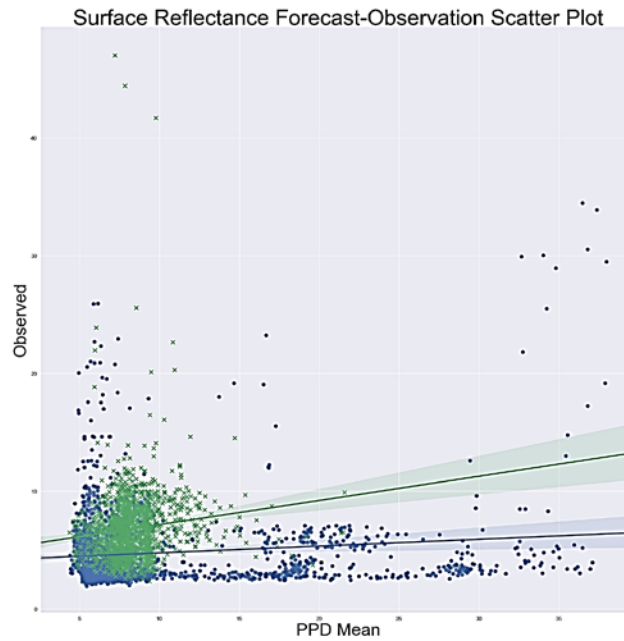
Figure 31. Absolute error of mean forecast value of a) set 6 b) set 7 c) set 8 d) set 9, and e) set 10

Table 4. MAE and Brier scores for each data set

	Total MAE	Sea MAE	Land MAE	Total Brier Score	Sea Brier Score	Land Brier Score
Set 6	6.7561	6.8833	6.5674	0.0832	0.0032	0.2018
Set 7	7.3071	7.6313	6.8264	0.0837	0.0032	0.2020
Set 8	11.6402	13.7798	8.4677	0.2421	0.2589	0.2172
Set 9	4.3779	5.6319	2.5185	0.1585	0.2311	0.0508
Set 10	5.5260	7.3719	2.7889	0.1345	0.1801	0.0668

The best (lowest) score for each metric is shown in blue. The worst (highest) score for each metric is shown in red.

a) Scatter plot of mean forecast value versus observed reflectance



b) Scatter plot of absolute error grouped by cluster

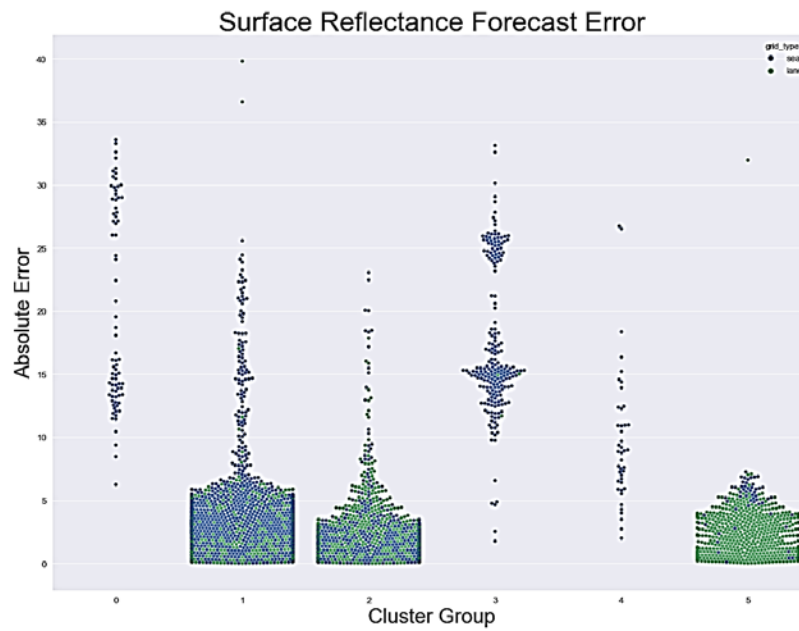


Figure 32. Data set 9 scatter plots of a) mean forecast value versus observed reflectance, and b) absolute error grouped by cluster

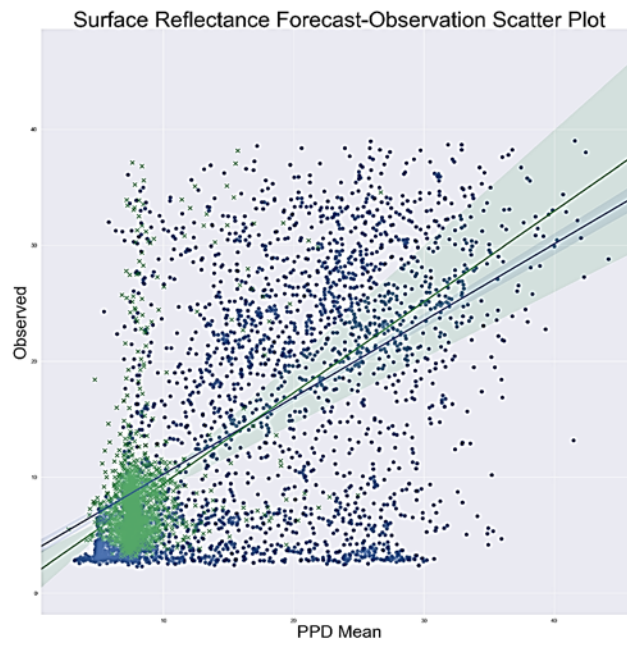


Figure 33. Data set 10 scatter plot of mean forecast value versus observed reflectance

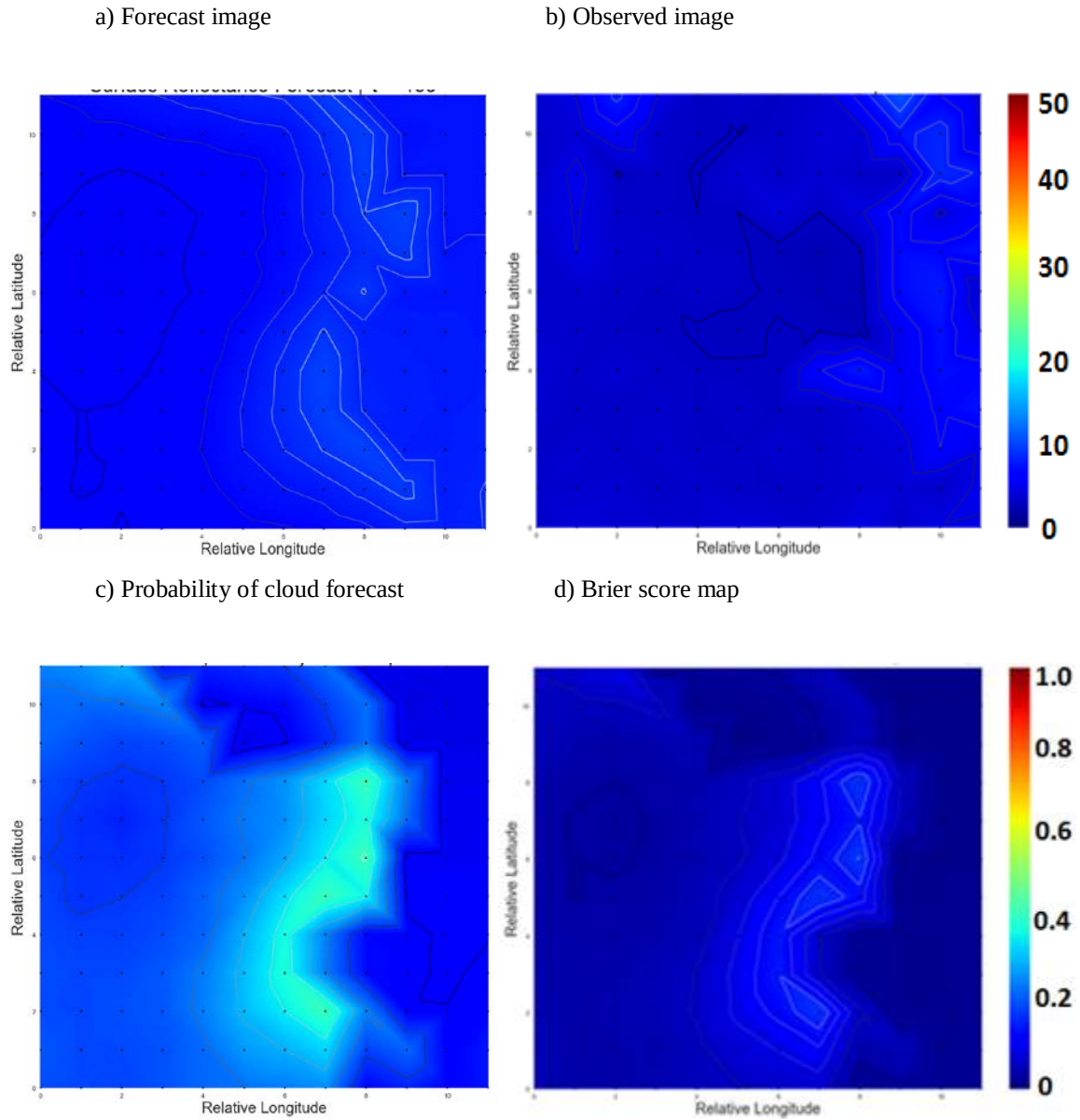


Figure 34. Data set 9 June 20 2016 1700Z a) forecast image, b) observed image, c) probability of cloud product, d) Brier score map, and e) reference map (continued on next page)

e) map included for reference

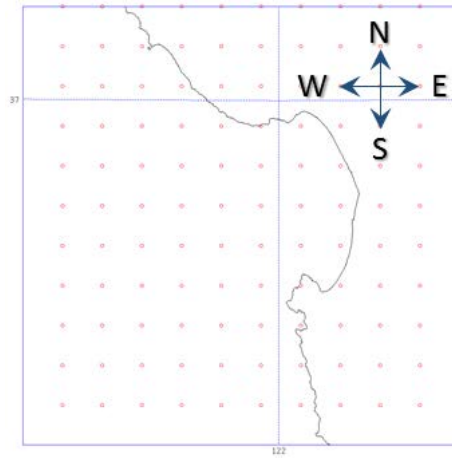


Figure 34. (Continued from previous page)

Also evident in Table 4 and Figure 30 (shown previously) is that data set 8 scored the worst (highest) on every scoring metric. One significant reason for this is that there was a large amount of variation in the observed cloud field over land and especially over sea grid points. This variability is shown in Figure 35, a plot of forecast mean values versus observed values. Another potential reason is that the training period data points did not match up well to the forecast period data points, and due to the variability in the forecast period cloud field, the *k*-means cluster scheme was not able to solve this problem as it did in sets 9 and 10. Lastly, there were several irregular cloud patterns observed during the forecast period that were not present during the training period. One example is shown in Figure 36. It is reasonable to assume that a longer training period—which would capture more variability—would have improved these results. The main reason that sets 6 and 7 performed well is likely that the training data and forecast data matched up so well. Figure 37 shows a grey-scale forecast image compared to the observed image for data set 7. The predicted pseudo-satellite image (Figure 37.a) shows similarity to the observed image (37.b) but lacks the finer scale structure. This lack of detail in the predicted image is most likely due to a lack of resolution in the model forecast fields.

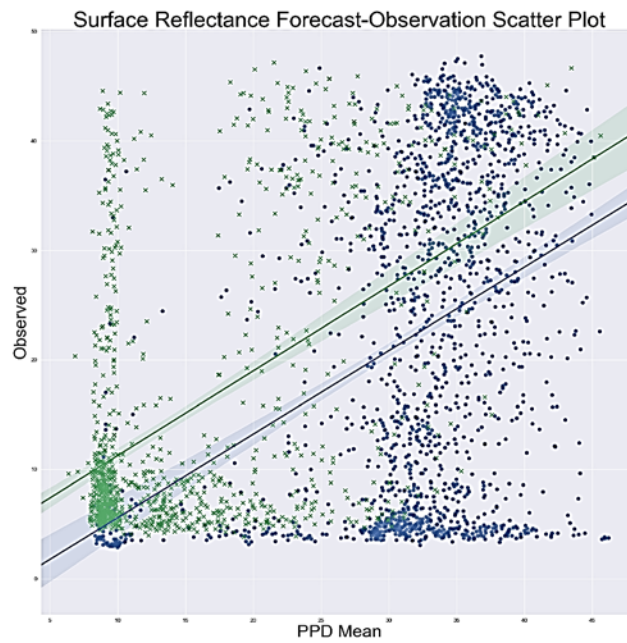


Figure 35. Data set 8 scatter plot of mean forecast value versus observed reflectance

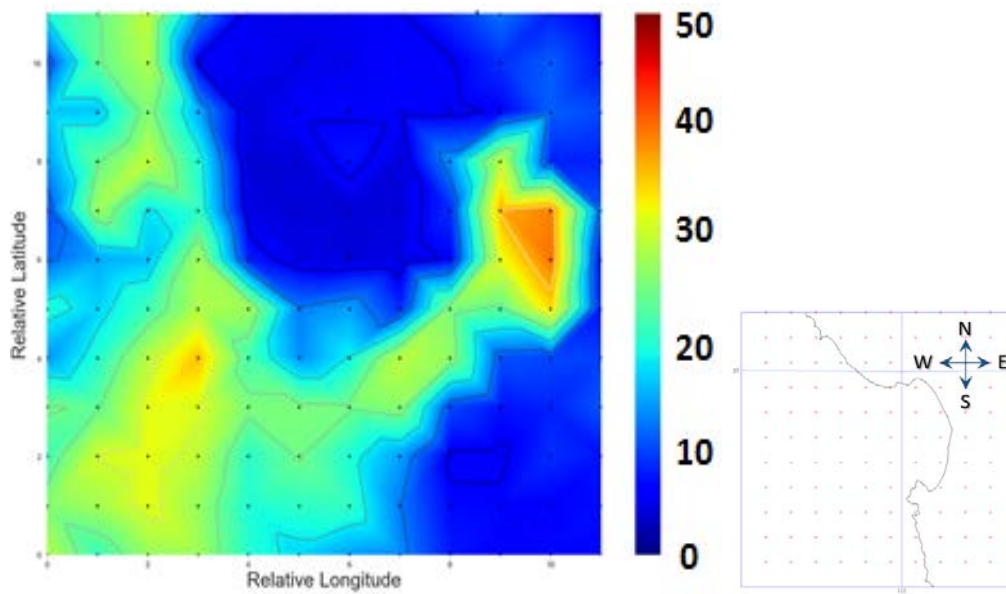


Figure 36. Data set 8 June 13 2016 2100Z observed image with reference map included

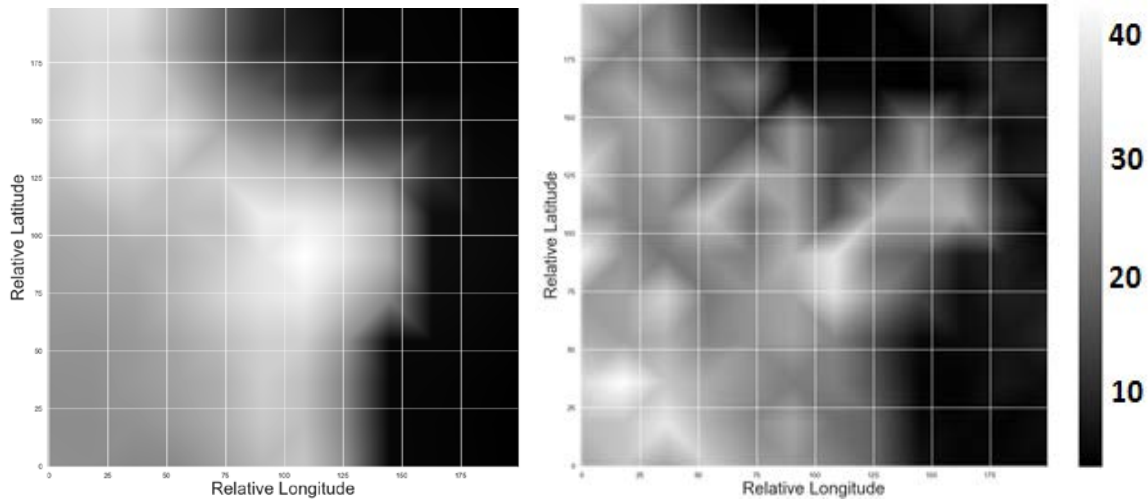


Figure 37. Data set 7 June 06 2200Z grey-scale image a) forecast and b) observed

D. GENERAL PERFORMANCE

In order to create skill scores for NPS nowcast, a climatology-based forecast method was created called “mean land-sea climatology.” “Land-sea climatology” scored fairly well in terms of absolute error, particularly over land and particularly over the smaller data sets that saw less variation in the observed cloud field. Figure 38 displays the land-sea climatology forecast calculated for data set 8. This “climocast” remains the same for each forecast hour in the data set.

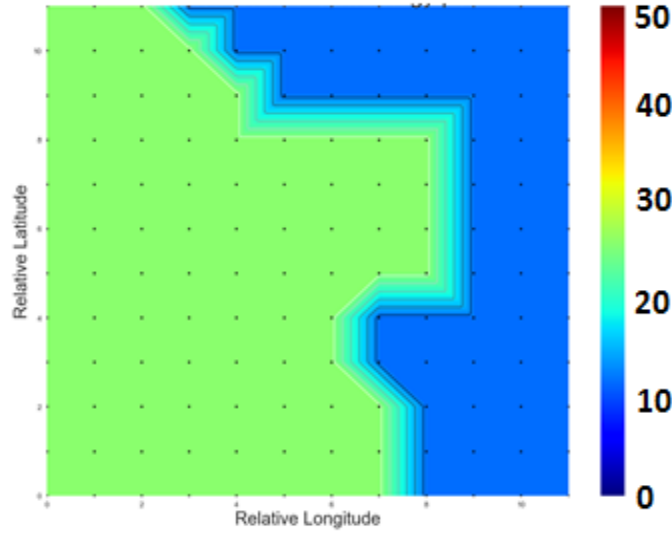


Figure 38. Data set 8 land-sea “climocast”

“Land-sea climatology” was used to compute skill scores for MAE, mean squared error (MSE) and Brier score. A skill score is used to determine the relative improvement of a forecast (NPS nowcast) over a reference forecast (“land-sea climatology”). The formula used to compute a skill score is: $skill\ score = \frac{score_{forecast} - score_{reference}}{score_{perfect} - score_{reference}}$. In this case, the $score_{perfect}$ equals zero because zero MAE, MSE and Brier scores indicate perfect forecasts. A positive skill score indicates that NPS nowcast outscored land-sea climatology. 1.0 is the highest attainable skill score. The MSE skill score is also known as a reduction of variance.

Table 5 shows the skill scores for each data set. In the shorter data sets—6, 7 and 8—NPS nowcast did not add significant skill in terms of MAE and MSE. However, NPS nowcast displayed much better distributions of absolute error than did land-sea climatology. The land-sea climatology distributions for each data set are shown in Figure 39. Notice that the mode, or the most frequently observed value, occurs toward the center of the land-sea climatology absolute error distribution for each data set; all of the NPS nowcast MAE distributions (shown previously in Figure 31) had the bulk of the error shifted to the left (toward zero), and so were much more desirable distributions.

Table 5. Skill scores

	MAE skill score	MSE skill score	Brier skill score
6	-0.0196	-0.0224	0.8621
7	-0.0644	-0.0620	0.7891
8	0.0323	-0.0716	0.0384
9	0.6717	0.8080	0.4659
10	0.2850	0.2305	0.3396

Data sets 9 and 10 added noticeable skill to climatology in all three skill categories. This demonstrates the ability of larger training periods to capture more variability and produce more accurate forecasts. The skill scores for data set 9 were also aided by the fact that the forecast period did not match up well to the training period; this caused very high error in land-sea climatology (which is derived from training period data) but did not negatively impact the NPS nowcast.

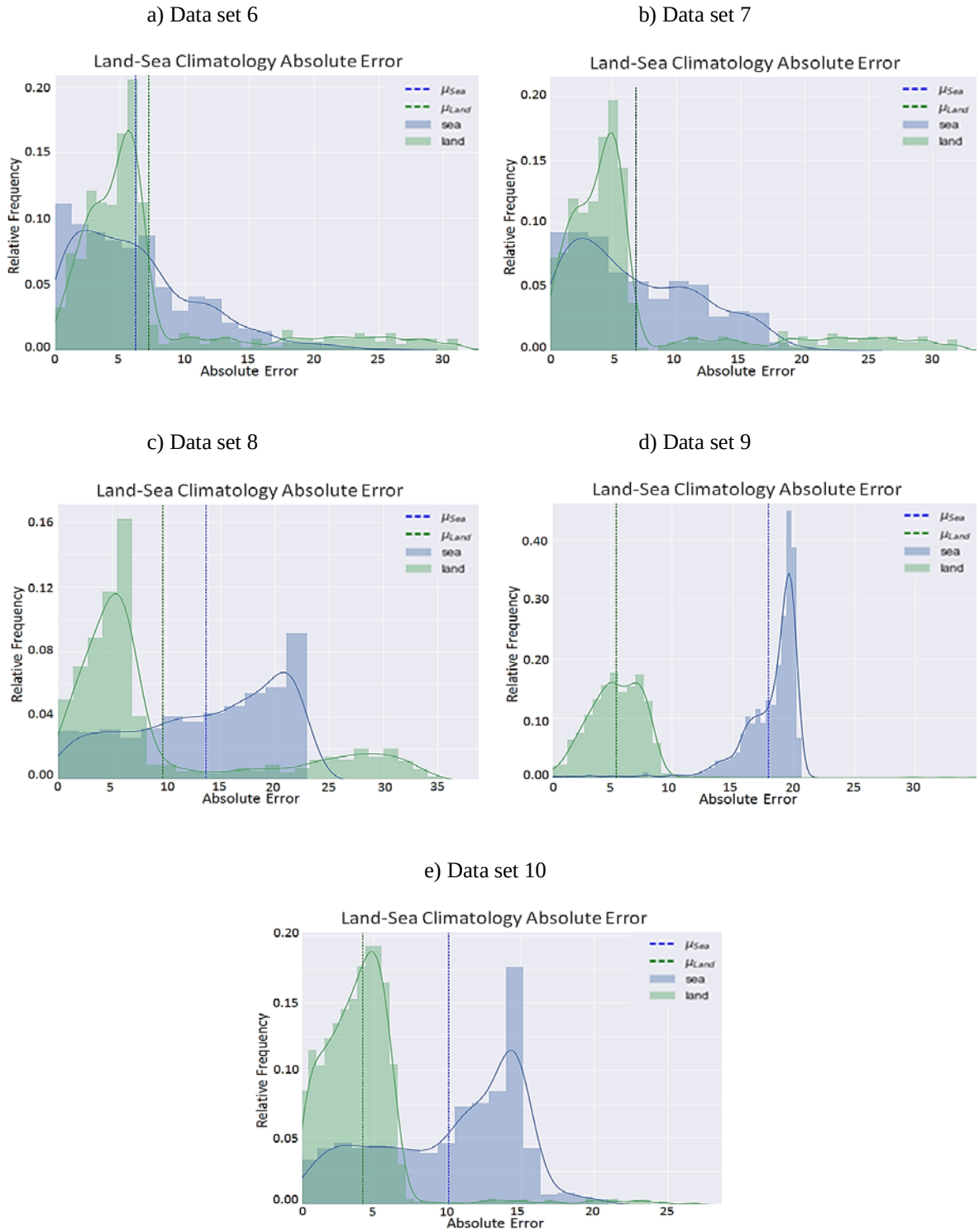


Figure 39. Land-sea climatology error distribution for each data set

Additional analysis of all results was conducted to assess the impacts of geographic location and valid time on forecast results. This analysis focuses on data sets 9 and 10 because cross validation confirmed that the longer training periods used in these sets yield better forecast results. Interestingly, in both data sets 9 and 10, the NPS nowcast identified very similar regions of distinctly higher cloud probability; these regions correspond roughly to the land and sea regions. Essentially, the NPS nowcast is able to identify the land and sea regions even though it did not know whether each grid point was a land or sea point (the *k*-means cluster scheme does not include this information). Thus, additional grouping into land and sea grid points is not necessary. This is evident in the heat map of mean cloud probability forecasts for data sets 9 and 10 at each of the 144 grid points, shown in Figure 40. On these plots, the increasing Y direction points toward the west; the mapped coastline is included for reference.

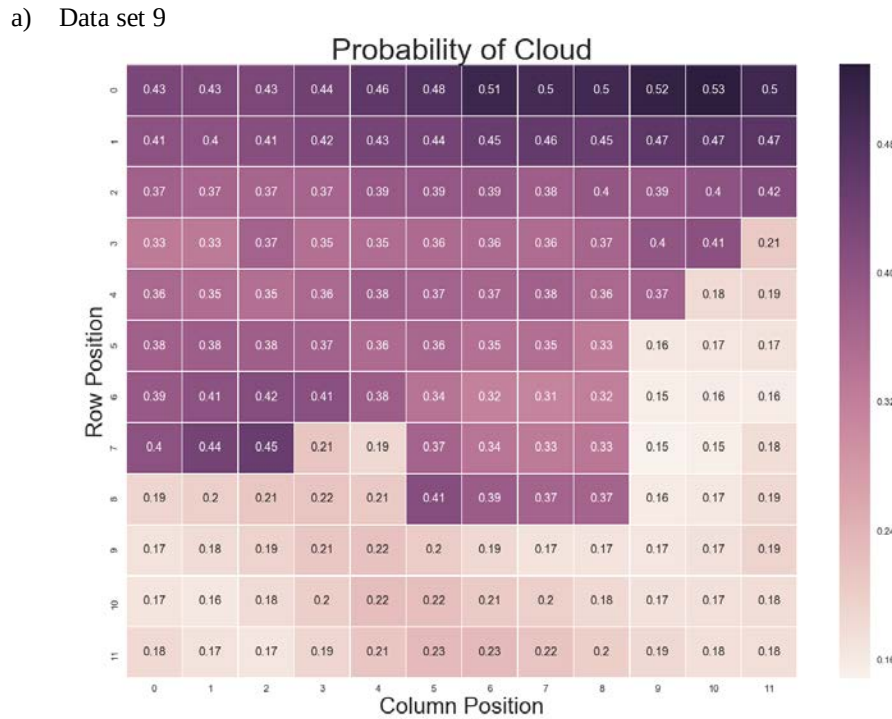
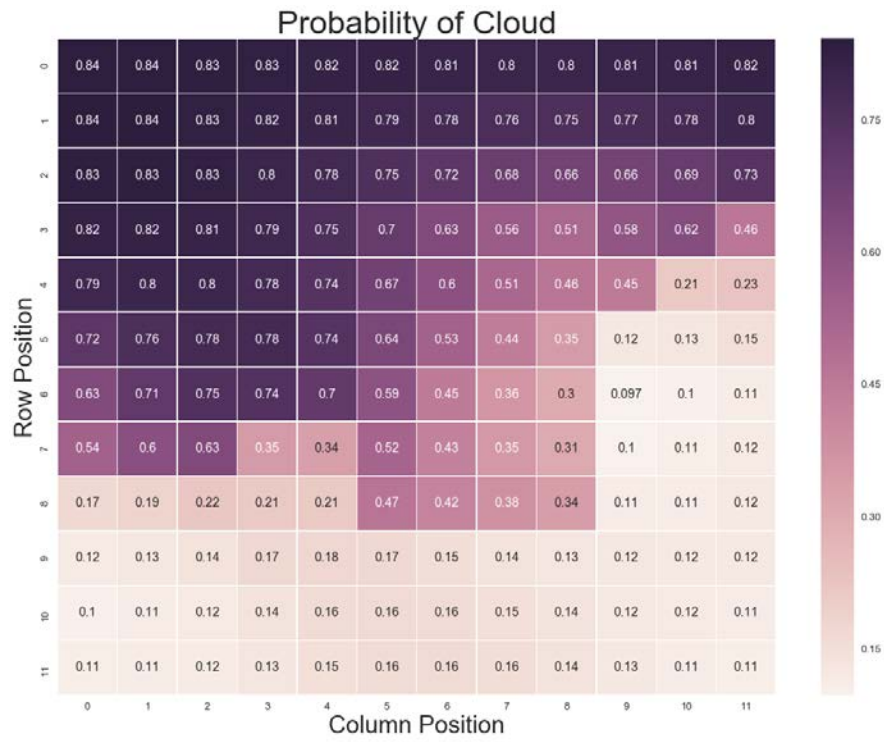


Figure 40. Probability of cloud heat map for a) data set 9 and b) data set 10 with c) reference map included (continued on next page)

b) Data set 10



c) Reference map

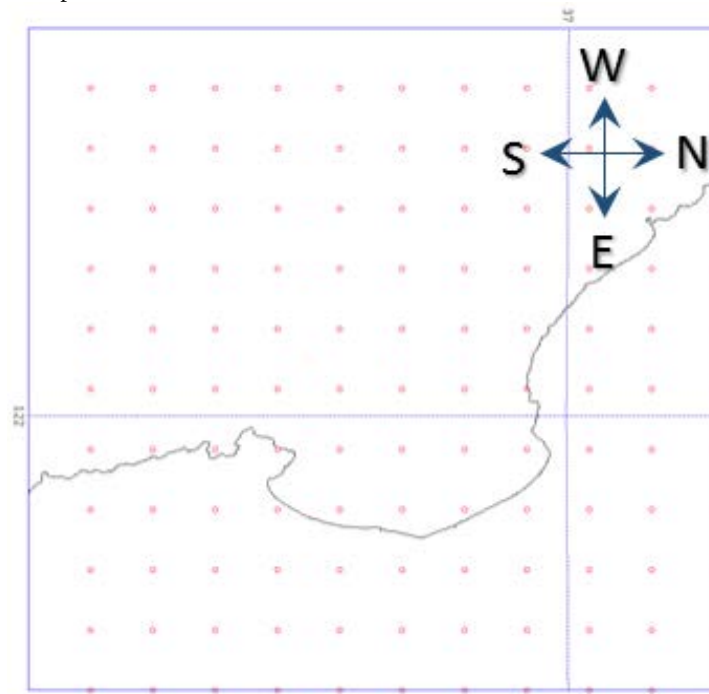


Figure 40. (Continued from previous page)

The difference in forecast accuracy between the land and sea regimes is also very evident. NPS nowcast was much more accurate over land than over sea. This was true for all 5 data sets using the *k*-means cluster scheme. This result was expected because there are more sea points. Also, the sea grid points tend to experience more variation in the low cloud field (i.e., the clouds are constantly building and clearing over the bay, but are only occasionally advected onto land). Figures 41 and 42 show heat maps of MAE and Brier scores calculated at each grid point over data sets 9 and 10. (Brier score is displayed here on a 1.0 to 2.0 scale; simply subtract 1.0 from each value to obtain the true Brier score.) These heat maps reveal the land-sea dichotomy in forecast accuracy. Careful examination of these plots also reveals higher error along the coastline; this is particularly evident for data set 10, during which there were more events of cloud being advected inland.

a) Data set 9 (note that Brier scores are displayed here on a 1.0 to 2.0 scale)

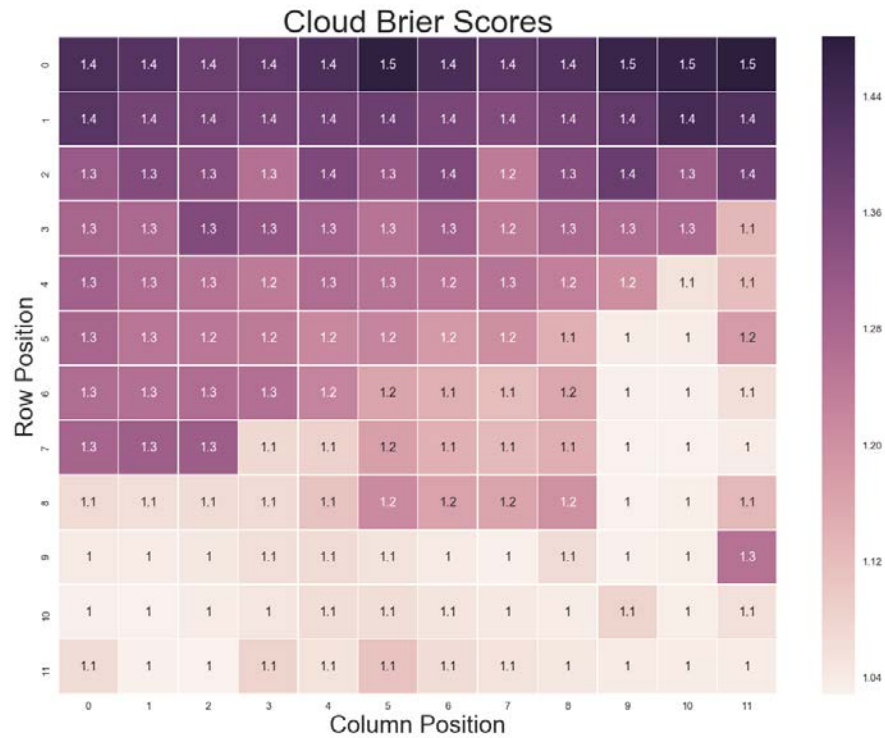
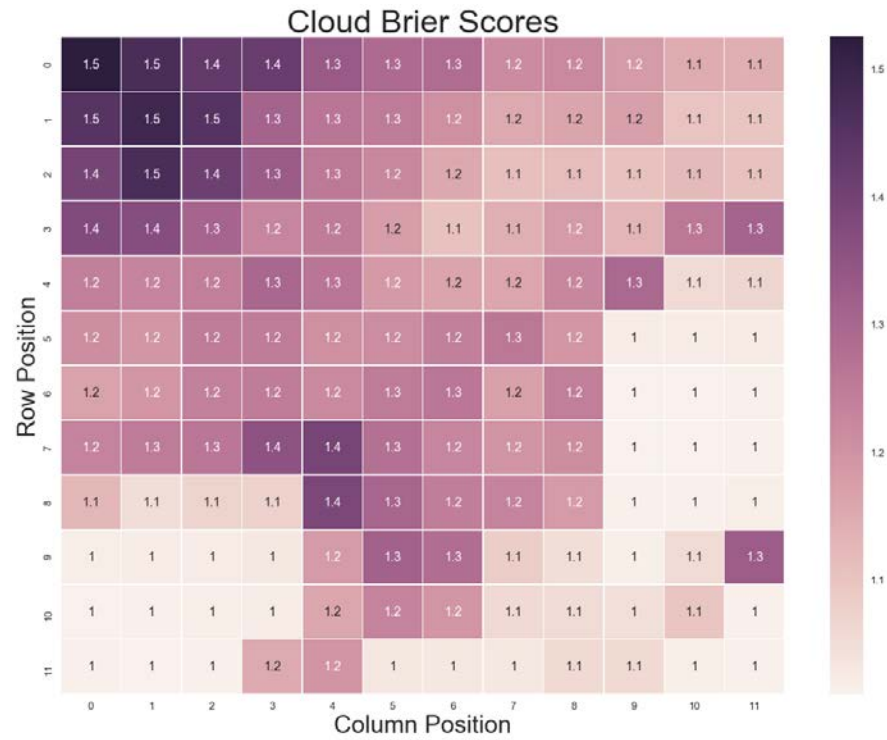


Figure 41. Brier score heat map for a) data set 9 and b) data set 10 with c) reference map included (continued on next page)

b) Data set 10 (note that Brier scores are displayed here on a 1.0 to 2.0 scale)



c) Reference map

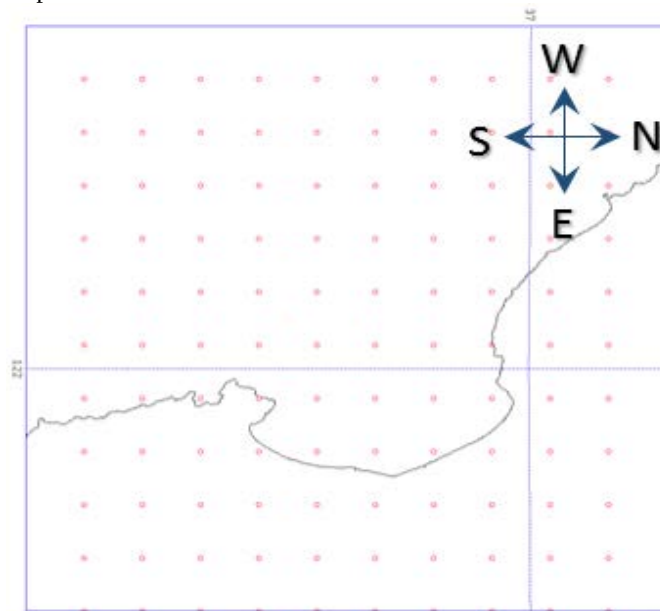
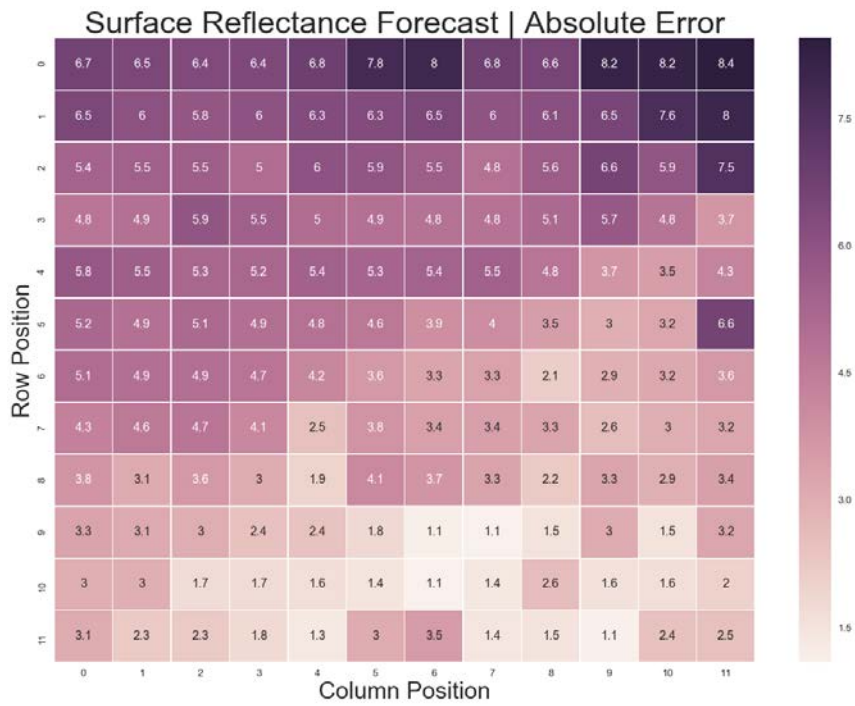


Figure 41. (Continued from previous page)

a) Data set 9



b) Data set 10

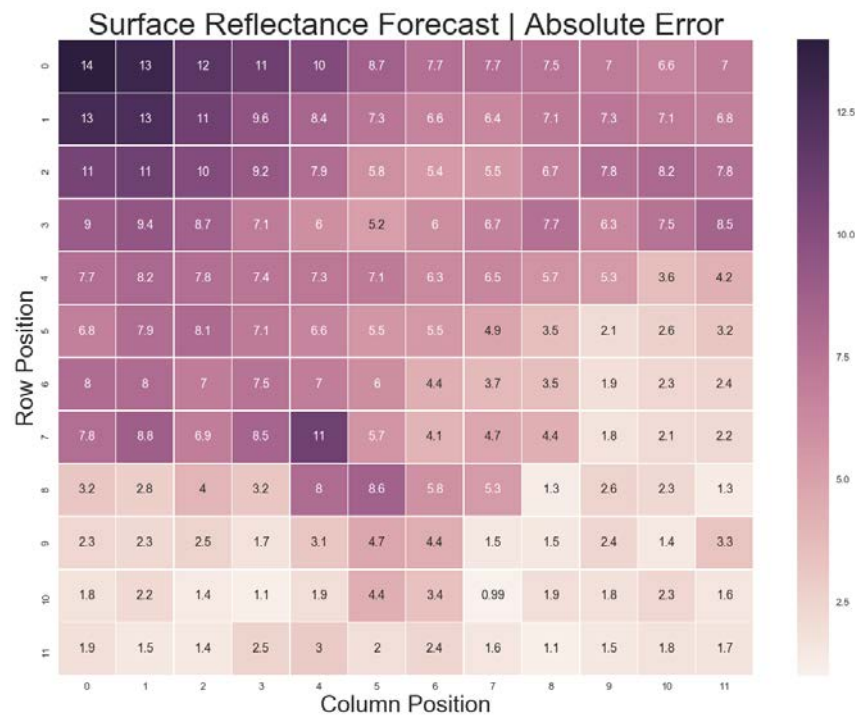
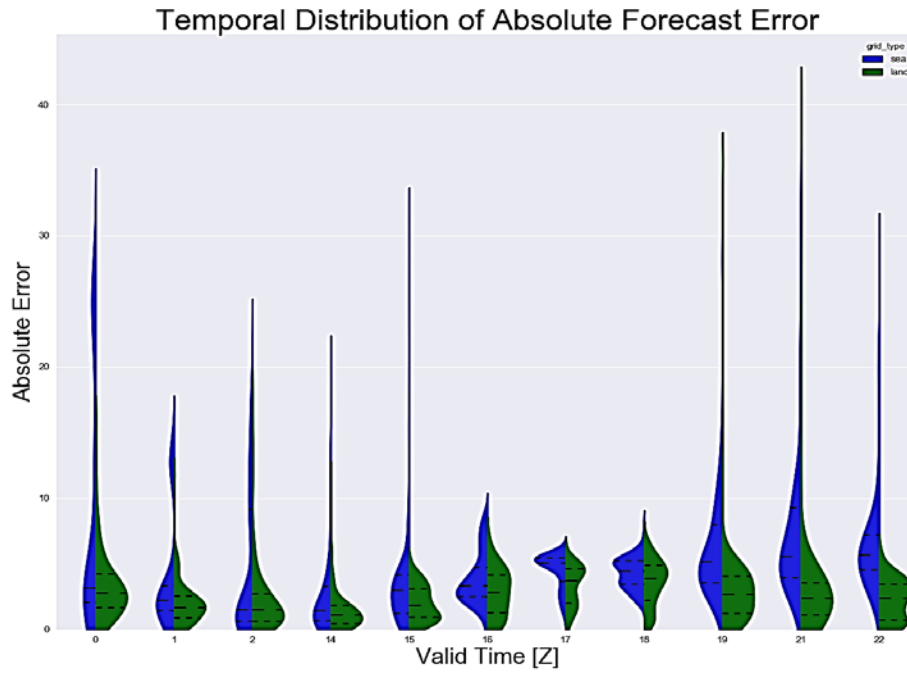


Figure 42. MAE heat map for a) data set 9 and b) data set 10

Analysis of forecast accuracy as a function of time was less conclusive. In each data set, there were certain valid times for which the NPS nowcast performed better or worse; however, these valid times were not consistent across multiple data sets. For data set 9, the midday hours exhibited greater absolute error. For data set 10, the morning hours exhibited greater absolute error. Figure 43 shows violin plots of absolute error for each forecast valid time. The body (i.e., the bulky part) of each “violin” describes the bulk of the data points that fell under each valid time. Therefore, if the body of the violin is contained below the 5.0 absolute error line, then the majority of data points had an absolute error less than 5.0 for that valid time. A desirable error distribution is one in which the majority of the violin is contained close to zero. The data points are also broken into land (green) and sea (blue). For data set 9: 19Z, 21Z and 22Z (1200, 1400 and 1500 local time) exhibit the least desirable distributions where the error is shifted towards higher values, especially over water—note that there were no 20Z forecasts present in data set 9 due to unavailability of 20Z satellite imagery. For data set 10: 14Z, 15Z and 16Z (0700, 0800 and 0900 local time) exhibit the least desirable distributions, again due to the broad distribution that is shifted toward higher values. Because the data set 10 forecast period contained more observed cloud, this may indicate that higher forecast error occurs during morning valid times when cloud is present, and that higher error occurs during midday when cloud is not present. If true, that would indicate that the *k*-means cluster method is not appropriately addressing model bias differences between valid times and should be adjusted. However, further testing is needed to make that conclusion. On average, data set 9(10) only contained 2(3) forecasts for each valid time.

a) Data set 9



b) Data set 10

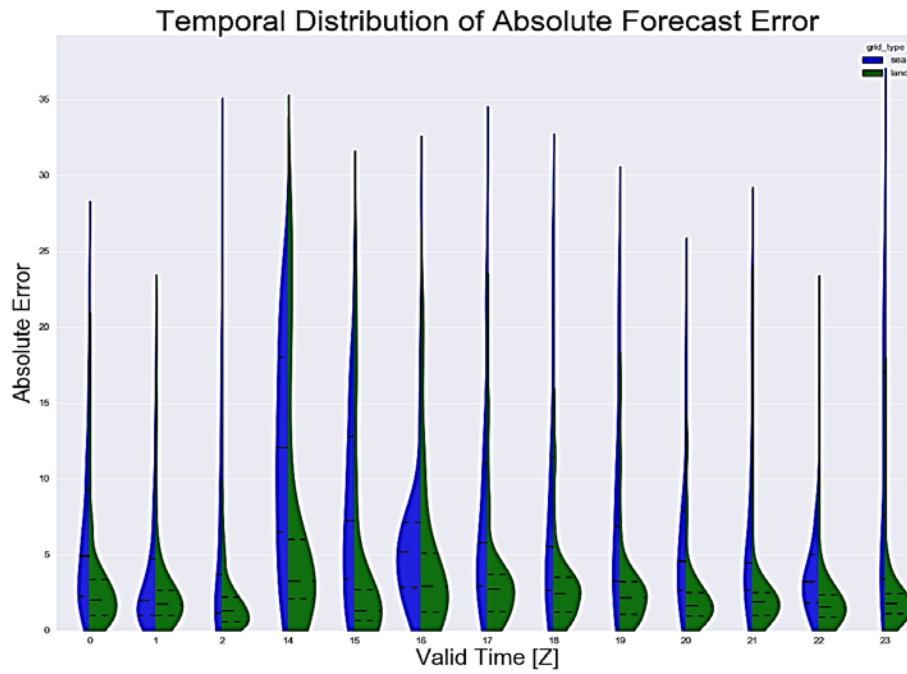


Figure 43. Violin plots of absolute error for each valid time in a) data set 9 and b) data set 10

V. CONCLUSION

A. SUMMARY AND CONCLUSIONS

In summary, NPS nowcast was created by performing Bayesian estimation over a 9 km resolution grid on a training period of observed GOES-15 visible channel imagery and 06-h RAP forecast fields. This analysis produces a generalized linear model (GLM) that inputs current RAP 06-h predictor variable fields to generate a probabilistic 06-h forecast of low cloud reflectance values. This information is used to produce “pseudo-satellite” and “probability of cloud” forecast products. NPS nowcast was tested on a case study of summertime low level stratus within a 100 by 100 km box around the Monterey Bay.

Testing and analysis yielded several significant results. Firstly, *k*-means clustering was demonstrated to be a valuable pre-processing technique. It improves forecast accuracy by grouping the forecast data with like training data so that the most applicable GLM is used for each forecast. As shown in the data set 9 results, it allows the nowcast system to accurately forecast anomalous events that would otherwise throw off the model. Also of note, the particular *k*-means cluster scheme applied in this study grouped data points using the three predictor variable values and the number of hours offset from 2100Z. It is very possible that a different application of *k*-means clustering could produce better results.

In addition, a longer training period was shown to improve accuracy for this particular forecast application, geographic location and time of year. It is likely that accuracy would increase with training period length as long as the period is contained within the same synoptic regime. A nowcast system operating over a different location or time of year would require cross validation testing to determine the ideal training period for that particular time and forecast area.

The computational cost of this nowcast approach is very reasonable. The longest data set took 25.5 minutes to perform the Bayesian processing on. Adding predictor variables and increasing resolution would increase this time; however, the Bayesian

processing step used to update the GLM does not need to be run every hour. It could be run once a day without any negative impacts to forecast results. The amount of time it takes this approach to actually generate the forecast for each hour is negligible.

The NPS nowcast system forecasted the general cloud area well but struggled with detail along the cloud edge as well as with fine structure within the cloud field. This can likely be improved with higher resolution mesoscale model data and by using the full set of 1 km resolution satellite reflectance data. NPS nowcast noticeably outperformed the land-sea climatology forecast that was created as a reference, particularly when a longer training period was used. This shows that the success of NPS nowcast is not simply due to an easy forecast challenge.

Results showed that NPS nowcast performs better over land than over sea. They also indicated slightly increased error along the coastline—the transition area between the land and sea regimes. Although all three predictor variables displayed some “meaningfulness” that varied with cluster, *relative humidity* (RH) and % *low cloud* (LCLD) were the most useful predictor variables across all data sets. Further testing may reveal that LCLD is more useful for longer training periods.

This study has shown that a successful nowcast system can be developed using Bayesian estimation to accomplish machine learning given prior geostationary satellite imagery and 06-h mesoscale model forecast fields. It has also shown the utility of GOES visible channel imagery in forecasting the low cloud field during the daytime. It is very reasonable to assume that these methods will be even more successful in producing 00–05-h forecasts. It is also very likely that forecast results can be improved by using a longer training period and a larger set of predictor variables. Finally, it is reasonable that these methods will also be successful using higher resolution model data and the full 1 km resolution data set, which will improve the level of detail in the cloud field structure and cloud edge.

B. RECOMMENDATIONS FOR FURTHER RESEARCH

The results indicate that additional research and testing of this nowcast application may produce more accurate results and broaden the scope of its use. Further research pertaining directly to this forecast application should investigate:

- Applying the NPS nowcast system to generate 00–05-h forecasts
- Using the full 10,000 grid points (1 km resolution) in the forecast process
- Using the most recent satellite imagery as a predictor variable, or to derive multiple predictive variables such as average reflectance or upstream reflectance
- Longer training periods
- Modifications of the *k*-means clustering scheme to determine the ideal application
- Expanding the predictor variable list to contain forecast fields from other numerical models
- Using other GOES channels and a multivariate scheme to forecast for low, middle, and high clouds

Further research that broadens the scope of this nowcast application should investigate:

- Using real time in-situ observations to “nudge” the analysis and subsequent forecast cloud fields
- Using higher resolution mesoscale model data
- Other forecast locations and seasons
- Cloud field nowcasting under a convective regime
- Related forecast applications such as turbulence and icing
- Additional forecast fields such as pressure, temperature, etc.

In conclusion, this study has laid down a foundation on which further research and testing can be built to eventually produce a highly accurate, high-resolution nowcast system that can be applied using any numerical weather model and satellite imagery combination over any geographical location in the world, no matter how sparse the observation network. This approach has the potential to be adapted to forecast not only low cloud fields, but high clouds, convection, turbulence, icing and a variety of other forecast fields relevant to special operations. Such a nowcast system will provide operators with the most accurate, updated and detailed information needed for mission success. Furthermore, it will reduce cost by preventing unsuccessful sorties and decreasing the number of forecasters needed.

APPENDIX: NRL GOES-15 IMAGERY DIMENSIONS

center_lat : 37 15.00 N
center_lon : 122 30.00 W
pixel_width : 1.0154
pixel_height : 1.0154
total_width : 812.3054
total_height : 811.7591

upper_edge_len : 810.8702
lower_edge_len : 810.4583
left_edge_len : 811.7551
right_edge_len : 811.7551

upper_left_lat : 40 48.79 N
upper_left_lon : 127 18.46 W
upper_right_lat : 40 48.79 N
upper_right_lon : 117 41.54 W
lower_left_lat : 33 30.45 N
lower_left_lon : 126 51.76 W
lower_right_lat : 33 30.45 N
lower_right_lon : 118 8.24 W

mid_left_lat : 37 9.67 N
mid_left_lon : 127 4.46 W
mid_right_lat : 37 9.67 N
mid_right_lon : 117 55.54 W
mid_upper_lat : 40 54.39 N
mid_upper_lon : 122 30.00 W
mid_lower_lat : 33 35.53 N
mid_lower_lon : 122 30.00 W

center_sun : 37.29
upper_left_sun : 41.04
upper_right_sun : 33.74
lower_left_sun : 40.62
lower_right_sun : 33.34
mid_left_sun : 40.94
mid_right_sun : 33.63
mid_lower_sun : 36.98
mid_upper_sun : 37.40

center_sat : 45.04
upper_left_sat : 42.37
upper_right_sat : 39.77

lower_left_sat	:	50.26
lower_right_sat	:	47.20
mid_left_sat	:	46.31
mid_right_sat	:	43.50
mid_lower_sat	:	48.88
mid_upper_sat	:	41.20
sat_sub_lat	:	0 0.00 N
sat_sub_lon	:	135 0.00 W

Data provided by Mr. Kim Richardson, a satellite specialist at Naval Research Laboratory (personal correspondence, Dec. 12, 2016).

LIST OF REFERENCES

- Benjamin et al., 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, 144, 1669–1694.
- Bishop, C. H., and K. T. Shanley, 2008: Bayesian model averaging’s problematic treatment of extreme weather and a paradigm shift that fixes it. *Mon. Wea. Rev.*, 136, 4641–4652, doi:10.1175/2008MWR2565.1.
- Bremnes, J. B., and S. C. Michaelides, 2007: Probabilistic visibility forecasting using neural networks. *Pure Appl. Geophys.*, 164, 1365–1382.
- Chmielecki, R. M., and A. E. Raftery, 2011: Probabilistic visibility forecasting using Bayesian model averaging. *Mon. Wea. Rev.*, 139, 1626–1636, doi:10.1175/2010mwr3516.1.
- English, S., J. Eyre, and J. Smith, 1999: A cloud-detection scheme for use with satellite sounding radiances in the context of data assimilation for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, 125, 2359–2378, doi:10.1002/qj.49712555902.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, 2013: *Bayesian Data Analysis*, Third Edition. Chapman & Hall/CRC Press, 675 pp.
- Hall, T. J., R. N. Thessin, G. J. Bloy, and C. N. Mutchler, 2010: Analog sky condition forecasting based on a k-nn algorithm. *Wea. Forecasting*, 25, 1463–1478, doi:10.1175/2010waf2222372.1.
- Hamill, T. M., 2007: Comments on “Calibrated surface temperature forecasts from the Canadian Ensemble Prediction System using Bayesian model averaging.” *Mon. Wea. Rev.*, 135, 4226–4230, doi:10.1175/2007mwr1963.1.
- Hilliker, J. L., and J. M. Fritsch, 1999: An observations-based statistical system for warm-season hourly probabilistic forecasts of low ceiling at the San Francisco International Airport. *Journal of Applied Meteorology*, 38, 1692–1705, doi:10.1175/1520-0450(1999)038<1692:aobssf>2.0.co;2.
- Hodyss, D., E. Satterfield, J. Mclay, T. M. Hamill, and M. Scheuerer, 2016: Inaccuracies with multimodel postprocessing methods involving weighted, regression-corrected forecasts. *Mon. Wea. Rev.*, 144, 1649–1668, doi:10.1175/mwr-d-15-0204.1.
- Housel, T., J. Mun, D. Ford, S. Hom, and M. Cornachio, 2016: Measuring the return on investment and real option value of weather sensor bundles for Air Force unmanned aerial vehicles. *Thirteenth Annual Acquisition Research Symposium*,

- Monterey, CA, NPS Acquisition Research Program, 35–46. [Available online at <http://www.acquisitionresearch.net/files/FY2016/SYM-AM-16-023.pdf>.]
- Jedlovec, G., S. Haines, and F. Lafontaine, 2008: Spatial and temporal varying thresholds for cloud detection in GOES imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 46, 1705–1717, doi:10.1109/tgrs.2008.916208.
- Kemp, E. M., and R. J. Alliss, 2007: Probabilistic cloud forecasting using logistic regression. *22nd Conference on Weather Analysis and Forecasting*. [Available online at <https://ams.confex.com/ams/pdfpapers/122886.pdf>.]
- Kruschke, J. K., 2010: *Doing Bayesian Data Analysis: A Tutorial Introduction with R and BUGS*. Academic Press Inc, s.l.
- Lockhart, M., 2015: Improving environmental situational awareness for UAVs — PEMDAS. *PEMDAS*. Accessed March 22 2017. [Available online at <https://pemdastechnologies.com/improving-environmental-situational-awareness-for-uav-operations/>.]
- Lorenz, E. N., 1963: Section of planetary sciences: The predictability of hydrodynamic flow*,†. *Transactions of the New York Academy of Sciences*, 25, 409–432, doi:10.1111/j.2164-0947.1963.tb01464.x.
- Marzban, C., S. M. Leyton, and B. Colman, 2007: Ceiling and visibility forecasts via neural networks. *Wea. Forecasting*, 22, 466–479.
- Mass, C., 2012: Nowcasting: the promise of new technologies of communication, modeling, and observation. *Bull. Amer. Met. Soc.*, 93, 797–809, doi:10.1175/bams-d-11-00153.1.
- Mueller, C., T. Saxen, R. Roberts, J. Wilson, T. Betancourt, S. Dettling, N. Oien, and J. Yee, 2003: NCAR Auto-Nowcast System. *Wea. Forecasting*, 18, 545–561, doi:10.1175/1520-0434(2003)018<0545:nas>2.0.co;2.
- NOAA ESRL, 2016: Rapid Refresh (RAP). Accessed March 22 2017. [Available online at <https://rapidrefresh.noaa.gov/>.]
- NOAA Satellite Information System, 2013: GOES imager instrument. Accessed March 1 2017. [Available online at <http://noaasis.noaa.gov/NOAASIS/ml/imager.html>.]
- Nuss, W. A., and D. W. Titley, 1994: Use of multiquadric interpolation for meteorological objective analysis. *Mon. Wea. Rev.*, 122, 1611–1631, doi:10.1175/1520-0493(1994)122<1611:uomifm>2.0.co;2.
- Nuss, W., and S. Drake, 1990: VISUAL meteorological and diagnostic display program, 51 pp. [Unpublished user guide available from Wendell A. Nuss at nuss@nps.edu.]

- Pasero, E., and W. Moniaci, 2004: Artificial neural networks for meteorological nowcast. *CIMSA 2004 - IEEE International Conference on Computational Intelligence for Measurement System and Applications*, Boston, MA, 36–39, doi:10.1109/CIMSA.2004.1397226.
- Pasini, A., V. Pelino, and S. Potesta, 2001: A neural network model for visibility nowcasting from surface observations: Results and sensitivity to physical input variables. *J. Geophys. Res.*, 106, 14 951–14 959.
- PEMDAS, 2015: PEMDAS Arctic UAV Hotwash. [Unpublished presentation available from Wendell A. Nuss at nuss@nps.edu.]
- Robert, C. P. and G. Casella, 2011: A short history of Markov Chain Monte Carlo: Subjective recollections from incomplete data. *Statistical Science*, **26**(1), 102–15.
- Roquelaure, S., and T. Bergot, 2008: A local ensemble prediction system for fog and low clouds: Construction, Bayesian model averaging calibration, and validation. *J. Appl. Meteor. Climatol.*, 47, 3072–3088.
- Roquelaure, S., and T. Bergot, 2009: Contributions from a Local Ensemble Prediction System (LEPS) for improving low cloud forecasts at airports. *Wea. Forecasting*, 24, 39–52.
- Roquelaure, S., R. Tardif, S. Remy, and T. Bergot, 2009: Skill of a ceiling and visibility Local Ensemble Prediction System (LEPS) according to fog-type prediction at Paris-Charles de Gaulle Airport. *Wea. Forecasting*, 24, 1511–1523.
- Schroeder, A. J., D. R. Stauffer, N. L. Seaman, A. Deng, A. M. Gibbs, G. K. Hunter, and G. S. Young, 2006: An automated high-resolution, rapidly relocatable meteorological nowcasting and prediction system. *Mon. Wea. Rev.* 134, 1237–1265, doi:10.1175/mwr3118.1.
- Vislocky, R. L., and J. M. Fritsch, 1997: An automated, observations-based system for short-term prediction of ceiling and visibility. *Wea. Forecasting*, 12, 31–43, doi:10.1175/1520-0434(1997)012<0031:aaobsf>2.0.co;2.
- Uddstrom, M. J., W. R. Gray, R. Murphy, N. A. Oien, and T. Murray, 1999: A Bayesian cloud mask for sea surface temperature retrieval. *J. Atmos. Oceanic Technol.*, 16, 117–132, doi:10.1175/1520-0426(1999)016<0117:abcmfs>2.0.co;2.
- Warner, M., n.d.: Stats & linux sandbox. *Stats & Linux Sandbox*. Accessed Feb 01 2017. [Available online at <http://mickstat.blogspot.com/>.]
- Wendt, R.D.T., 2017: Statistical post-processing of ensemble forecasts with Bayesian estimation and Markov Chain Monte Carlo sampling methods, Ph.D. dissertation proposal, Naval Postgraduate School, 54 pp. [Available from LCDR Robert Wendt (rdwendt@nps.edu).]

Wilks, D. S., 2006: Comparison of ensemble-MOS methods in the Lorenz '96 setting. *Meteorological Applications*, 13, 243, doi:10.1017/s1350482706002192.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
Ft. Belvoir, Virginia
2. Dudley Knox Library
Naval Postgraduate School
Monterey, California