

Sky Cover

By

Jordan J. Gerth

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Atmospheric and Oceanic Sciences)

at the

UNIVERSITY OF WISCONSIN-MADISON

2013

Date of final oral examination: 12/18/13

The dissertation is approved by the following members of the Final Oral Committee:

Steven Ackerman, Professor, Atmospheric and Oceanic Sciences

Andrew Heidinger, Adjunct Assistant Professor, Atmospheric and Oceanic Sciences

Tristan L'Ecuyer, Assistant Professor, Atmospheric and Oceanic Sciences

Jonathan Martin, Professor, Atmospheric and Oceanic Sciences

Gregory Tripoli, Professor, Atmospheric and Oceanic Sciences

Sky Cover

Jordan J. Gerth¹

Department of Atmospheric and Oceanic Sciences
Cooperative Institute for Meteorological Satellite Studies
University of Wisconsin, Madison, Wisconsin

¹ *Corresponding author address:* Jordan Gerth, Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin, 1225 W. Dayton St., Madison, WI 53706.
E-mail: jordang@ssec.wisc.edu

“Time works changes, brings into existence new conditions and purposes. Therefore a principle, to be vital, must be capable of wider application than the mischief which gave it birth.”

Joseph McKenna, Associate Justice of the Supreme Court of the United States

Weems vs. United States, 217 U.S. 349 (1910)

Abstract

Of all of the standard meteorological parameters collected and observed daily, sky cover is not only one of the most complex, but the one that is fairly ambiguously defined and difficult to quantify. Despite that, the implications of how cloud fraction and sky cover are understood not only impact daily weather forecasts, but also present challenges to assessing the state of the earth’s climate system. Part of the reason for this is the lack of observational methods for verifying the skill of clouds represented and parameterized in numerical models.

While human observers record sky cover as part of routine duties, the spatial coverage of such observations in the United States is relatively sparse. There is greater spatial coverage of automated observations, and essentially complete coverage from geostationary weather satellites that observe the Americas. A good analysis of sky cover reconciles differences between manual observations, automated observations, and satellite observations, through an algorithm that accounts for the strengths and weaknesses of each dataset. This work describes the decision structure for trusting and weighting these similar observations. Some of the issues addressed include: human and instrument error resulting from approximations and estimations, a deficiency in high cloud detectability using surface-based ceilometers, poorly resolved low cloud

using infrared channels on space-based radiometers during overnight hours, and decreased confidence in satellite-detected cloud during stray light periods.

Using the blended sky cover analysis as the best representation of cloudiness, it is possible to compare the analysis to numerical model fields in order to assess the performance of the model and the parameterizations therein, as well as confirm or uncover additional relationships between sky cover and pertinent fields using an optimization methodology. The optimizer minimizes an affine expression of adjusted fields to the “truth” sky cover analysis. Results include discussion about how the blended sky cover analysis correlates with the cloud ice, cloud water, rain, snow, and other analysis fields from the High-Resolution Rapid Refresh (HRRR). The intent is to suggest a reasonable operational and scientific definition for sky cover and demonstrate an observational method that can bring consistency to analyses and forecasts of sky cover.

Executive Summary

There is the lack of an observational method through which to verify the behavior of cloud parameterizations in climate and weather models, which are useful in examining cloud feedbacks. The proposed solution to this problem is a sky cover approach. Sky cover is an hour-averaged quantity of cloud coverage within the celestial dome. The celestial dome is the amount of sky, from horizon to horizon, visible to a human or instrument observer at a single point.

Pursuant to the problem, this project achieves the following:

- It devises a blended sky cover analysis that incorporates the advantageous properties of surface observations of sky cover and geostationary satellite cloud products.
- It defines a framework for optimizing the blended analysis based on the current near-term human-produced forecasts from the National Digital Forecast Database (NDFD).
- It constructs an affine expression of High-Resolution Rapid Refresh (HRRR) relative humidity, mixing ratio, and vorticity analysis fields with adjustable coefficients and scalars that is optimized to decrease the absolute error compared to the “truth” analysis.

The primary conclusions are:

- The combination of surface observations and the satellite sky cover product improves the detection of nocturnal low cloud and general high cloud compared to a single source.
- Relative humidity and cloud water mixing ratio are closely correlated with sky cover, particularly in the lowest levels of the troposphere.
- The linear optimization approach produces an optimal sky cover product with decreased mean error, mean absolute error, and root-mean-square error when validated against the NDFD one-hour forecast, compared to the current operational HRRR output.

Acknowledgements

Dr. Steven Ackerman, a professor of atmospheric and oceanic sciences at the University at Wisconsin, provided meaningful guidance and oversight during the course of this research. The worth of his advice and interest in a thorough scientific process is immeasurable. The University of Wisconsin's Space Science and Engineering Center (SSEC) is a valuable location to perform work of this kind due to its academic wealth, integral science staff, and superior computing resources.

I would like to thank the National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA), the National Environmental Satellite, Data, and Information Service (NESDIS), the Geostationary Operational Environmental Satellite R-Series (GOES-R) Program Office, the Joint Polar Satellite System (JPSS) Program Office, the National Weather Service (NWS), and the SSEC for the ability and support to conduct this work, as well as the researchers at the Cooperative Institute for Meteorological Satellite Studies (CIMSS) for their unending support in guiding my academic and professional pursuits. Specifically, thanks are extended to Robert Aune, James Nelson III, Timothy Schmit, and Anthony Schreiner for their active role in assisting with the development of this research topic. Finally, the faculty members in the Department of Atmospheric and Oceanic Sciences at the University of Wisconsin deserve recognition for building a world-class program for undergraduate and graduate meteorological studies. Committee members Dr. Andrew Heidinger, Dr. Tristan L'Ecuyer, Dr. Jonathan Martin, and Dr. Gregory Tripoli are thanked for their careful and honest review of this dissertation.

Table of Contents

Abstract	ii
Executive Summary	iv
Acknowledgements.....	v
List of Tables	viii
List of Figures	ix
1. Introduction.....	1
a. Early history	2
b. Challenges linger today.....	2
c. Goals	4
d. Definitions.....	4
e. Types of observations and analyses	5
Surface observations	5
Satellite observations	7
Real-Time Mesoscale Analysis.....	8
National Digital Forecast Database	8
Mean value.....	9
2. Relationships.....	9
a. From literature.....	10
b. Approximations.....	12
3. Creating a blended analysis	14
a. Gridding surface observations.....	16
b. Basic methodology	17
c. Low cloud.....	18
d. Blending logic	18
Stray light tests.....	19
Spatial continuity	20
e. Comparisons.....	21
f. Cases	21
g. Efficacy of blended analysis	24
4. Optimization methodology	26
a. Preparation model	26
b. Primary model.....	28
Data variables.....	33
Model variables.....	33
c. Secondary model.....	38

5. Cases	43
a. 18 UTC on 4 October 2013	43
b. 3 UTC on 19 October 2013	47
c. 12 UTC on 5 November 2013	50
6. Results	54
a. Relative humidity	56
b. Cloud water mixing ratio	57
c. Cloud ice mixing ratio	58
d. Rain water mixing ratio	58
e. Snow mixing ratio	59
f. 200 hPa	59
g. 300 hPa	60
h. 500 hPa	60
i. 700 hPa	61
j. 800 hPa through 950 hPa	61
k. 1000 hPa	62
l. Overall performance	63
m. Shear and curvature vorticity	65
7. Conclusion	66
Bibliography	69
Tables	74
Figures	80
Appendix A	139
List of Acronyms	140

List of Tables

Table 1. Summary of logic for comparing a grid point containing a satellite observation of sky cover to a like in-situ surface observation with a collocated satellite observation	74
Table 2. Frequency of runs between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013 with a relative humidity threshold adjustment to the primary model output	74
Table 3. Mean and frequency of non-zero relative humidity coefficients produced from optimizer runs between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	75
Table 4. Frequency of non-zero coefficients corresponding to relative humidity at a level, produced from optimizer runs between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	75
Table 5. Mean and frequency of non-zero cloud water mixing ratio coefficients and scalars produced from optimizer runs between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	76
Table 6. Frequency of non-zero coefficient and/or scalar corresponding to cloud water mixing ratio at a level, produced from optimizer runs between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	76
Table 7. Mean and frequency of non-zero cloud ice mixing ratio coefficients and scalars produced from optimizer runs between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	77
Table 8. Frequency of non-zero coefficient and/or scalar corresponding to cloud ice mixing ratio at a level, produced from optimizer runs between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	77
Table 9. Mean and frequency of non-zero rain water mixing ratio coefficients and scalars produced from optimizer runs between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	78
Table 10. Frequency of non-zero coefficient and/or scalar corresponding to rain water mixing ratio at a level, produced from optimizer runs between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	78
Table 11. Mean and frequency of non-zero snow mixing ratio coefficients and scalars produced from optimizer runs between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	79
Table 12. Frequency of non-zero coefficient and/or scalar corresponding to snow mixing ratio at a level, produced from optimizer runs between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	79

List of Figures

Figure 1. GOES-East infrared window image valid at 11:02 UTC on 20 October 2013	80
Figure 2. Composite geostationary satellite effective cloud amount analysis at 11:00 UTC on 20 October 2013.....	81
Figure 3. Composite geostationary satellite cloud top pressure analysis at 11:00 UTC on 20 October 2013.....	82
Figure 4. Composite geostationary satellite celestial dome effective cloud amount analysis at 11:00 UTC on 20 October 2013.....	83
Figure 5. Composite geostationary satellite sky cover product for hour beginning at 11:00 UTC on 20 October 2013.....	84
Figure 6. Sky cover reports from automated and manned surface observing stations for hour beginning at 11:00 UTC on 20 October 2013.....	85
Figure 7. Blended sky cover analysis for hour beginning at 11:00 UTC on 20 October 2013.....	86
Figure 8. Four-panel plot containing histograms of the celestial dome effective cloud amount, sky cover product, surface observations, and blended sky cover analysis valid at or within the one hour after 11:00 UTC on 20 October 2013	87
Figure 9. Four-panel plot comparing the sky cover product to the blended sky cover analysis for hour beginning at 11:00 UTC on 20 October 2013	88
Figure 10. Geographical comparison of the sky cover product and blended sky cover analysis for hour beginning at 11:00 UTC on 20 October 2013	89
Figure 11. Optimal sky cover analysis for hour beginning at 11:00 UTC on 20 October 2013 ..	90
Figure 12. Two-panel plot comparing the optimal sky cover analysis to the blended sky cover analysis for hour beginning at 11:00 UTC on 20 October 2013.....	91
Figure 13. Geographical comparison of the optimal sky cover analysis to the blended sky cover analysis for hour beginning at 11:00 UTC on 20 October 2013.....	92
Figure 14. GOES-East visible image valid at 18:25 UTC on 4 October 2013	93
Figure 15. Blended sky cover analysis for hour beginning at 18:00 UTC on 4 October 2013.....	94
Figure 16. NDFD total cloud cover one-hour forecast for hour beginning at 18:00 UTC on 4 October 2013.....	95

Figure 17. HRRR total cloud cover analysis valid at 18:00 UTC on 4 October 2013.....96

Figure 18. Initial-hour HRRR optimal sky cover product for hour beginning at 18:00 UTC on 4 October 2013.....97

Figure 19. Four-panel plot containing histograms of the blended sky cover analysis, NDFD total cloud cover one-hour forecast, HRRR total cloud cover analysis, and initial-hour HRRR optimal sky cover product valid at or within the one hour after 18:00 UTC on 4 October 201398

Figure 20. Four-panel plot containing difference histograms comparing the HRRR total cloud cover analysis and initial-hour HRRR optimal sky cover product to the blended sky cover analysis and NDFD total cloud cover one-hour forecast for hour beginning at 18:00 UTC on 4 October 2013.....99

Figure 21. Geographical comparison of the initial-hour HRRR optimal sky cover product to the blended sky cover analysis for hour beginning at 18:00 UTC on 4 October 2013.....100

Figure 22. Geographical comparison of the initial-hour HRRR optimal sky cover product to the NDFD total cloud cover one-hour forecast for hour beginning at 18:00 UTC on 4 October 2013101

Figure 23. GOES-East infrared window image valid at 3:00 UTC on 19 October 2013102

Figure 24. Blended sky cover analysis for hour beginning at 3:00 UTC on 19 October 2013...103

Figure 25. NDFD total cloud cover one-hour forecast for hour beginning at 3:00 UTC on 19 October 2013.....104

Figure 26. HRRR total cloud cover analysis valid at 3:00 UTC on 19 October 2013.....105

Figure 27. Initial-hour HRRR optimal sky cover product for hour beginning at 3:00 UTC on 19 October 2013.....106

Figure 28. Four-panel plot containing histograms of the blended sky cover analysis, NDFD total cloud cover one-hour forecast, HRRR total cloud cover analysis, and initial-hour HRRR optimal sky cover product valid at or within the one hour after 3:00 UTC on 19 October 2013107

Figure 29. Four-panel plot containing difference histograms comparing the HRRR total cloud cover analysis and initial-hour HRRR optimal sky cover product to the blended sky cover analysis and NDFD total cloud cover one-hour forecast for hour beginning at 3:00 UTC on 19 October 2013.....108

Figure 30. Geographical comparison of the initial-hour HRRR optimal sky cover product to the blended sky cover analysis for hour beginning at 3:00 UTC on 19 October 2013.....109

Figure 31. Geographical comparison of the initial-hour HRRR optimal sky cover product to the NDFD total cloud cover one-hour forecast for hour beginning at 3:00 UTC on 19 October 2013	110
Figure 32. GOES-East infrared window image valid at 12:15 UTC on 5 November 2013	111
Figure 33. Blended sky cover analysis for hour beginning at 12:00 UTC on 5 November 2013	112
Figure 34. NDFD total cloud cover one-hour forecast for hour beginning at 12:00 UTC on 5 November 2013	113
Figure 35. HRRR total cloud cover analysis valid at 12:00 UTC on 5 November 2013	114
Figure 36. Initial-hour HRRR optimal sky cover product for hour beginning at 12:00 UTC on 5 November 2013	115
Figure 37. Four-panel plot containing histograms of the blended sky cover analysis, NDFD total cloud cover one-hour forecast, HRRR total cloud cover analysis, and initial-hour HRRR optimal sky cover product valid at or within the one hour after 18:00 UTC on 5 November 2013	116
Figure 38. Four-panel plot containing difference histograms comparing the HRRR total cloud cover analysis and initial-hour HRRR optimal sky cover product to the blended sky cover analysis and NDFD total cloud cover one-hour forecast for hour beginning at 18:00 UTC on 5 November 2013	117
Figure 39. Geographical comparison of the initial-hour HRRR optimal sky cover product to the blended sky cover analysis for hour beginning at 18:00 UTC on 5 November 2013	118
Figure 40. Geographical comparison of the initial-hour HRRR optimal sky cover product to the NDFD total cloud cover one-hour forecast for hour beginning at 18:00 UTC on 5 November 2013	119
Figure 41. Trend in mean of non-zero coefficients applied to 200 hPa quantities between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	120
Figure 42. Trend in percentage of non-zero coefficients applied to 200 hPa quantities between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	121
Figure 43. Trend in mean of non-zero coefficients applied to 300 hPa quantities between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	122
Figure 44. Trend in percentage of non-zero coefficients applied to 300 hPa quantities between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	123

Figure 45. Trend in mean of non-zero coefficients applied to 500 hPa quantities between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	124
Figure 46. Trend in percentage of non-zero coefficients applied to 500 hPa quantities between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	125
Figure 47. Trend in mean of non-zero coefficients applied to 700 hPa quantities between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	126
Figure 48. Trend in percentage of non-zero coefficients applied to 700 hPa quantities between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	127
Figure 49. Trend in mean of non-zero coefficients applied to 800 hPa quantities between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	128
Figure 50. Trend in percentage of non-zero coefficients applied to 800 hPa quantities between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	129
Figure 51. Trend in mean of non-zero coefficients applied to 850 hPa quantities between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	130
Figure 52. Trend in percentage of non-zero coefficients applied to 850 hPa quantities between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	131
Figure 53. Trend in mean of non-zero coefficients applied to 900 hPa quantities between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	132
Figure 54. Trend in percentage of non-zero coefficients applied to 900 hPa quantities between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	133
Figure 55. Trend in mean of non-zero coefficients applied to 950 hPa quantities between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	134
Figure 56. Trend in percentage of non-zero coefficients applied to 950 hPa quantities between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	135
Figure 57. Trend in mean of non-zero coefficients applied to 1000 hPa quantities between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	136
Figure 58. Trend in percentage of non-zero coefficients applied to 1000 hPa quantities between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	137
Figure 59. Trend in mean 1000 hPa relative humidity threshold between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013	138

1. Introduction

Clouds are a fundamental part of the earth system. Their impacts on the earth are far-reaching, with both large and small implications for humans. Clouds, cloud evolution, and cloud feedbacks are correlated with radiation, latent heating, temperature, moisture, precipitation processes and efficiency, and aerosols. They are parameterized in climate and short-term weather prediction models. Cloudiness impacts global industries, such as air travel and the renewable energy sector. And it even has an impact on the outcome of professional baseball games (Kent and Sheridan 2011).

Despite the implications on everything from a daily local weather forecast to a global climate prediction, and multiple observing platforms capable of sensing or detecting cloud, including geostationary satellites, polar-orbiting satellites, and a surface-based observation network of humans and automated instrumentation, the analysis and forecast of cloud fraction and other cloud-related quantities remains a quandary in the atmospheric sciences (Young 1967; Wylie and Menzel 1989). The National Weather Service (NWS) has had a long-standing requirement to forecast sky cover as part of the weather information that is supplied to the public routinely. However, meteorologists have long complained about the lack of good analyses and forecasts from the numerical weather prediction (NWP) models for cloud fraction and sky cover, with Gerth (2011) demonstrating significant error in cloud cover forecasts from experimental regional numerical model runs adequately initialized, even within 12 hours. The problem is compounded as a result of an unclear and inconsistent operational and scientific definition for sky cover. As a result, consistent observations, forecasts, and scientific understanding of sky cover remain elusive despite progress elsewhere in the field.

a. Early history

Cloud cover was a topic of interest after the first satellites launched, as evinced by work from Arking in 1964. Malberg (1973) was among the first to compare satellite cloud observations with those from the surface. Malberg used the Environmental Science Services Administration (ESSA) 8 satellite to compare imagery of the Europe and North Atlantic area between December 1967 and November 1970. The United States Air Force and National Environmental Satellite Service produced a global atlas of relative cloud cover using data collected from satellites between 1967 and 1970.

Early studies of cloud cover from satellites faced a number of challenges. First, techniques had not been developed to discriminate between cloud and surface features with a high albedo, such as sand and snow. Second, the satellite imagery of the day offered less spatial and temporal resolution than is available today. Combined with better geolocation capabilities afforded from satellite stabilization, as well as multiple spectral bands on modern radiometers, the current state of the science in this area has improved over the preceding several decades, though issues remain.

b. Challenges linger today

Sky cover presents a different formulation for considering cloudiness compared to traditional definitions. The greatest strength in using sky cover is that it is an observable quantity and much more characteristic of the surrounding atmosphere for validation purposes than a single cloud observation, or observation at a fixed point. However, this is different than the point-based output of cloud fraction from current NWP methods. Thus, the science is well served in assessing how weather and climate models handle clouds and their coverage. Due to

the complex nature of cloud development and decay, numerical models provide one of the only ways to understand cloud processes and how they impact the rest of the atmosphere. However, global circulation and NWP models are weakened through the use of cloud parameterizations. Cloud parameterizations are built on assumptions and empirical formulations that are difficult to evaluate (Stephens 2005). The science currently lacks observational methods for assessing the performance of cloud parameterizations.

Accordingly, the lack of a single analysis to represent observed cloudiness is the motivating factor behind this work. The limited knowledge behind what quantities and how the distribution of those quantities in the atmosphere contribute to cloudiness is the motivating factor for the methodology. This work will demonstrate cases in which single-source observation platforms fail to make the proper cloud detections. Furthermore, as a consequence of the limited number of spectral bands on the current geostationary satellite imagers, and radiative properties of certain types of clouds, current satellite-based detection methods for clouds can sometimes fail to adequately discriminate cloud from the surface.

It is hypothesized that a reasonable diagnostic formulation of cloud cover, as aforementioned, can be calculated from prognostic cloud and moisture variables, and potentially other atmospheric variables which are related to cloudiness, and that such a formulation can consistently perform better than traditional cloud schemes used in operational weather models today. Cloud schemes in such models fit into one of three different categories. The older, simplest schemes are diagnostic schemes, where cloud quantities are diagnosed from other model variables during model execution or during post-processing. In comparison, prognostic schemes are those where cloud cover, water vapor, and condensate variables are interconnected, dependent, and advanced/calculated during model execution. Finally, hybrid schemes are those

where condensate variables are prognostic, but cloud cover is diagnostic. The Global Forecast System (GFS) and Weather Research and Forecast (WRF) models employ hybrid schemes.

While relative humidity schemes – those diagnostic schemes that relate cloud cover to relative humidity – are described subsequently, there is also significant use of statistical approaches, where a probability density function (PDF) is specified for the total water within a numerical model grid cell. Most of the PDFs are unimodal, though the shape and skewness of the PDFs vary by approach (Tompkins 2005). Regardless of the approach and scheme, the common assumption is that the numerical model allows cloud in a subsaturated grid cells, whereas, in nature, cloud forms in a supersaturated environment. As such, there is either a critical relative humidity threshold or assumption about the subgrid-scale temperature and/or humidity behavior that is a central component of the approach/scheme in all implementations.

c. Goals

There is the lack of an observational method through which to verify the behavior of cloud parameterizations in climate and weather models, which are useful in examining cloud feedbacks. There are two goals to solving this problem. The first goal is to produce a sky cover analysis that is representative of current conditions and suitable for use as validation. Thus, the second goal is to determine the existence of and quantify the relationship between sky cover as purported by the analysis and related atmospheric quantities in a cloud-resolving NWP model.

d. Definitions

There are two definitions of sky cover. The Federal Meteorological Handbook No. 1 defines sky cover as “the amount of the celestial dome hidden by clouds and/or obscurations” (“Federal Meteorological Handbook No. 1, Surface Weather Observations and Reports” 2005).

The NWS defines sky cover as the “amount of opaque clouds ... (in percent) covering the sky” over a one-hour period (“National Weather Service Instruction 10-201, National Digital Forecast Database and Local Database Description and Specifications” 2012). Sky cover is similar to cloud cover, which is traditionally defined as the fraction of entire sky visible to an observer standing at a point. The entire sky bounded by the horizon is known as the celestial dome. For the purposes of this research, sky cover is a time-average of cloud cover (within the celestial dome) over a one-hour window.

In contrast, cloud fraction is defined as the amount of cloud covering a sampled area. In the case of a satellite sensor, the cloud fraction for a given pixel is the fraction of cloudiness within the pixel or amongst a group of adjacent pixels. Since the satellite is unable to resolve sub-pixel cloudiness, the term effective cloud amount (or effective cloud emissivity, in some circles) is employed to describe the product of the emissivity of the cloud and the cloud fraction. Unfortunately, such quantities are not routinely output from NWP models.

e. Types of observations and analyses

A central component to the subsequent verification and improvement of NWP model output is developing an adequate analysis of sky cover. Developing such an analysis requires the use of both in-situ and remote observations. For the purposes of such an analysis, an hour or less is the preferred frequency of observation. Two main sources of sky cover observations meet the requisite criteria: surface stations and geostationary satellite radiometers.

SURFACE OBSERVATIONS

In the United States, human weather observers at surface observing stations routinely make cloud observations over the celestial dome. Most surface stations are near or collocated at

airports or military facilities. At sites with a staffed human observer, the observer typically has a ceilometer available to help assess the ceiling. In some instances, there is evidence that the human observer trusts the observing equipment without modification. At some observing locations, human observers are not available at all hours because there is no need for a “manned” observation once commercial air traffic has ceased for the evening. For stations with observers overnight, there are additional challenges with overnight cloud observations, particularly high cloud, due to the lack of natural light. This is a bias difficult to correct because diurnal processes in the atmosphere can lead to changes in cloud cover between day and night. There are also a significant number of stations automated to produce an observation in the absence of a human observer.

Unfortunately, examples of poor observations are readily available. A comparison of surface observations to satellite imagery, as shown in Schreiner et al. (1993), reveals that the most common error of cloud cover occurs with areas of high cloud, which automated observing systems do not report, and inattentive human observers occasionally fail to correct. Human observations are also imprecise. In addition, observations are only required once per hour, and cloud reporting generally consists of one of five classifications. Meteorological software packages, such as the GEneral Meteorology PAcKage (GEMPAK), converts the classification code to a sky cover amount, in units of percent, where clear is 0% sky cover, a few clouds is 25% sky cover, scattered cloudiness is 40% sky cover, broken cloudiness is 75% sky cover, and overcast or obscured is 100% sky cover (DesJardins et al. 1991). As such, surface observations often sky cover lack precision. In addition, short-term changes in cloud cover can be missed.

Surface observations are most reliable when reporting low, overcast cloud. The ceilometers used at surface sites maintain a fixed position. Consequently, the reported cloud

amount is a result of a 30-minute temporal average of cloud, double-weighted within the ten minutes immediately prior to the observation time (“Sky Condition Algorithm for Vaisala Ceilometer’s” 2010; “Sky Condition” 2013). Currently, ceilometers are programmed to only report cloud coverage and heights detected within 12 kft (3700 m) above the instrument (“Sky Condition” 2013), which is approximately 600 hPa. A skilled human observer has the option to augment the ceilometer height report in the final statement.

SATELLITE OBSERVATIONS

A complementary source of observations is geostationary satellite imagery. The United States operates a constellation of two geostationary satellites, which collectively provide coverage of the country with a scan approximately every 15 minutes, occasionally more during periods of severe weather. Geostationary weather satellite imager scan strategies enable the detection of changes in cloud cover with a greater temporal frequency than surface observations, and are not subject to human error or subjectivity. Such imagers also are able to collect observations over a large spatial area. Since a weather satellite observes the atmosphere from above, a number of issues arise with cloud detection from satellite and comparing those remote observations to surface observations (Malberg 1973; Schreiner et al. 1993). First, satellites do not resolve discrete clouds smaller than the scanning resolution of the sensor. In the case of the current Geostationary Operational Environmental Satellite (GOES), the resolution, at nadir, of the infrared channels is 4 km. Issues also arise when thin high cloud is above low cloud, and when low cloud exhibits similar properties to the surface temperature and emissivity it overlays.

The easiest method for distinguishing a cloud in the terrestrial infrared window from the surrounding atmosphere is to locate a spatial contrast in either the emissivity or the emitting

temperature. Fortunately, a general decrease of the temperature of the troposphere with height makes determining middle and upper tropospheric cloud relatively easy, particularly in the mid-latitudes. Cloud existing near the surface or within or beneath an inversion, with a warm, terrestrial temperature, is more difficult to detect; also challenging are ice clouds over snow and ice fields.

REAL-TIME MESOSCALE ANALYSIS

The current cloud fraction analysis in the operational Real-Time Mesoscale Analysis (RTMA) comes from a GOES sounder cloud composite available once per hour (De Pondeva et al. 2011). Information from the GOES sounder also helps to produce the operational satellite cloud product (Schreiner, Schmit, and Aune 2002), which is widely used today to report current sky conditions over the National Oceanic and Atmospheric Administration (NOAA) weather radio. Since the next generation of GOES satellites, the GOES-R series (Schmit et al. 2005), is not expected to have a sounder (Schmit et al. 2008), a new analysis based on the imager bands is required. In addition, the operational satellite cloud product is site-based; it is not gridded.

NATIONAL DIGITAL FORECAST DATABASE

The NWS has an operational requirement to forecast sky cover. This is done through the National Digital Forecast Database (NDFD). The NDFD is a gridded forecast for most portions of the United States. Though the input for the NDFD comes from NWP models, operational meteorologists within the NWS routinely monitor the quality of the current conditions and forecast, and make adjustments as necessary (Glahn and Ruth 2003). Unfortunately, the NDFD struggles from accuracy issues and potentially from the forecast process that results from offices

insufficiently collaborating and failing to produce a consistent forecast across political boundaries. Sharp non-meteorological boundaries are evident. When compared to the clear areas identified with GOES, the problem is even more substantial. A number of forecasters overpredict cloud in clear or nearly clear areas as assessed from satellite imagery.

MEAN VALUE

Depending on the observation type or related sky cover analysis or product, the mean value differs substantially, typically 15% or more. A portion of the difficulty in establishing a sky cover parameter of record is having a benchmark to compare it against. In this case, an exact validating analysis that matches in time and space does not exist.

One complication resulting from determining a mean value for surface observation sky cover reports is that surface stations are not distributed uniformly across the domain, nor is their reporting interval always standard. Consequently, cloudy conditions across one portion of the grid with a high number of surface observations could bias the mean away from a large area of clear skies with relatively few surface observations. In many situations, however, the mean value of the surface observations is significantly less than the satellite sky cover mean. This results from the surface observation network's limited detectability of high cloud, and possibly the correspondence between cloud classifications and sky cover used in this study.

2. Relationships

Beyond the different standalone analyses of sky cover based on single sources, NWP models also output an analysis and forecast of cloud cover. However, cloud cover output from a NWP model is particularly unique because of its relationship to one or more prognostic model

variables, as well as certain approximations. Previous studies have investigated relationships between cloud cover and other atmospheric quantities.

a. From literature

Accurately representing cloudiness in output from NWP models is an ongoing issue that originates from the early days of NWP science (Sims 1973). In 1960, Smagorinsky suggested an affine formation between relative humidity and cloud fraction. Since that time, other empirical formulations of cloud fraction have appeared in the literature, generally expressed as a function of relative humidity (Slingo 1980; Williamson et al. 1987; Walcek 1994; Xu and Randall 1996; Teixeira 2001). The current computation for cloud fraction output from operational NWP models, including the WRF model, is the Xu and Randall (1996) scheme, which defines cloud fraction, C , in terms of condensate mixing ratio and relative humidity, RH :

$$C = \begin{cases} RH^k [1 - \exp\left(\frac{-\beta_0 \bar{q}_l}{[(1 - RH)q_{vs}]^\tau}\right)], & \text{if } RH < 1 \\ 1, & \text{if } RH \geq 1 \end{cases}$$

where \bar{q}_l is the large-scale liquid water mixing ratio and q_{vs} is the saturation water vapor mixing ratio. The empirical values of k , β_0 , and τ are 0.25, 100, and 0.49, respectively.

Diagnostic schemes predate prognostic schemes. A prognostic equation for cloud fraction was first proposed in Tiedtke (1993). The Tiedtke (1993) scheme contained a formulation of the time tendency of cloud cover controlled by source terms representing increased cloudiness via convection, boundary-layer turbulence, and stratiform condensation processes; and a sink term for the decrease in cloudiness as a result of evaporation.

Beyond what is known about cloud formation and sustenance at a very basic level, previous studies have examined relationships between atmospheric quantities and clouds. There

are two general approaches. One approach is to consider those quantities that are correlated with cloud and cloud fraction but the degree to which is unclear. This includes the class of condensate variables and potentially other terms in the total water budget by extension. For example, in order to form a cloud, water is extracted from vapor form in the surrounding atmosphere into a liquid (including the supercooled state) or ice form. Or, the converse, in order for a cloud to disperse, water is extracted from liquid or ice state back into the atmosphere as vapor. The other approach is to consider those quantities that portend, support, or hinder clouds. This includes a number of kinematic and stability quantities. While some of these quantities, such as vertical motion, may have some interaction with cloud and moisture quantities, through diabatic heating for example, their relationship to cloudiness depends on how the atmosphere has been preconditioned to the development or decay of cloudiness. Quantities that can force cloudiness but require a source of water vapor are considered secondary quantities.

While it is possible that secondary quantities are important to understanding the cloud lifecycle, the formulation presented here makes it difficult to gauge the impact of those secondary quantities, particularly kinematic quantities and related parameters that depend on atmospheric motion. This is in part because of some complications with using NWP models. For example, NWP models require the use of diffusion terms to dampen numerical noise that results from time-stepping and other interfaces within the model. In addition, the predictive skill of the model for some quantities, those adjusted by a single thunderstorm, for instance, is not ideal for an exercise such as this. This is an inherent problem with using cloud-resolving models and small grid spacing where the initial conditions are not on the same scale. Other quantities are impacted, some strongly so, through the use of parameterizations. With large-scale models, there is previous evidence that cloudiness decreased with an increase in horizontal spatial

resolution (Tiedtke 1993). However, Tiedtke (1993) did not find a drastic dependency on horizontal resolution using his proposed prognostic cloud fraction scheme.

Zhang (2003) investigated hourly satellite cloud data between June and July 1997 to assess the relationship of cloud properties to atmospheric quantities as resolved in the Rapid Update Cycle (RUC) model. Zhang (2003) found correlations between high cloud and upward vertical motion as well as low cloud and subsidence, though the results are from the United States Central Plains during the summer. In addition, while, consistent with other studies (Smagorinsky 1960; Williamson et al. 1987), Zhang (2003) found evidence of a clear relationship between relative humidity and cloudiness, he did not find a threshold value for cloudiness that is sometimes found in other studies on this topic. For example, Smith (1990) developed a diagnostic cloud fraction formulation with two relative humidity thresholds, depending on the vertical location in the troposphere. Even Tiedtke (1993) used a humidity threshold as a regulator for forming clouds. This confirms previous work that suggests for a high relative humidity value, there is not necessarily a trend toward high cloud fraction, nor is there a threshold of relative humidity where cloud is consistently absent (Teixeira 2001).

b. Approximations

For the mixing ratio terms, Wood and Field (2000) proposed a differential equation in which the change of cloud fraction with respect to cloud mixing ratio is related to the cloud fraction and the saturation mixing ratio. An integration of this differential equation with the proper bounds, namely that cloud fraction is zero when there is no cloud mixing ratio, provides an equation of the form $f(x) = 1 - e^{-Vx}$ where $f(x)$ is the cloud fraction and V is a constant related to saturation mixing ratio and fit to observational data.

In developing optimal sky cover output from NWP output, it is important to provide for affine relationships between sky cover and the condensate terms. Though the exponential function is nonlinear, the Maclaurin series for the function e^{-x} is

$$e^{-x} = \sum_{n=0}^{\infty} \frac{(-x)^n}{n!} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \dots$$

where $0 \leq x \leq 1$ for this application. If higher order terms, where $n \geq 2$, are disregarded, $f(x) = 1 - e^x$ can be reduced to $f(x) \approx 1 - (1 - x) = x$, which is linear. Such an adjustment is possible because significant error between the approximated value and the actual value arises for $x > 1$, where the value of x becomes much larger than the approximated function. Since $0 \leq x \leq 1$, it is not necessary to define and/or center a Taylor series at a higher x value.

Furthermore, doing so would introduce nonlinear terms.

The approximation of the additive property of cloud fraction between levels is considered next. The exponential relationship between cloud cover and mixing ratio assumes the same form as absorbance absent scattering, $1 - t$, where transmittance $t = e^{-\tau}$, and τ is the optical depth. The Beer-Lambert law, or Beer's law, provides that there is a linear relationship between absorbance and concentration of the absorber. Though clouds have scattering properties, it is conceptually viable that the cloud fraction within a column would be linearly proportional to the absorbance of light through the column¹. The transmittance through two adjacent atmospheric layers is the product of the transmittance through each layer individually. The geometric argument for cloud fraction is construed similarly. The clear sky through an atmospheric column is the product of the clear sky fraction for the layers within the column, such that

¹ However, for this application, the light source would be at the surface, and clouds could only absorb or transmit.

$$A_Z = 1 - C_Z = \prod_{i=0}^N (1 - C_i) = \prod_{i=0}^N (A_i)$$

for $i \in N$ layers in the column, where A_i is the clear sky fraction for layer i or the entire column for A_Z , and C_i is the cloud fraction for layer i or the entire column for C_Z .

Since optical depth is a function of the extinction coefficient, the extinction coefficient is proportional to mixing ratio, and given there exists a direct relationship between absorbance and cloud fraction, it is possible to define the cloud cover as the sum of adjusted relative humidity values and adjusted mixing ratio values within a column. Therefore, the sum of adjusted mixing ratio quantities is approximated using the same Maclaurin series of e^{-x} as previously, such that the total cloud fraction for a given column $g = e^{-\tau} \approx 1 - (1 - \tau) = \tau$ where $g = \sum f_i(x)$ for all i , and τ defined as the column optical depth. Based on the aforementioned assumptions, this is a sum of i affine functions $f_i(x)$, and consequently, is also affine. These approximations make it easier to correlate a blended analysis with multiple output fields from a NWP model.

3. Creating a blended analysis

In this project, there are several types of new satellite analyses produced. Generally, each is derived from the one previous. One of the primary bands for cloud detection is the infrared window, though other bands can and do play an important role in discriminating clouds from atmospheric and terrestrial features. An example of the infrared window is found in Figure 1. All of the subsequent figures in this section are valid at approximately, or within the one-hour window after, 11 UTC on 20 October 2013. The original product input is the satellite effective cloud amount (ECA) output from Andrew Heidinger (Heidinger 2011a), an example of which is shown in Figure 2. Once a cloud is identified, radiative techniques exist to compute the cloud

top pressure (CTP), shown in Figure 3, assessed from spectral imagery collected with the satellite radiometer (Heidinger 2011b; Heidinger and Pavolonis 2009). The CTP is useful for determining the height of the cloud and whether it is detectable with surface-based automated instrumentation. Each point of the ECA product is then averaged over a point-centered 11 by 11 box representing the celestial dome to produce the celestial dome effective cloud amount (CDECA). The CDECA also includes corrections which increase the similarity of the product to sky cover. These corrections are noted in Appendix A, and an example of the CDECA product is found in Figure 4. The most notable correction involves updating the average emissivity when the ECA is less than 50% but the cloud fraction and cloud probability are sufficiently high, and the cloud phase consists of ice. The CDECA products for multiple times within a one-hour window are then averaged to produce a satellite sky cover product, shown in Figure 5. A celestial dome cloud top pressure (CDCTP) product² is also created for use in conjunction with the CDECA product, which contains the mean of the CTPs within a point-centered 11 by 11 box for corresponding non-clear points in the ECA where each contributing point of CTP is weighted by the ECA such that, for a given point,

$$CDCTP = \frac{\sum_{non-clear} CTP \times ECA}{\sum_{non-clear} ECA}$$

All satellite cloud products are computed independently at each scan time prior to combination with multiple times and multiple satellites. There is a blend zone between GOES-East and GOES-West between 100° West and 105° West longitude where product values from both satellites are averaged. Products are averaged within a one-hour window and all products are given a valid time at the beginning of window in which the average begins.

² Anthony Schreiner developed the methodology for calculating the CDCTP product.

a. Gridding surface observations

A component of the project, pursuant to the problem at hand, is to combine the satellite sky cover product and surface observations through leveraging the strengths of each, and producing a single analysis, herein referred to as the blended sky cover analysis. The intent is to ameliorate the weaknesses compared to a single platform analysis. The first step in the process is to create an analysis of sky cover from surface observations. For this, all of the observations within a one-hour window are analyzed to a 10 km Lambert Conformal grid, as shown in Figure 6. Gridding the surface observations and creating a spatially contiguous grid of surface reports is a multi-step process. For the continental United States domain, the initial step is to bin all of the observations from within a one-hour window to the nearest grid cell. This includes multiple observations from the same station or observations from adjacent stations residing within the same grid cell. Both automated and manual observations are used.

The subsequent step is to fill the portions of the grid without a resident surface station. For a candidate grid point, a search radius is expanded to adjacent grid points until a match is found. When a match is found, the observation is checked against the satellite sky cover product value at that point. If the grid cell with a non-missing value and corresponding satellite sky cover product value match within a pre-established tolerance, 25% coverage, and the satellite observation is clear at the candidate point, the pixel is considered clear at this stage. Otherwise, the candidate point assumes the non-missing sky cover amount of the closest point with a resident observation. If there are two or more points found of equal distance to the candidate point, the mean of those closest points is taken.

For grid cells without a resident observation, another weighted average is performed based upon the distance to the nearest grid cell with a resident observation, depending on the

characteristics of the sky cover amount at that grid cell. The weighted average covers a broader area for those points that are further away from an observing station. The square area of weighted points is centered on each eligible point and has a side edge equal to twice the distance to the closest surface observation, less one. The center point and the edge points of each concentric square are weighted the same. Therefore, for the progressively larger interior squares, less weight is given to each edge point. Each point is weighted except when the center point, containing the value of the nearest surface observation consistent with the aforementioned logic, is valued at 0% coverage, the satellite sky cover product exceeds the tolerance for clear skies, 5% coverage, and the satellite CDCTP is above 750 hPa, or more than half of the observations in the box match these criteria. In such cases, a sky cover amount of 0% is assigned. The CTP requirement helps to prevent false satellite cloud detections from impacting subsequent analyses. Such potential cases are discussed subsequently.

b. Basic methodology

The essence of the methodology behind producing the blended analysis hinges on the need to preserve not only the location of cloud edges, but also the magnitude of the sky cover gradient across them. This is particularly important in areas under and near a stratocumulus deck, for example, which occurs frequently on the Pacific Coast of the United States. Furthermore, the non-clear surface observation sky cover reports are the most important, since they are the best representation of truth. The basic logic behind the blended sky cover analysis is as follows for a given grid cell:

- Where the observation of sky cover from the surface is clear, with a sky cover amount of less than 5%, the blended analysis will assume the value from the satellite sky cover

product depending on the corresponding satellite CTP or CDCTP. If the CDCTP value is sufficiently high (low cloud), then the satellite cloud detection is considered false.

- Where the aforementioned case does not apply and the value of the nearest surface observation of sky cover is greater than that of the satellite sky cover product, the sky cover value of the surface observation is used in the blended analysis.
- In all other instances where both observations are available, a weighted average is employed, based on the proximity to the nearest observing station.
- If one observation type is available for a point, but not the other, the available observation is used in the blended analysis. For example, over ocean, there are no nearby surface observations. In such cases, the analysis uses solely the value from the satellite sky cover product.

c. Low cloud

In the case of geostationary satellites, detectability of low cloud is the most difficult, particularly in the seasons of spring and fall. This is evident in the case from 11 UTC on 20 October 2013, where morning fog in the Seattle, Washington, metropolitan area, evaded detection from satellite due to an emissivity and brightness temperature similar to that of the background land and water surface. Such cases are relatively common when radiation fog forms beneath an inversion on cool fall nights with otherwise little cloudiness. Fortunately, surface observations typically report overcast skies or a vertical obstruction in these situations.

d. Blending logic

When the observations are blended, a decision structure is used to determine the final sky cover amount for each pixel. The decision structure is based on the strengths of the contributing

observation types, or where only one observation type exists. The creation of the blended analysis is predominantly guided in two ways. First, surface observations of cloudy skies are the least likely to be inaccurate. Second, for a given satellite cloud algorithm detection over an area identified as clear by a surface observation, the satellite observation is usually adequate, especially for situations with middle and high cloud. There are also cases in which the sky cover product indicates cloudiness but the detection is false. The logic works around such situations.

STRAY LIGHT TESTS

At night, geostationary satellite imagery is subject to stray light at certain times. Stray light occurs when sunlight enters the optics of the instrument without a reflection from the earth's surface. There are approaches to mitigate stray light (Matthews 2013; Zhenping et al. 2013), but impacts on satellite algorithms remain. The first test involves areas of the filled surface observation grid containing values less than the tolerance value for clear skies, 5% coverage. In such cases, if the satellite CDCTP value is greater than the threshold of 750 hPa (between the surface and 750 hPa), the filled surface observation grid value is used. In addition, the filled surface observation grid value is used when a CDECA value is available for a point from both GOES-East and GOES-West and certain conditions are met to discredit the integrity of the satellite sky cover product coverage amount. In such cases, significant differences in cloud detection are possible between GOES-East and GOES-West when stray light is impacting the imager on one of the satellites but not the other. This is a possibility when there are two conditions met. First, the satellite sky cover product value for the point falls between the a third of the CDECA value at the top of the hour, less 5% coverage, and the tolerance of 40% coverage. The 40% coverage threshold results from the observation that the stray light impact

lasts for one quarter or one third of the total time window, depending on the number of other scans that are part of the window averaging. Second, the satellite sky cover must be greater than 10% and less than a quarter of the CDECA value including a tolerance of 5%, at the top of the hour. These conditions assure that the point is not partly cloudy across multiple scans.

When the conditions are met and both GOES-East and GOES-West CDECA and CDCTP values are available, the filled surface observation grid value is applied to the final blended sky cover analysis if the CDCTP is between the surface and 750 hPa. If only one satellite is available, then the CDCTP must fall between the surface and 750 hPa. 750 hPa, approximately 7 kft (2100 m), is used as the upper limit due to the detectability range of the surface station ceilometer. In all other cases, the satellite sky cover product value is trusted and advanced to the final blended sky cover analysis.

SPATIAL CONTINUITY

The remaining logic details spatial transitions from surface observation values to satellite sky cover product values. An iterative process is applied to all grid points. In general, when the filled surface observation value is higher than the sky cover product value for a given candidate point, the filled surface observation value is used within two-thirds of the number of grid cells between the candidate point and the nearest surface station. When within the remaining one-third of the distance nearest the candidate point, the value of the candidate point is scaled with the surface observation, weighted linearly depending on the distance to the two-thirds extent and the limit of influence, which is 14 grid points or approximately 140 km. This is predominantly a cosmetic effect to deter irregular boundaries or artificial sharp edges where surface stations are not close together. Lastly, if the filled surface observation grid value is less than the satellite sky

cover product value, or they are equal, then an average is performed, weighted linearly depending on the distance from the surface station pixel.

There are two reasons for trusting the surface observation grid over the satellite sky cover product when both are reporting some degree of cloudiness. First, surface observations are closer to the traditionally thicker clouds, such as stratus, which are responsible for overcast conditions. Second, the single field of view (FOV) ECA values from satellite are for the highest cloud only. Consequently, the calculated ECA may be less than the ECA if the higher cloud is not present. This presents challenges for assessing sky cover over the ocean or in situations where multi-layer cloud exists but there are no surface observations.

e. Comparisons

Upon initial comparisons between surface observations and the satellite sky cover product, the lack of precision in the surface observations is evident and the correlation appears poor. There are relatively few sky cover amounts falling between the primary classifications. This suggests that non-routine surface station reports with adjustments to the cloud condition are relatively rare. However, the histogram shown in Panel B of Figure 8 indicates a greater frequency of coverage values above 70% compared to lesser coverage values that are not clear.

f. Cases

The simplest and highest confidence cases are ones in which the satellite observation at the test point, the satellite observation at the closest surface station, and the surface observation at the closest site are all clear or all cloudy. All possible cases are summarized in Table 1. In this section, satellite observation refers to the sky cover amount from the satellite sky cover product. The surface observation refers to an in-situ sky cover report from automated

instrumentation or a human observer. Cloudy cases are those that are not clear. In many cases, the sky cover amounts for cloudy cases are reconciled between the observing platforms.

The case in which the satellite observation at the test point and the satellite observation at the closest surface station are clear, but the surface observation at the closest site is cloudy, occurs when the cloud properties are too similar to the underlying surface. This is usually the case for low cloud and fog during the overnight hours, where visible channels cannot be used to determine high reflectance values indicative of cloud. In addition, in the infrared window, diurnal heating of exposed land surfaces heightens contrast to low clouds, which typically erode during the morning hours. This case may also occur in a scenario where there is a local, non-cloud obstruction, such as smoke, or when it is difficult to differentiate between snow and cloud using remote sensing techniques. The latter situation is most likely to occur when there is a human observer to view cloud given the limitations of the ceilometers in detecting high cloud. The result in the blended analysis is to trust the surface observation value of cloudy to override the clear observation from the satellite sky cover product.

The case in which the satellite observation at the test point is clear, the satellite observation at the closest surface station is cloudy, and the surface observation at the closest site is clear most likely occurs in a situation where there is high cloud over the observing site that the ceilometer does not detect. It is possible that this may be a scenario in which the satellite detection of cloud is false at the closest surface station. However, since it is clear at the test point, via the satellite sky cover product, it is assumed that this is a natural cloud gradient, most likely involving high cloud, and the result is therefore to maintain clear in the blended analysis.

When the satellite observation at the test point is clear, but the satellite observation at the closest surface station and the surface observation at that station are both cloudy, a natural cloud

gradient is assumed and the result in the blended analysis for the test point is to maintain the clear observation from the satellite product output. The converse case, where the satellite observation at the test point is cloudy, but the satellite observation at the closest surface station and the surface observation at that station are both clear, is another situation of a natural cloud gradient. In both situations, the agreement between the closest surface observation and the corresponding satellite observation provide confidence in the satellite sky cover product depiction of the surrounding sky cover regime. Therefore, the satellite observation at the test point is trusted.

The case in which the satellite observation at the test point and the surface observation at the closest site is cloudy, but the satellite observation at the closest site is clear, is a situation where there is typically a gradient in the type of cloud, or limited detectability of the cloud using remote sensing techniques. The most likely scenario is that there is low cloud of emissivity similar to the surface over the observing station, but high cloud at the test point. Such occurrences make it difficult to quantify the magnitude of sky cover because it is uncertain if the low cloud extends beneath the high cloud deck.

The case in which the satellite observation at the test point and the satellite observation at the nearest surface station are cloudy, but the surface observation at that nearest station is clear, is a situation where there is most likely high cloud that is beyond the height of detectability for the instrumentation at the station. The satellite observation is trusted in this scenario as long as it is high or mid-level cloud. In some situations, where the satellite CDCTP value is indicative of low cloud, with a CTP greater than or equal to 750 hPa, the clear sky surface observation at the nearest station is trusted because of the heightened likelihood of false cloud detection on the satellite sky cover product.

The blended sky cover analysis, shown in Figure 7, captures more cloudiness than when compared to each of the inputs for the analysis independently. This results from limitations of the surface instrumentation or remote sensing techniques in quantifying cloudiness. Validating the blended analysis is difficult because there are limited single sources that provide a unique analog of sky cover, not already incorporated into the blended analysis, which can easily be geo-referenced with the blended analysis grid.

g. Efficacy of blended analysis

Certain comparison techniques utilized assure the blended analysis is representative of the inputs. Figure 8 shows the distribution of cloud/sky cover amounts in 10% increments for the CDECA product, in Panel A, sky cover product, in Panel B, surface observations, in Panel C, and the blended sky cover analysis, in Panel D. Similarities between the satellite sky cover product and blended sky cover analysis are evident. First, in both cases, sky cover amounts greater than 70% are more prevalent than the other cloud classifications. However, the number of points with a clear or nearly clear sky cover amount, at or below 10% coverage, is substantial in this case. This differs from summertime cases in which mid-coverage cloudiness is more common during the day as diurnal heating supports increased growth of the cumulus field over a large portion of the non-mountainous United States.

The comparison between the blended analysis and the surface sky cover product is a bit more involved. However, the vast number of points in the blended sky cover analysis matches the output from the surface sky cover product, as shown in Figure 9. Figure 9, Panel A, shows the scatterplot of collocated points between the blended sky cover analysis and the sky cover product. Panel B is density plot of Panel A. The greatest deviation from the diagonal occurs for

corresponding points where both the blended sky cover analysis and satellite sky cover product amounts are mostly cloudy. There are two effects at play in this sector of the plot. First, there are areas where the gridded surface observation sky cover amounts are higher than the corresponding amounts in the satellite sky cover product. In such cases, the logic trusting the higher surface observation value prevails. These adjustments typically involve areas where cirrus clouds are above lower stratus clouds, and are often modest. In cases where the satellite sky cover product value is higher than the gridded surface observation value, the satellite sky cover product values are adjusted to the mean value of the two inputs in the final blended analysis. Such areas are likely where surface stations report a lesser sky cover amount than the corresponding surface sky cover product value, likely due to a bias in the sky cover product.

The predominant manner for assuring the ideal methodology behind the blended analysis is to evaluate scatterplots matching the independent observation sources. In Panel C of Figure 9, a scatterplot of matched surface observations and satellite sky cover product values is shown. The correspondence between the surface observations and satellite sky cover product is relatively poor, with no evident correlation along the dashed diagonal axis. However, in preparing the blended analysis, the surface observations are trusted. This decreases the number of points in the blended analysis where an average is employed.

In Panel D of Figure 9, a histogram showing the frequency of differences, at increments, between the sky cover product and blended sky cover analysis is shown. In general, the blended sky cover analysis increases the sky cover product value because the surface observation value is trusted over the sky cover product when at least some cloud exists. In addition, as previously discussed, the blended sky cover product contains cloud where the sky cover product may not due to the addition of surface observations, as shown in Figure 10. There is significant

cloudiness added as a result of the surface observations in the Pacific Northwest and over parts of Appalachia. There is also an adjustment to the degree of cloudiness over the Midwest. With this blended analysis, it is possible to compare it to the NDFD cloud cover one-hour forecast and prepare an optimal sky cover analysis.

4. Optimization methodology

As computer and data storage capabilities have increased over the past two decades, optimization has become a formidable method for exploring and quantifying relationships between multiple datasets. Optimization is the process of solving a problem through systematically trying to obtain the best objective value via changes to variables, subject to constraints. The constraints form a feasible region, in the case of a linear program, a polyhedron, in which a solution is possible. The point in the feasible region with the best objective value is the solution.

In this stage of the project, there are three separate optimization procedures executed. Each one is explained in depth in this section. The first produces an optimal sky cover analysis from the blended sky cover analysis. This is referred to as the preparation model. The second correlates NWP model variables at the analysis time to the optimal sky cover analysis. This is referred to as the primary model. The third refines the output from the previous step to further pursue the objective if possible. This is referred to as the secondary model.

a. Preparation model

The purpose of running the preparation model is to establish for what coverage categories that the blended analysis did well, and those that it did poorly, and then provide a non-negative affine adjustment (coefficient and scalar) for fixed increments, with anchors at 0% and 100% sky

cover. Generally, when the slope of the line (coefficient) is one, the distribution of values in the blended sky cover analysis is best. The constraints in the model require that the slope of the line be non-negative. The output from the preparation model is the optimal sky cover analysis.

Assuming that the accuracy of the near-term grids is best, the blended sky cover analysis is adjusted based on a minimization of mean absolute error (Willmott 2005) between the NDFD total cloud cover one-hour forecast r_I and the blended sky cover analysis d_I every three hours. An optimization model performs the adjustment. One out of every 15 points is provided to the optimization model as representative of the entire grid. In order to prevent inflated cloud amounts from adversely impacting the optimization model, areas identified as clear in the satellite sky cover product are reset to clear in the adjusted blended sky cover analysis provided to the optimization model. The six coefficient p_J and six scalar q_J solved variable values are then applied back to the blended sky cover analysis to produce the optimal sky cover analysis. Accordingly, the formulation is as follows with objective function δ :

$$\begin{aligned} \min \delta &= \sum_{i=1}^n |p_J d_I^J + q_J - r_I| \\ 0 &\leq p_J d_I^J + q_J \leq 100 \\ q_1 &= 0 \\ p_1 5 + q_1 &= p_2 5 + q_2 \\ p_2 25 + q_2 &= p_3 25 + q_3 \\ p_3 50 + q_3 &= p_4 50 + q_4 \\ p_4 75 + q_4 &= p_5 75 + q_5 \\ p_5 95 + q_5 &= p_6 95 + q_6 \\ p_6 100 + q_6 &= 100 \\ p_J &\geq 0 \end{aligned}$$

where

$$\begin{aligned} J &\in \{K \in \mathbb{Z}: 1 \leq K \leq 6\}, \\ d_I^1 &\in \{d_I \in \mathbb{R}: 0 \leq d_I < 5\} \text{ (clear)}, \\ d_I^2 &\in \{d_I \in \mathbb{R}: 5 \leq d_I < 25\} \text{ (mostly clear)}, \\ d_I^J &\in \{d_I \in \mathbb{R}: 25(I-2) \leq d_I < 25(I-1)\}, 3 \leq I \leq 4 \text{ (partly cloudy)}, \end{aligned}$$

$$d_l^5 \in \{d_l \in \mathbb{R}: 75 \leq d_l < 95\} \text{ (mostly cloudy), and}$$

$$d_l^6 \in \{d_l \in \mathbb{R}: 95 \leq d_l \leq 100\} \text{ (cloudy).}$$

An example of how the preparation model produces the optimal sky cover product is evident in the case from 11 UTC on 20 October 2013. The case is fairly generic. An example of the optimal sky cover output for this date and time is found in Figure 11. Figure 12 shows a correspondence between the optimal sky cover amount and the blended analysis. According to Panel A, the optimal sky cover product value is slightly higher than the blended sky cover analysis value for blended analysis values of less than 50%. For blended sky cover analysis values of at least 65%, the optimal sky cover product value is a comparative decrease. For example, according to Panel A, a 95% sky cover on the blended analysis is approximately 75% after optimized. Panel B shows the difference of the blended sky cover analysis from the optimal sky cover product. This indicates that approximately 21% of points experienced an overall decrease in value, with 10% receiving an increase. Figure 13 shows geographically where the optimal sky cover is an increase or decrease from the blended analysis. This occurs along gradients of sky cover.

b. Primary model

Once the blended sky cover analysis is optimized, the optimal sky cover analysis is available for use in comparison with NWP model output. One goal of this study focuses on the use of optimization to find a set of coefficients and scalar adjustments to NWP model output cloud and moisture quantities. The High-Resolution Rapid Refresh (HRRR) model, based on the WRF model framework (Ikeda et al. 2013), provides the numerical output for input into the optimization model. Since the model assimilates GOES cloud products, it is ideal for providing an analysis and output which closely matches the blended and optimal analyses produced as part

of the first portion of this study. The spatial resolution and hourly availability further increase the utility of the HRRR model for this project, compared to other operational and research NWP models that are initialized only four times per day at spatial resolutions well above cloud scale.

The HRRR model is a 3 km model run hourly as an experimental cloud-resolving atmospheric model (Benjamin 2013). It is based on the Advanced Research WRF (ARW) core and uses the Mellor-Yamada-Nakanishi-Niino (MYNN) planetary boundary layer (PBL) scheme and the RUC nine-level land surface model (Benjamin 2013). The HRRR model employs the Thompson et al. (2008) microphysics scheme, where hydrometeors are binned into one of five classifications: cloud water, cloud ice, rain, snow, and graupel. The model assimilates data from rawinsondes, profilers, radars, lightning detectors, aircraft, surface stations, and buoys, as well as GOES atmospheric motion vectors; Advanced Microwave Sounding Unit (AMSU), High-resolution Infrared Radiation Sounder (HIRS), and Microwave Humidity Sounder (MHS) radiances; GOES CTP and temperature retrievals, Global Positioning System (GPS) precipitable water amounts, and WindSat scatterometer data, among others (Benjamin 2013). Temperatures for HRRR model grid points beneath the surface are calculated through an adiabatic adjustment of the mean temperature at the second and third model layers above the surface, with relative humidity maintained for underground points from the mean of the same second and third model layers (Chuang, DiMego, and Baldwin 2004). For all points, relative humidity in the HRRR model is calculated according to Schlatter and Baker (1981) using a polynomial approximation.

For the primary model, a binary (or “zero-one”) optimization problem, a case of the mixed integer optimization problems is formulated. While a mixed integer formulation is nonconvex by definition, and therefore difficult to solve, relaxation procedures can produce a convex linear optimization problem that can be solved relatively easily. However, because

binary optimization problems fall in computational complexity case “NP-complete”³ (Karp 1972), there is not an efficient method for finding the solution. The larger the problem, the longer it takes to solve.

In this portion of the project, two commercial optimizers/solvers are used to produce a solution: CPLEX (<http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/>) and Gurobi (<http://www.gurobi.com/>). Both were provided free of charge for use in this project under an academic license. The primary reason for the use of these optimizers is performance. Since this project requires running the optimizer every hour, consistent with the run frequency of the HRRR model, a slow optimizer is computationally inhibitive. Even with these commercial optimizers, the time to reach a solution for a model with 110 representative sample points, as developed here, can exceed 1000 seconds on a multi-threaded server. Two different optimizers are used to ensure an ideal solve time, compare results, and ensure consistency.

The objective of the primary model is to minimize the mean absolute error (Willmott 2005) between the optimal sky cover analysis and an affine sum of relative humidity, cloud water mixing ratio, rain water mixing ratio, cloud ice mixing ratio, and snow mixing ratio, at nine pressure surfaces: 200 hPa, 300 hPa, 500 hPa, 700 hPa, 800 hPa, 850 hPa, 900 hPa, 950 hPa, and 1000 hPa, as well as absolute vorticity at 200 hPa. As part of staging the HRRR model analysis output for the optimization model, each of the HRRR model output fields is converted to a new grid that matches the domain (where defined) and projection of the blended analysis and is modified point-wise such that the updated center point value is a mean value for the celestial dome, consisting of gridded values at and surrounding each grid point. The celestial dome mean is the mean of the values for the field within a square of side length 40 km (two points left, right,

³ The “NP” in “NP-complete” refers to “nondeterministic polynomial time”.

up, and down). Lastly, the mean celestial dome values from the initialization time output and the subsequent forecast hour output are averaged with equal weight. The formulation of the optimization model is careful to address and limit the impact of the discrepancy that results from a linear approximation of a nonlinear relationship. In addition, per the definition, cloud/sky cover, as it counts toward the objective value, cannot exceed completely overcast skies (100%), and cloud/sky cover is constrained from becoming negative (less than 0%). The mixed integer optimization implementation does allow the sum of all of the products of the coefficient and the mixing ratio quantities, and attendant scalars, within a column to exceed or fall below the defined cloud cover range (0% to 100% sky cover) depending on whether there is condensate within the column. However, when this occurs, the error is calculated as if the quantity is at that bound (0% or 100% sky cover).

Lastly, absolute vorticity is an odd presence in this formulation. Though Kvamstø (1991) found no relationship between cloud cover and other dynamical quantities, including vertical velocity, advection of thickness, advection of equivalent potential temperature, relative vorticity, and stratification, absolute vorticity is added here as a pre-conditioner for the other terms. That is, the main purpose of absolute vorticity is to adjust the cloud cover based on the favorability of the environment for cloudiness from a dynamical perspective. Though there is not a bright-line relationship for absolute vorticity and cloud cover, and the thermodynamic environment plays a controlling role in cloud cover, areas of positive vorticity (in the Northern Hemisphere) are more dynamically active, invoke diabatic processes, and therefore can be cloudier, whereas the converse is true in areas of negative vorticity.

Due to the substantial number of grid points, a subset of 110 points is chosen for use with the primary optimization model. These grid points are selected randomly at each cycle based on

their general representativeness of all of the points on the grid, and their quality for forming meaningful relationships between the selected atmospheric quantities and sky cover. The distribution of sky cover amounts for the selected points is similar to the distribution for all points. All grid points are binned depending on the corresponding “truth” value from the optimal sky cover analysis and whether or not the point contains a non-zero mixing ratio amount. The bins have widths of 10% sky cover with a separate bin for test points with no condensate but an optimal sky cover amount less than the mean value rounded to the nearest 10% interval. For optimal sky cover amounts less than 10%, the binned points must contain no cloud condensate. This makes 11 bins in total, with one bin for each 10% increment in sky cover, and then one bin for high optimal sky cover amounts but no condensate. Points with no condensate and an optimal sky cover amount of at least the mean value rounded to the nearest 10% interval are not considered.

In addition, fields from the HRRR model output with few non-zero values are removed. If the number of randomly selected points with a non-zero value is less than six for a given field, then the values of all of the points for that field are set to zero. This eliminates uncharacteristic coefficient and scalar values in the solution that are calculated from only a few points.

Along with the aforementioned physical and mathematical assumptions, there are a few basic assumptions that are part of the optimization process. First, the assumption is that the distribution and magnitude of the atmospheric quantities in the model analysis are correct. For that reason, it is desirable to use a NWP model that assimilates a large and diverse number of observations. Second, the assumption is that the relationship between the quantities in the analysis is the same as in subsequent forecast hours. For example, if there is a deficiency in the model or one of the parameterizations that decreases overall water content with time, the

accuracy of the solutions formulated using the optimization model adjustments will likewise decrease with time. Third, due to the complexity of the problem, slight alterations in the data and constraints may lead to notably contrasting solutions despite similar objective values. For that reason, results must be considered over a substantial time interval containing multiple runs and scenarios.

DATA VARIABLES

The data variables for the primary model are defined prior to the execution of the model. They are as follows for this subsection. a_j represents the “truth” value for the j th randomly chosen point. In this case, a_j is drawn from optimal blended sky cover grid, where the notation used for any such point on the grid, not just a chosen point, is \dot{a}_j . $b_{i,j}$ represents the HRRR model analysis value for the matching j th point in the i th field, whereas $o_{i,j}$ is equal to one when $b_{i,j}$ is greater than zero, and zero otherwise, for the condensate fields. The notation $\dot{b}_{i,j}$ refers to all of the non-missing points on the grid. ζ_j^+ represents the positive values of absolute vorticity from the HRRR model analysis, at corresponding j th points, and is zero otherwise, whereas ζ_j^- contains the negative values of absolute vorticity for the HRRR model analysis and is zero otherwise. Thus, absolute vorticity $\zeta_j = \zeta_j^+ + \zeta_j^-$. Likewise, $\dot{\zeta}_j^+$ and $\dot{\zeta}_j^-$ contain all non-missing grid points of positive and negative absolute vorticity, respectively, not just those corresponding to a contributing a_j . The data variable c_j is defined subsequently.

MODEL VARIABLES

The model variables are those variables that are solved through the execution of the primary model in seeking the best objective value. All of the optimization model variables are

interconnected through constraints and the objective. The mean absolute error (MAE) value used in the objective is $y_j \in (-\infty, \infty) \subset \mathbb{R}$ and the variable that carries the slack as appropriate is $t_j \in (-\infty, \infty) \subset \mathbb{R}$. Other real-valued model variables are constrained such that

$$B_L, g_j, s_j, w, z_i \in [0, 100] \subset \mathbb{R},$$

where B_L is the relative humidity threshold at 1000 hPa above which a coefficient and possible scalar apply, and $Nw = \sum_j s_j$ where N is equal to the number of j points and w is therefore the mean of all s_j points. g_j is defined subsequently as part of the threshold procedure for 1000 hPa.

Otherwise, no scalar values z_i apply to relative humidity values. In addition,

$$c_j, f_j, v^+, v^-, x_i \in [0, \infty) \subset \mathbb{R},$$

where v^+ is the variable applied to positive absolute vorticity and v^- is the variable applied to negative absolute vorticity. The variable f_j is defined subsequently as part of the threshold procedure for 1000 hPa. The general variable x_i is applied as a coefficient to field i . Lastly, $h_i \in [-1, 0] \subset \mathbb{R}$, where h_i is the variable applied to the negative relative humidity values and 100% such that the lesser the relative humidity value, the greater the decrease in sky cover depending on the magnitude of h_i . The h_i coefficient decreases the entire cloud fraction independent from the corresponding x_i variable for relative humidity values, but h_i does not apply to condensate or other non-relative humidity variables. The full contribution of relative humidity to the final sky cover coverage is therefore $(x_i - h_i)b_{i,j} + 100h_i$ for a given i th relative humidity field, remembering that h_i is negative or zero.

The mixed integer (binary) variables in the optimization model, controlling the decision structure, are n_j and u_j , such that $n_j, u_j \in [0, 1] \subset \mathbb{Z}$. These variables are valued at each point, with u_j activating the slackness variable and n_j controlling points subject to the threshold.

If there is condensed water in the integrated column, where $x_i \leq 1$ for fields $i \leq L$, in the cases of the relative humidity levels, including 1000 hPa, and $z_i = 0$ for fields $i < L$, in the cases of the relative humidity levels, excluding 1000 hPa, then for such a point j ,

$$s_j + t_j = \left(\sum_i b_{i,j} x_i \right) + \zeta_j^+ v^+ + \zeta_j^- v^- + \sum_i o_{i,j} z_i + \sum_i (100 - b_{i,j}) h_i + b_{L,j} f_j + g_j,$$

and scalars only apply when there is a non-zero quantity at that point.

The following constraints assure that $s_j + t_j$ is at least valued at 1% sky cover, and t_j assumes a positive value when s_j is at its upper limit. Since s_j appears in the objective, and it is physically possible for the sum of cloud coverage through multiple layers to exceed 100%, the variable t_j allows for growth of the left-hand side of the equation without impact to the objective value, since the value of s_j cannot exceed 100%. Below, the integer variable u_j assures that t_j is only non-zero when s_j is at its maximum.

$$\begin{aligned} s_j &\geq 1 \\ t_j &\geq 0 \\ 100(1 - u_j) + s_j &\geq 100 \\ 10000u_j &\geq t_j \end{aligned}$$

If there is no condensed water in the integrated column (only water vapor), then for such a point j ,

$$s_j + t_j = \left(\sum_i b_{i,j} x_i \right) + \zeta_j^+ v^+ + \zeta_j^- v^- + \sum_i (100 - b_{i,j}) h_i + b_{L,j} f_j + g_j,$$

where the following constraints assure that $s_j + t_j$ does not exceed 100%, and t_j assumes a negative value when s_j is at its lower limit. In such scenarios, t_j assumes the role of the slackness variable where $s_j + t_j$ is non-positive when s_j is zero, and these constraints apply.

$$\begin{aligned} t_j &\leq 0 \\ s_j &\leq 100(1 - u_j) \\ -10000u_j &\leq t_j \end{aligned}$$

The following constraints apply for all j points to minimize the absolute error between the “truth” value and the optimal formulation.

$$\begin{aligned} a_j - s_j &\leq y_j \\ s_j - a_j &\leq y_j \end{aligned}$$

The following two constraints establish whether n_j is equal to zero or one for each j th point. This is based on the threshold variable B_L . If $b_{L,j}$ is greater than B_L , then n_j must be set to zero according to the constraint $b_{L,j}n_j \leq B_L$ where B_L is the threshold within field L , which is relative humidity at 1000 hPa. In addition, $B_L - b_{L,j} \leq 100n_j$ is also a necessary constraint in order to assure n_j assumes the value of one in the circumstances where $b_{L,j}$ is less than or equal to B_L . The following three constraints set f_j equal to x_L if n_j is zero. If $n_j = 1$, then $f_j = 0$.

$$\begin{aligned} f_j &\leq x_L \\ f_j &\leq 100(1 - n_j) \\ f_j &\geq x_L - 100n_j \end{aligned}$$

The following three constraints set g_j equal to z_L if $n_j = 0$. If $n_j = 1$, then $g_j = 0$.

$$\begin{aligned} g_j &\leq z_L \\ g_j &\leq 100(1 - n_j) \\ g_j &\geq z_L - 100n_j \end{aligned}$$

For field $i > L$, in situations where there is condensate within the column, individual scalar adjustments are constrained such that, for a point j in a set J of all point values \dot{a}_j ,

$$z_i \leq \min_{j \in J} c_j$$

where c_j is a point assuming the value of \dot{a}_j when $\dot{a}_j \geq 10$ and $b_{L,j} > 0$, and is otherwise zero.

In addition, in order to assure some upper bound on a coefficient that may promote sky cover well beyond overcast (due to the presence of the slackness variable t_j), the following constraint is set such that no value $b_{i,j}^{\dot{}}$ exists whose product with the solution coefficient x_i is greater than 200% sky cover.

$$\left(\max_{j \in J} b_{i,j}^{\dot{}} \right) x_i \leq 200$$

There is also a constraint to assure that the maximum positive absolute vorticity on the grid does not produce more than complete cloud/sky coverage.

$$\left(\max_{j \in J} \zeta_j^{\dot{+}} \right) v^+ \leq 100$$

For all j points, $b_{i,j}^{\ddot{}}$ is equal to $b_{i,j}^{\dot{}}$ except when $b_{i,j}^{\dot{}} = 0$. In such a case when $b_{i,j}^{\dot{}} = 0$, $b_{i,j}^{\ddot{}}$ is effectively infinity. For a given field i , if there is at least one point j for which $b_{i,j}^{\dot{}} \neq 0$, then the following constraint is set:

$$\left(\min_{j \in J} b_{i,j}^{\ddot{}} \right) x_i + z_i + \left(\min_{j \in J} \zeta_j^{\dot{-}} \right) v^- + \sum_i \left(100 - \min_{j \in J} b_{i,j}^{\ddot{}} \right) h_i + \sum_i \left(\min_{j \in J} b_{i,j}^{\ddot{}} \right) x_i \geq 1.$$

This constraint assures that for a minimal amount of condensate, the output provides for a positive, non-zero sky cover amount. The distribution of the values in the output is also controlled, such that $\gamma - \beta \leq w \leq \gamma + \beta$, where γ is the mean from the optimal sky cover product and $\beta = 5$ (%) is the tolerance.

There are also constraints to assure the integrity of the values at the extremes of the range. This promotes areas that are completely clear or nearly clear, and completely cloudy or nearly cloudy. The constraint to maintain the mean of the points with “truth” values in the lowest 10% coverage range, such that $0 \leq W^M \leq 2\beta$, is $N_S^M W^M = \sum_j s_j^M$, where $s_j^M \in S^M$, S^M

contains only and all s_j^M which have a corresponding $a_j \leq \beta$ for a given j , and N_S^M is the number of objects in S^M . Similarly, for the highest 10% coverage range, where $s_j^X \in S^X$, S^X contains only and all s_j^X which have a corresponding $a_j \geq 100 - \beta$ for a given j , and N_S^X is the number of objects in S^X , $100 - 2\beta \leq W^X \leq 100$ where $N_S^X W^X = \sum_j s_j^X$.

Lastly, the objective, a minimization function, is

$$\min \frac{1}{N} \sum_j y_j$$

where N , the same as referenced previously, is the number of objects in Y , and every and only $y_j \in Y$. From a physical perspective, the optimizer is seeking a solution where a combination of certain input fields most closely matches the sky cover for the incorporated points. In order to do this, it is providing for a relationship between the chosen field and the cloudiness, or lack thereof. The input fields are expected to contain the physical basis for cloudiness. Those fields that are consistently chosen, over multiple runs, are most likely to have a direct physical role in the existence of cloudiness, and possibly, the creation or decay of it.

c. Secondary model

There is one final optimization step done following the completion of the primary optimization. A second optimization model is run to determine within a given test field, for chosen points with no condensate, whether a model-determined threshold exists such that setting the sky cover to zero for points not greater than the threshold decreases the MAE of the entire grid, and for chosen points with condensate, whether an equal-valued or greater threshold exists, at and under which the output is scaled per a model-determined coefficient, which decreases the MAE of the entire grid. During this stage, the number of test points in each bin increases to ten

times the number used for the primary optimization model. The initial set of points is retained in the bins as new points are added so the total size of the bin is 11 times of that previous.

Then, a series of models are run. The first model execution tests the primary model sky cover output to assess if thresholds exist within the sky cover output to decrease the MAE. Each subsequent model execution tests a single level of relative humidity. The one with the lowest MAE is used as the final. The model is formulated similarly to the 1000 hPa relative humidity threshold test in the primary model. The significant difference is that here, the selection of the relative humidity threshold for one level impacts the entire result – the contribution of each level to the sky cover is negated at certain points. In addition, columns with cloud condensate are treated differently than those without. This assures that cloudy regimes are preserved from the HRRR model. The MAE is always at least no worse than the output from the primary model.

The secondary model employs the following new model variables, with constrained ranges, that are discussed subsequently. Two variables, s^L and s^M , are thresholds, and are constrained such that $s^L \leq s^M$. m is a model variable coefficient for scaling sky cover at points with condensate and/or positive absolute vorticity.

$$\begin{aligned} y_j^A &\in (-\infty, \infty) \subset \mathbb{R} \\ s_j^A, s^L, s^M &\in [0, 100] \subset \mathbb{R} \\ m &\in \left[\frac{\beta}{100}, 1\right] \subset \mathbb{R} \\ n_j^A &\in [0, 1] \subset \mathbb{Z} \end{aligned}$$

For all points, the following constraints are set, where a_j^A is the “truth” value from the optimal sky cover analysis, as also used in the primary model (but now with more points), s_j^A is a model variable for the updated sky cover amount, and y_j^A is the model variable representing the MAE, similar to the primary model.

$$\begin{aligned} a_j^A - s_j^A &\leq y_j^A \\ s_j^A - a_j^A &\leq y_j^A \end{aligned}$$

In addition, s_j^O is the s_j value from the primary model and n_j^A is a binary integer model variable for controlling the decision structure discussed in this subsection.

$$s_j^A \geq s_j^O - 100n_j^A$$

For any j point where there is no condensate in the column, and absolute vorticity is non-positive, the constraint $s_j^A \leq s_j^O$ must hold such that the adjusted value at point j cannot exceed the value s_j^O from the primary model, and the constraint $s_j^A \leq 100(1 - n_j^A)$ must hold such that the model variable n_j^A controls the adjusted value model variable s_j^A . If $n_j^A = 1$, then $s_j^A = 0$. The initial run of the secondary model uses only the output from the primary model. Individual fields are not examined until the second and subsequent runs. Collectively, for the initial run, these two constraints force original point values s_j^O greater than a threshold s^L to assume a value of $n_j^A = 0$. This drives the subsequent decision structure involving n_j^A .

$$\begin{aligned} s_j^O n_j^A &\leq s^L \\ s^L - s_j^O &\leq 100n_j^A \end{aligned}$$

Subsequent runs examine relative humidity at the levels specified in the primary model, where λ_j^O is the generic original relative humidity variable value at a level from the HRRR model at point j . The constraints for those tests are formulated as $\lambda_j^O n_j^A \leq s^L$ and $s^L - \lambda_j^O \leq 100n_j^A$.

For points where cloud condensate and/or positive absolute vorticity exist, the following constraints apply a coefficient adjustment m to the original value s_j^O when s_j^O is less than the threshold s^M . This is to ensure integrity of the placement of model cloud condensate.

$$s_j^A \leq s_j^O + 100n_j^A$$

$$\begin{aligned} s_j^A &\leq 100(1 - n_j^A) + ms_j^O \\ s_j^A &\geq ms_j^O - 100(1 - n_j^A) \end{aligned}$$

For the initial run containing such condensate points, the following constraints apply.

They are similar to those previous with the exception that a different threshold, s^M , is used.

$$\begin{aligned} s_j^O n_j^A &\leq s^M \\ s^M - s_j^O &\leq 100n_j^A \end{aligned}$$

Otherwise, for subsequent runs using point values λ_j^O from a relative humidity field for a given level, the constraints are as follows for condensate and positive absolute vorticity points.

$$\begin{aligned} \lambda_j^O n_j^A &\leq s^M \\ s^M - \lambda_j^O &\leq 100n_j^A \end{aligned}$$

Finally, the constraint $\gamma - \beta \leq w \leq \gamma + \beta$, where $\beta = 5$ (%), controls the mean w where $N^A w = \sum_j s_j^A$ and N^A is the number of j points contributing to the sum. There are also constraints to assure the integrity of the values at the extremes of the range. This promotes the integrity of areas that are completely clear or nearly clear, and completely cloudy or nearly cloudy. The constraints to maintain the mean of the points with “truth” values in the lowest 10% coverage range are as follows, where $(s_j^A)^M \in (S^A)^M$, $(S^A)^M$ contains only and all $(s_j^A)^M$ which have a corresponding $a_j^A \leq \beta$ for a given j , and $(N_S^A)^M$ is the number of objects in $(S^A)^M$.

$$\begin{aligned} 0 &\leq (W^A)^M \leq 2\beta \\ (N_S^A)^M (W^A)^M &= \sum_j (s_j^A)^M \end{aligned}$$

Similarly, where $(s_j^A)^X \in (S^A)^X$, $(S^A)^X$ contains only and all $(s_j^A)^X$ which have a corresponding $a_j^A \geq 100 - \beta$ for a given j , and N_S^X is the number of objects in $(S^A)^X$, the following constraints apply.

$$100 - 2\beta \leq (W^A)^X \leq 100$$

$$(N_S^A)^X (W^A)^X = \sum_j (s_j^A)^X$$

Lastly, where N^A , as aforementioned, is the number of objects in Y^A , and every and only $y_j^A \in Y^A$, the objective minimizes the MAE:

$$\min \frac{1}{N^A} \sum_j y_j^A.$$

As formulated, this model attempts to characterize and refine the column cloudiness based on the sky cover, or relative humidity field at a single level. Although the model solution occasionally includes one of the levels, run-to-run consistency is not evident. In general, for approximately 54% of the runs during the demonstration, discussed subsequently, the optimizer running the secondary model could not find a level that improved on the MAE better than an adjustment based on thresholds applied to the primary model sky cover output itself. For the cases in which a relative humidity level decreased the MAE, that level most commonly fell in the middle or upper troposphere. These results are summarized in Table 2. This suggests up to three possible conclusions. First, it may not be possible to represent the relationship between sky cover and relative humidity as a threshold in this manner. Second, assuming some relationship between relative humidity and sky cover, it is possible that middle and upper tropospheric cloudiness is a larger contributor to sky cover than lower tropospheric cloudiness. Third, it is possible that cloudiness in the middle and upper troposphere is not strongly correlated to lower tropospheric relative humidity. While it is possible that the first two potential conclusions can be attributed to the aforementioned assumptions and the design of the secondary model, the third conclusion is examined subsequently in the summarized results of the primary model solutions.

5. Cases

Three cases are presented that demonstrate the performance of the optimization process and the three stages of optimization. The first case is a synoptic pattern similar to one characteristic of summer, with a large field of cumulus clouds and convective processes, some deep. The second case is a large mid-latitude cyclone over the contiguous United States. Finally, the third case is a relatively clear period. The adaptive nature of the algorithm allows for a different but ideal result in each of these cases. However, some cloud features that exist on climatic or seasonal scales, such as marine stratocumulus off the West Coast of the United States, will always play into the model logic for assuring the best formulation.

a. 18 UTC on 4 October 2013

The first case is 18 UTC on 4 October 2013. This case demonstrates the skill of the technique for a case with a mid-latitude cyclone, a tropical cyclone, and a large diurnal cumulus cloud field. On this day, there is surface high pressure across the southeastern United States with tropical storm Karen south of New Orleans, Louisiana. A surface stationary front from New York City, New York, to Chicago, Illinois, to Dodge City, Kansas, separates an above average warm and moist day for early October in the Ohio Valley from cooler, seasonal weather to the north. A low pressure system is strengthening over western Kansas. The visible satellite image in Figure 14 indicates the approximate type and coverage of clouds over various areas in the contiguous United States. Scattered clouds are evident in the southeastern United States with deep convection associated with tropical storm Karen over the Gulf of Mexico. The Dakotas and Great Lakes are overcast. There is also low cloud apparent in the Willamette Valley of western Oregon.

The blended sky cover analysis is shown in Figure 15, depicting the overcast skies across the northern United States and partial cloudiness over the southeastern United States. The NDFD total cloud cover one-hour forecast in Figure 16 is similar to blended sky cover analysis with some differences in the extent of overcast skies. For example, the magnitude of the sky cover resulting from the low cloud in the Willamette Valley is much less than in the blended analysis. The blended analysis also suggests more substantial sky cover across portions of the Lower Mississippi Valley. The NDFD one-hour forecast contains mostly clear skies across New Mexico and Nevada that is not as widespread in the blended sky cover analysis.

The initial-hour HRRR total cloud cover analysis is shown in Figure 17. The deep convection in the Gulf of Mexico is resolved as overcast. Sky cover across the cumulus field is patchy. There are portions of the cumulus field across western Arkansas and northeastern Missouri that are output as clear in the HRRR total cloud cover analysis. The cloud cover output for the low cloud in the Willamette Valley is resolved.

The HRRR optimal sky cover product shown in Figure 18 contains a broad area of overcast skies across the Northern Plains, with additional overcast skies related to the convection in the Gulf of Mexico. An enhanced area in the cumulus field runs along Lower Mississippi Valley and Ohio Valley. A patch of low cloud in the Willamette Valley is resolved. There is a gradient from mostly clear skies to mostly cloudy skies along the border between Montana and North Dakota, extending northward into Saskatchewan, Canada. There are clear skies from California extending through Arizona and New Mexico.

The histograms in Figure 19 demonstrate the distribution of certain sky cover increments for the four products previously discussed. There are some evident features. First, comparing Panel C to the others, the HRRR total cloud cover analysis contains significantly more clear and

nearly clear sky coverage areas than the others. This comes mainly at the expense of cloud coverage areas between 10% and 90%, which are fairly evenly distributed in 10% increments. Second, the tri-modal nature of the distribution in Panel D is most similar to Panel B. The blended sky cover analysis histogram in Panel A shows a greater amount of sky coverage values between 70% and 90% than between 40% and 70%.

Figure 20 shows the distribution of difference values for the HRRR total cloud cover analysis and HRRR optimal sky cover product compared to the blended sky cover analysis and NDFD total cloud cover one-hour forecast. Panel A is the HRRR total cloud cover analysis less the blended sky cover analysis. Panel B is the HRRR optimal sky cover product less the blended sky cover analysis. Both have negative bias. The mean error for the HRRR optimal sky cover product compared to the blended sky cover analysis is -8.3% sky cover. This compares to a mean error of -17.4% sky cover for the HRRR total cloud cover analysis compared to the same analysis. The improvement in mean error for the HRRR optimal sky cover product is 9.1% compared to blended sky cover analysis. Panel C is the HRRR total cloud cover analysis less the NDFD total cloud cover one-hour forecast, and Panel D is the HRRR optimal sky cover product less the NDFD total cloud cover one-hour forecast. The mean error for the HRRR optimal sky cover product compared to the NDFD total cloud cover one-hour forecast is -4.1% sky cover. This is better than a mean error of -13.3% sky cover for the HRRR total cloud cover analysis compared to the same analysis. The improvement in the mean error for the HRRR optimal sky cover product is 9.2% compared to the NDFD total cloud cover one-hour forecast.

The geographical distribution of the difference between the HRRR optimal sky cover product and the blended analysis is shown in Figure 21. The areas of best agreement are those that are either completely clear or completely cloudy. Results are mixed in the diurnal cumulus

field area. Significant areas of mean error include southeastern Utah and northern Maine. In these areas, the blended sky cover analysis value is much higher than the HRRR optimal sky cover product output. An investigation into these areas finds that there is little to no cloud water condensate and relative humidity values from the correlated fields of 200 hPa, 850 hPa, 900 hPa, 950 hPa, and 1000 hPa, over 90% relative humidity, are generally low compared to the rest of the grid. For southeastern Utah, there is non-zero 700 hPa cloud water mixing ratio. However, the optimizer running the primary model did not strongly correlate the quantity with sky cover. As a result, cloudiness is underrepresented there. For northern Maine, the optimizer did not have any evidence of cloud condensate indicating a potential fault in the assimilation and pre-execution stage of the HRRR model run providing the input for the optimization model.

In comparison, the difference of the NDFD total cloud cover one-hour forecast from the HRRR optimal sky cover product, depicted in Figure 22, demonstrates similar issues. The HRRR optimal sky cover product indicates a particularly high bias compared to the blended sky cover analysis over Montana and a large portion of Idaho. This figure also indicates that the NDFD one-hour forecast does not delineate the region of more significant cloudiness over the Lower Mississippi Valley and the Ohio Valley, with evidence of a high bias in that general corridor. There is a slight overestimate, compared to the NDFD one-hour forecast, of the low cloud in the Willamette Valley, with an underestimate surrounding it. There is also an underestimate of sky cover over New Mexico. In some cases of underestimated sky cover, the fault for the difference lies more with the NDFD one-hour forecast than the optimal sky cover product. This example demonstrates that forecasters may not currently have sufficient skill or dedicate enough attention to delineate between denser fields of cumulus clouds despite potential model skill in this area.

b. 3 UTC on 19 October 2013

The next case is from 3 UTC on 19 October 2013. This case demonstrates the skill of the technique for a situation with a large area of clear skies, a large area of non-overcast skies, and a stratocumulus deck off the California coast. The synoptic pattern across the contiguous United States consists of surface high pressure across the western and central United States with three weak low pressure systems in succession from over the Great Lakes to the Deep South. The first system is moving from Manitoba, Canada, into Minnesota. The second system extends from a weak area of low pressure over Lake Huron along a cold front to a secondary low over Arkansas, and the front continuing to extend into the Gulf of Mexico near Galveston, Texas. The third system is a weak depression over southeastern Georgia. The geographical distribution of cloudiness across the contiguous United States is shown through the infrared window satellite image in Figure 23. The satellite image reveals patchy low cloud across the Northern Plains, more widespread mid-level and low cloud across the Middle and Lower Mississippi Valley, and some convective clouds in the vicinity of New Orleans, Louisiana. There is also a small patch of cloud on the leeside of the ridge over central Colorado.

The blended sky cover analysis appears in Figure 24. There are a number of local minima in and surrounding the state of Minnesota as a result of variations in the sky cover reports from the dense surface observation network in that state. One contiguous overcast cloud area is in central Missouri with another in the Dakotas. This is in contrast to the NDFD total cloud cover one-hour forecast, as shown in Figure 25, which depicts relatively few overcast areas. The NDFD one-hour forecast also lacks any definition to the sky cover field over Colorado, with generally inflated sky cover amounts further south across eastern New Mexico, compared to the blended analysis and subjective analysis of the infrared window imagery.

The HRRR total cloud cover analysis shown in Figure 26 indicates patches of clear sky within the cloud deck over the Northern Plains, and similarly over the Deep South. However, these patches do not correspond to areas that are void of cloud as evident using the blended sky cover analysis. Areas of overcast cloud are likewise patchy, with the exception of the marine stratocumulus deck off the coast of California.

The HRRR optimal sky cover product shown in Figure 27 differs from the HRRR total cloud cover analysis in that there is a more expansive area of overcast clouds from Missouri into central Texas, and also along the Gulf Coast into the Carolinas. The clear area in the northeastern United States is well defined with a complex sky cover regime over the Northern Plains. The magnitude of sky cover resulting from the stratocumulus deck appears appropriate as well, with a large clear region over the western United States.

The histograms shown in Figure 28 depict a diversity of distributions and balance between clear and overcast areas. The distribution of values in the histogram of the blended sky cover analysis that is shown in Panel A reveals a large percentage of clear and nearly clear areas. Panel A also indicates a relatively even distribution of sky cover amounts between 10% and 70%. The distribution of Panel B for the NDFD total cloud cover one-hour forecast reveals a slight increase in the number of sky cover amounts between 60% and 90%. Panel D differs from Panel C in that there is a marked decrease in the number of clear and nearly clear values in the initial-hour HRRR optimal sky cover product compared to the HRRR total cloud cover analysis, and a marked increase in the overcast and nearly overcast increment for sky cover amounts between 90% and 100%. There is a decrease in the number of sky cover amounts within 10% increments between 10% and 90% for the HRRR optimal sky cover product. There is little change over this range for the HRRR total cloud cover analysis.

The HRRR total cloud cover analysis and HRRR optimal sky cover product are verified against the blended sky cover analysis and NDFD total cloud cover one-hour forecast, and the histograms of these difference comparisons is shown in Figure 29. All four panels show a negative bias, but Panel B, the difference of the blended sky cover analysis from the HRRR optimal sky cover product shows a particularly good relationship compared to Panel A, the difference of the blended sky cover analysis from the HRRR total cloud cover analysis. The mean error for the HRRR total cloud cover analysis is -13.2% sky cover, whereas the mean error for the initial-hour HRRR optimal sky cover product is -6.3% sky cover, when both are compared to the blended analysis. The same relationship is evident when comparing the HRRR total cloud cover analysis and HRRR optimal sky cover product to the NDFD total cloud cover one-hour forecast, shown in Panels C and D, respectively. The mean error for the HRRR total cloud cover analysis is -13.7% sky cover, whereas the mean error for the HRRR optimal sky cover product is -6.8% sky cover, when both are compared to the NDFD one-hour forecast. When verifying the HRRR total cloud cover analysis and HRRR optimal sky cover product against either the blended analysis or the NDFD one-hour forecast, the HRRR optimal sky cover product improves the mean error by 6.9% sky cover over the HRRR total cloud cover analysis.

The geographical difference of the blended sky cover analysis from the HRRR optimal sky cover product is depicted in Figure 30. Areas of the most significant disagreement extend from central Montana through Iowa. There is also an overestimate in Colorado. In both cases, the significant difference is the result of lacking cloud water mixing ratio at the levels provided to the optimizer, and competing relationships between relative humidity and sky cover. The optimizer found relationships between relative humidity and sky cover at 300 hPa, 500 hPa, 900 hPa, 950 hPa, and 1000 hPa over 88% relative humidity. The relationship between 950 hPa

relative humidity and sky cover is the strongest. In this case, there is particularly high relative humidity at 950 hPa in eastern Colorado, and relative minima in relative humidity at that level over southern Iowa, southwestern South Dakota, and much of Montana, largely responsible for the inflated sky cover and insufficient sky cover, respectively. This case indicates that despite the magnitude of the relative humidity, the level of cloudiness can vary considerably. It also demonstrates a pitfall with attributing cloud/sky cover solely to relative humidity.

Interestingly, the comparison of the HRRR optimal sky cover product with the NDFD total cloud cover one-hour forecast, as shown in Figure 31, reveals some of the same issues as seen with the aforementioned comparison to the blended sky cover analysis. This confirms that the blended sky cover analysis is an adequate validating analysis for the NDFD total cloud cover one-hour forecast. The blended sky cover analysis is also ideal because it lacks political boundaries resulting from different forecasts between NWS office areas of responsibility. It also has the potential to adequately represent the conditions without unintentional human error such as failing to update the forecast grid routinely. For example, the underestimate in sky cover over New Mexico is likely false considering the blended sky cover analysis.

c. 12 UTC on 5 November 2013

The last case is 12 UTC on 5 November 2013. This case demonstrates the skill of the optimization method for a situation with a limited degree of clear skies and a wide range of cloud types and sky cover amounts evident on satellite imagery. On this day, an occluded low pressure system is south of Hudson Bay with developing low pressure in southeastern Colorado. High pressure at the surface and ridging aloft extends across the eastern United States. There is another low pressure system approaching the Pacific Northwest. The geographical distribution

of cloudiness across the contiguous United States is shown through the infrared window satellite image in Figure 32. A large amount of thick cloud is apparent across the central Plains, with additional cloudiness across the Lower Mississippi Valley extending into the Ohio Valley. There is also expansive partial cloudiness over the Atlantic Ocean.

The blended sky cover analysis in Figure 33 shows a wide range of sky cover conditions. Overcast skies are evident in the Central Plains and Pacific Northwest. There are also pockets of lesser sky cover across parts of Minnesota and the Northern Plains. The NDFD total cloud cover one-hour forecast shown in Figure 34 depicts the range in cloudiness that NWS meteorologists predicted. The general pattern is similar to the blended sky cover analysis, but the details are much different. For example, there are differences in forecaster opinion between NWS office areas of responsibility that lead to gradients of sky cover along political boundaries. This is most evident on the border between Alabama and Mississippi, but also is apparent in Lake Michigan.

The current HRRR total cloud cover analysis for sky cover is not indicative of the sky condition at all portions of the domain. For example, the HRRR total cloud cover analysis, shown in Figure 35, does not suggest any cloudiness over portions of Appalachia, where it is evident from satellite imagery and the blended analysis that cloudiness is present. In addition, there are a large number of relative maxima and minima, sometimes in close proximity, making for unnatural sky cover gradients.

The initial-hour HRRR optimal sky cover product is shown in Figure 36. A large area of overcast skies is captured across the Central Plains and Pacific Northwest. Clear skies are evident offshore from California, with adequate partial cloudiness over the Atlantic Ocean. However, there is also cloudiness in areas that are evidently clear on satellite imagery and the blended sky cover analysis as depicted in Figure 33. This is most evident over interior Mexico.

Figure 37 is a four-panel plot showing histograms of the frequency of values on the respective domains plotted in Figure 33, Figure 34, Figure 35, and Figure 36. Panel A of Figure 37 shows the blended sky cover analysis distribution, with 39% of the domain covered by sky cover amounts of 90% or greater, with clear or nearly clear skies representing only 11% of the domain. Sky cover amounts greater than or equal to 70% but less than 90% collectively represent 25% of the points on the domain, whereas other amounts between 10% and 70% are approximately evenly distributed. This is in contrast to Panel B, which shows the NDFD total cloud cover one-hour forecast distribution. For this distribution, the percentage of partly cloudy and mostly cloudy skies between 40% and 90% coverage are approximately evenly distributed, with a more significant number of nearly overcast or overcast skies with sky cover amounts of at least 90%. Panel C confirms the distribution of the HRRR total cloud cover analysis that is suspected from Figure 35, with nearly even distribution between 10% and 90% coverage, and approximately balanced clear or nearly clear skies and overcast or nearly overcast skies. Lastly, Panel D depicts the distribution of the HRRR optimal sky cover product with a substantially less amount of clear sky cover amounts compared to the HRRR total cloud cover analysis in Panel C, and with a distribution closer to Panels A and B, despite a somewhat bimodal appearance.

Figure 38 is a four-panel plot showing histograms of the difference of values between two grids where corresponding, non-missing values from two products are matched. Panel A shows the distribution of values for the difference of the blended sky cover analysis from the HRRR total cloud cover analysis. Panel B shows the distribution of the values for the difference of the blended sky cover analysis from the initial-hour HRRR optimal sky cover product. Comparing Panels A and B, a much smaller bias is noted for the initial-hour HRRR optimal sky cover product as opposed to HRRR total cloud cover analysis. The mean error for the HRRR

total cloud cover analysis is -21.5% sky cover, compared to -5.4% sky cover for the HRRR optimal sky cover product, when both are validated against the blended analysis. The same bias is noted when comparing the HRRR total cloud cover analysis and the HRRR optimal sky cover product to the NDFD total cloud cover one-hour forecast, as the difference histograms indicate in Panels C and D. The mean error for the HRRR total cloud cover analysis is -16.3% sky cover, compared to -0.2% sky cover for the HRRR optimal sky cover product, when both are compared to the NDFD one-hour forecast. In both validations, the mean error of the HRRR total cloud cover analysis is 16.1% sky cover less than the HRRR optimal sky cover product in this case.

Figure 39 shows the geographical areas where there is a difference between the HRRR optimal sky cover product and blended sky cover analysis. The best performance is noted in clear and cloudy areas. The worst performance occurs in areas where the cloud distribution is likely multi-layered, or the geographic distribution of cloud condensate is not adequately captured in the HRRR model during the pre-processing and analysis steps⁴. One area of particularly poor performance, compared to the blended sky cover analysis, is in eastern Tennessee and western North Carolina. In this case, the significant underestimation is the result of low relative humidity values over that geographic area within the fields chosen for the correlation. The optimizer identified the 500 hPa, 900 hPa, and 950 hPa relative humidity fields as contributing to sky cover. The optimizer also found that 1000 hPa relative humidity values above 94% contribute to the HRRR optimal sky cover product in this case. High relative humidity values likely supported an overestimate of cloudiness over Nevada.

The comparison of the HRRR optimal sky cover product to the NDFD total cloud cover one-hour forecast, shown in Figure 40, demonstrates approximately the same issues. Like in

⁴ The initial analysis is typically recycled from the previous one-hour forecast (Benjamin 2013).

comparison to the blended sky cover analysis, overcast and clear skies are the areas where there is little difference between the HRRR optimal sky cover product and the NDFD total cloud cover one-hour forecast. Furthermore, the HRRR optimal sky cover product is low in bias over Appalachia. However, the difference is not as substantial over Nevada. Over Wisconsin and Minnesota, there is both positive and negative bias. Part of the cause for this bias, as well as the spatial variations, is likely a poor one-hour forecast or overall lack of predictability.

6. Results

In this section, results from all of the completed primary and secondary model optimization runs between 1 UTC on 21 September 2013 and 23 UTC on 1 November 2013 are compiled and interpreted. Results are arranged in three ways. First, the characteristics of relationships of the tested quantities to sky cover throughout the test period are assessed. Second, the behavior of quantities relative to others on individual levels is investigated. Third, scores are compared for the HRRR optimal sky cover product and the legacy HRRR total cloud cover analysis when verified against the blended sky cover analysis and NDFD total cloud cover one-hour forecast.

One strong benefit of the optimization approach, particularly with respect to the primary model, is that the output not only provides the best solution as sought through the objective function, but the solution of the variables related to the quantities provide insights. Though it may seem fairly direct that cloud water mixing ratio and other condensates correlate strongly to cloud, the optimization procedure tests that assumption in part to validate the model analysis. In past studies, indicators of cloud fraction were evaluated in a test atmosphere, or with data collected during the course of a field campaign. But this discounts seasonal impacts and en masse application as the cloud scale transitions from sub-mesoscale to the mesoscale and

synoptic scale. The optimization approach provides insight into which cloud-related fields have the best predictability, or at least which atmospheric quantity and level most closely match cloudy points. In general, areas where a tested or incorporated field has a high value outside of a cloud deck are going to minimize the impact of that field on the solution. The primary model generally treats each level of each quantity independently.

Not all primary model formulations provided to the optimizer are solved in the 30-minute time allotted. In addition, some models are unable to run because of a lack of input. Usually the lack of input is the result of no HRRR model output. The optimization models are run in real-time beginning approximately two hours after the analysis time, usually as soon as the output from the corresponding HRRR model run is available. It is important to not only look at how the contributing fields to the final solution are scaled, but also how frequently the field is used as a partial indicator of sky cover throughout multiple runs.

There are two predominant types of correlations that are referenced. A scalar correlation is one where for any non-zero value of the quantity, a non-zero scalar adjustment is provided to increase the overall value of sky cover at that point regardless of the magnitude of the non-zero field value it is applied to. A coefficient correlation adjusts the distribution of the sky cover based on the distribution of the field to which the coefficient is applied. A coefficient correlation adjusts spatial gradients.

The quantities tested in this model are all expected to correlate with sky cover based on previous literature and implicit relationships. This is necessary because the use of optimization models to determine relationships without an expectation of one can yield results without significant depth. Particularly when it comes to optimization models with a large number of variables, the increase in variables also increases the number of potential solutions within the set.

It is possible to foresee scenarios in which random patterning may yield an acceptable solution if enough such spatial patterns are provided.

Finally, there is a seasonal dependence of the correlations based on the formulation of the primary model. This is demonstrated in the test period, where there is a seasonal transition from summer to fall. In the summer months, cloud ice mixing ratio and snow mixing ratio are generally associated with areas of deep convection, which are difficult to predict with numerical models, and such quantities exist only in the upper troposphere. In the winter months of the mid-latitudes, clouds more frequently consist of ice, as snow becomes a dominant precipitation type, particularly over the northern half of the United States, and cloud ice mixing ratio and snow mixing ratio are more prevalent throughout the troposphere.

a. Relative humidity

Examining all of the runs within the period, coefficients to the relative humidity fields are the most frequent output from the end-to-end optimization procedure. The mean value of the coefficient responsible for the correlation is highest at 300 hPa, excluding 1000 hPa, which has the added advantage of a threshold. The mean value of the relative humidity field coefficient at each level is shown in Table 3. In addition, correlation is most frequent at 300 hPa, with approximately 59% of the runs producing a non-zero coefficient. Combining 200 hPa and 300 hPa, a non-zero coefficient is produced in approximately 77% of the runs. Looking at the related 100% less relative humidity quantity, an indication of the degree of dryness, a 300 hPa coefficient is also most common. This suggests that when there is cloudiness in this layer, relative humidity is correlated with it. In areas where relative humidity values are small, the optimizer running the primary model decreases the degree of cloudiness.

Table 4 summarizes the frequency of at least one coefficient at each level, applied to either the relative humidity field or the 100% less relative humidity field. The most reliable correlation, other than at 1000 hPa, is at 300 hPa, where at least one correlation occurs approximately 78% of the time, and both occur approximately 25% of the time. The least common correlation is with 850 hPa, where a relative humidity coefficient or 100% less relative humidity coefficient is found in only 33% of the runs. For 1000 hPa, all (100%) of the runs produce a non-zero relative humidity coefficient, non-zero 100% less relative humidity coefficient, or both, for values greater than the threshold.

b. Cloud water mixing ratio

Cloud water mixing ratio in the lower troposphere is also routinely correlated with sky cover where it exists. Considering the runs during the demonstration, approximately 96% of the runs contained at least one non-zero coefficient for correlation to 800 hPa, 850 hPa, and 900 hPa cloud water mixing ratio. The optimizer running the primary model produced a scalar adjustment in approximately the same percentage of runs, considering 800 hPa, 850 hPa, and 900 hPa collectively. This indicates that the spatial distribution of cloud water mixing ratio from the HRRR model is at least somewhat similar to sky cover. This information is summarized in Table 5, which also indicates the mean value of the coefficient for cloud water mixing ratio at the specified levels incorporated into the primary model.

The optimizer running the primary model found a coefficient or scalar correlation for the 950 hPa cloud water mixing ratio field in approximately 97% of runs, and a coefficient or scalar correlation for the 900 hPa cloud water mixing ratio field in approximately 94% of runs. In approximately two-thirds of the runs in the demonstration, the optimizer found a coefficient and

scalar adjustment for 950 hPa. In fact, a coefficient or scalar correlation exists for over 80% of runs at 700 hPa, 800 hPa, and 850 hPa individually. This information is summarized in Table 6.

c. Cloud ice mixing ratio

In contrast to cloud water mixing ratio, there is a lesser degree and consistency of correlation for the cloud ice mixing ratio quantity. Mean values for the cloud ice mixing ratio coefficient are highest at 500 hPa, though there are only 30 such cases of a non-zero coefficient at that level, as shown in Table 7. At 300 hPa, there are a much greater number of runs with a non-zero coefficient correlation for cloud ice mixing ratio, but the mean value is approximately an order of magnitude smaller. Correlation between cloud ice mixing ratio and sky cover is most significantly noted at 300 hPa, not only judging the percentage of cases with a non-zero coefficient, but also the percentage of cases with a non-zero scalar. This information is summarized in Table 8. A coefficient or scalar is produced for non-zero 300 hPa cloud ice mixing ratio amounts in approximately 58% of all runs.

d. Rain water mixing ratio

Rain water mixing ratio is the variable least correlated with sky cover. The mean value for the coefficient, as indicated in Table 9, does not show strong dependence to a level for those levels with a sufficient number of runs contributing to the mean. For an individual level, 700 hPa produced the greatest number of coefficient and scalar correlations. However, one of the two only appeared at this level in 24% of model runs. This information is summarized in Table 10. It is likely that the distribution and magnitude of rain water mixing ratio made it less ideal for correlation compared to other variables such as cloud water mixing ratio. Areas with positive rain water mixing ratio are also areas that are most likely cloudy.

e. Snow mixing ratio

The mean value for the snow mixing ratio coefficient at individual levels is approximately on par with values of rain water mixing ratio, as listed in Table 11, though the coefficient is somewhat smaller in magnitude when applied to the 300 hPa snow mixing ratio field values. This suggests either a higher snow mixing ratio value at 300 hPa compared to the other levels, or more correlated levels or other quantities within the column to adjust the cloudiness when snow mixing ratio is present. Snow mixing ratio as a quantity for correlation to sky cover did best in the middle troposphere, with approximately 45% of runs supporting a coefficient correlation at 500 hPa, and 59% of runs supporting a scalar correlation. This information is summarized in Table 12. Based on the demonstration period early in the fall months, the lack of correlation with sky cover in the lowest levels is not surprising.

f. 200 hPa

Figure 41 demonstrates the trend in the mean value of the coefficient applied to the relative humidity, condensate, and vorticity variables at 200 hPa. There is little overall trend with time throughout the period. Cloud water mixing ratio and rain water mixing ratio are missing because there is no correlation with these variables due to the lack of precipitating liquid cloud high in the troposphere in the HRRR model runs. Figure 42 shows that the percentage of runs where the optimizer applied a non-zero coefficient to each field during the demonstration. Positive absolute vorticity is the most frequently correlated at 200 hPa, with a coefficient found in approximately 73% of runs, compared to a negative absolute vorticity coefficient produced in 54% of runs. Snow mixing ratio and cloud ice mixing ratio are the least correlated. In the case of the latter two quantities, the limited correlation is likely due to the limited spatial coverage.

The mean value of the non-zero positive absolute vorticity coefficient at 200 hPa is 7.77×10^4 %•s, while the mean value of the non-zero absolute negative vorticity coefficient at 200 hPa is 2.60×10^7 %•s. Because the negative vorticity coefficient is applied to a negative value, the coefficient works to decrease the sky cover amount. The higher mean value of the absolute negative vorticity coefficient, by three orders of magnitude, is due to the smaller absolute mean value of negative absolute vorticity.

g. 300 hPa

The trend over time of relative humidity and condensate variables at 300 hPa is shown in Figure 43. In general, the mean value of the non-zero coefficients is relatively steady over time. However, the percent of time over the demonstration in which they are non-zero is not as steady, as shown in Figure 44. The most significant contributor to sky cover, at this level, is cloud ice mixing ratio. The mean contribution of the cloud ice mixing ratio coefficient to sky cover at a point is approximately 19%. This compares to the relative humidity coefficient and snow mixing ratio coefficient, each which contribute 7% on average to the sky cover at a point where the respective field value is non-zero.

h. 500 hPa

The trend over time of coefficients corresponding to relative humidity and condensate quantities at 500 hPa is shown in Figure 45. In general, there is no discernable trend. Some of the local trends on the plot are the result of a limited sample size. Examining Figure 46, it is evident that early in the period, the optimizer found non-zero coefficients of cloud water mixing ratio and snow mixing ratio at approximately the same frequency. Late in October, the frequency of a coefficient correlation with cloud water mixing ratio decreases, likely as a result

of cooler seasonal temperatures. The distribution of cloud ice mixing ratio is not frequently correlated, likely due to the limited amount of cloud ice at 500 hPa. In general, the lack of any field at 500 hPa significantly correlated with sky cover well over 50% indicates the lack of substantial skill in correlating with cloud in this layer.

i. 700 hPa

Figure 47 shows the steady trend of solved relative humidity and condensate coefficients correlated with sky cover at 700 hPa. The degree of the contribution to sky cover, when the coefficient is non-zero, is about the same for cloud water mixing ratio, rain water mixing ratio, and snow mixing ratio at this level. Based on the distribution of cloud water mixing ratio relative to sky cover, cloud water mixing ratio is the most significant contributor at 700 hPa, particularly late in the demonstration, as the trend in Figure 48 shows. Cloud water mixing ratio contributes to the sky cover formulation with a regularity of better than 50% of runs, and the distribution of rain water mixing ratio contributes to sky cover in approximately 15% of runs.

j. 800 hPa through 950 hPa

Similar to 700 hPa, cloud water mixing ratio is the most dominant contributor for the lower tropospheric levels between 800 hPa and 950 hPa, with limited contribution from snow mixing ratio and cloud ice mixing ratio. The steady trend of the mean non-zero coefficient value is shown in Figure 49 for 800 hPa, Figure 51 for 850 hPa, Figure 53 for 900 hPa, and Figure 55 for 950 hPa. The mean value cloud water mixing ratio is lowest overall at 850 hPa despite the lack of other routine predictors. At 800 hPa and 850 hPa, the most frequent contributor to sky cover is cloud water mixing ratio, as indicated in Figure 50 and Figure 52, respectively, though the frequency of the contribution is somewhat variable at between approximately 50% and 80%

of the runs. At 900 hPa and 950 hPa, there is a marked increase in the frequency of the contribution of cloud water mixing ratio to sky cover, as indicated in Figure 54 and Figure 56, respectively. This suggests that low, liquid cloud is amongst the most predictable fields using the HRRR model. Another conclusion evident in the data is that the 950 hPa cloud water mixing ratio is the strongest routine contributor to sky cover across all variables at all levels.

There is also a contribution to sky cover from relative humidity at 950 hPa at a greater value than the other lower levels. At 950 hPa, relative humidity factors into the sky cover formulation for greater than 30% of runs in the demonstration. Despite that, the mean value of the contribution is not notably more substantial than at the other levels.

k. 1000 hPa

Unlike at the other levels, at 1000 hPa, the primary optimization model allows a coefficient and scalar to correlate to relative humidity above a threshold. The primary optimization model allows for a range of relative humidity values between 0% and the threshold for which there is no contribution to the final sky cover. The purpose of the implementation is to allow for sharp gradients in 1000 hPa relative humidity to correlate with cloud edges. There are some regimes where this is useful, particularly in marine environments, which provide perpetual challenges due to the typically dry troposphere above a relatively shallow cloud deck, and against terrain boundaries.

Figure 57 shows the trend in the mean value of the non-zero coefficient when correlated to relative humidity, cloud water mixing ratio, and rain water mixing ratio. Compared to 950 hPa, there is no significant change to the mean value or the trend of the optimizer solution for condensate variables throughout the demonstration. However, the mean value of the relative

humidity coefficient is higher than evident at other levels. This coefficient is only applied above a certain threshold, and Figure 58 shows the trend in the frequency of a non-zero coefficient over all runs for relative humidity, cloud water mixing ratio, and rain water mixing ratio during the demonstration. The trend in the mean threshold is depicted in Figure 59. On average, the threshold is approximately 70% or greater; the mean value is 73.3%.

1. Overall performance

Metrics for judging performance of this approach include mean error, MAE, and root-mean-square error (RMSE). Validated against the optimal sky cover analysis over the continental United States, the adaptive optimizer solution, produced from the primary model and secondary model, is a consistent outperformer of the current operational cloud cover output from the HRRR model. However, it is evident that the degree and trend of the error over time is similar. This suggests that part of the reason for the error is in fact a deficiency in the skill of the NWP model itself. That is, the placement of the atmospheric parameters necessary for diagnosing cloud is not always accurate.

That given, the MAE and mean error are closer to zero for the HRRR optimal sky cover product compared to the routine HRRR total cloud cover analysis. For the initial-hour HRRR optimal sky cover product, the mean error over 793 runs within the aforementioned test window, but beginning at 0 UTC, using the blended sky cover analysis as truth, is -7.6% sky cover, MAE is 17.5% sky cover, and RMSE is 26.5% sky cover. This compares to the HRRR total cloud cover analysis, where the mean error over 823 runs within the same test window, using the blended sky cover analysis as truth, is -13.1% sky cover, MAE is 17.3% sky cover, and RMSE is 27.2% sky cover. The difference in the number of compared runs is a result of data availability.

While the mean error for the HRRR optimal sky cover is markedly less, the other metrics do not indicate substantial improvement. This suggests that both the HRRR optimal sky cover product and the legacy total cloud cover analysis suffer from significant differences for some points, despite the HRRR optimal sky cover product slightly more balanced in bias for some of those differences. In addition, much of the competitive performance of the HRRR total cloud cover analysis is likely due to the expansive clear area.

Examining the three-hour forecast of the HRRR total cloud cover quantity and the HRRR optimal sky cover product, a different result is realized. For the three-hour forecast of the HRRR optimal sky cover product, the mean error over 784 runs where the output is valid within the same test window beginning at 0 UTC, using the blended sky cover analysis as truth, is -9.6% sky cover, MAE is 20.3% sky cover, and RMSE is 30.0% sky cover. This compares to the HRRR total cloud cover three-hour forecast, where the mean error over 810 runs and the output is valid within the same test window, using the blended sky cover analysis as truth, is -12.4% sky cover, MAE is 23.9% sky cover, and RMSE is 35.7% sky cover. Similar results are evident in comparing the six-hour and nine-hour forecasts. This suggests that predictability increases using the optimization approach in the near-term forecast range. That is, the HRRR model has better skill for sky cover than the current total cloud cover formulation allows it to.

Using the NDFD one-hour forecast as truth, the mean error of the initial-hour HRRR total cloud cover analysis over 281 runs is -11.9% sky cover, MAE is 20.6% sky cover, and RMSE is 28.4% sky cover. The mean error of the initial-hour HRRR optimal sky cover product over 270 runs is -6.4% sky cover, MAE is 16.1% sky cover, and RMSE is 22.4% sky cover. For all metrics, the initial-hour HRRR optimal sky cover product is the best performer. The initial-hour HRRR optimal cloud cover product is much more competitive due to the lack of the significant

clear areas in the NDFD one-hour forecast. In addition, performance is better in part because the input into the optimizer is optimized and therefore adjusted to minimize MAE compared to the NDFD one-hour forecast. However, despite the adequate performance, the skill scores for the optimizer output prove a bit more variable than the routine model output, despite attempts to constrain the optimizer from returning solutions that did not match the nature of the input data.

The probability of detection (POD) metrics for clear skies, less than or equal to 5% coverage, and cloudy skies, greater than or equal to 95% coverage, are also examined in assessing skill. The routine and optimal HRRR analyses are compared to the blended analysis and NDFD one-hour forecast as truth. For both validating analyses, due to the bias of the routine HRRR total cloud cover analysis to underrepresent cloud, the routine output did regularly prove more adequate in compared POD metrics for clear skies. However, considering the POD metrics for cloudy skies, the HRRR optimal sky cover product is generally better at capturing such conditions.

m. Shear and curvature vorticity

The performance of a change to the 200 hPa vorticity terms in the primary optimization model is assessed to further the inquiry on the relationship between vorticity and clouds. The primary model contains two vorticity terms, one term containing 200 hPa positive absolute vorticity values, with zero otherwise, and the other term containing 200 hPa negative absolute vorticity values, with zero otherwise. The primary model contains four vorticity terms after the revision, where two terms are 200 hPa positive relative vorticity, and the other two are 200 hPa negative relative vorticity. At each grid point, relative vorticity is split into a shear vorticity component and a curvature vorticity component, consistent with Bell and Keyser (1993). There

are 19 cases examined between 6 UTC on 29 October 2013 and 6 UTC on 13 November 2013. For these cases, the impact of this change on the MAE is mixed. While there are some cases where the additional vorticity variables prove beneficial in decreasing the MAE, there are also a number of cases where the MAE increased. More importantly, there is no evidence of any consistent correlation between the positive shear and curvature vorticity terms, and sky cover, though a coefficient for negative curvature vorticity appears to occur more frequently in the optimizer output, to decrease cloud cover, than negative shear vorticity, for the limited number of cases.

7. Conclusion

The work conducted here forms the initial basis for understanding sky cover and quantifying it in a routine manner using an approach which balances observations from both terrestrial and space-based observing systems. Beyond formalizing sky cover as a quantity with value in scientific studies, this work examines how the forecast of short-term cloud and moisture quantities in NWP models can improve. It also provides an avenue through which to conduct subsequent validation studies of clouds in models. It is evident that cloud and moisture quantities contribute to the cloud and sky cover output that models provide. Compared to the blended sky cover analysis, the optimizer produces a sky cover analysis and short-term forecasts which are better than what the HRRR model currently outputs using the Xu and Randall (1996) formulation. The added benefit of the optimization approach is the additional information provided about the relative relationships between different HRRR model quantities.

The optimization approach reveals strong correlations between cloud water mixing ratio and sky cover at 950 hPa and 1000 hPa, as well as correlations between relative humidity and

sky cover at almost all tested levels. This result demonstrates the predictability of low cloud but continues to suggest challenges for resolving middle and upper tropospheric cloud. Furthermore, the sky cover approach demonstrated here is a more realistic representation of the cloudiness problem because it provides a more consistent field when individual clouds are on a smaller scale than are resolved using satellites and grid-based numerical models.

One natural extension of this research is to assimilate the optimal sky cover back into the model for the subsequent time step. For a real-time weather model, this may prove difficult on all but the highest-performing supercomputers due to the amount of time required to solve even a basic formulation of an optimization problem. An additional challenge includes assessing where to place the clouds vertically within the model. A significant error in the height placement has the potential to substantially alter the NWP model's water and energy budgets at some grid points, which could prove detrimental as the model executes for subsequent time steps.

There is the potential for several subsequent avenues of research related to the creation of the analysis. One such potential avenue is adding further logic to the blended sky cover analysis for handling low cloud trapped as a result of terrain. Low cloud in river valleys is spatially evident in visible satellite imagery when not obscured with high-level cloudiness. A high-resolution topographical dataset could prove valuable in checking the elevation of surface stations and parent grid cells against adjacent grid cells to determine the likelihood of cloudiness beyond the parent grid cell through the spatial gradients in the dew point depression and wind regime. There are similar applications that may apply to stratiform cloudiness at the top of the marine boundary layer.

It is also possible to apply optimization to solve other problems in the atmospheric sciences where existing theory does not provide a clear path forward, relationships are assumed,

and better relationships between observable quantities are sought. Thunderstorm predictability is one such area. This could ultimately lead to additional blended products and predictions that are a part of an operational meteorologist's toolbox.

Despite consistently weak relationships between the value of relative humidity and the extent of cloud cover, as shown in previous literature and through this work, there is likely no abandoning parameterizations using relative humidity for real-time weather models until computer performance improves even more to allow for better resolution of turbulent processes behind individual clouds. To this end, it is important to see advancements not only in data assimilations and parameterizations, but also to maintain the integrity of our observational platforms. There are exciting new observing capabilities available and coming soon that have the possibility to assist with longstanding challenges. For example, nighttime visible imagery courtesy of the day-night band on the Suomi National Polar-orbiting Partnership (NPP) satellite is providing opportunities for nocturnal cloud detection with reflected sunlight from the moon. The active lidar instrument on the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) satellite has the potential to help with cloud detection and resolving the height of multiple cloud decks in multi-layer cloud regimes.

As future weather satellites increase our clarity of cloudiness, the dexterity of the blended sky cover analysis can only increase. This will provide subsequent opportunities not only to learn about cloud impacts and feedbacks on short-term and climatic time scales, but also help to improve numerical models and their parameterizations. Combined with techniques such as optimization, it is possible to find and test assumptions, and bring confidence to existing relationships that are known to the field. This study is a first step in demonstrating such a concept.

Bibliography

- Arking, A. 1964. "Latitudinal Distribution of Cloud Cover from Tiros III Photographs." *Science (New York, N.Y.)* 143: 569–572. doi:10.1126/science.143.3606.569.
- Bell, Gerald D., and Daniel Keyser. 1993. "Shear and Curvature Vorticity and Potential-Vorticity Interchanges: Interpretation and Application to a Cutoff Cyclone Event." *Monthly Weather Review*. doi:10.1175/1520-0493(1993)121<0076:SACVAP>2.0.CO;2.
- Benjamin, Stan. 2013. "Data Assimilation and Model Updates in the 2013 Rapid Refresh (RAP) and High-Resolution Rapid Refresh (HRRR) Analysis and Forecast Systems." http://ruc.noaa.gov/pdf/NCEP_HRRR_RAPv2_6jun2013-Benj-noglob.pdf.
- Chuang, Hui-Ya, Geoff DiMego, and Michael Baldwin. 2004. "NCEP's WRF Post Processor and Verification Systems." In *Papers Presented at the 5th WRF / 14th MM5 Users' Workshop*.
- De Pondeva, Manuel S F V, Geoffrey S Manikin, Geoff DiMego, Stanley G Benjamin, David F Parrish, R James Purser, Wan-Shu Wu, et al. 2011. "The Real-Time Mesoscale Analysis at NOAA's National Centers for Environmental Prediction: Current Status and Development." *Weather and Forecasting* 26 (5) (April 27): 593–612. doi:10.1175/WAF-D-10-05037.1. <http://dx.doi.org/10.1175/WAF-D-10-05037.1>.
- DesJardins, Mary L, Keith F Brill, United States. National Aeronautics and Space Administration. Scientific and Technical Information Division, and Steven S Schotz. 1991. *GEMPAK5 Version 5*. [Washington, D.C.]: [Springfield, Va.: National Aeronautics and Space Administration, Office of Management, Scientific and Technical Information Division ; For sale by the National Technical Information Service]. <http://forward.library.wisconsin.edu/catalog/ocm48985808>.
- "Federal Meteorological Handbook No. 1, Surface Weather Observations and Reports." 2005. Washington, D.C. <http://www.ofcm.gov/fmh-1/fmh1.htm>.
- Gerth, Jordan J. 2011. "Improving Cloud and Moisture Representation by Assimilating GOES Sounder Products into Numerical Weather Prediction Initial Conditions." <http://forward.library.wisconsin.edu/catalog/ocn794186932>.
- Glahn, Harry R, and David P Ruth. 2003. "The New Digital Forecast Database of the National Weather Service." *Bulletin of the American Meteorological Society* 84 (2) (February 1): 195–201. doi:10.1175/BAMS-84-2-195. <http://dx.doi.org/10.1175/BAMS-84-2-195>.
- Heidinger, Andrew K. 2011a. "Algorithm Theoretical Basis Document: ABI Cloud Mask." http://www.goes-r.gov/products/ATBDs/baseline/Cloud_CldMask_v2.0_no_color.pdf.

- . 2011b. “Algorithm Theoretical Basis Document: ABI Cloud Height.” http://www.goes-r.gov/products/ATBDs/baseline/Cloud_CldHeight_v2.0_no_color.pdf.
- Heidinger, Andrew K, and Michael J Pavolonis. 2009. “Gazing at Cirrus Clouds for 25 Years through a Split Window. Part I: Methodology.” *Journal of Applied Meteorology and Climatology* 48 (6) (June 1): 1100–1116. doi:10.1175/2008JAMC1882.1. <http://dx.doi.org/10.1175/2008JAMC1882.1>.
- Ikeda, Kyoko, Matthias Steiner, James Pinto, and Curtis Alexander. 2013. “Evaluation of Cold-Season Precipitation Forecasts Generated by the Hourly Updating High-Resolution Rapid Refresh Model.” *Weather and Forecasting* 28 (4) (May 7): 921–939. doi:10.1175/WAF-D-12-00085.1. <http://dx.doi.org/10.1175/WAF-D-12-00085.1>.
- Karp, Richard M. 1972. “Reducibility Among Combinatorial Problems.” In *Complexity of Computer Computations SE - 9*, edited by Raymond E. Miller, James W. Thatcher, and Jean D. Bohlinger, 85–103 LA – English. Springer US. doi:10.1007/978-1-4684-2001-2_9. http://dx.doi.org/10.1007/978-1-4684-2001-2_9.
- Kent, Wes P, and Scott C Sheridan. 2011. “The Impact of Cloud Cover on Major League Baseball.” *Weather, Climate, and Society* 3 (1) (January 1): 7–15. doi:10.1175/2011WCAS1093.1. <http://dx.doi.org/10.1175/2011WCAS1093.1>.
- Kvamstø, Nils Gunnar. 1991. “An Investigation of Diagnostic Relations Between Stratiform Fractional Cloud Cover and Other Meteorological Parameters in Numerical Weather Prediction Models.” *Journal of Applied Meteorology* 30 (2) (February 1): 200–216. doi:10.1175/1520-0450(1991)030<0200:AIODRB>2.0.CO;2. [http://dx.doi.org/10.1175/1520-0450\(1991\)030<0200:AIODRB>2.0.CO](http://dx.doi.org/10.1175/1520-0450(1991)030<0200:AIODRB>2.0.CO).
- Malberg, Horst. 1973. “Comparison of Mean Cloud Cover Obtained By Satellite Photographs and Ground-Based Observations Over Europe and the Atlantic.” *Monthly Weather Review* 101 (12) (December 1): 893–897. doi:10.1175/1520-0493(1973)101<0893:COMCCO>2.3.CO;2. [http://dx.doi.org/10.1175/1520-0493\(1973\)101<0893:COMCCO>2.3.CO](http://dx.doi.org/10.1175/1520-0493(1973)101<0893:COMCCO>2.3.CO).
- Matthews, Grant. 2013. “Recursive Bessel Filter for Stray Light Corrections in an Imager”. United States of America: United States Patent and Trademark Office. <http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/PTO/search-adv.htm&r=1&f=G&l=50&d=PALL&S1=08484271&OS=PN/08484271&RS=PN/08484271> 1.
- “National Weather Service Instruction 10-201, National Digital Forecast Database and Local Database Description and Specifications.” 2012. <http://www.nws.noaa.gov/directives/sym/pd01002001curr.pdf>.

- Schlatter, T.W., and D.V. Baker. 1981. "Algorithms for Thermodynamic Calculations". Boulder, CO. <http://icoads.noaa.gov/software/other/profs>.
- Schmit, Timothy J, Mathew M Gunshor, W Paul Menzel, James J Gurka, Jun Li, and A Scott Bachmeier. 2005. "Introducing the Next-Generation Advanced Baseline Imager on GOES-R." *Bulletin of the American Meteorological Society* 86 (8) (August 1): 1079–1096. doi:10.1175/BAMS-86-8-1079. <http://dx.doi.org/10.1175/BAMS-86-8-1079>.
- Schmit, Timothy J., Jun Li, Jinlong Li, Wayne F. Feltz, James J. Gurka, Mitchell D. Goldberg, and Kevin J. Schrab. 2008. "The GOES-R Advanced Baseline Imager and the Continuation of Current Sounder Products." *Journal of Applied Meteorology and Climatology* 47 (10) (October): 2696–2711. doi:10.1175/2008JAMC1858.1. <http://journals.ametsoc.org/doi/abs/10.1175/2008JAMC1858.1>.
- Schreiner, Anthony J, Timothy J Schmit, and Robert M Aune. 2002. "Maritime Inversions and the GOES Sounder Cloud Product." *National Weather Digest* 26 (1, 2): 27–38. <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Maritime+Inversions+and+the+GOES+Sounder+Cloud+Product#0>.
- Schreiner, Anthony J, Kathy I Strabala, David A Unger, W Paul Menzel, Gary P Ellrod, and Jackson L Pellet. 1993. "A Comparison of Ground and Satellite Observations of Cloud Cover." *Bulletin of the American Meteorological Society* 74 (10) (October 1): 1851–1861. doi:10.1175/1520-0477(1993)074<1851:ACOGAS>2.0.CO;2. [http://dx.doi.org/10.1175/1520-0477\(1993\)074<1851:ACOGAS>2.0.CO](http://dx.doi.org/10.1175/1520-0477(1993)074<1851:ACOGAS>2.0.CO).
- Sims, James R. 1973. "An Objective Method of Preparing Cloud Cover Forecasts." *Journal of Applied Meteorology* 12 (4) (June 1): 726–728. doi:10.1175/1520-0450(1973)012<0726:AOMOPC>2.0.CO;2. [http://dx.doi.org/10.1175/1520-0450\(1973\)012<0726:AOMOPC>2.0.CO](http://dx.doi.org/10.1175/1520-0450(1973)012<0726:AOMOPC>2.0.CO).
- "Sky Condition." 2013. Accessed November 6. <http://www.nws.noaa.gov/asos/sky.htm>.
- "Sky Condition Algorithm for Vaisala Ceilometer's." 2010. [http://www.vaisala.com/Vaisala/Documents/Brochures and Datasheets/Sky-Condition-CL31-Datasheet-B210507EN-B-LoRes.pdf](http://www.vaisala.com/Vaisala/Documents/Brochures%20and%20Datasheets/Sky-Condition-CL31-Datasheet-B210507EN-B-LoRes.pdf).
- Slingo, Julia M. 1980. "A Cloud Parametrization Scheme Derived from GATE Data for Use with a Numerical Model." *Quarterly Journal of the Royal Meteorological Society* 106 (450): 747–770. <http://dx.doi.org/10.1002/qj.49710645008>.
- Smagorinsky, Joseph. 1960. "On the Dynamical Prediction of Large-Scale Condensation by Numerical Methods." In *Physics of Precipitation: Proceedings of the Cloud Physics Conference, Woods Hole, Massachusetts, June 3–5, 1959*, 71–78. American Geophysical Union. doi:10.1029/GM005p0071. <http://dx.doi.org/10.1029/GM005p0071>.

- Smith, R N B. 1990. "A Scheme for Predicting Layer Clouds and Their Water Content in a General Circulation Model." *Quarterly Journal of the Royal Meteorological Society* 116 (492): 435–460. <http://dx.doi.org/10.1002/qj.49711649210>.
- Stephens, Graeme L. 2005. "Cloud Feedbacks in the Climate System: A Critical Review." *Journal of Climate* 18 (2) (January 1): 237–273. doi:10.1175/JCLI-3243.1. <http://dx.doi.org/10.1175/JCLI-3243.1>.
- Teixeira, João. 2001. "Cloud Fraction and Relative Humidity in a Prognostic Cloud Fraction Scheme." *Monthly Weather Review* 129 (7) (July 1): 1750–1753. doi:10.1175/1520-0493(2001)129<1750:CFARHI>2.0.CO;2. [http://dx.doi.org/10.1175/1520-0493\(2001\)129<1750:CFARHI>2.0.CO](http://dx.doi.org/10.1175/1520-0493(2001)129<1750:CFARHI>2.0.CO).
- Thompson, Gregory, Paul R Field, Roy M Rasmussen, and William D Hall. 2008. "Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part II: Implementation of a New Snow Parameterization." *Monthly Weather Review* 136 (12) (December 1): 5095–5115. doi:10.1175/2008MWR2387.1. <http://dx.doi.org/10.1175/2008MWR2387.1>.
- Tiedtke, M. 1993. "Representation of Clouds in Large-Scale Models." *Monthly Weather Review* 121 (11) (November 1): 3040–3061. doi:10.1175/1520-0493(1993)121<3040:ROCILS>2.0.CO;2. [http://dx.doi.org/10.1175/1520-0493\(1993\)121<3040:ROCILS>2.0.CO](http://dx.doi.org/10.1175/1520-0493(1993)121<3040:ROCILS>2.0.CO).
- Tompkins, A. M. 2005. "The Parametrization of Cloud Cover." http://www.ecmwf.int/newsevents/training/lecture_notes/pdf_files/PARAM/Cloudcover.pdf
- Walcek, Chris J. 1994. "Cloud Cover and Its Relationship to Relative Humidity During a Springtime Midlatitude Cyclone." *Monthly Weather Review* 122 (6) (June 1): 1021–1035. doi:10.1175/1520-0493(1994)122<1021:CCAIRT>2.0.CO;2. [http://dx.doi.org/10.1175/1520-0493\(1994\)122<1021:CCAIRT>2.0.CO](http://dx.doi.org/10.1175/1520-0493(1994)122<1021:CCAIRT>2.0.CO).
- Williamson, David L, Jeffrey T Kiehl, Veerabhadran Ramanathan, Robert E Dickinson, and James J Hack. 1987. "Description of the NCAR Community Climate Model (CCM1)". University Corporation for Atmospheric Research. doi:10.5065/D6TB14WH. <http://nldr.library.ucar.edu/repository/collections/TECH-NOTE-000-000-000-560>.
- Willmott, Cort J. 2005. "Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance." *Climate Research* 30 (1): 79–82. <http://www.int-res.com/abstracts/cr/v30/n1/p79-82/>.

- Wood, Robert, and Paul R Field. 2000. "Relationships Between Total Water, Condensed Water, and Cloud Fraction in Stratiform Clouds Examined Using Aircraft Data." *Journal of the Atmospheric Sciences* 57 (12) (June 1): 1888–1905. doi:10.1175/1520-0469(2000)057<1888:RBTWCW>2.0.CO;2. [http://dx.doi.org/10.1175/1520-0469\(2000\)057<1888:RBTWCW>2.0.CO](http://dx.doi.org/10.1175/1520-0469(2000)057<1888:RBTWCW>2.0.CO).
- Wylie, D P, and W P Menzel. 1989. "Two Years of Cloud Cover Statistics Using VAS." *Journal of Climate* 2 (4) (April 1): 380–392. doi:10.1175/1520-0442(1989)002<0380:TYOCCS>2.0.CO;2. [http://dx.doi.org/10.1175/1520-0442\(1989\)002<0380:TYOCCS>2.0.CO](http://dx.doi.org/10.1175/1520-0442(1989)002<0380:TYOCCS>2.0.CO).
- Xu, Kuan-Man, and David A Randall. 1996. "A Semiempirical Cloudiness Parameterization for Use in Climate Models." *Journal of the Atmospheric Sciences* 53 (21) (November 1): 3084–3102. doi:10.1175/1520-0469(1996)053<3084:ASCPFU>2.0.CO;2. [http://dx.doi.org/10.1175/1520-0469\(1996\)053<3084:ASCPFU>2.0.CO](http://dx.doi.org/10.1175/1520-0469(1996)053<3084:ASCPFU>2.0.CO).
- Young, Murray J. 1967. "Variability in Estimating Total Cloud Cover from Satellite Pictures." *Journal of Applied Meteorology* 6 (3) (June 1): 573–579. doi:10.1175/1520-0450(1967)006<0573:VIETCC>2.0.CO;2. [http://dx.doi.org/10.1175/1520-0450\(1967\)006<0573:VIETCC>2.0.CO](http://dx.doi.org/10.1175/1520-0450(1967)006<0573:VIETCC>2.0.CO).
- Zhang, Guang J. 2003. "Lagrangian Study of Cloud Properties and Their Relationships to Meteorological Parameters over the U.S. Southern Great Plains." *Journal of Climate* 16 (16) (August 1): 2700–2716. doi:10.1175/1520-0442(2003)016<2700:LSOCPA>2.0.CO;2. [http://dx.doi.org/10.1175/1520-0442\(2003\)016<2700:LSOCPA>2.0.CO](http://dx.doi.org/10.1175/1520-0442(2003)016<2700:LSOCPA>2.0.CO).
- Zhenping, Li, Michael Grotenhuis, Timothy J. Schmit, Xiangqian Wu, James P. Nelson III, Anthony J. Schreiner, Fangfang Yu, and Hyre Bysal. 2013. "GOES Imager IR Channel to Channel Co-Registration Correction Program in the GOES Ground System." [http://satelliteconferences.noaa.gov/2013/docs/Tuesday Poster Session Final Posters/T7_NOAA_SATCONF_Poster_ZhenpingLi-2.pdf](http://satelliteconferences.noaa.gov/2013/docs/Tuesday%20Poster%20Session%20Final%20Posters/T7_NOAA_SATCONF_Poster_ZhenpingLi-2.pdf).

Tables

Satellite Observation at Test Point	Satellite Observation at Closest Surface Station	Surface Observation at Closest Site	Blended Result	Deference
Clear	Clear	Clear	Clear	Both
Clear	Clear	Cloudy	Cloudy	Surface
Clear	Cloudy	Clear	Clear	Satellite
Clear	Cloudy	Cloudy	Clear	Satellite
Cloudy	Clear	Clear	Cloudy	Satellite
Cloudy	Clear	Cloudy	Cloudy	Both
Cloudy	Cloudy	Clear	Cloudy	Satellite
Cloudy	Cloudy	Cloudy	Cloudy	Both

Table 1. This table summarizes the logic for comparing a grid point containing a satellite observation of sky cover to a like in-situ surface observation nearby with a collocated satellite observation. A cloudy point is one that is not clear. The point result for the blended sky cover analysis appears in the second to rightmost column. Deference of the case result to the observing platform is noted in the rightmost column.

Level	Number of occurrences	Percentage of all runs
None	446	54.1%
200 hPa	113	13.7%
300 hPa	59	7.2%
500 hPa	58	7.0%
700 hPa	33	4.0%
800 hPa	37	4.5%
850 hPa	24	2.9%
900 hPa	19	2.3%
950 hPa	24	2.9%
1000 hPa	11	1.3%

Table 2. This table contains a summary of the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The number of runs for which the relative humidity field at the level listed in the leftmost column is used through the secondary model to adjust the result in part or in full from the primary model is shown in the middle column. The percentage in the rightmost column is the frequency of occurrence out of 824 runs in the time window. The “None” row is an adjustment based purely on the output from the primary model with no filtering based on a relative humidity field.

Rel. Hum.	Coefficient (RH)					Coefficient (100-RH)					
	Level	Mean	Number		Percentage		Mean	Number		Percentage	
200 hPa	0.254	340	636		41.3%	77.2%	-0.189	174	463	21.1%	56.2%
300 hPa	0.270	485			58.9%		-0.228	359		43.6%	
500 hPa	0.221	440			53.4%		-0.166	304		36.9%	
700 hPa	0.174	320			38.8%		-0.153	222		26.9%	
800 hPa	0.155	232	540		28.2%	65.5%	-0.130	153	406	18.6%	49.3%
850 hPa	0.156	186			22.6%		-0.128	128		15.5%	
900 hPa	0.184	295			35.8%		-0.182	209		25.4%	
950 hPa	0.229	351	513		42.6%	62.3%	-0.209	274		33.3%	
1000 hPa	0.386	296			35.9%		NA	NA	NA	NA	NA

Table 3. This table contains a summary of the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The coefficient referenced in leftmost columns two through six, is the multiplier applied to the relative humidity (RH) field value at each grid point, in units of %/%. The coefficient reference in the rightmost five columns is the multiplier applied to the 100% less relative humidity (100-RH) field value at each grid point, in units of %/%. The means are calculated from all non-zero values. The number represents the number of runs that the optimizer found a non-zero coefficient, and the percentage is the frequency of occurrence out of 824 runs in the time window. Where the column is split across multiple rows, the number or percentage is calculated based one whether at least one of the covered levels is non-zero. The model is not designed to calculate a coefficient for the 100% less relative humidity field value at 1000 hPa.

Rel. Hum.	At least one coefficient (RH, 100-RH)		Two coefficients (RH, 100-RH)		
	Level	Number	Percentage	Number	Percentage
200 hPa		416	50.5%	98	11.9%
300 hPa		640	77.7%	204	24.8%
500 hPa		602	73.1%	142	17.2%
700 hPa		473	57.4%	69	8.4%
800 hPa		336	40.8%	49	5.9%
850 hPa		272	33.0%	42	5.1%
900 hPa		410	49.8%	94	11.4%
950 hPa		500	60.7%	125	15.2%
Rel. Hum.	Coefficient or Scalar		Coefficient and Scalar		
Level	Number	Percentage	Number	Percentage	
1000 hPa	824	100.0%	134	16.3%	

Table 4. This table contains a summary of the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. For levels between 200 hPa and 950 hPa, the second column from the left contains the number of runs in which the optimizer found a non-zero coefficient for application to the relative humidity (RH) field values, 100% less relative humidity (100-RH) field values, or both, at the respective levels in the leftmost column. For those same levels, the number of cases in which the optimizer returned a non-zero coefficient for the quantities of relative humidity (RH) and 100% less relative humidity (100-RH) is listed in the fourth column. The third and fifth columns contain a percentage representing the frequency of occurrence out of 824 runs in the time window, based on the number in the column immediately to its left. For 1000 hPa, the second column from the left contains the number of runs in which the optimizer found a non-zero coefficient or scalar for application to the relative humidity field values. Cases in which the optimizer returned both a non-zero coefficient and non-zero scalar for the same run, counted exclusively in the fourth column, are also included in the number. These coefficients and scalars are only applied above a certain threshold, as defined by the optimizer during the execution of each run.

Cl. Water Level	Coefficient				Scalar					
	Mean	Number		Percentage		Mean	Number		Percentage	
200 hPa	NA	0	0	0.0%	0.0%	NA	0	0	0.0%	0.0%
300 hPa	NA	0		0.0%		NA	0		0.0%	
500 hPa	4.60×10^5	246		29.9%		8.8	270		32.8%	
700 hPa	2.41×10^5	542		65.8%		8.2	583		70.8%	
800 hPa	1.75×10^5	540	792	65.5%	96.1%	7.0	520	790	63.1%	95.9%
850 hPa	1.22×10^5	484		58.7%		7.1	509		61.8%	
900 hPa	1.38×10^5	629		76.3%		8.2	649		78.8%	
950 hPa	2.33×10^5	758	779	92.0%	94.5%	7.8	608	677	73.8%	82.2%
1000 hPa	2.45×10^5	277		33.6%		9.3	294		35.7%	

Table 5. This table contains a summary of the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The coefficient is the multiplier applied to the cloud water mixing ratio field value at each grid point, in units of %•kg/kg. The scalar is the value added to the non-zero cloud water mixing ratio field value at each grid point, regardless of its magnitude, in units of %. The means are calculated from all non-zero values. The number represents the number of runs that the optimizer found a non-zero coefficient or scalar, and the percentage is the frequency of occurrence out of 824 runs in the time window. Where the column is split across multiple rows, the number or percentage is calculated based on whether at least one of the covered levels is non-zero. The notation “NA” indicates the lack of output to use in calculating the mean.

Cl. Water Level	Coefficient or Scalar		Coefficient and Scalar	
	Number	Percentage	Number	Percentage
200 hPa	0	0.0%	0	0.0%
300 hPa	0	0.0%	0	0.0%
500 hPa	344	41.7%	172	20.9%
700 hPa	708	85.9%	417	50.6%
800 hPa	721	87.5%	339	41.1%
850 hPa	691	83.9%	302	36.7%
900 hPa	775	94.1%	503	61.0%
950 hPa	800	97.1%	566	68.7%
1000 hPa	324	39.3%	247	30.0%

Table 6. This table contains a summary of the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The second column from the left contains the number of runs in which the optimizer found a non-zero coefficient or scalar for application to the cloud water mixing ratio field values at the respective levels in the leftmost column. Cases in which the optimizer returned both a non-zero coefficient and non-zero scalar for the same run, counted exclusively in the fourth column, are also included in the number. The third and fifth columns contain a percentage representing the frequency of occurrence out of 824 runs in the time window, based on the number in the column immediately to its left.

Cloud Ice Level	Coefficient				Scalar					
	Mean	Number		Percentage		Mean	Number		Percentage	
200 hPa	5.27×10^6	115	427	14.0%	51.8%	9.2	107	417	13.0%	50.7%
300 hPa	4.21×10^6	389		47.2%		9.3	391		47.5%	
500 hPa	1.08×10^7	30		3.6%		12.5	26		3.2%	
700 hPa	7.19×10^6	83		10.1%		11.0	72		8.7%	
800 hPa	7.24×10^6	89	130	10.8%	15.8%	10.5	81	111	9.8%	13.5%
850 hPa	4.99×10^6	71		8.6%		9.9	46		5.6%	
900 hPa	2.88×10^6	7		0.8%		11.8	4		0.5%	
950 hPa	NA	0	0	0.0%	0.0%	NA	0	0	0.0%	0.0%
1000 hPa	NA	0		0.0%		NA	0		0.0%	

Table 7. This table contains a summary of the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The coefficient is the multiplier applied to the cloud ice mixing ratio field value at each grid point, in units of %•kg/kg. The scalar is the value added to the non-zero cloud ice mixing ratio field value at each grid point, regardless of its magnitude, in units of %. The means are calculated from all non-zero values. The number represents the number of runs that the optimizer found a non-zero coefficient or scalar, and the percentage is the frequency of occurrence out of 824 runs in the time window. Where the column is split across multiple rows, the number or percentage is calculated based on whether at least one of the covered levels is non-zero. The notation “NA” indicates the lack of output to use in calculating the mean.

Cloud Ice Level	Coefficient or Scalar		Coefficient and Scalar	
	Number	Percentage	Number	Percentage
200 hPa	150	18.2%	72	8.7%
300 hPa	482	58.5%	298	36.2%
500 hPa	40	4.9%	16	1.9%
700 hPa	96	11.7%	59	7.2%
800 hPa	98	11.9%	72	8.7%
850 hPa	77	9.3%	40	4.9%
900 hPa	7	0.8%	4	0.5%
950 hPa	0	0.0%	0	0.0%
1000 hPa	0	0.0%	0	0.0%

Table 8. This table contains a summary of the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The second column from the left contains the number of runs in which the optimizer found a non-zero coefficient or scalar for application to the cloud ice mixing ratio field values at the respective levels in the leftmost column. Cases in which the optimizer returned both a non-zero coefficient and non-zero scalar for the same run, counted exclusively in the fourth column, are also included in the number. The third and fifth columns contain a percentage representing the frequency of occurrence out of 824 runs in the time window, based on the number in the column immediately to its left.

Rain Level	Coefficient				Scalar					
	Mean	Number		Percentage		Mean	Number		Percentage	
200 hPa	NA	0	0	0.0%	0.0%	NA	0	0	0.0%	0.0%
300 hPa	NA	0		0.0%		NA	0		0.0%	
500 hPa	7.32×10^5	2		0.2%		10.3	2		0.2%	
700 hPa	2.66×10^5	121		14.7%		7.0	155		18.8%	
800 hPa	2.48×10^5	91	144	11.0%	17.5%	6.6	139	273	16.9%	33.1%
850 hPa	2.93×10^5	64		7.8%		5.7	103		12.5%	
900 hPa	3.11×10^5	50		6.1%		5.2	117		14.2%	
950 hPa	2.51×10^5	56	124	6.8%	15.0%	6.3	98	199	11.9%	24.2%
1000 hPa	2.91×10^5	102		12.4%		6.5	141		17.1%	

Table 9. This table contains a summary of the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The coefficient is the multiplier applied to the rain water mixing ratio field value at each grid point, in units of %•kg/kg. The scalar is the value added to the non-zero rain water mixing ratio field value at each grid point, regardless of its magnitude, in units of %. The means are calculated from all non-zero values. The number represents the number of runs that the optimizer found a non-zero coefficient or scalar, and the percentage is the frequency of occurrence out of 824 runs in the time window. Where the column is split across multiple rows, the number or percentage is calculated based on whether at least one of the covered levels is non-zero. The notation “NA” indicates the lack of output to use in calculating the mean.

Rain Level	Coefficient or Scalar		Coefficient and Scalar	
	Number	Percentage	Number	Percentage
200 hPa	0	0.0%	0	0.0%
300 hPa	0	0.0%	0	0.0%
500 hPa	2	0.2%	2	0.2%
700 hPa	194	23.5%	82	10.0%
800 hPa	188	22.8%	42	5.1%
850 hPa	140	17.0%	27	3.3%
900 hPa	144	17.5%	23	2.8%
950 hPa	131	15.9%	23	2.8%
1000 hPa	182	22.1%	61	7.4%

Table 10. This table contains a summary of the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The second column from the left contains the number of runs in which the optimizer found a non-zero coefficient or scalar for application to the rain water mixing ratio field values at the respective levels in the leftmost column. Cases in which the optimizer returned both a non-zero coefficient and non-zero scalar for the same run, counted exclusively in the fourth column, are also included in the number. The third and fifth columns contain a percentage representing the frequency of occurrence out of 824 runs in the time window, based on the number in the column immediately to its left.

Snow Level	Coefficient				Scalar					
	Mean	Number		Percentage		Mean	Number		Percentage	
200 hPa	8.49×10^4	62	243	7.5%	29.5%	9.6	72	431	8.7%	52.3%
300 hPa	9.61×10^4	206		25.0%		7.8	406		49.3%	
500 hPa	9.78×10^4	270		32.8%		8.4	466		56.6%	
700 hPa	2.12×10^5	372		45.1%		10.2	484		58.7%	
800 hPa	3.43×10^5	192	262	23.3%	31.8%	10.5	242	341	29.4%	41.4%
850 hPa	3.86×10^5	105		12.7%		9.9	134		16.3%	
900 hPa	4.23×10^5	73		8.9%		9.6	116		14.1%	
950 hPa	5.43×10^5	60	60	7.3%	7.3%	8.7	61	61	7.4%	7.4%
1000 hPa	NA	0		0.0%		NA	0		0.0%	

Table 11. This table contains a summary of the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The coefficient is the multiplier applied to the snow mixing ratio field value at each grid point, in units of %•kg/kg. The scalar is the value added to the non-zero snow mixing ratio field value at each grid point, regardless of its magnitude, in units of %. The means are calculated from all non-zero values. The number represents the number of runs that the optimizer found a non-zero coefficient or scalar, and the percentage is the frequency of occurrence out of 824 runs in the time window. Where the column is split across multiple rows, the number or percentage is calculated based on whether at least one of the covered levels is non-zero. The notation “NA” indicates the lack of output to use in calculating the mean.

Snow Level	Coefficient or Scalar		Coefficient and Scalar	
	Number	Percentage	Number	Percentage
200 hPa	95	11.5%	39	4.7%
300 hPa	445	54.0%	167	20.3%
500 hPa	540	65.5%	196	23.8%
700 hPa	577	70.0%	279	33.9%
800 hPa	314	38.1%	120	14.6%
850 hPa	177	21.5%	62	7.5%
900 hPa	140	17.0%	49	5.9%
950 hPa	79	9.6%	42	5.1%
1000 hPa	0	0.0%	0	0.0%

Table 12. This table contains a summary of the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The second column from the left contains the number of runs in which the optimizer found a non-zero coefficient or scalar for application to the snow mixing ratio field values at the respective levels in the leftmost column. Cases in which the optimizer returned both a non-zero coefficient and non-zero scalar for the same run, counted exclusively in the fourth column, are also included in the number. The third and fifth columns contain a percentage representing the frequency of occurrence out of 824 runs in the time window, based on the number in the column immediately to its left.

Figures

GOES-East CONUS IR Window Image

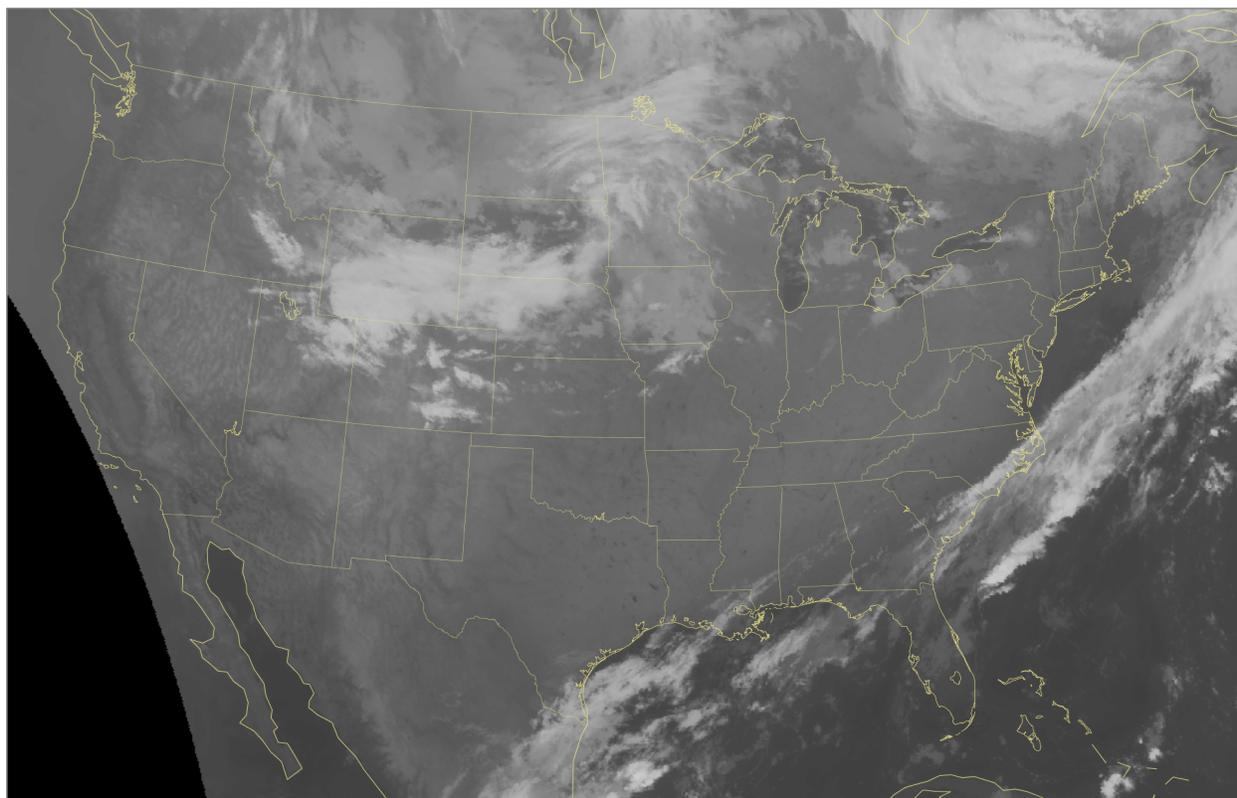


Figure 1. This figure depicts the infrared window satellite image from GOES-East valid at 11:02 UTC on 20 October 2013. A traditional enhancement is applied to the satellite image. Cold cloud tops are white.

Sun Oct 20 11:00:00 UTC 2013

GOES Imager Effective Cloud Amount (%)

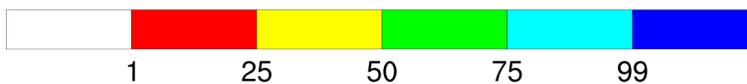
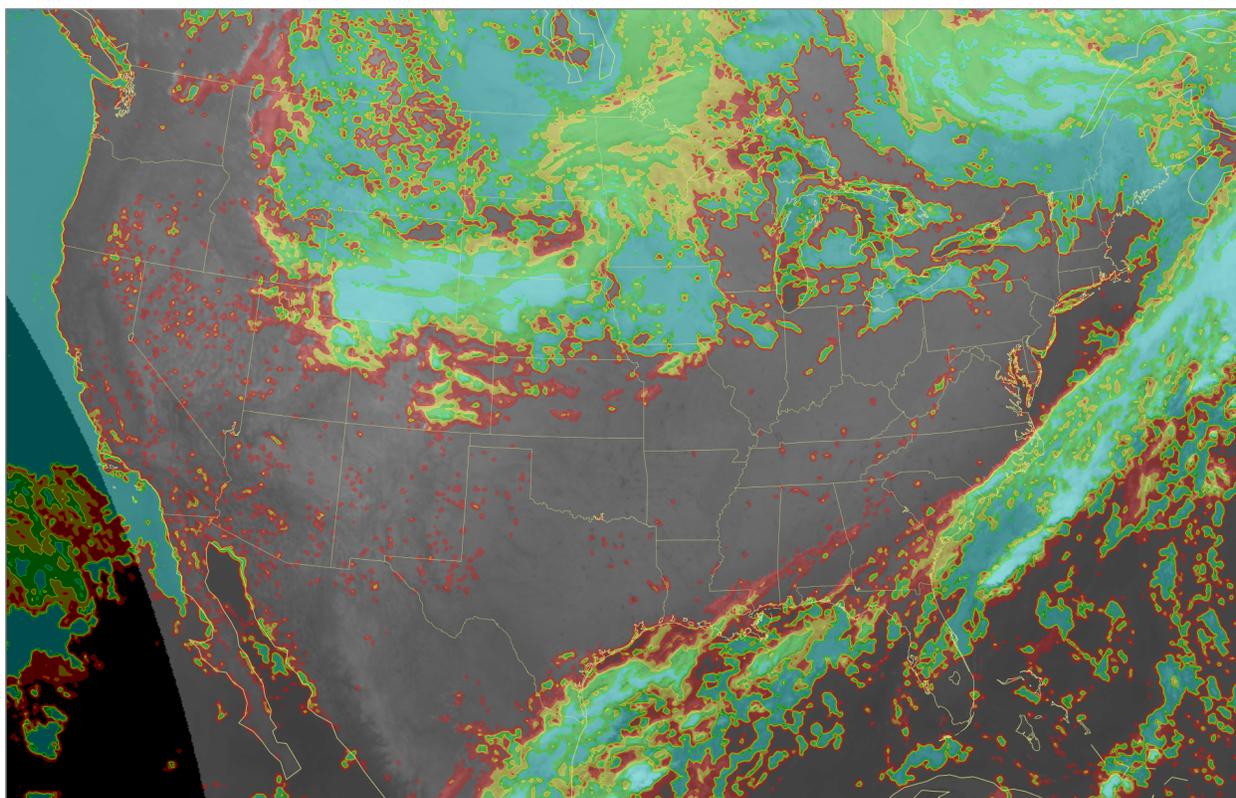


Figure 2. This figure depicts the composite geostationary satellite effective cloud amount analysis, in units of percent (%) effective emissivity, at 11:00 UTC on 20 October 2013. Cloudy areas are in cyan. The infrared window satellite image from GOES-East in the background is valid at 11:02 UTC on 20 October 2013. A traditional enhancement is applied to the satellite image. Cold cloud tops are white. The overlay is partially transparent, except clear areas, which are not filled.

Sun Oct 20 11:00:00 UTC 2013

GOES Imager Cloud Top Pressure (hPa)

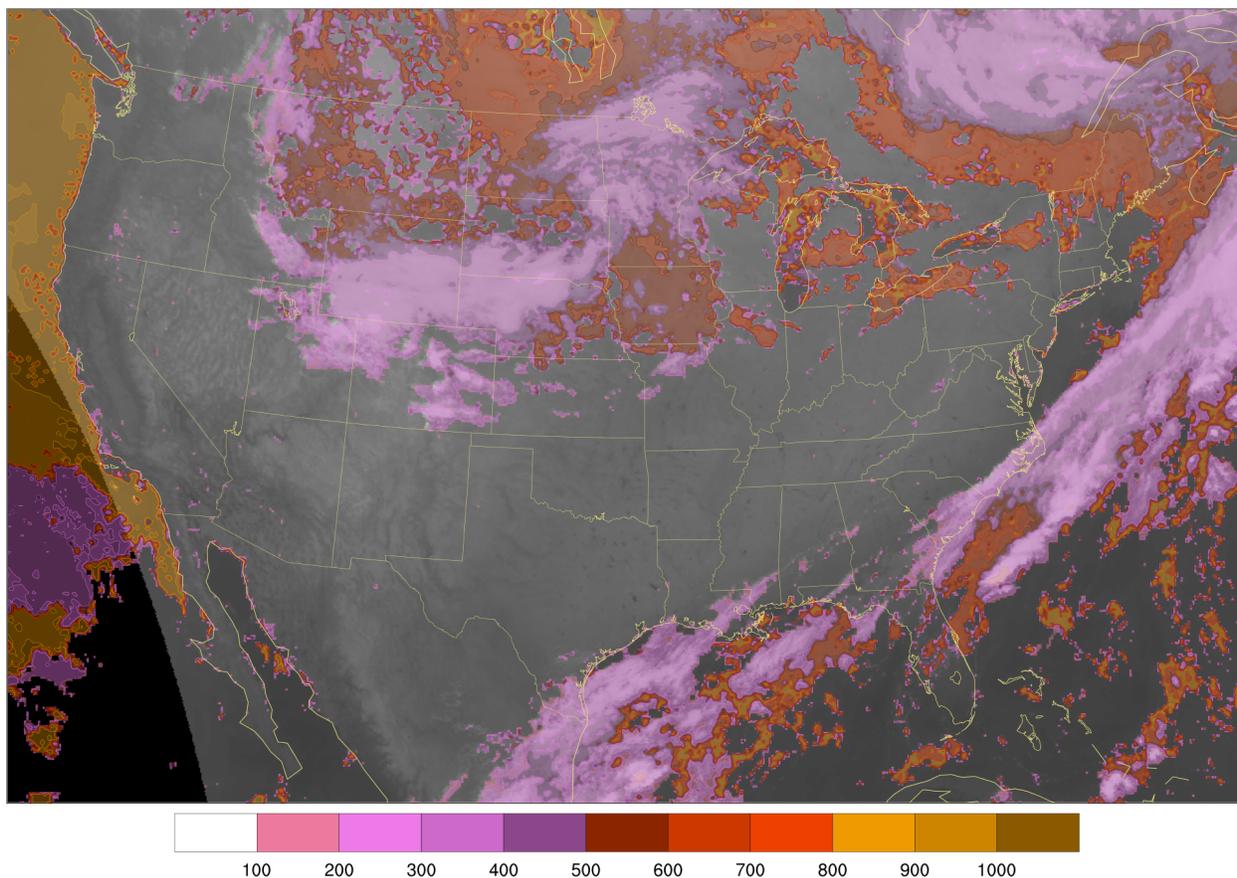


Figure 3. This figure depicts the composite geostationary satellite cloud top pressure analysis, in units of hPa, at 11:00 UTC on 20 October 2013. Low clouds, with heights below 800 hPa, are in orange and brown, mid-level clouds, with heights between 800 hPa and 500 hPa, are in shades of red, and high clouds, with heights above 500 hPa, are in shades of magenta. The infrared window satellite image from GOES-East in the background is valid at 11:02 UTC on 20 October 2013. A traditional enhancement is applied to the satellite image. Cold cloud tops are white. The overlay is partially transparent, except clear areas, which are not filled.

Sun Oct 20 11:00:00 UTC 2013

GOES Imager Celestial Dome ECA (%)

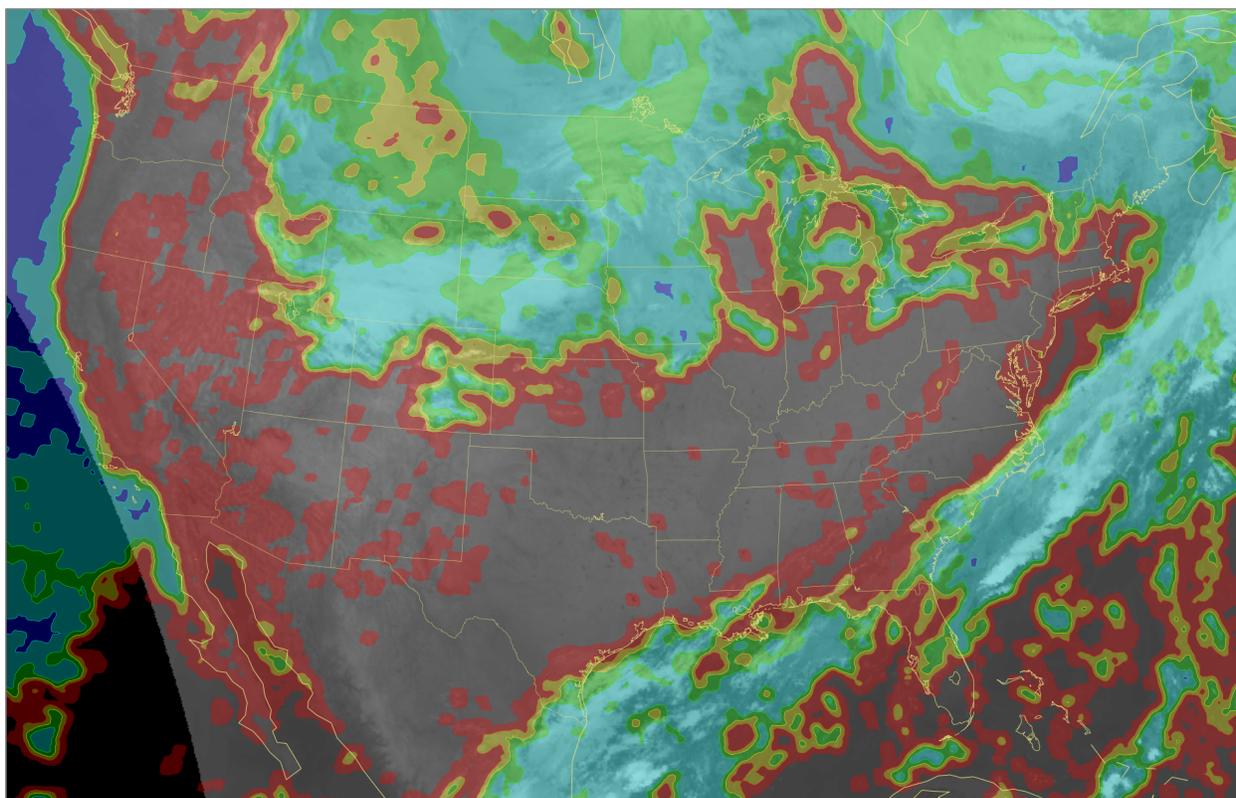


Figure 4. This figure depicts the composite geostationary satellite celestial dome effective cloud amount analysis, in units of percent (%) sky cover, at 11:00 UTC on 20 October 2013. Overcast areas are in blue. The infrared window satellite image from GOES-East in the background is valid at 11:02 UTC on 20 October 2013. A traditional enhancement is applied to the satellite image. Cold cloud tops are white. The overlay is partially transparent, except clear areas, which are not filled.

Sun Oct 20 11:00:00 UTC 2013

GOES Imager Sky Cover Product (%)

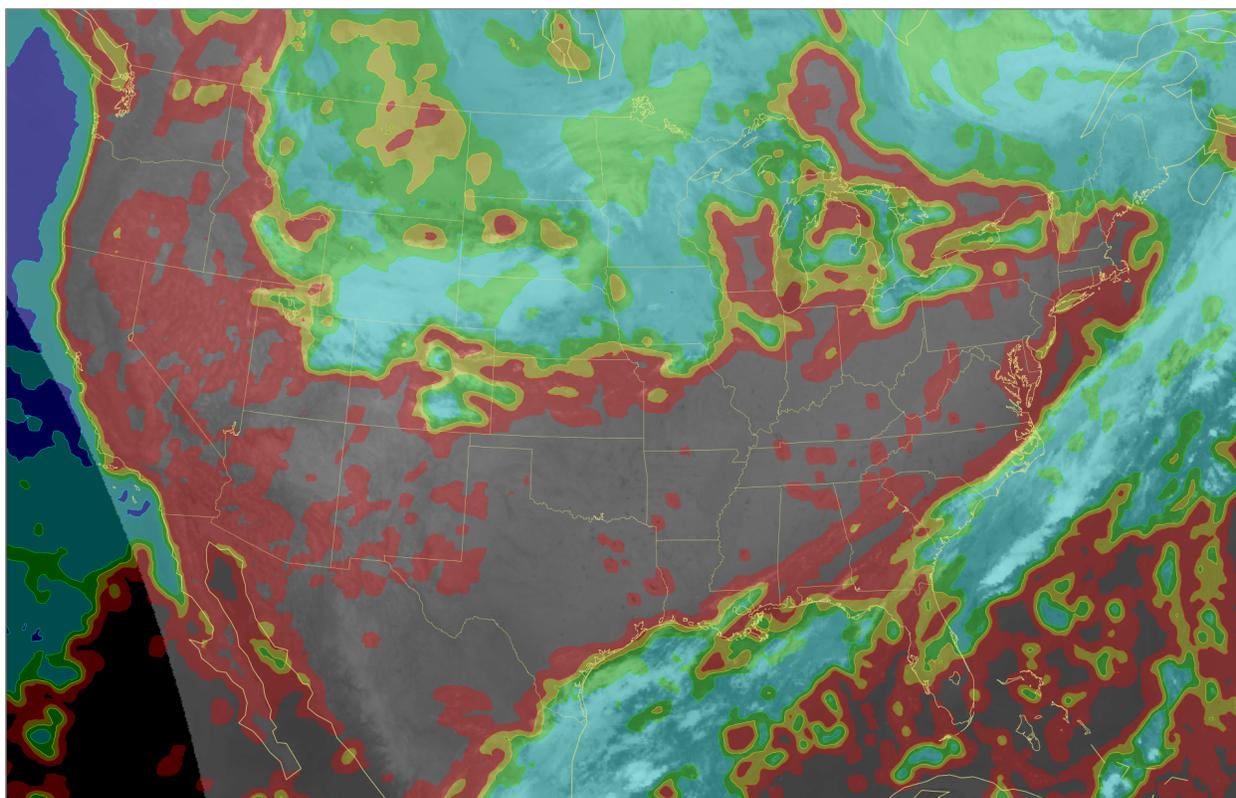


Figure 5. This figure depicts the composite geostationary satellite sky cover product, in units of percent (%) sky cover, for the one-hour window beginning at 11:00 UTC on 20 October 2013. Overcast areas are in blue. The infrared window satellite image from GOES-East in the background is valid at 11:02 UTC on 20 October 2013. A traditional enhancement is applied to the satellite image. Cold cloud tops are white. The overlay is partially transparent, except clear areas, which are not filled.

Sun Oct 20 11:00:00 UTC 2013

Surface Sky Observations/METARs (%)

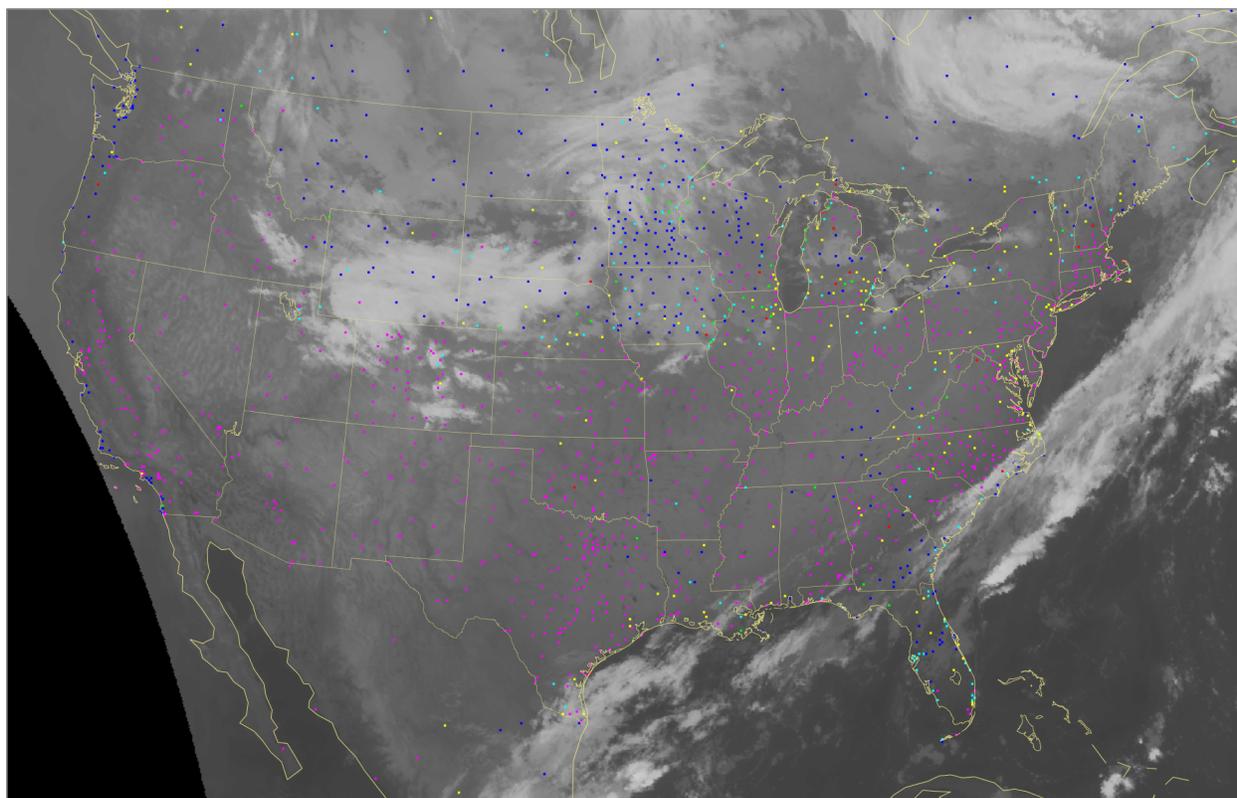


Figure 6. This figure depicts all automated and manned surface observing stations and their sky cover report, in units of percent (%) sky cover, for the one-hour window beginning at 11:00 UTC on 20 October 2013. Overcast areas are in blue. The infrared window satellite image from GOES-East in the background is valid at 11:02 UTC on 20 October 2013. A traditional enhancement is applied to the satellite image. Cold cloud tops are white.

Sun Oct 20 11:00:00 UTC 2013

Satellite/Surface Blended Sky Cover (%)

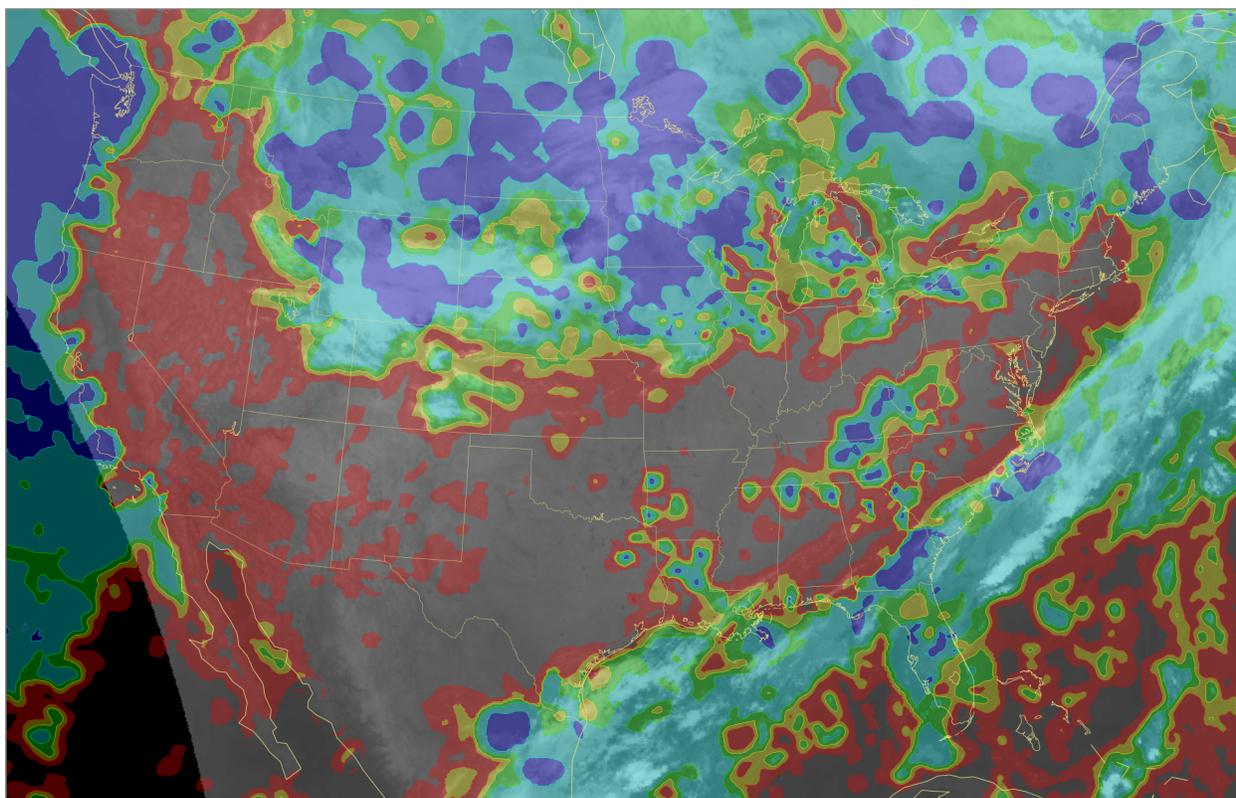


Figure 7. This figure depicts the blended sky cover analysis, in units of percent (%) sky cover, for the one-hour window beginning at 11:00 UTC on 20 October 2013. Overcast areas are in blue. The infrared window satellite image from GOES-East in the background is valid at 11:02 UTC on 20 October 2013. A traditional enhancement is applied to the satellite image. Cold cloud tops are white. The overlay is partially transparent, except clear areas, which are not filled.

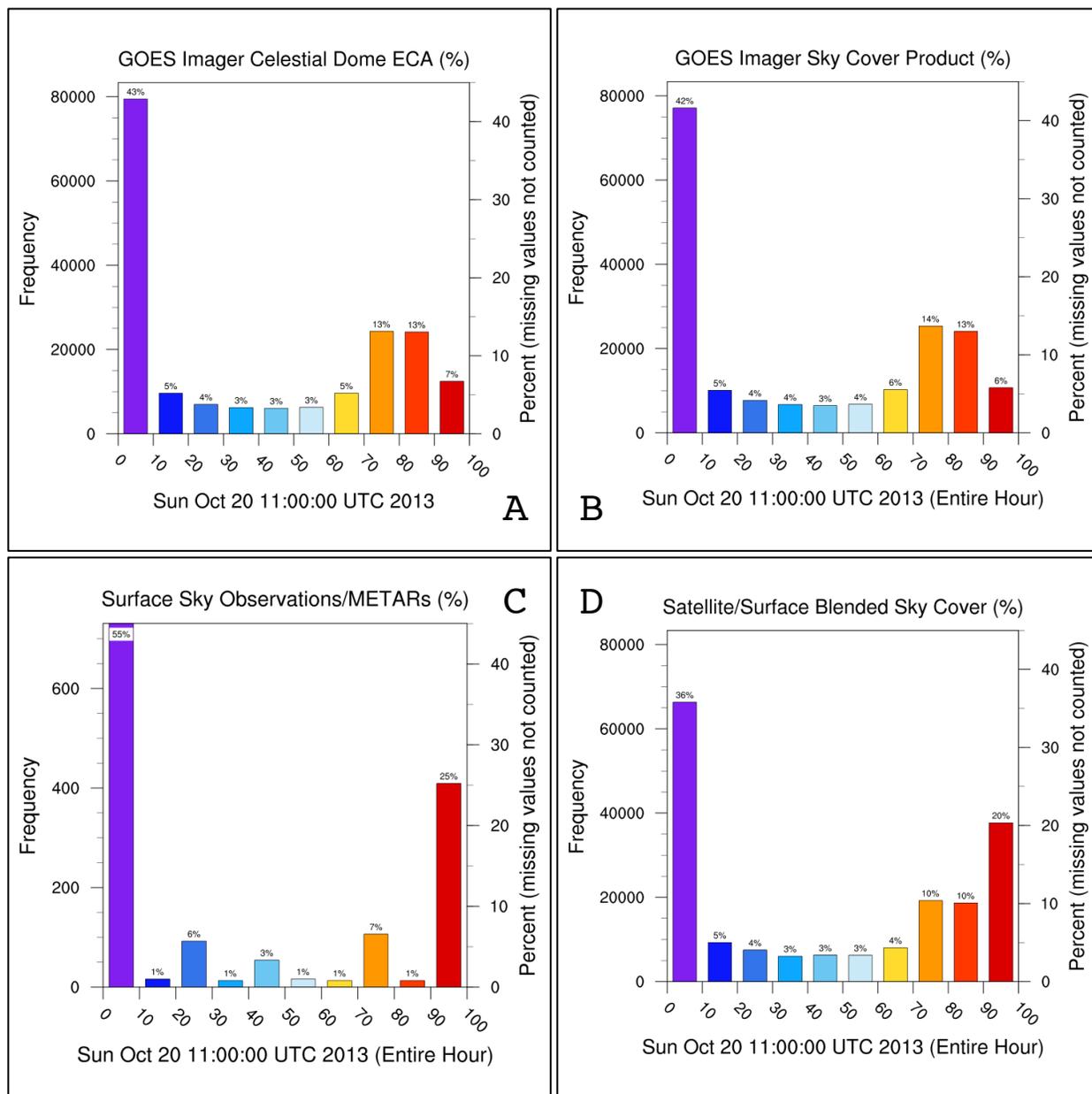


Figure 8. All of the histograms in this four-panel plot are valid for the one-hour window beginning at 11:00 UTC on 20 October 2013, except Panel A, which is valid at the top of the hour. Each histogram shows the frequency at which the value range, as indicated on the abscissa, occurs among all non-missing points in the indicated product. Each increment includes the range of values greater than or equal to the value to the left of the column and less than the value to the right of the column, except for the rightmost bin, which includes points with a value of 100%. The left ordinate axis indicates the frequency in terms of the number of points. The right ordinate axis indicates the percentage of the frequency compared to all non-missing points. This percentage also appears immediately above each bar. Panel A depicts the distribution for the composite geostationary satellite celestial dome effective cloud amount analysis. Panel B depicts the distribution for the composite geostationary satellite sky cover product. Panel C depicts the distribution for the surface sky observations. Panel D depicts the distribution for the blended sky cover analysis.

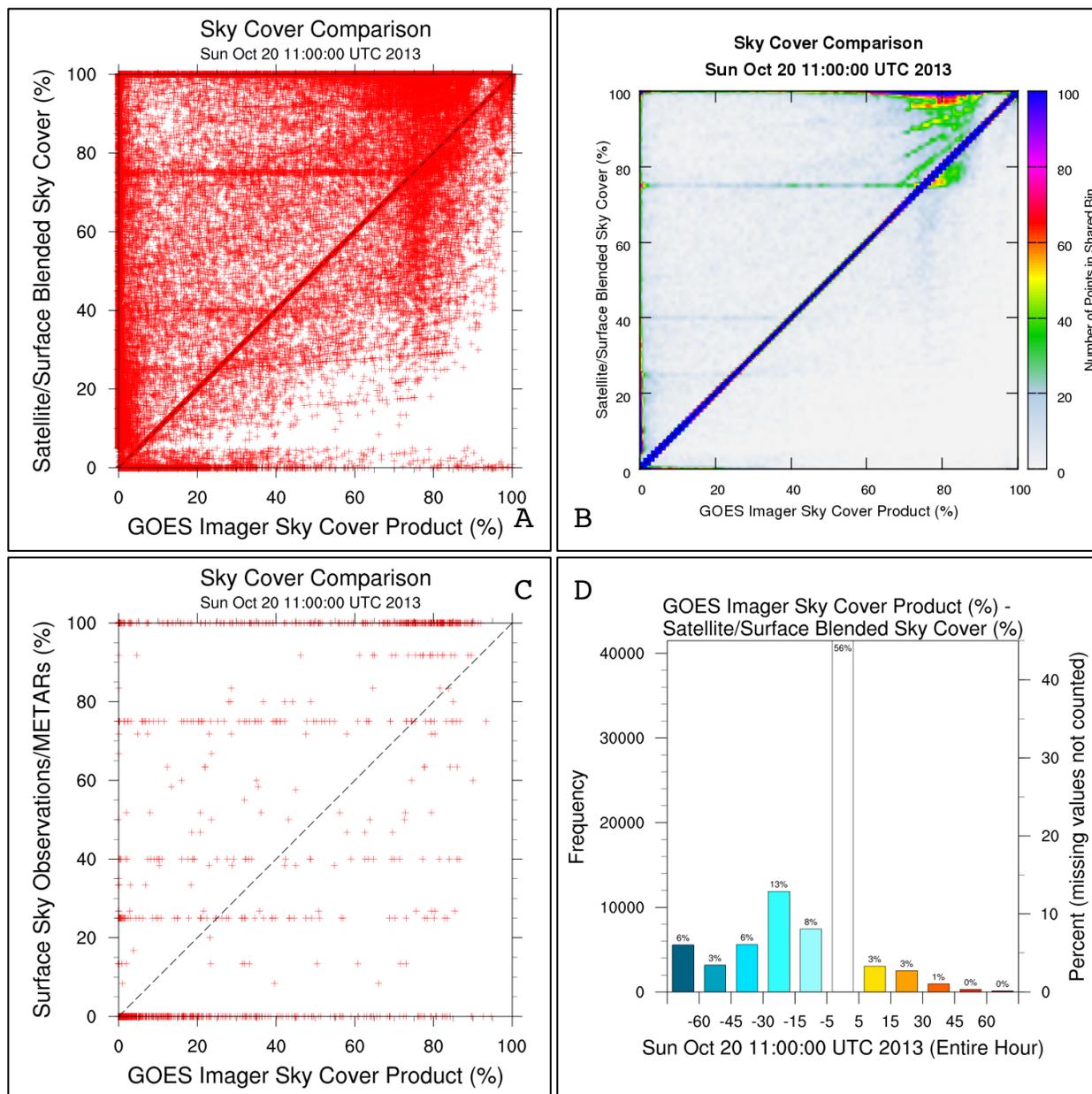


Figure 9. All of the panels in this plot are valid for the one-hour window beginning at 11:00 UTC on 20 October 2013. Panel A is a scatterplot with the composite geostationary satellite sky cover product on the abscissa and the blended sky cover analysis on the ordinate. Collocated points are plotted according to their value, in units of percent (%) sky cover, on each axis. Panel B is a binned density plot of Panel A. Panel C is a scatterplot with the composite geostationary satellite sky cover product on the abscissa and surface observations of sky cover on the ordinate. Both have units of percent (%) sky cover. Panel D is a histogram showing the occurrence frequency of the difference of the blended sky cover analysis from the composite geostationary satellite sky cover product between collocated, non-missing points within the increments as indicated on the abscissa. Each increment includes the range of values greater than or equal to the value to the left of the column, if present, and less than the value to the right of the column, if present. The left ordinate axis indicates the frequency in terms of the number of points. The right ordinate axis indicates the percentage of the frequency compared to all non-missing differenced points. This percentage also appears immediately above each bar.

Sun Oct 20 11:00:00 UTC 2013

GOES Imager Sky Cover Product (%) - Satellite/Surface Blended Sky Cover (%)

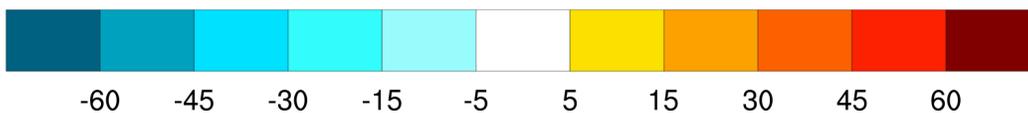
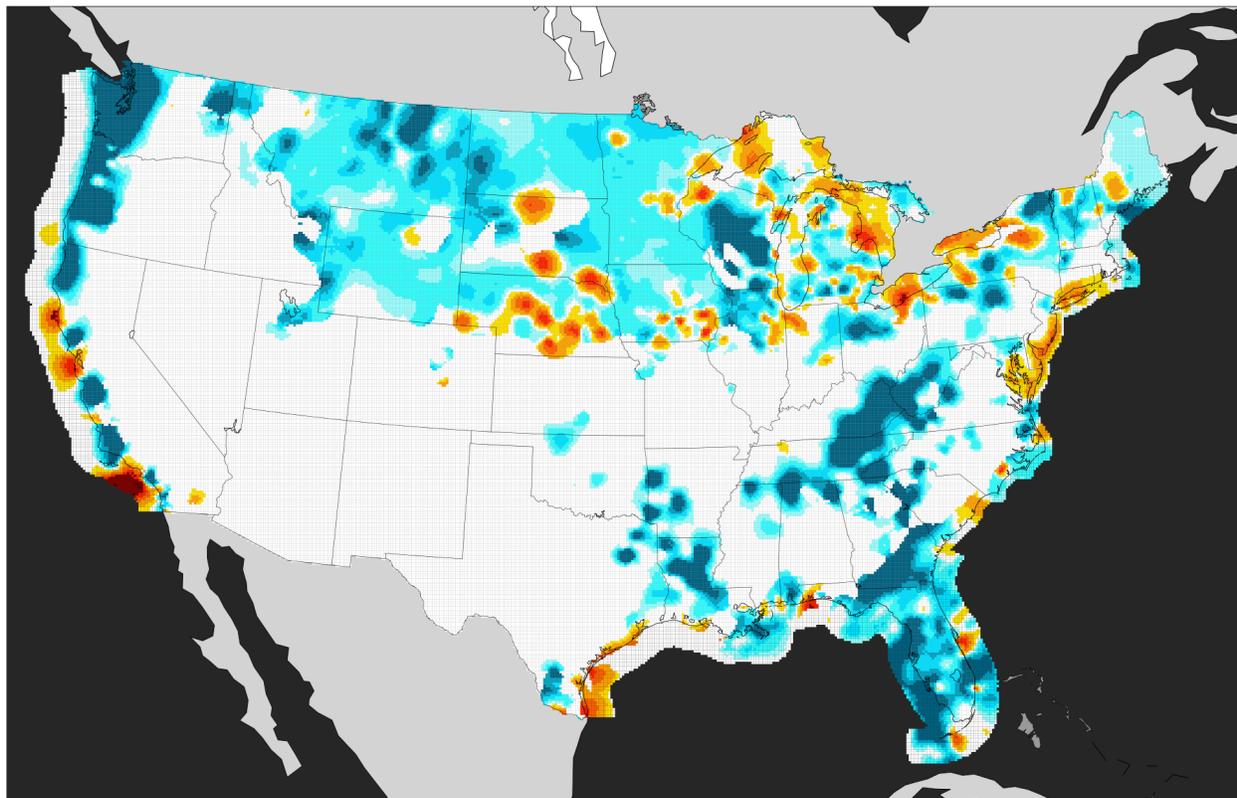


Figure 10. This figure of the difference of the blended sky cover analysis from the composite geostationary satellite sky cover product, in units of percent (%) sky cover, is valid for the one-hour window beginning at 11:00 UTC on 20 October 2013. Blue shades indicate a negative difference, where the blended sky cover analysis value for the shaded point is greater than the corresponding sky cover product value; red shades indicate a positive difference, where the sky cover product value for the shaded point is greater than the corresponding blended sky cover analysis value.

Sun Oct 20 11:00:00 UTC 2013

Satellite/Surface Optimal Sky Cover (%)

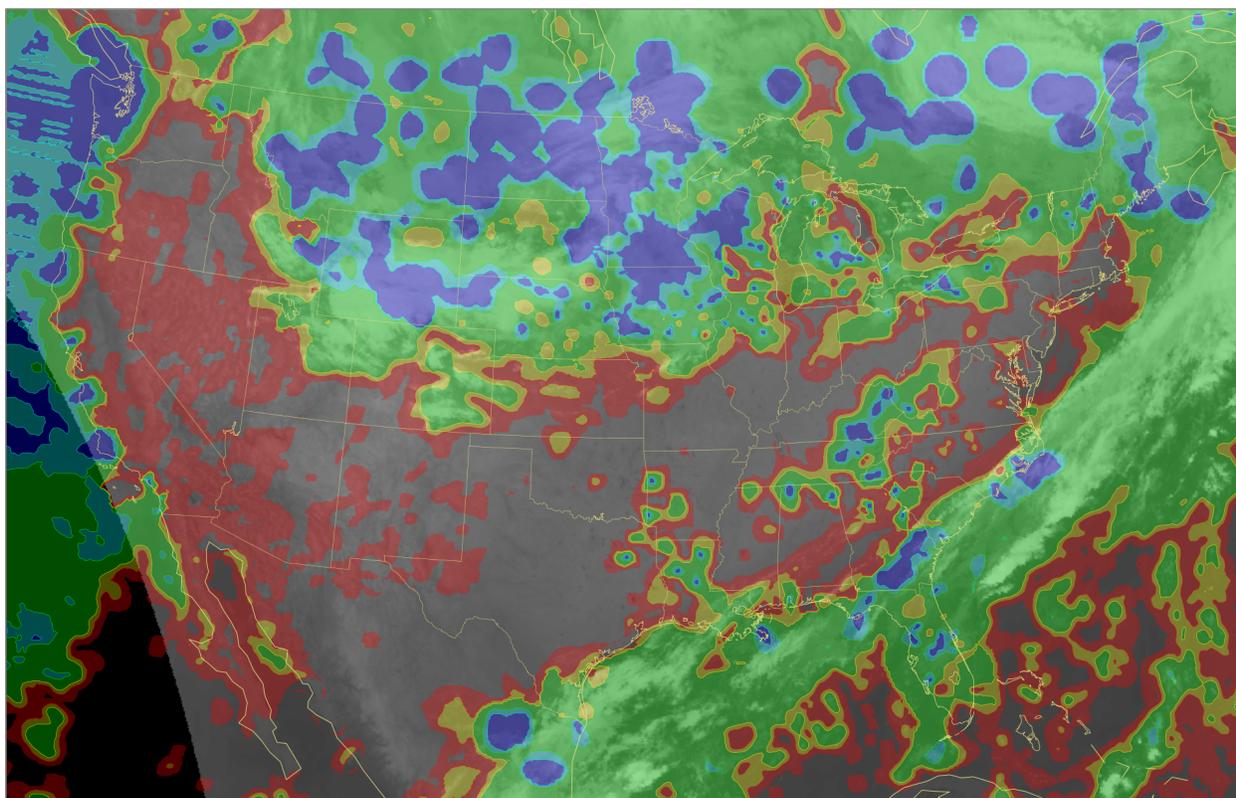


Figure 11. This figure depicts the optimal sky cover analysis, in units of percent (%) sky cover, for the one-hour window beginning at 11:00 UTC on 20 October 2013. Overcast areas are in blue. The infrared window satellite image from GOES-East in the background is valid at 11:02 UTC on 20 October 2013. A traditional enhancement is applied to the satellite image. Cold cloud tops are white. The overlay is partially transparent, except clear areas, which are not filled.

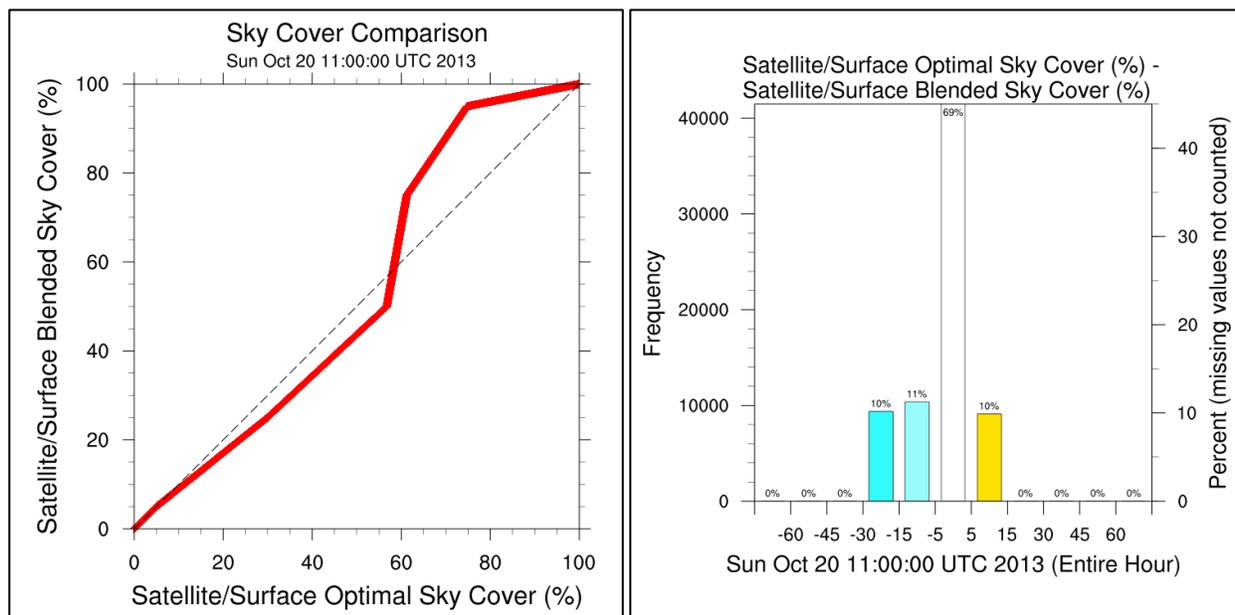


Figure 12. The two panels in this plot are valid for the one-hour window beginning at 11:00 UTC on 20 October 2013. The left panel is a scatterplot with the optimal sky cover analysis on the abscissa and blended sky cover analysis on the ordinate. Both have units of percent (%) sky cover. The right panel is a histogram showing the occurrence frequency of the difference of the blended sky cover analysis from the optimal sky cover analysis between collocated, non-missing points within the increments as indicated on the abscissa. Each increment includes the range of values greater than or equal to the value to the left of the column, if present, and less than the value to the right of the column, if present. The left ordinate axis indicates the frequency in terms of the number of points. The right ordinate axis indicates the percentage of the frequency compared to all non-missing differenced points. This percentage also appears immediately above each bar.

Sun Oct 20 11:00:00 UTC 2013

Satellite/Surface Optimal Sky Cover (%) - Satellite/Surface Blended Sky Cover (%)

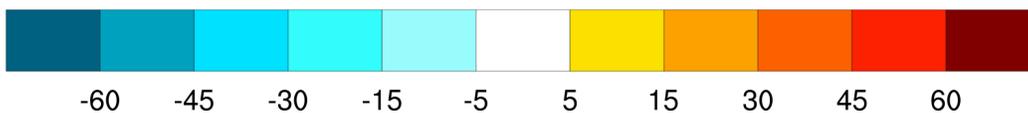
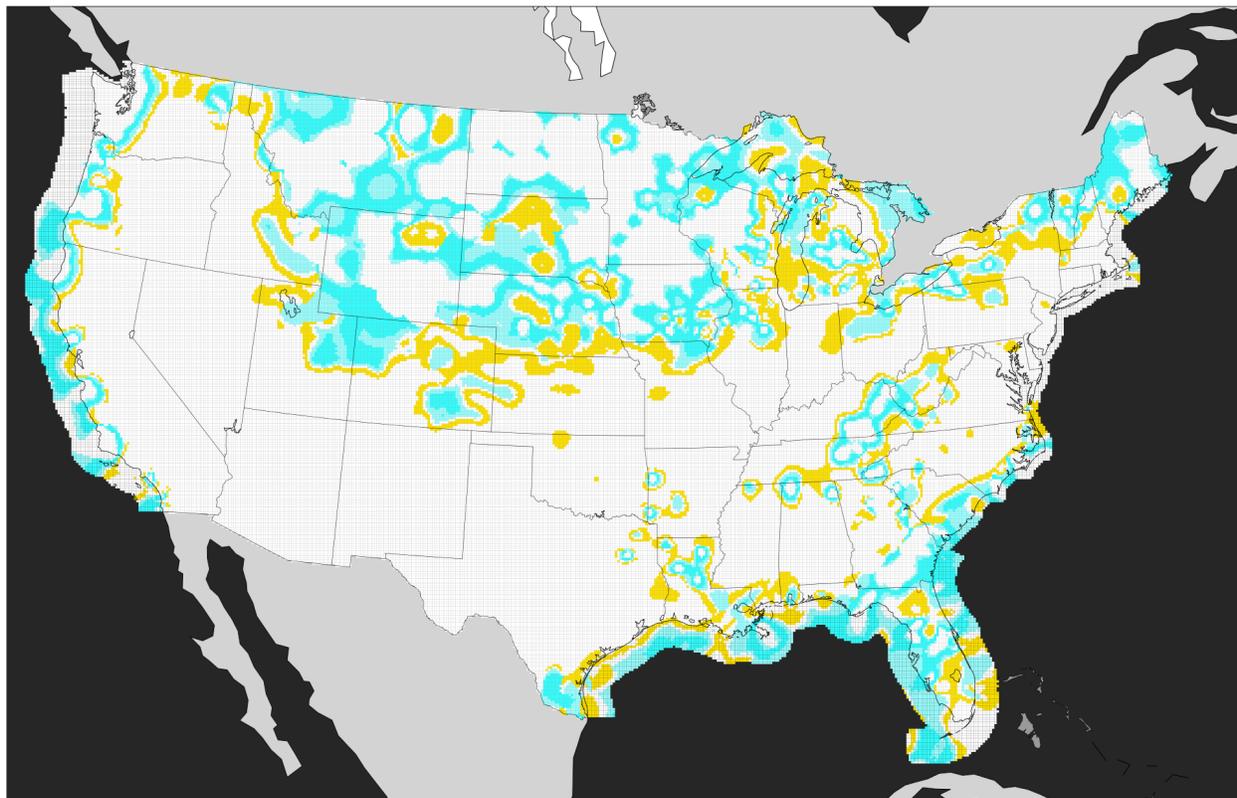


Figure 13. This figure of the difference of the blended sky cover analysis from the optimal sky cover analysis, in units of percent (%) sky cover, is valid for the one-hour window beginning at 11:00 UTC on 20 October 2013. Blue shades indicate a negative difference, where the blended sky cover analysis value for the shaded point is greater than the corresponding optimal sky cover analysis value; red shades indicate a positive difference, where the optimal sky cover analysis value for the shaded point is greater than the corresponding blended sky cover analysis value.

GOES-East Visible Imagery

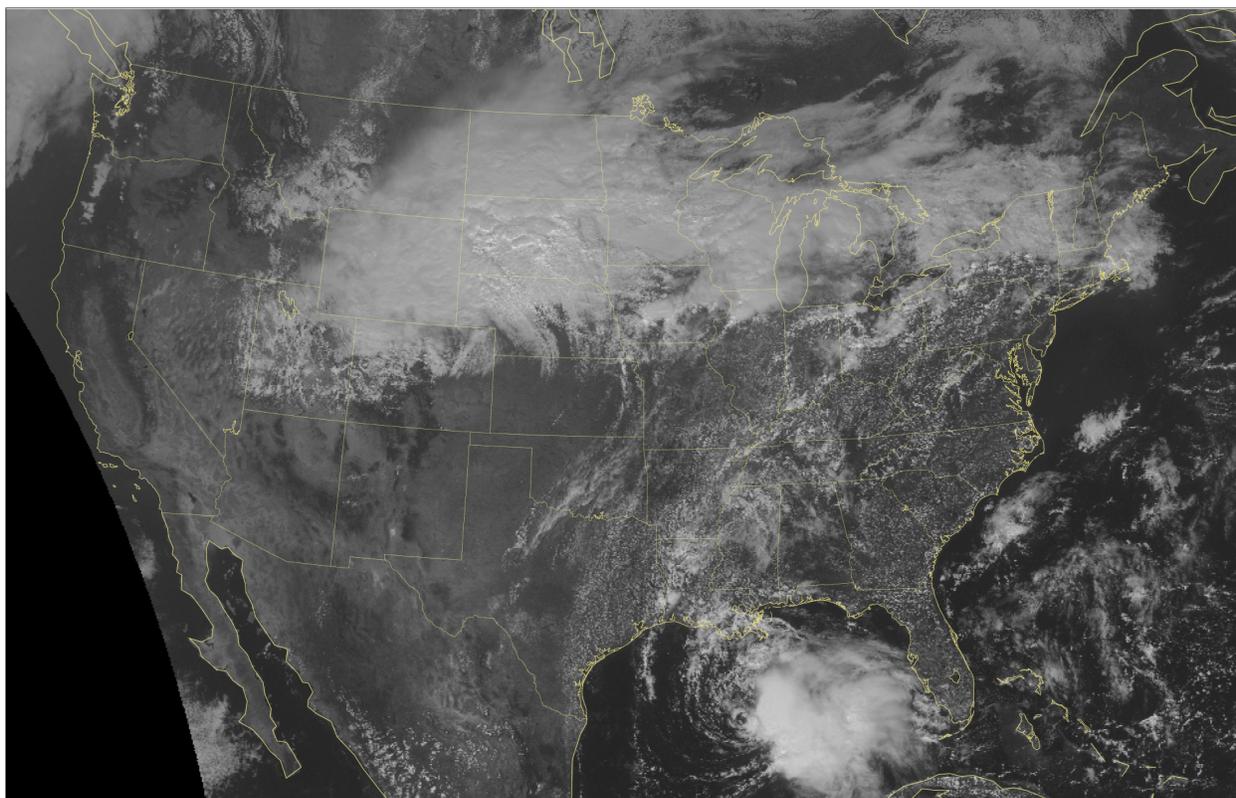


Figure 14. This figure depicts the visible satellite image from GOES-East valid at 18:25 UTC on 4 October 2013. A traditional enhancement is applied to the satellite image. Clouds are white.

Fri Oct 4 18:00:00 UTC 2013

Satellite/Surface Blended Sky Cover (%)

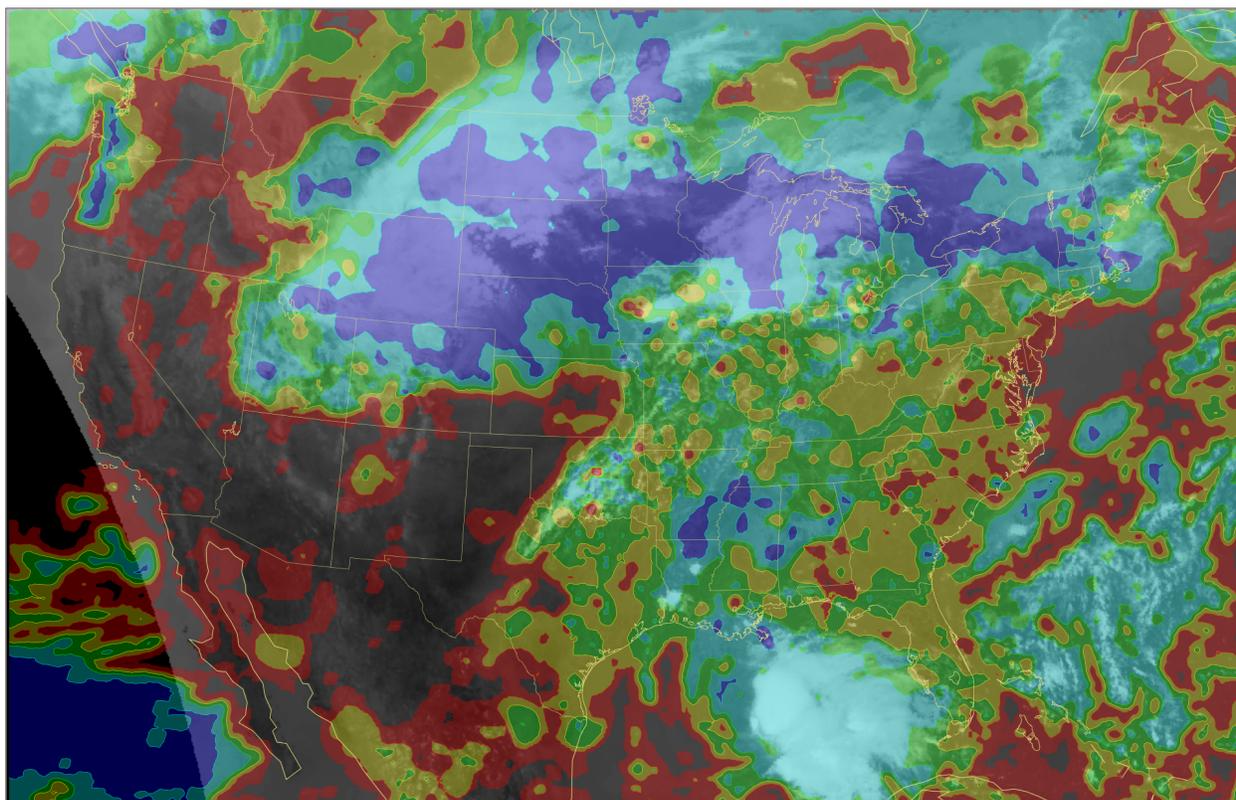


Figure 15. This figure depicts the blended sky cover analysis, in units of percent (%) sky cover, for the one-hour window beginning at 18:00 UTC on 4 October 2013. Overcast areas are in blue. The infrared window satellite image from GOES-East in the background is valid at 18:25 UTC on 4 October 2013. A traditional enhancement is applied to the satellite image. Cold cloud tops are white. The overlay is partially transparent, except clear areas, which are not filled.

Fri Oct 4 18:00:00 UTC 2013

NDFD Total Cloud Cover (%)

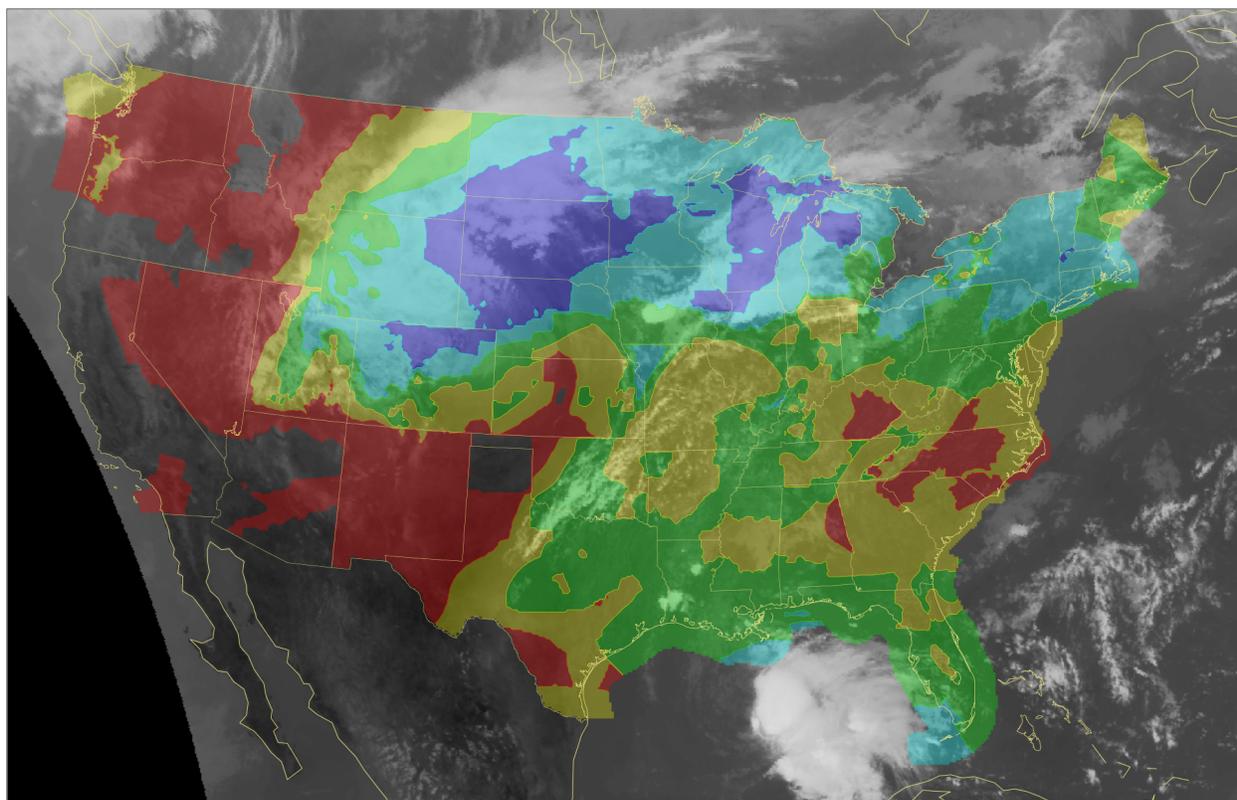


Figure 16. This figure depicts the NDFD total cloud cover one-hour forecast, in units of percent (%) sky cover, for the one-hour window beginning at 18:00 UTC on 4 October 2013. Overcast areas are in blue. The infrared window satellite image from GOES-East in the background is valid at 18:25 UTC on 4 October 2013. A traditional enhancement is applied to the satellite image. Cold cloud tops are white. The overlay is partially transparent, except clear areas, and areas outside of the domain, which are not filled.

Fri Oct 4 18:00:00 UTC 2013

HRRR Total Cloud Cover (%)

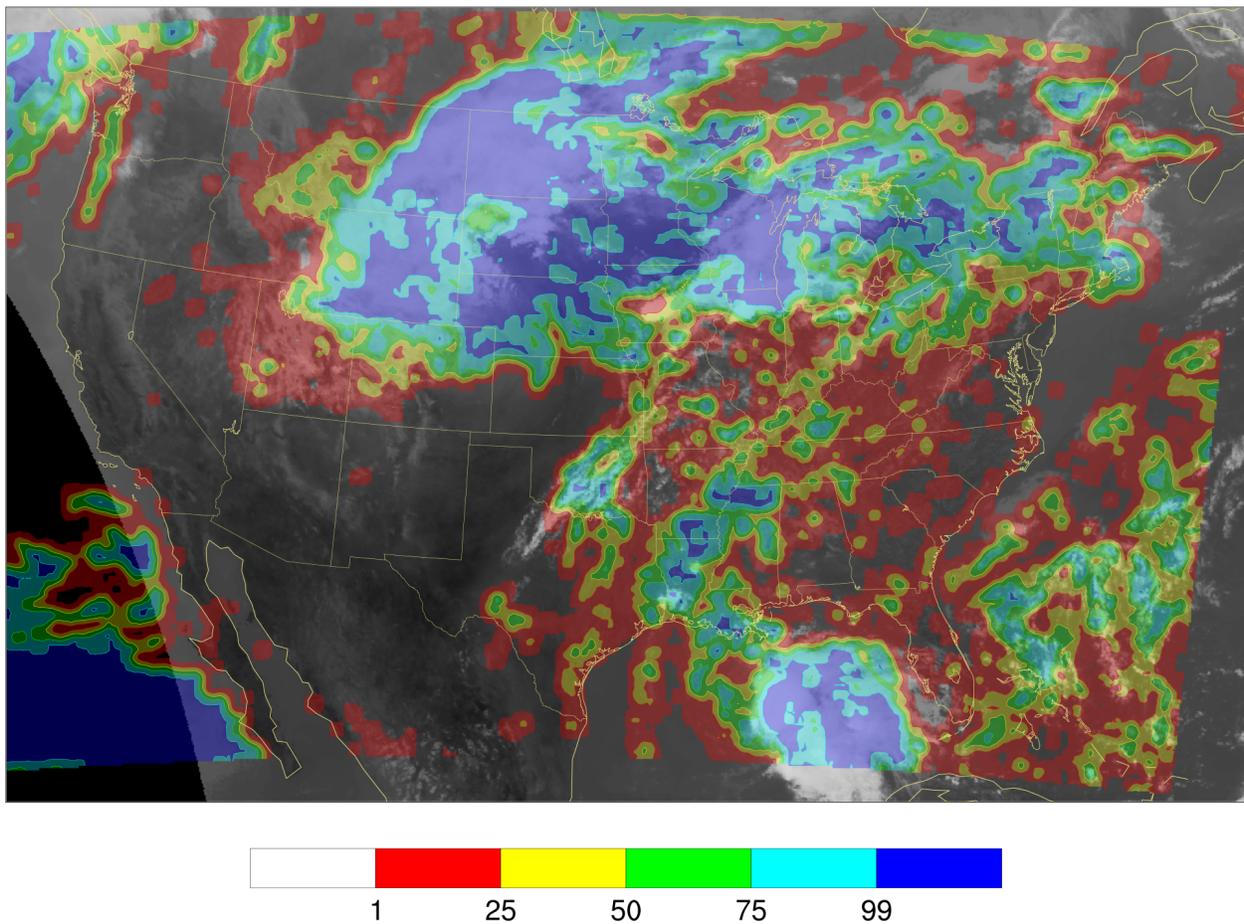


Figure 17. This figure depicts the HRRR total cloud cover analysis, in units of percent (%) cloud cover, valid at 18:00 UTC on 4 October 2013. Overcast areas are in blue. The infrared window satellite image from GOES-East in the background is valid at 18:25 UTC on 4 October 2013. A traditional enhancement is applied to the satellite image. Cold cloud tops are white. The overlay is partially transparent, except clear areas, and areas outside of the domain, which are not filled.

Fri Oct 4 18:00:00 UTC 2013

HRRR Optimal Sky Cover (%)

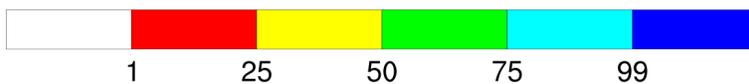
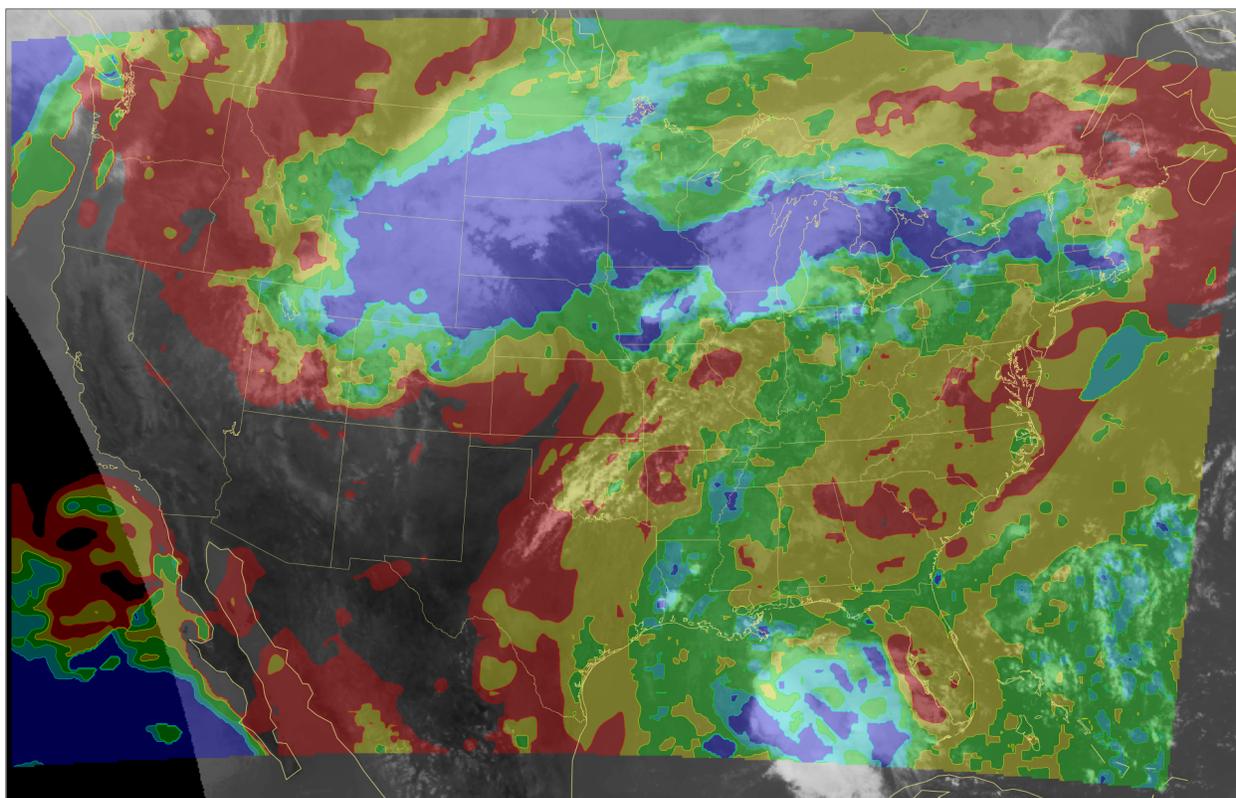


Figure 18. This figure depicts the initial-hour HRRR optimal sky cover product, in units of percent (%) sky cover, for the one-hour window beginning at 18:00 UTC on 4 October 2013. Overcast areas are in blue. The infrared window satellite image from GOES-East in the background is valid at 18:25 UTC on 4 October 2013. A traditional enhancement is applied to the satellite image. Cold cloud tops are white. The overlay is partially transparent, except clear areas, and areas outside of the domain, which are not filled.

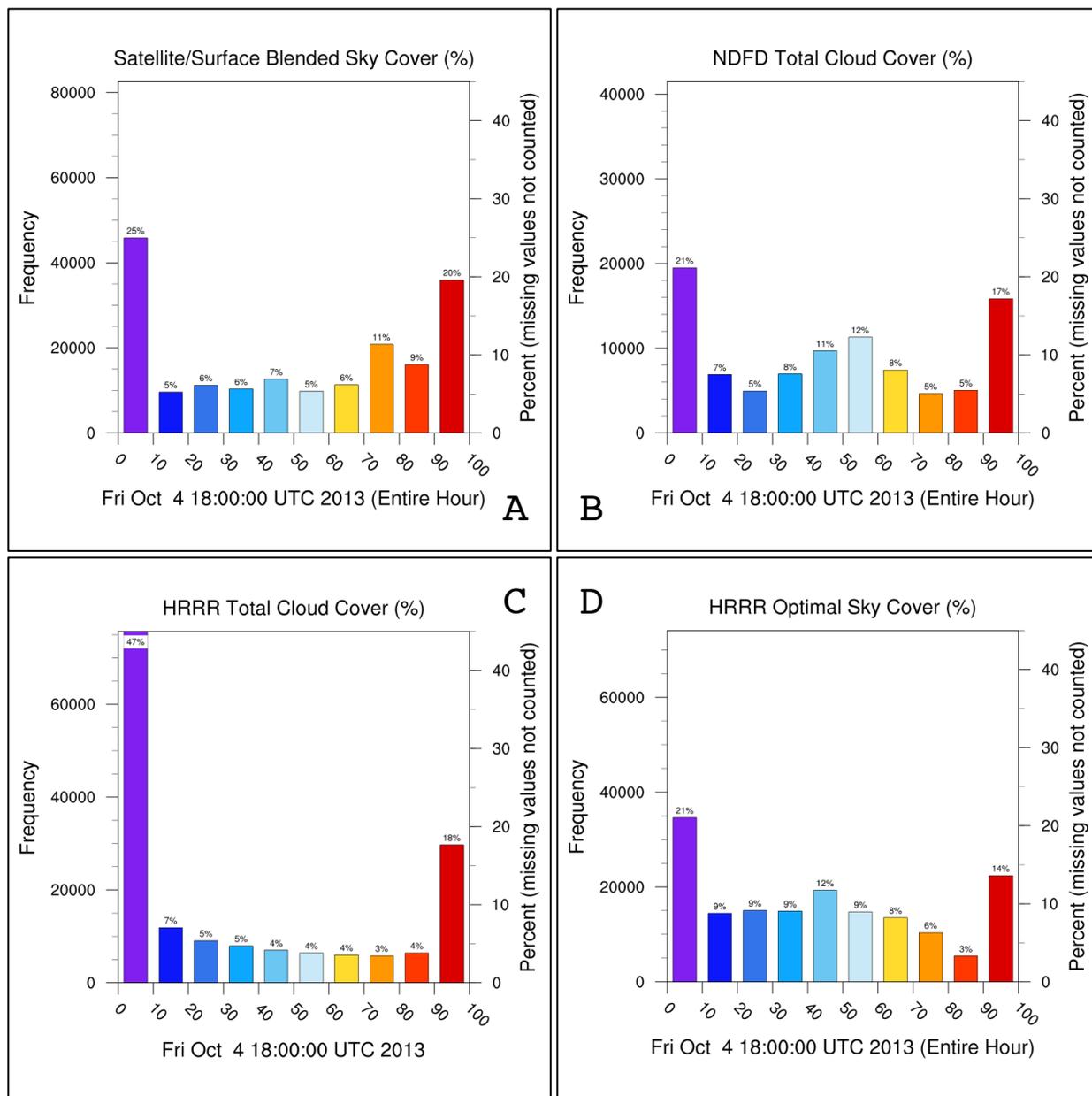


Figure 19. All of the histograms in this four-panel plot are valid for the one-hour window beginning at 18:00 UTC on 4 October 2013, except Panel C, which is valid at the top of the hour. Each histogram shows the frequency at which the value range, as indicated on the abscissa, occurs among all non-missing points in the indicated product. Each increment includes the range of values greater than or equal to the value to the left of the column and less than the value to the right of the column, except for the rightmost bin, which includes points with a value of 100%. The left ordinate axis indicates the frequency in terms of the number of points. The right ordinate axis indicates the percentage of the frequency compared to all non-missing points. This percentage also appears immediately above each bar. Panel A depicts the distribution for the blended sky cover analysis. Panel B depicts the distribution for the NDFD total cloud cover one-hour forecast. Panel C depicts the distribution for the HRRR total cloud cover analysis. Panel D depicts the distribution for the initial-hour HRRR optimal sky cover product.

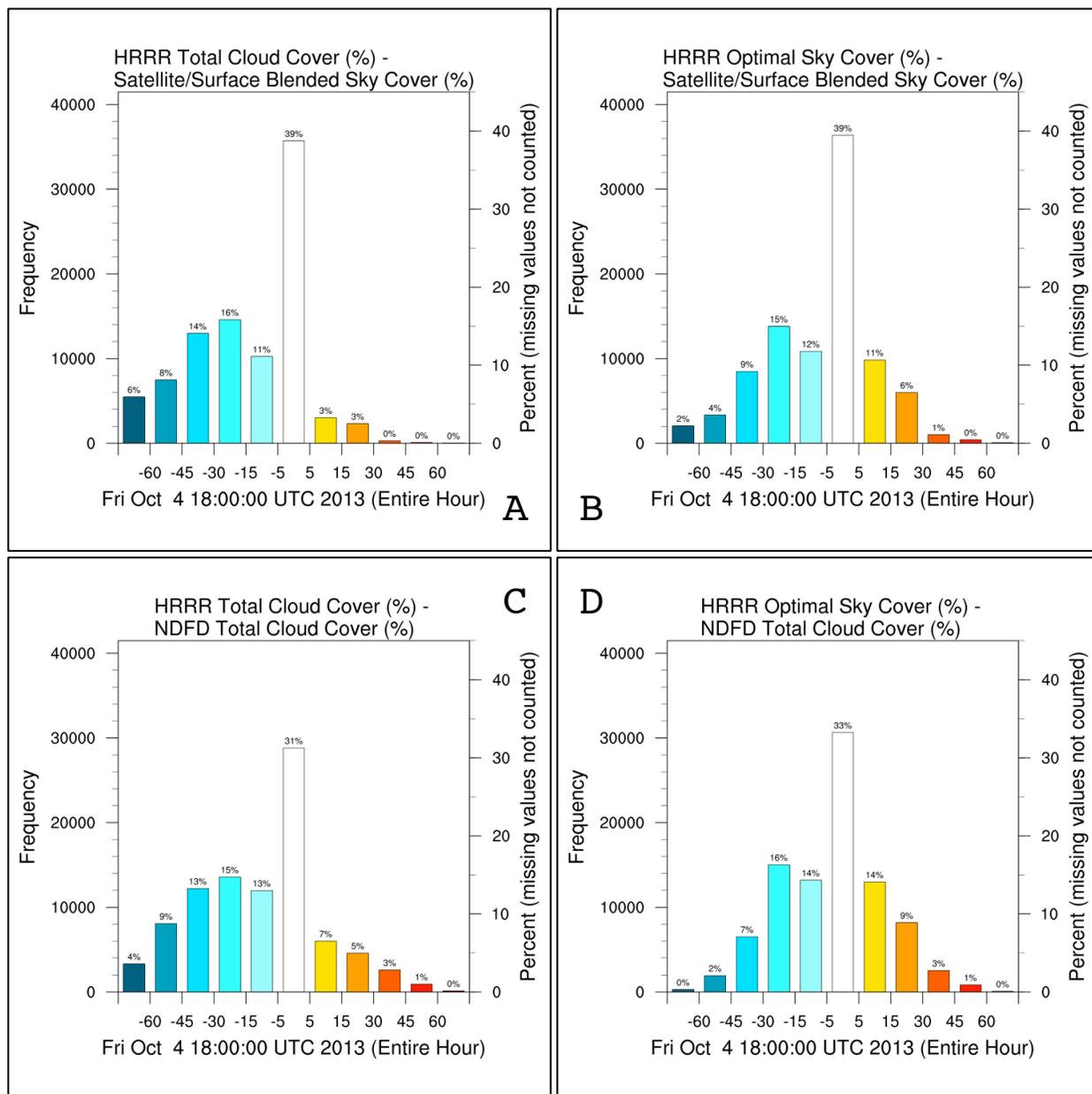


Figure 20. All of the histograms in this four-panel plot are valid for the one-hour window beginning at 18:00 UTC on 4 October 2013. Each histogram shows the occurrence frequency of differences between collocated, non-missing points within the increments as indicated on the abscissa. Each increment includes the range of values greater than or equal to the value to the left of the column, if present, and less than the value to the right of the column, if present. The left ordinate axis indicates the frequency in terms of the number of points. The right ordinate axis indicates the percentage of the frequency compared to all non-missing differenced points. This percentage also appears immediately above each bar. Panel A is the difference of the blended sky cover analysis from the HRRR total cloud cover analysis. Panel B is the difference of the blended sky cover analysis from the initial-hour HRRR optimal sky cover product. Panels C and D are the same as for panels A and B, respectively, except for using the NDFD total cloud cover one-hour forecast instead of the blended sky cover analysis.

Fri Oct 4 18:00:00 UTC 2013

HRRR Optimal Sky Cover (%) - Satellite/Surface Blended Sky Cover (%)

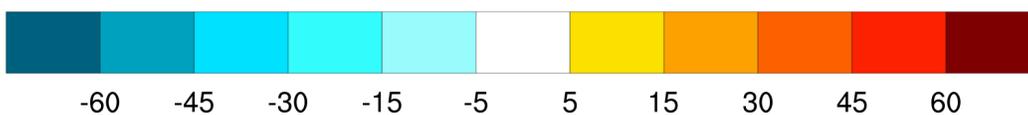
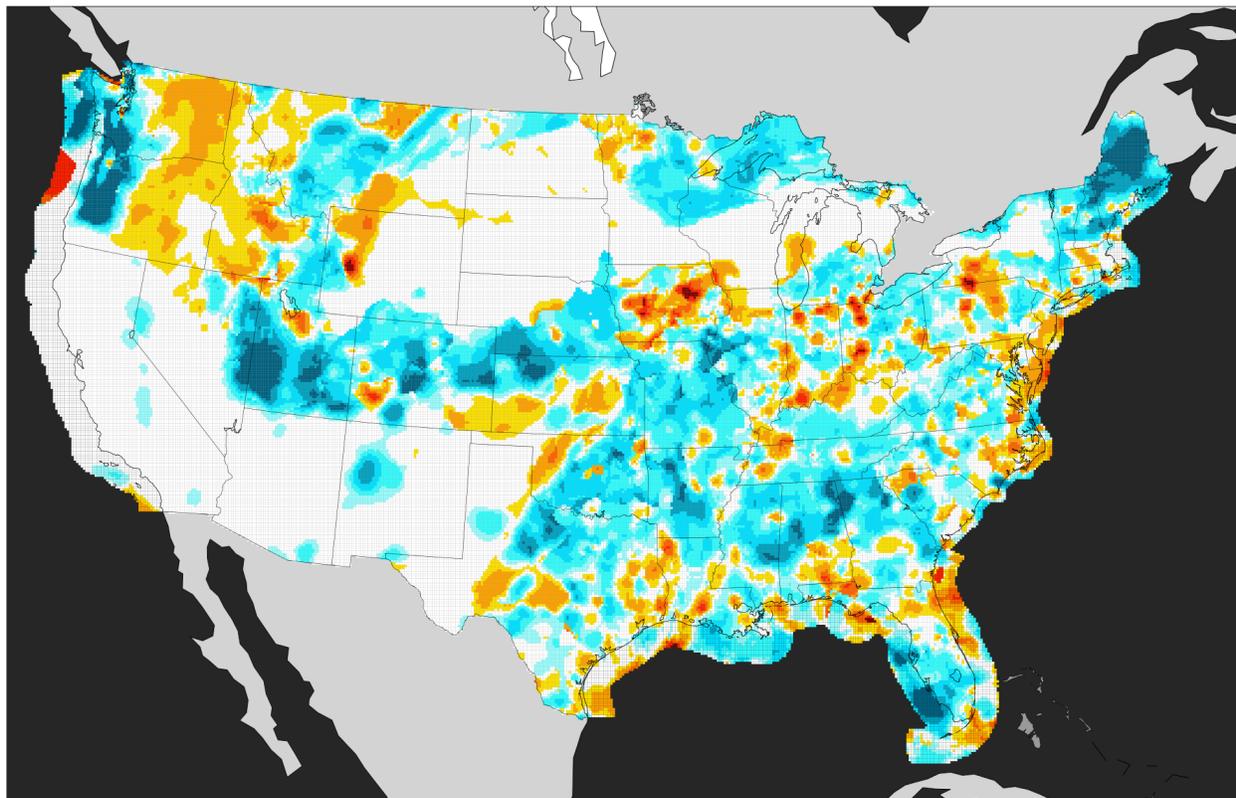


Figure 21. This figure of the difference of the blended sky cover analysis from the initial-hour HRRR optimal sky cover product, in units of percent (%) sky cover, is valid for the one-hour window beginning at 18:00 UTC on 4 October 2013. Blue shades indicate a negative difference, where the blended sky cover analysis value for the shaded point is greater than the corresponding HRRR optimal sky cover product value; red shades indicate a positive difference, where the HRRR optimal sky cover product value for the shaded point is greater than the corresponding blended sky cover analysis value.

Fri Oct 4 18:00:00 UTC 2013

HRRR Optimal Sky Cover (%) - NDFD Total Cloud Cover (%)

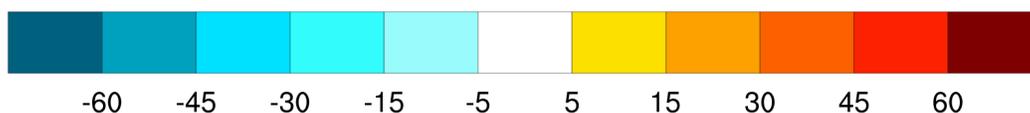
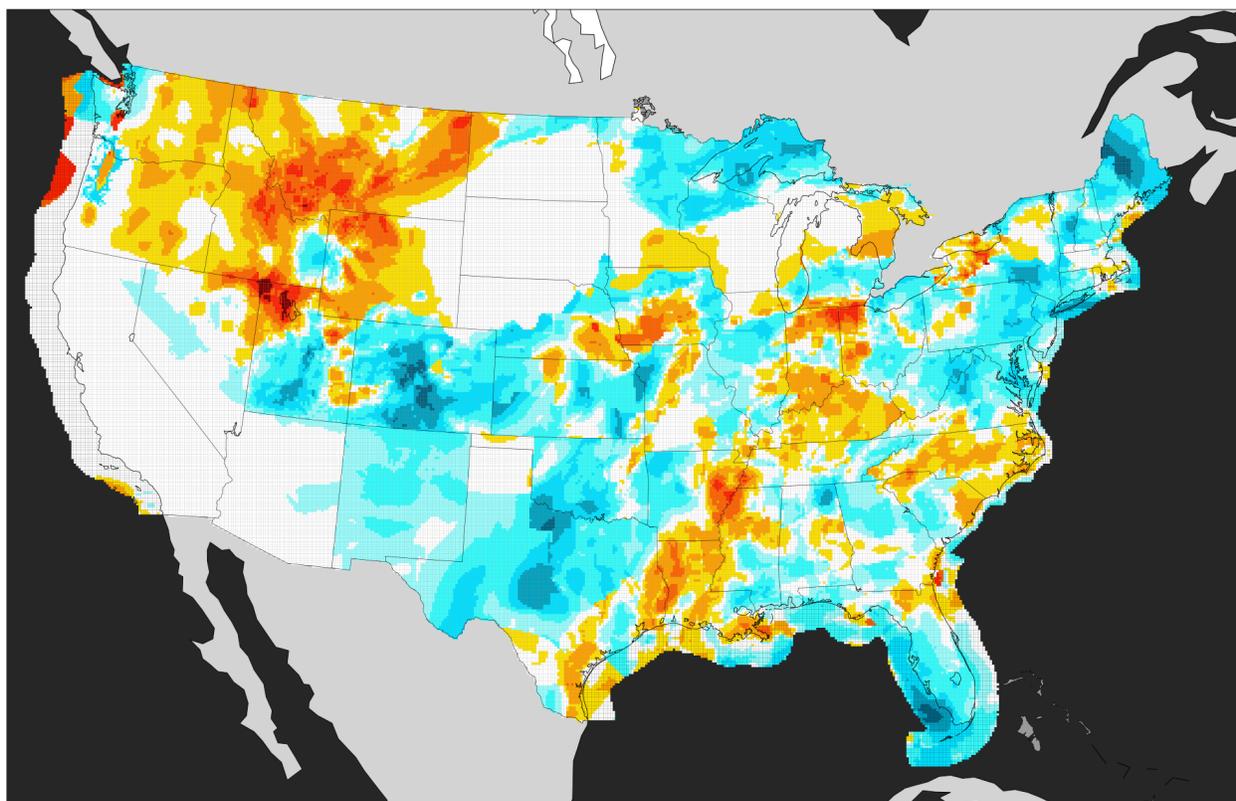


Figure 22. This figure of the difference of the NDFD total cloud cover one-hour forecast from the initial-hour HRRR optimal sky cover product, in units of percent (%) sky cover, is valid for the one-hour window beginning at 18:00 UTC on 4 October 2013. Blue shades indicate a negative difference, where the NDFD total cloud cover one-hour forecast value for the shaded point is greater than the corresponding HRRR optimal sky cover product value; red shades indicate a positive difference, where the HRRR optimal sky cover product value for the shaded point is greater than the corresponding NDFD total cloud cover one-hour forecast value.

GOES-East CONUS IR Window Image

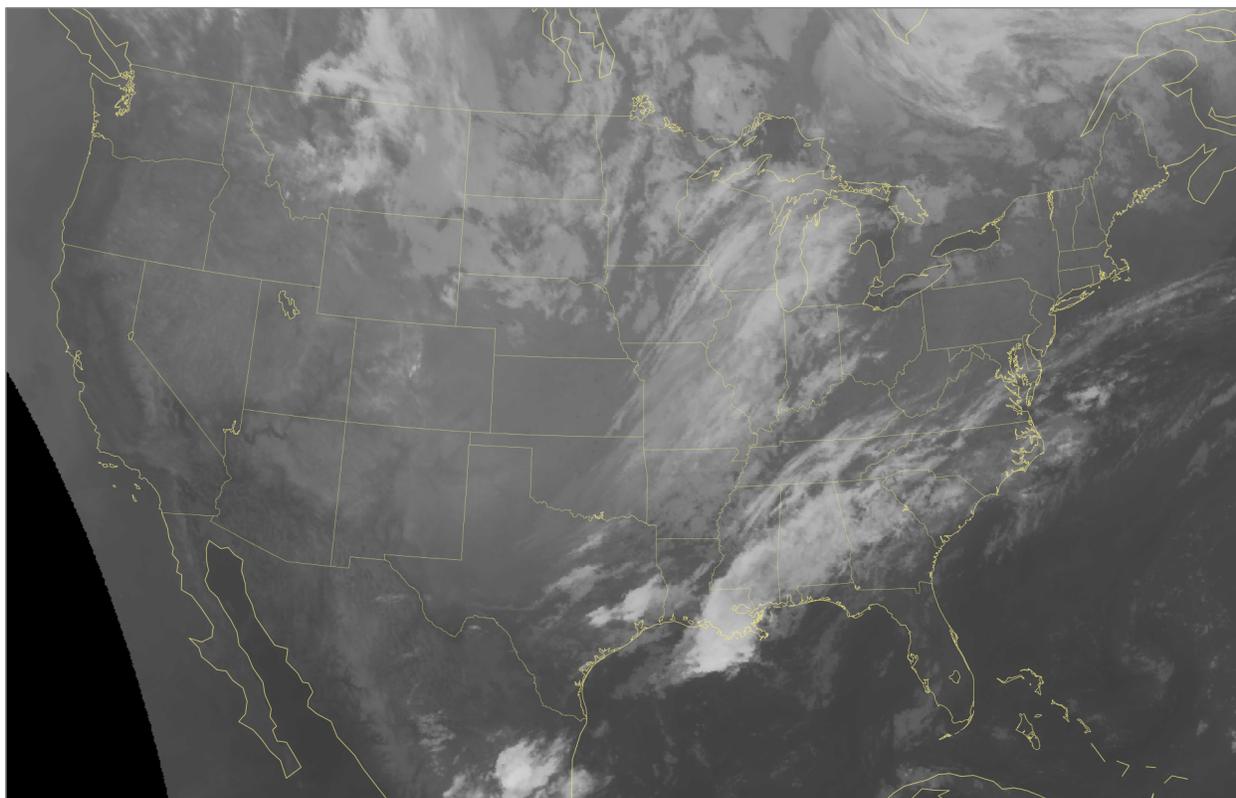


Figure 23. This figure depicts the infrared window satellite image from GOES-East valid at 3:00 UTC on 19 October 2013. A traditional enhancement is applied to the satellite image. Cold cloud tops are white.

Sat Oct 19 03:00:00 UTC 2013

Satellite/Surface Blended Sky Cover (%)

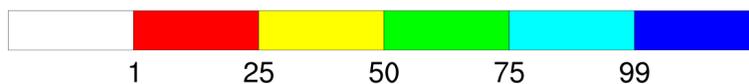
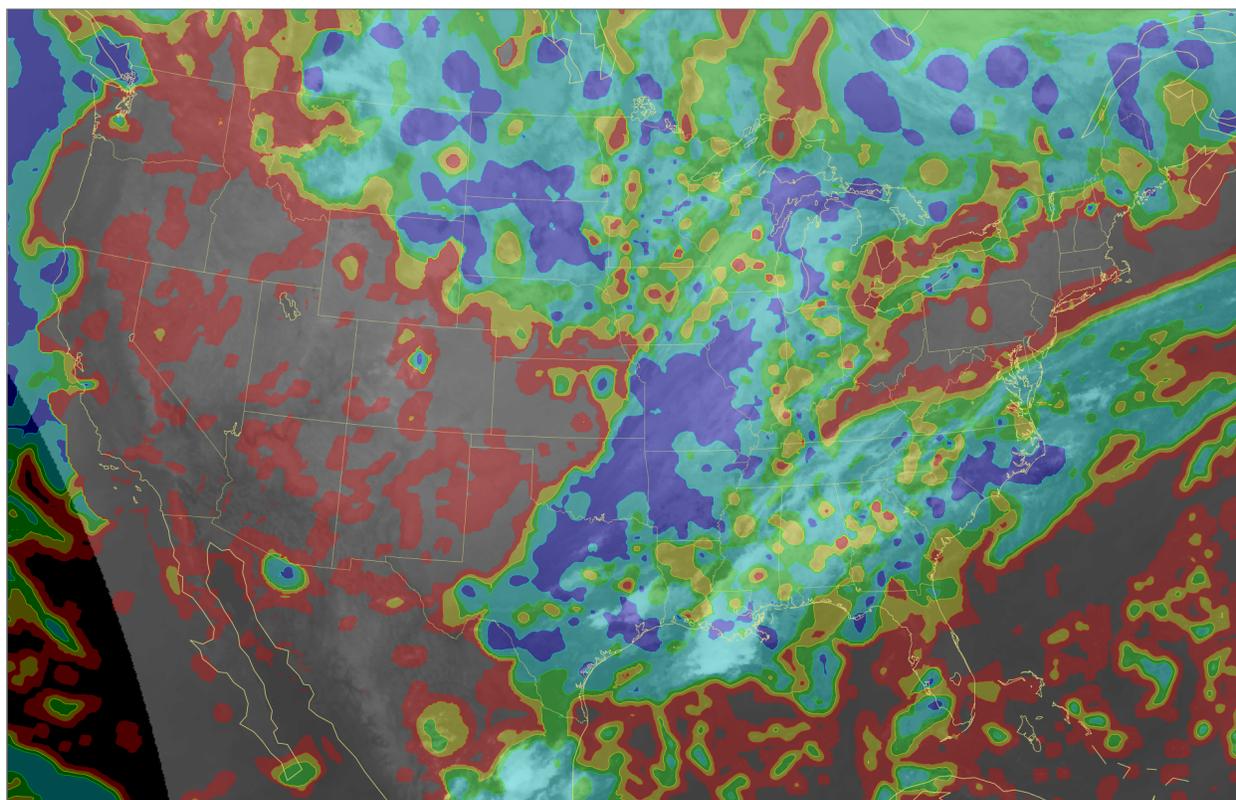


Figure 24. This figure depicts the blended sky cover analysis, in units of percent (%) sky cover, for the one-hour window beginning at 3:00 UTC on 19 October 2013. Overcast areas are in blue. The infrared window satellite image from GOES-East in the background is valid at 3:00 UTC on 19 October 2013. A traditional enhancement is applied to the satellite image. Cold cloud tops are white. The overlay is partially transparent, except clear areas, which are not filled.

Sat Oct 19 03:00:00 UTC 2013

NDFD Total Cloud Cover (%)

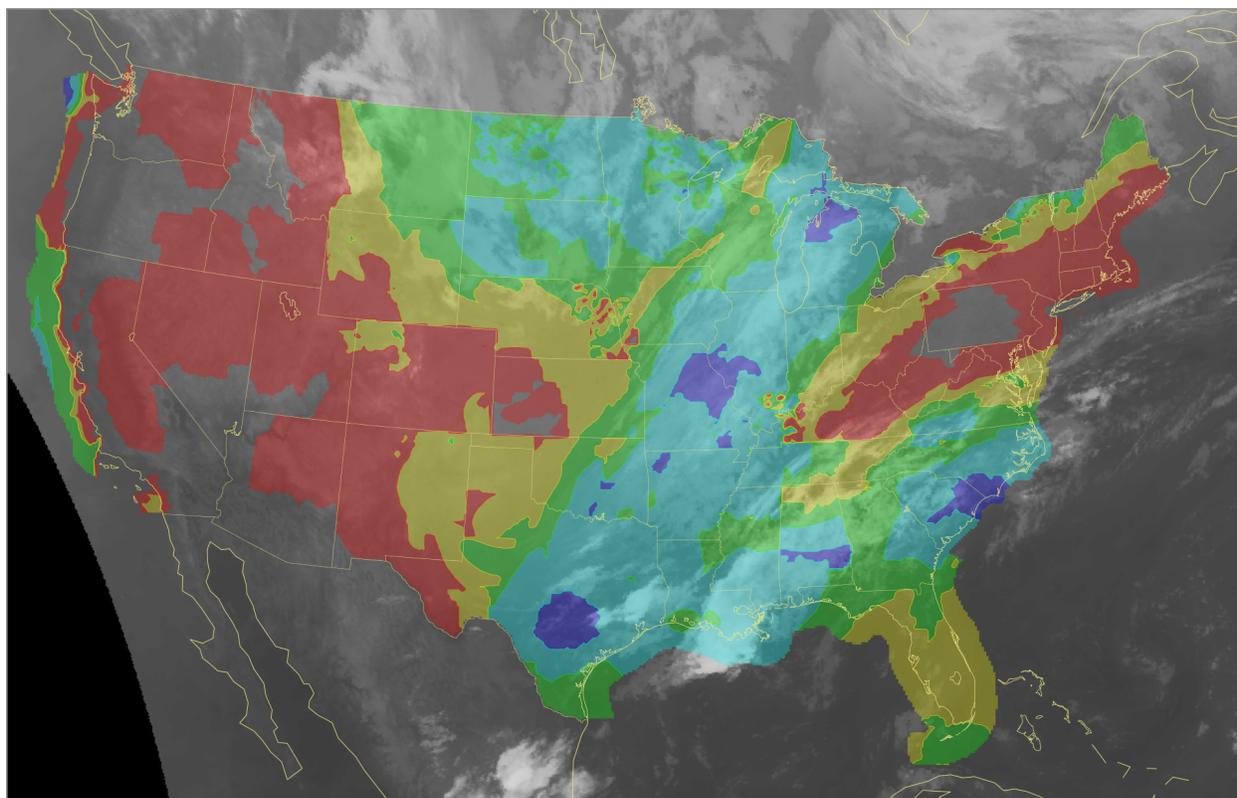


Figure 25. This figure depicts the NDFD total cloud cover one-hour forecast, in units of percent (%) sky cover, for the one-hour window beginning at 3:00 UTC on 19 October 2013. Overcast areas are in blue. The infrared window satellite image from GOES-East in the background is valid at 3:00 UTC on 19 October 2013. A traditional enhancement is applied to the satellite image. Cold cloud tops are white. The overlay is partially transparent, except clear areas, and areas outside of the domain, which are not filled.

Sat Oct 19 03:00:00 UTC 2013

HRRR Total Cloud Cover (%)

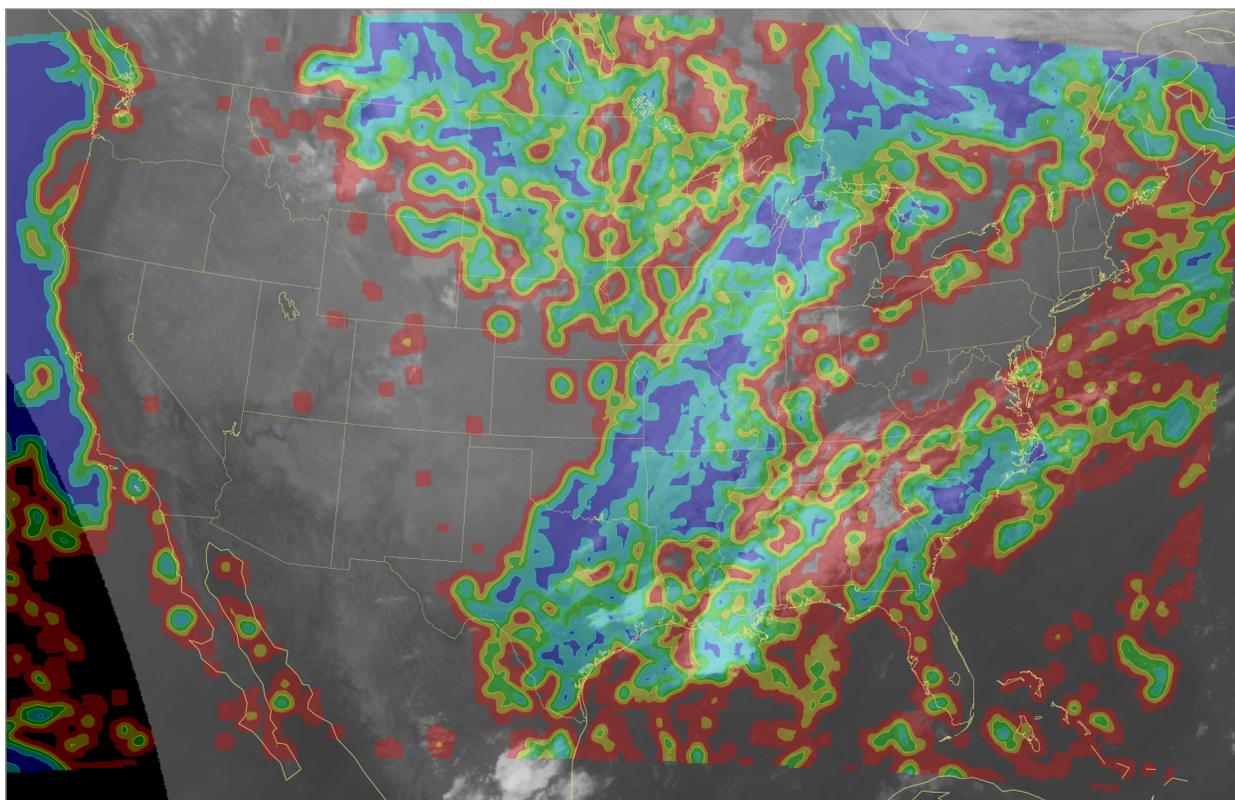


Figure 26. This figure depicts the HRRR total cloud cover analysis, in units of percent (%) cloud cover, valid at 3:00 UTC on 19 October 2013. Overcast areas are in blue. The infrared window satellite image from GOES-East in the background is valid at 3:00 UTC on 19 October 2013. A traditional enhancement is applied to the satellite image. Cold cloud tops are white. The overlay is partially transparent, except clear areas, and areas outside of the domain, which are not filled.

Sat Oct 19 03:00:00 UTC 2013

HRRR Optimal Sky Cover (%)

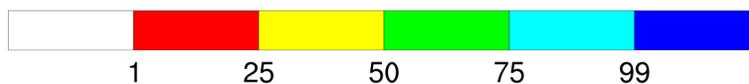
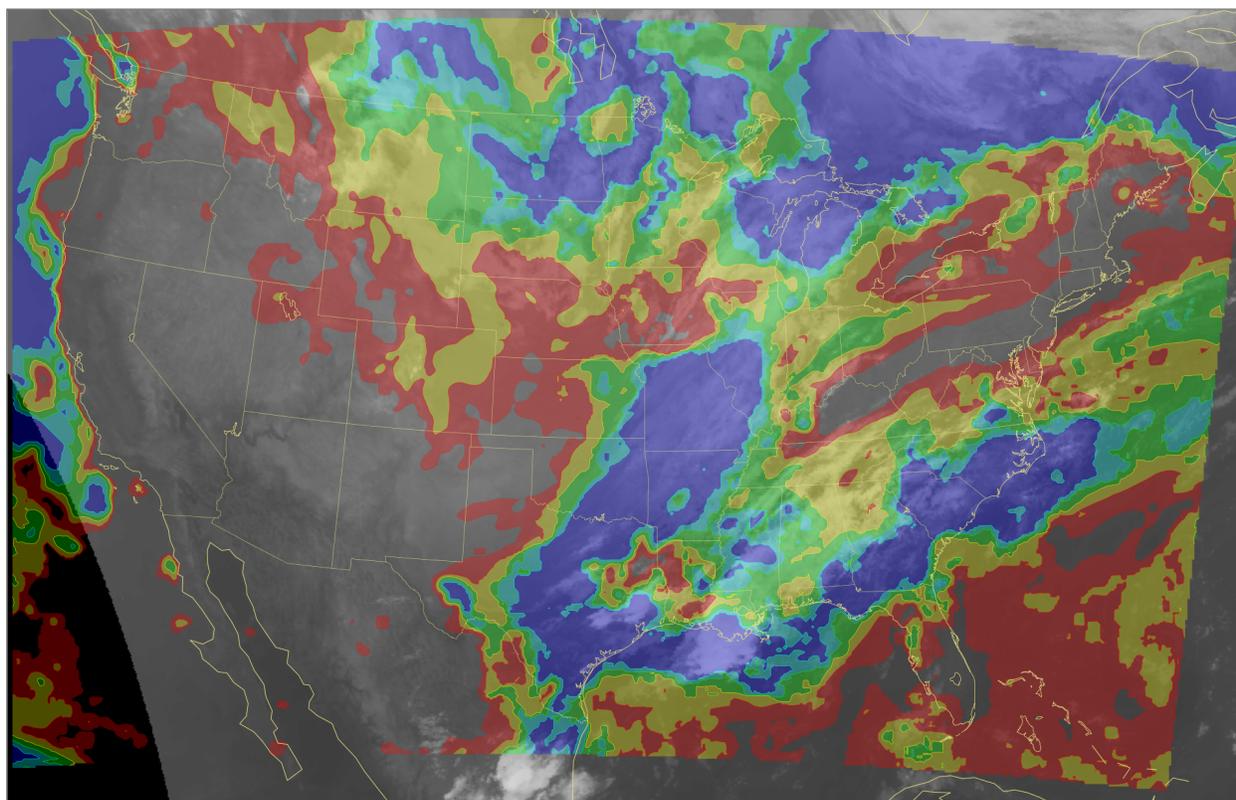


Figure 27. This figure depicts the initial-hour HRRR optimal sky cover product, in units of percent (%) sky cover, for the one-hour window beginning at 3:00 UTC on 19 October 2013. Overcast areas are in blue. The infrared window satellite image from GOES-East in the background is valid at 3:00 UTC on 19 October 2013. A traditional enhancement is applied to the satellite image. Cold cloud tops are white. The overlay is partially transparent, except clear areas, and areas outside of the domain, which are not filled.

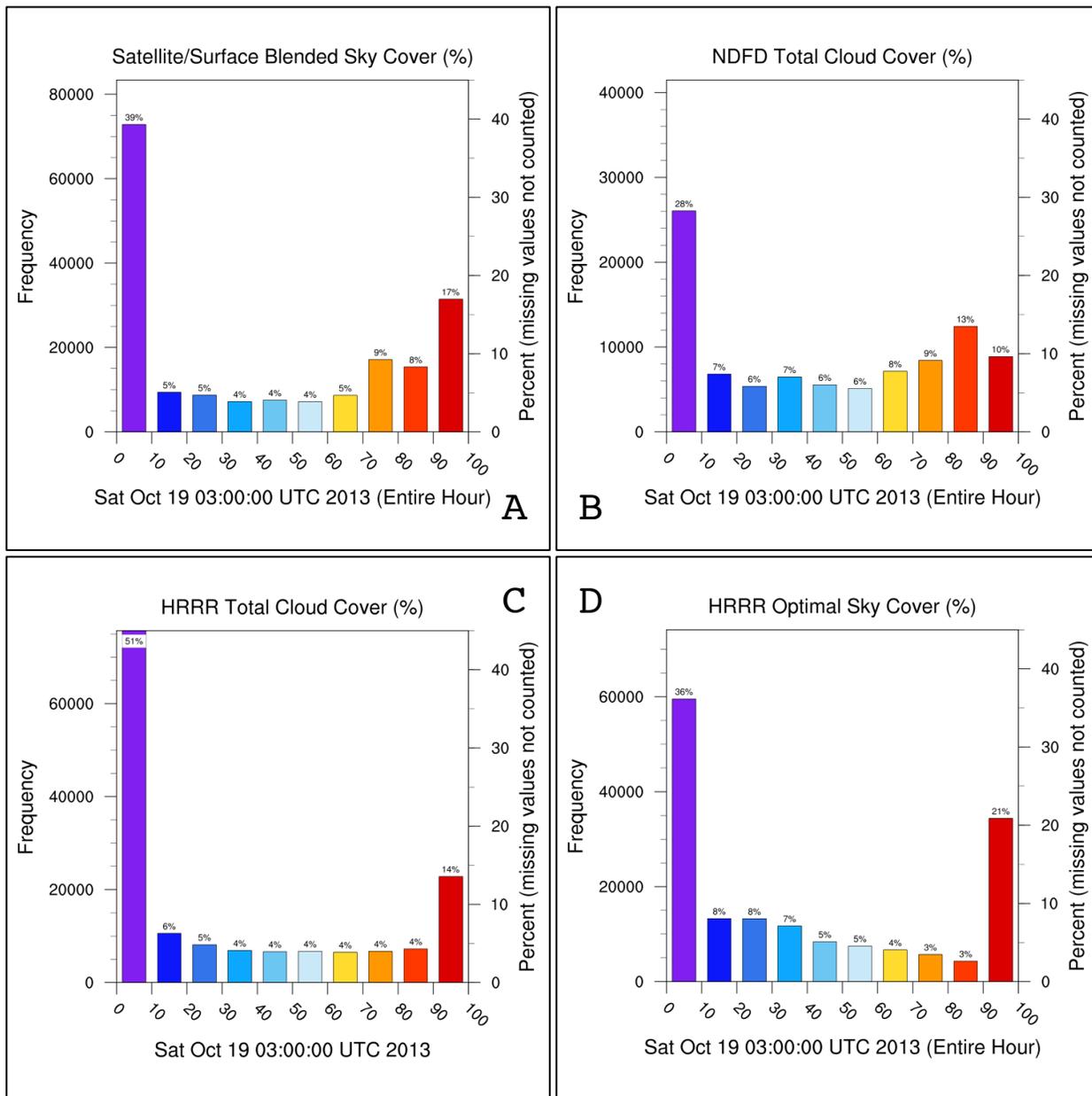


Figure 28. All of the histograms in this four-panel plot are valid for the one-hour window beginning at 3:00 UTC on 19 October 2013, except Panel C, which is valid at the top of the hour. The description of the axes and panels is found in the caption of Figure 19.

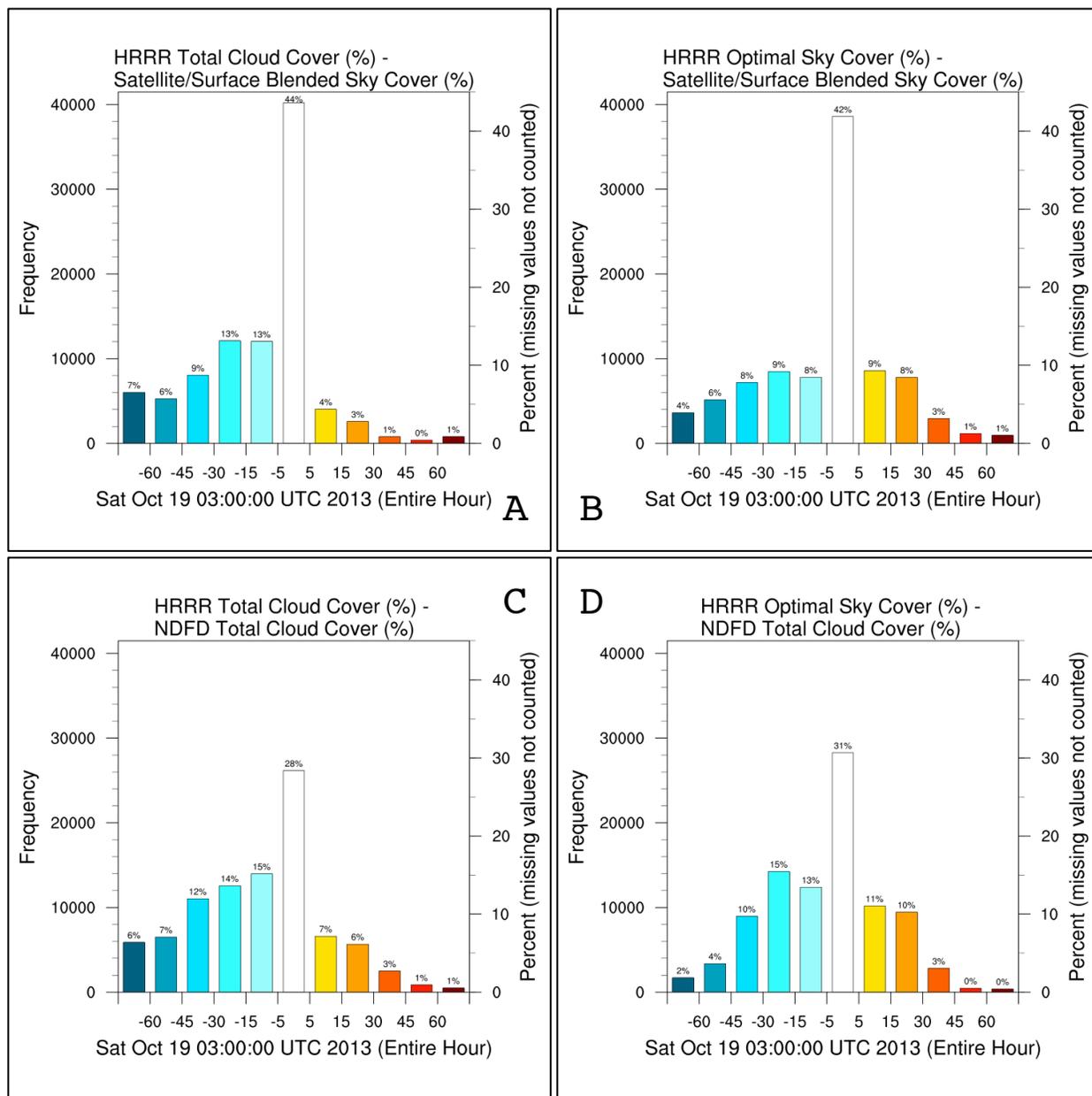


Figure 29. All of the histograms in this four-panel plot are valid for the one-hour window beginning at 3:00 UTC on 19 October 2013. The description of the axes and panels is found in the caption of Figure 20.

Sat Oct 19 03:00:00 UTC 2013

HRRR Optimal Sky Cover (%) - Satellite/Surface Blended Sky Cover (%)

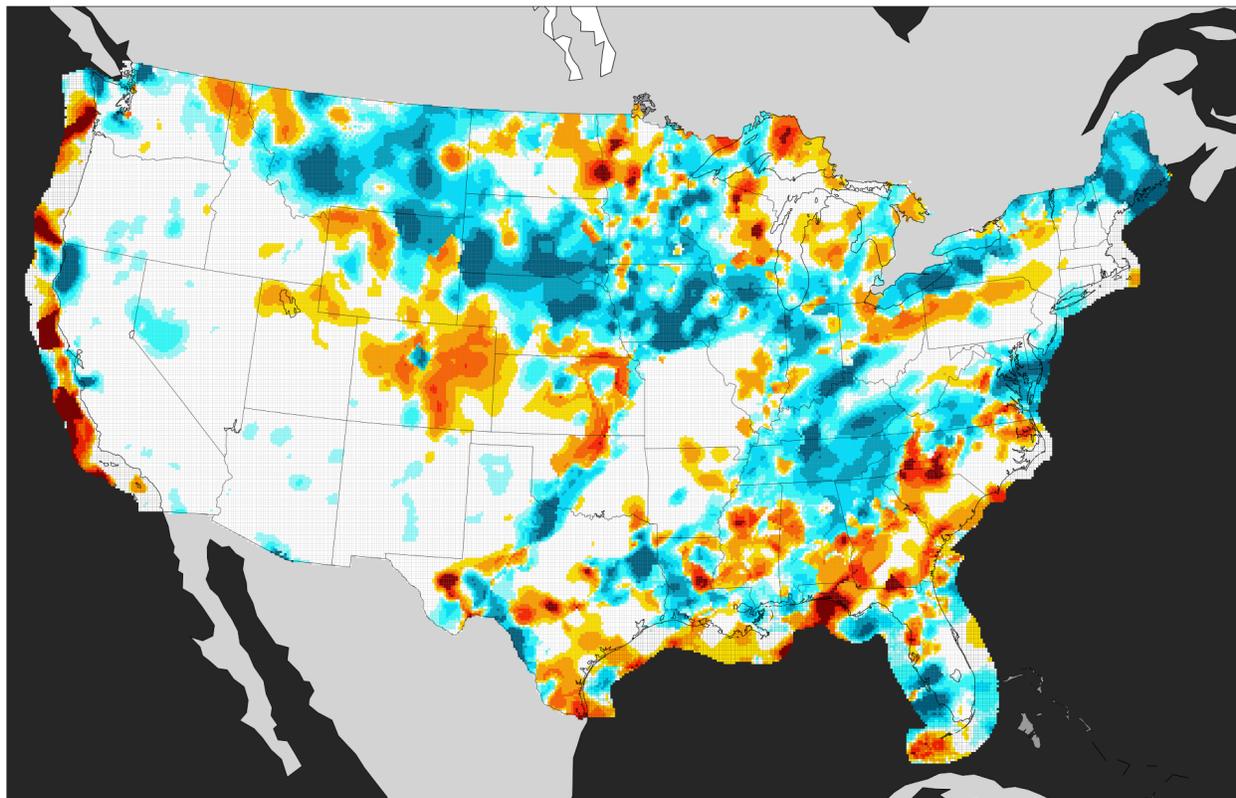


Figure 30. This figure of the difference of the blended sky cover analysis from the initial-hour HRRR optimal sky cover product, in units of percent (%) sky cover, is valid for the one-hour window beginning at 3:00 UTC on 19 October 2013. Blue shades indicate a negative difference, where the blended sky cover analysis value for the shaded point is greater than the corresponding HRRR optimal sky cover product value; red shades indicate a positive difference, where the HRRR optimal sky cover product value for the shaded point is greater than the corresponding blended sky cover analysis value.

Sat Oct 19 03:00:00 UTC 2013

HRRR Optimal Sky Cover (%) - NDFD Total Cloud Cover (%)

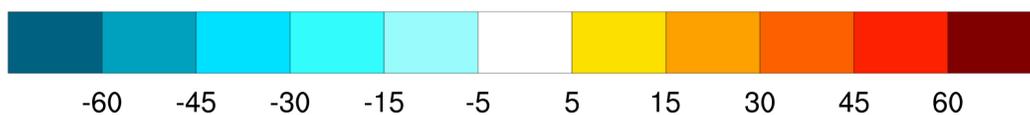
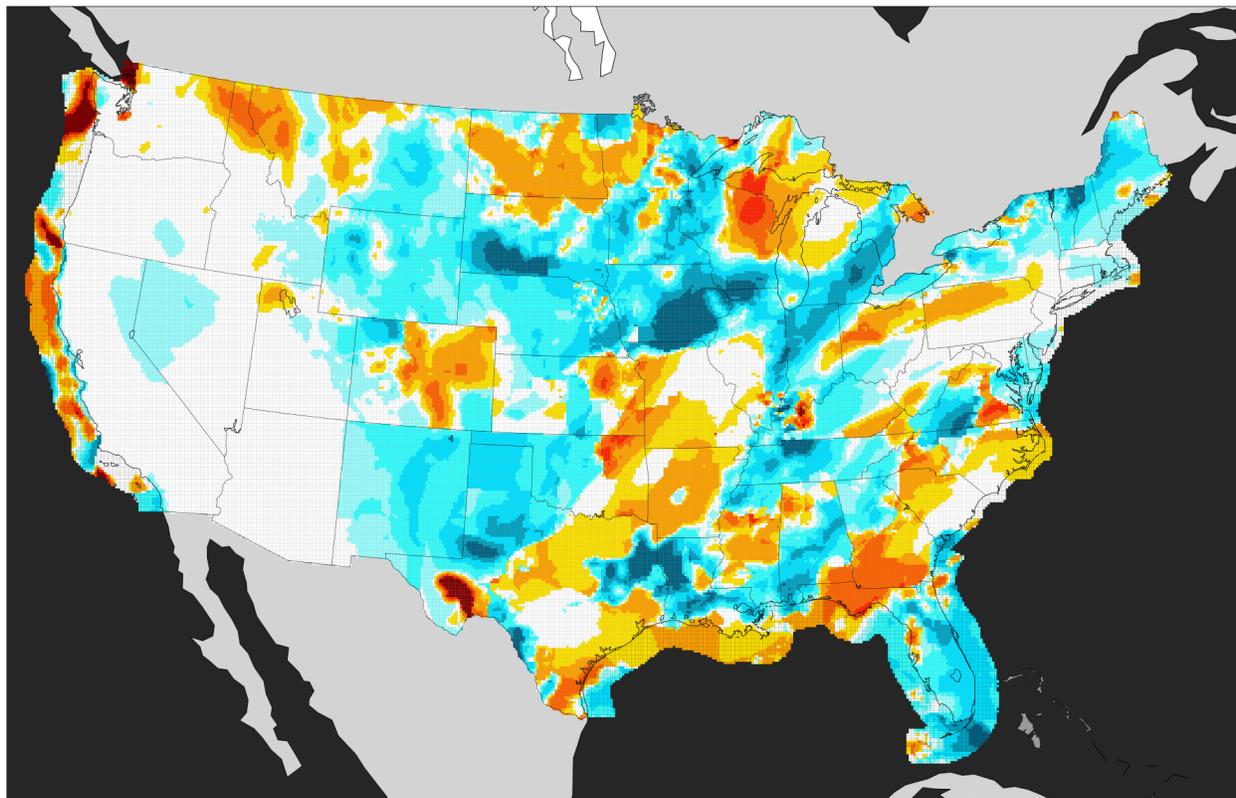


Figure 31. This figure of the difference of the NDFD total cloud cover one-hour forecast from the initial-hour HRRR optimal sky cover product, in units of percent (%) sky cover, is valid for the one-hour window beginning at 3:00 UTC on 19 October 2013. Blue shades indicate a negative difference, where the NDFD total cloud cover one-hour forecast value for the shaded point is greater than the corresponding HRRR optimal sky cover product value; red shades indicate a positive difference, where the HRRR optimal sky cover product value for the shaded point is greater than the corresponding NDFD total cloud cover one-hour forecast value.

GOES-East CONUS IR Window Image

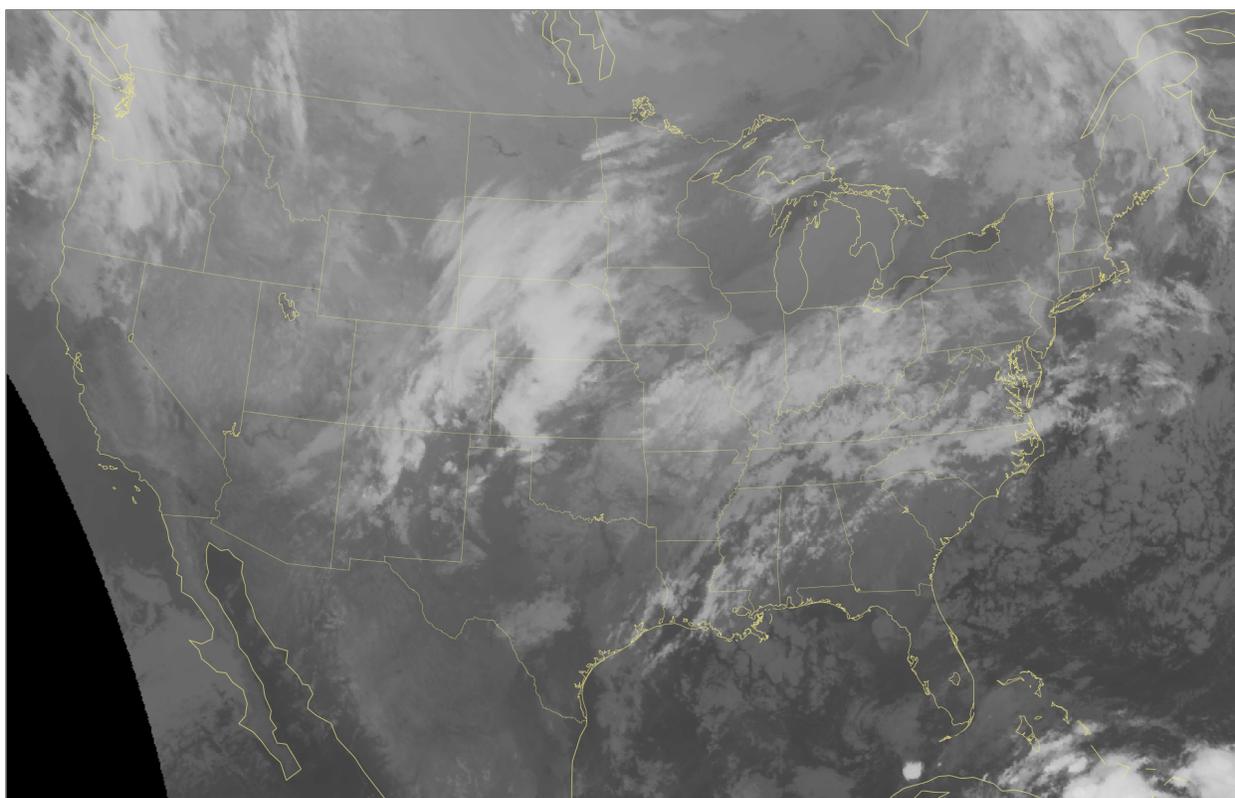


Figure 32. This figure depicts the infrared window satellite image from GOES-East valid at 12:15 UTC on 5 November 2013. A traditional enhancement is applied to the satellite image. Cold cloud tops are white.

Tue Nov 5 12:00:00 UTC 2013

Satellite/Surface Blended Sky Cover (%)

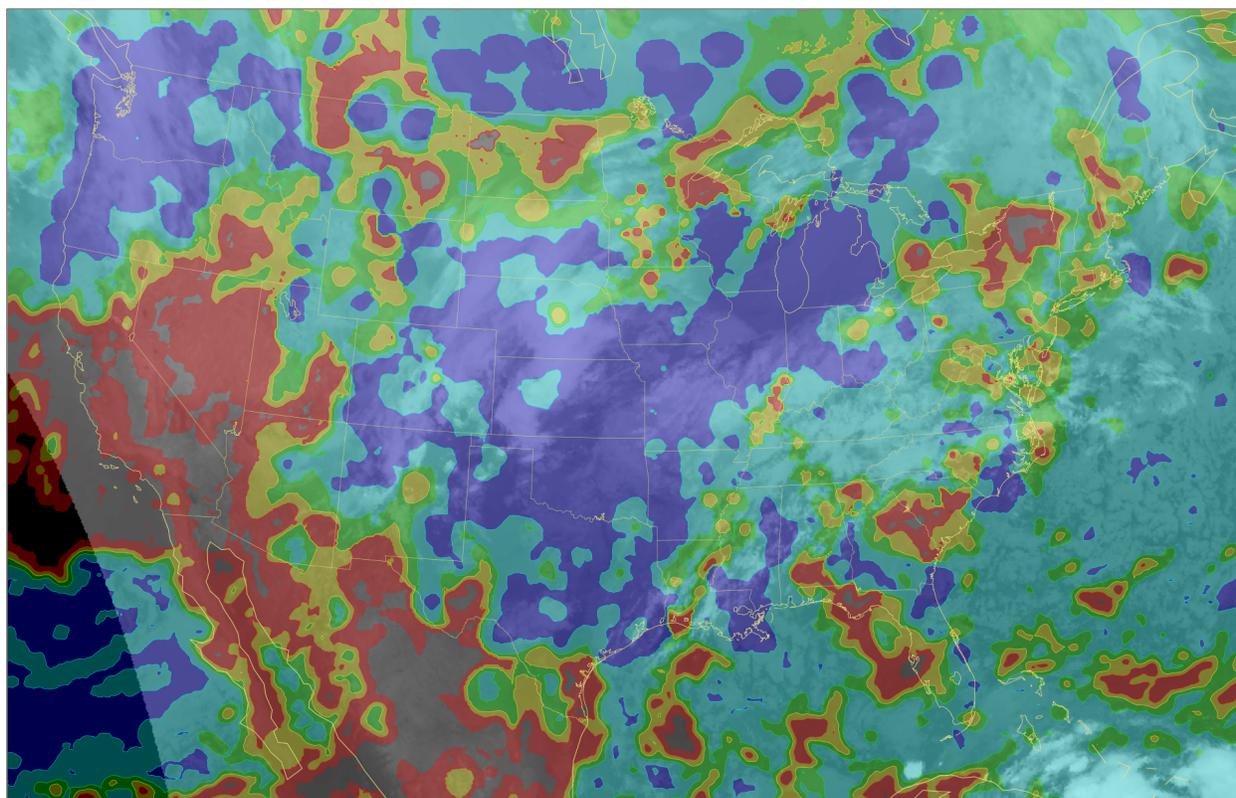


Figure 33. This figure depicts the blended sky cover analysis, in units of percent (%) sky cover, for the one-hour window beginning at 12:00 UTC on 5 November 2013. Overcast areas are in blue. The infrared window satellite image from GOES-East in the background is valid at 12:15 UTC on 5 November 2013. A traditional enhancement is applied to the satellite image. Cold cloud tops are white. The overlay is partially transparent, except clear areas, which are not filled.

Tue Nov 5 12:00:00 UTC 2013

NDFD Total Cloud Cover (%)

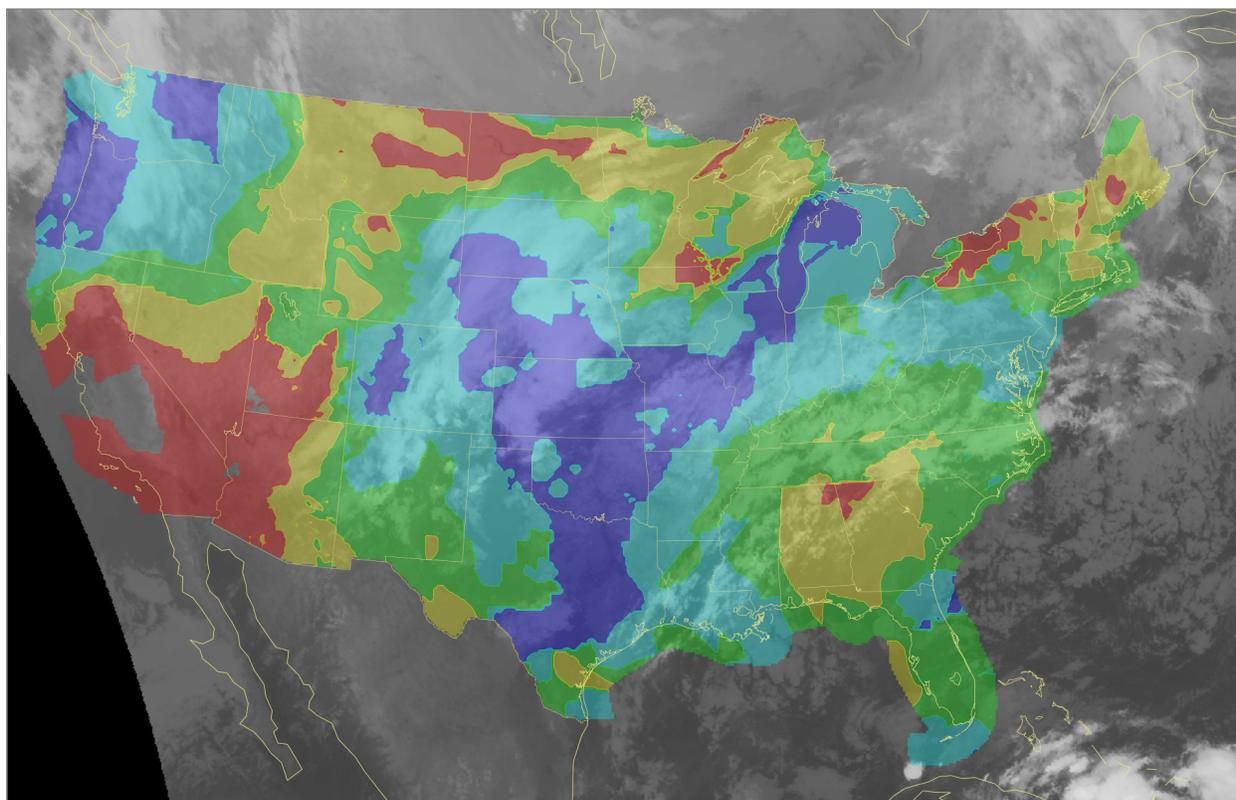


Figure 34. This figure depicts the NDFD total cloud cover one-hour forecast, in units of percent (%) sky cover, for the one-hour window beginning at 12:00 UTC on 5 November 2013. Overcast areas are in blue. The infrared window satellite image from GOES-East in the background is valid at 12:15 UTC on 5 November 2013. A traditional enhancement is applied to the satellite image. Cold cloud tops are white. The overlay is partially transparent, except clear areas, and areas outside of the domain, which are not filled.

Tue Nov 5 12:00:00 UTC 2013

HRRR Total Cloud Cover (%)

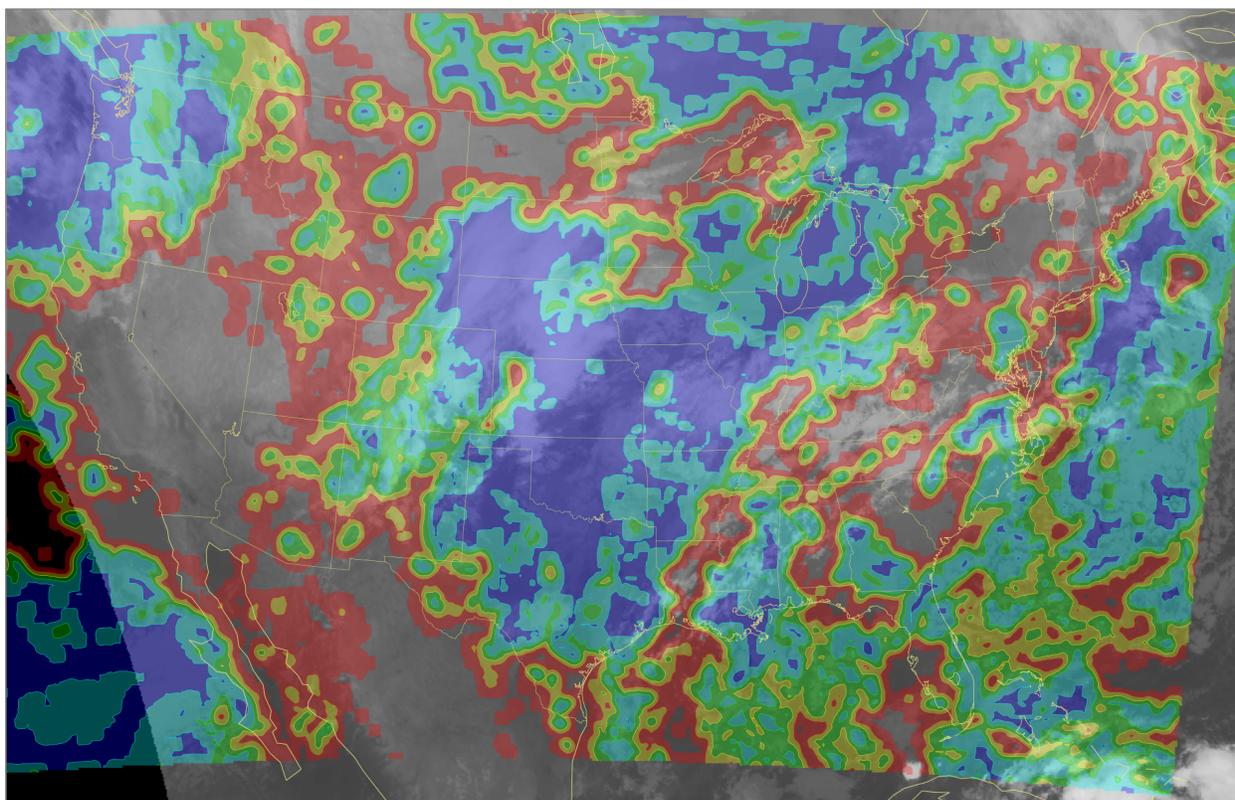


Figure 35. This figure depicts the HRRR total cloud cover analysis, in units of percent (%) cloud cover, valid at 12:00 UTC on 5 November 2013. Overcast areas are in blue. The infrared window satellite image from GOES-East in the background is valid at 12:15 UTC on 5 November 2013. A traditional enhancement is applied to the satellite image. Cold cloud tops are white. The overlay is partially transparent, except clear areas, and areas outside of the domain, which are not filled.

Tue Nov 5 12:00:00 UTC 2013

HRRR Optimal Sky Cover (%)

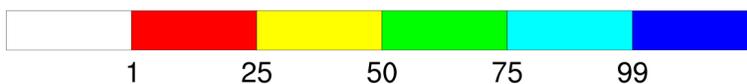
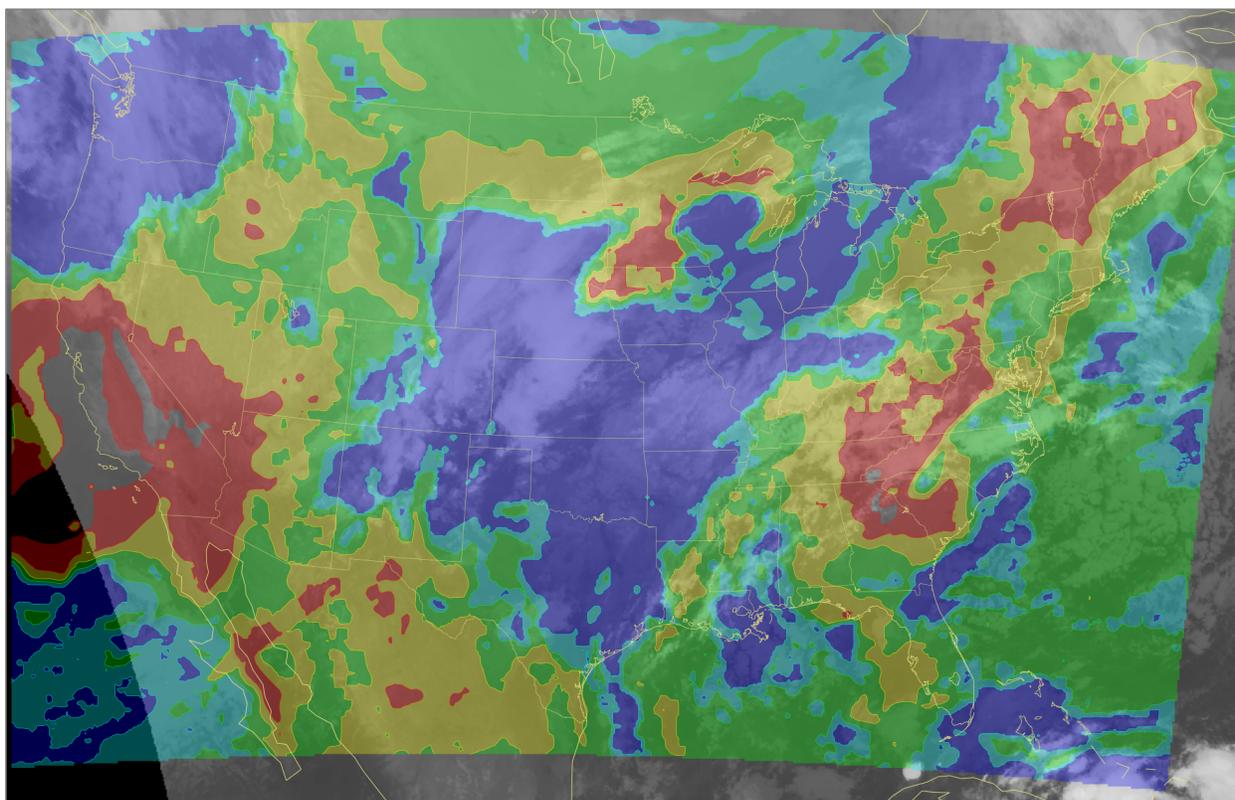


Figure 36. This figure depicts the initial-hour HRRR optimal sky cover product, in units of percent (%) sky cover, for the one-hour window beginning at 12:00 UTC on 5 November 2013. Overcast areas are in blue. The infrared window satellite image from GOES-East in the background is valid at 12:15 UTC on 5 November 2013. A traditional enhancement is applied to the satellite image. Cold cloud tops are white. The overlay is partially transparent, except clear areas, and areas outside of the domain, which are not filled.

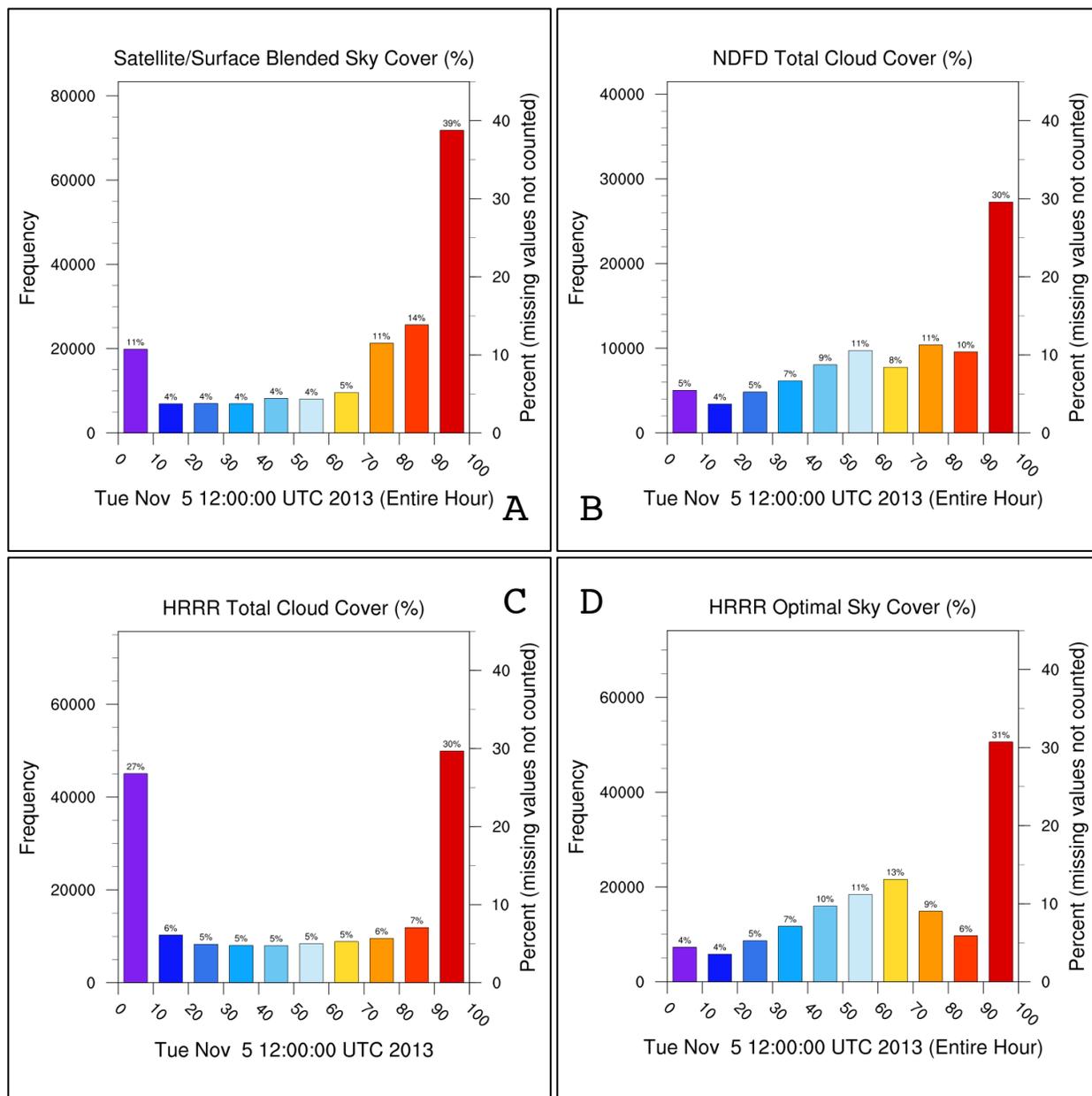


Figure 37. All of the histograms in this four-panel plot are valid for the one-hour window beginning at 12:00 UTC on 5 November 2013, except Panel C, which is valid at the top of the hour. The description of the axes and panels is found in the caption of Figure 19.

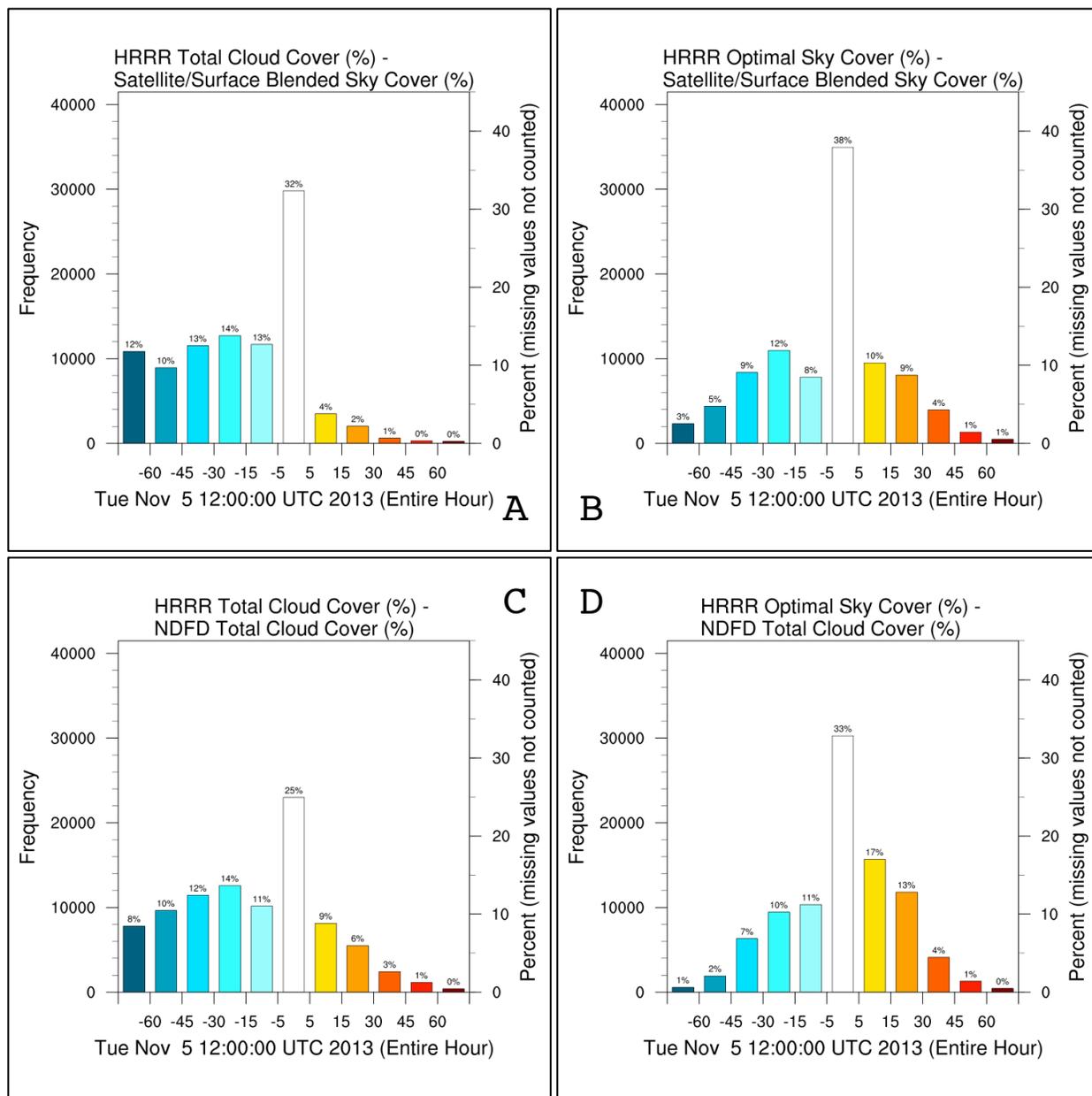


Figure 38. All of the histograms in this four-panel plot are valid for the one-hour window beginning at 12:00 UTC on 5 November 2013. The description of the axes and panels is found in the caption of Figure 20.

Tue Nov 5 12:00:00 UTC 2013

HRRR Optimal Sky Cover (%) - Satellite/Surface Blended Sky Cover (%)

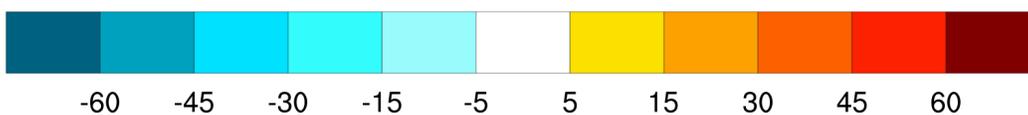
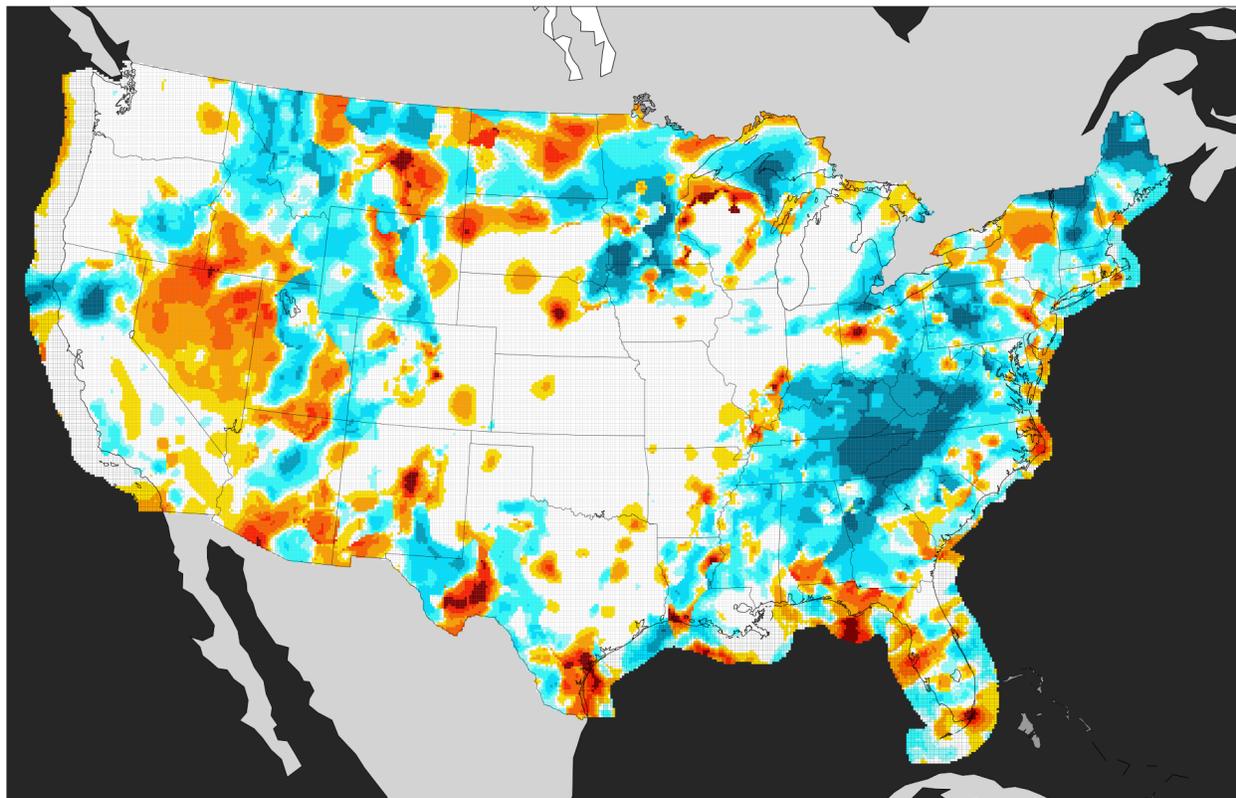


Figure 39. This figure of the difference of the blended sky cover analysis from the initial-hour HRRR optimal sky cover product, in units of percent (%) sky cover, is valid for the one-hour window beginning at 12:00 UTC on 5 November 2013. Blue shades indicate a negative difference, where the blended sky cover analysis value for the shaded point is greater than the corresponding HRRR optimal sky cover product value; red shades indicate a positive difference, where the HRRR optimal sky cover product value for the shaded point is greater than the corresponding blended sky cover analysis value.

Tue Nov 5 12:00:00 UTC 2013

HRRR Optimal Sky Cover (%) - NDFD Total Cloud Cover (%)

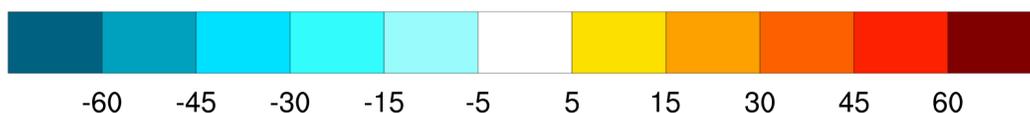
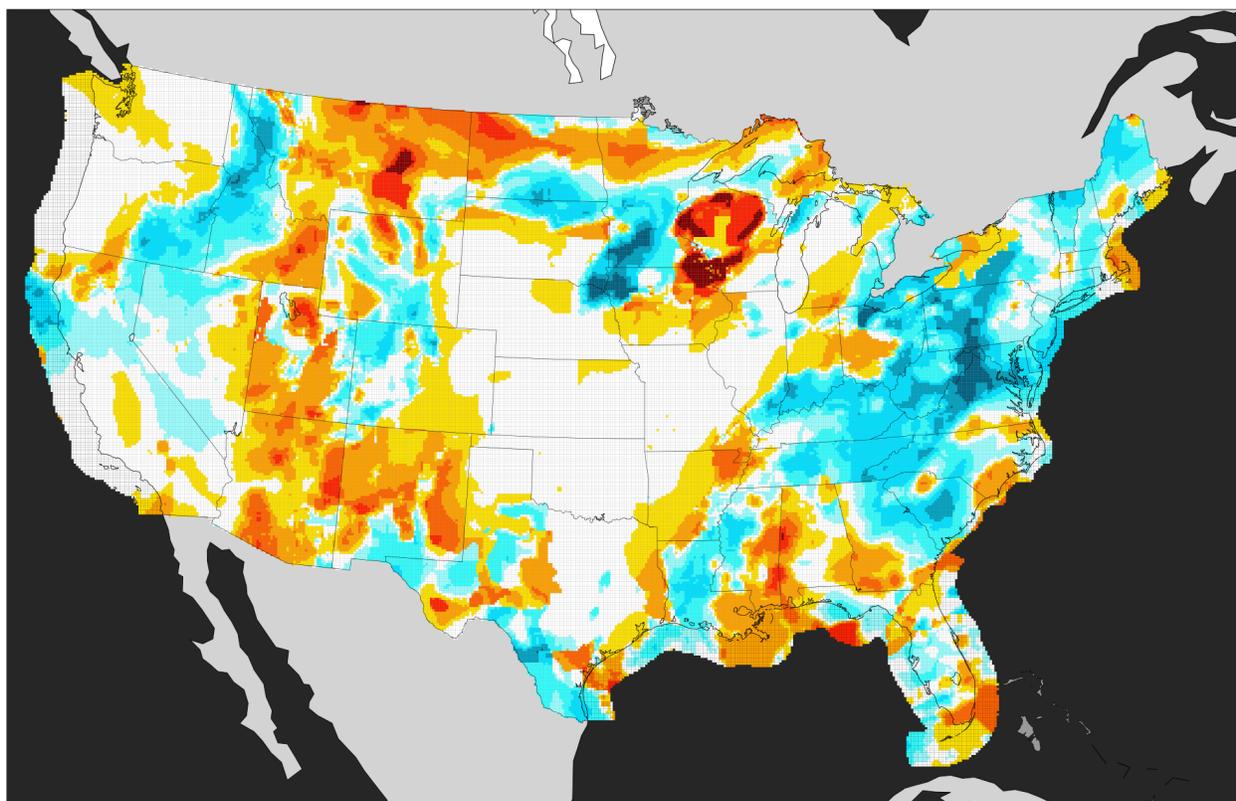


Figure 40. This figure of the difference of the NDFD total cloud cover one-hour forecast from the initial-hour HRRR optimal sky cover product, in units of percent (%) sky cover, is valid for the one-hour window beginning at 12:00 UTC on 5 November 2013. Blue shades indicate a negative difference, where the NDFD total cloud cover one-hour forecast value for the shaded point is greater than the corresponding HRRR optimal sky cover product value; red shades indicate a positive difference, where the HRRR optimal sky cover product value for the shaded point is greater than the corresponding NDFD total cloud cover one-hour forecast value.

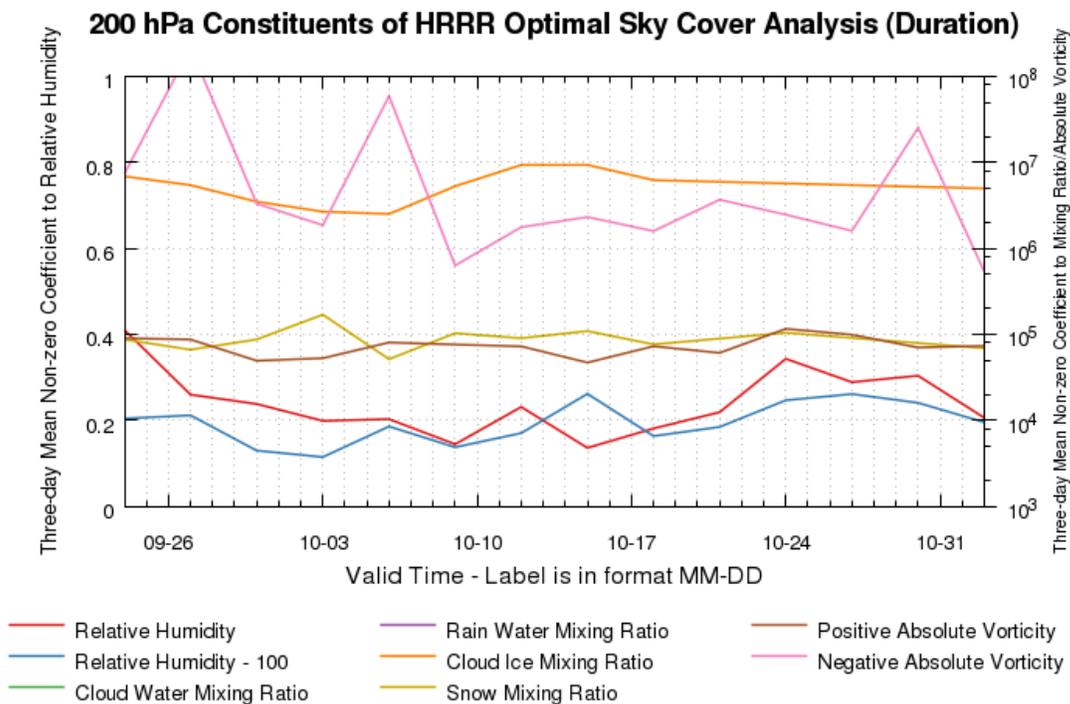


Figure 41. This figure shows the trend in the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The three-day mean non-zero coefficient value from runs where the optimizer produced a solution, is shown for the following quantities at the 200 hPa level: relative humidity, in units of %/%, relative humidity less 100%, in units of %/%, cloud water mixing ratio, in units of %•kg/kg (absent), rain water mixing ratio, in units of %•kg/kg (absent), cloud ice mixing ratio, in units of %•kg/kg, snow mixing ratio, in units of %•kg/kg, positive absolute vorticity, in units of %•s, and negative absolute vorticity, in units of %•s. The scale on the left for the relative humidity coefficients is linear. The scale on the right for the condensate and vorticity coefficients is logarithmic. Increments on the abscissa are one day. A discontinuity in a line, or missing line on the graph, indicates that there are no non-zero coefficient values for that quantity within the open interval.

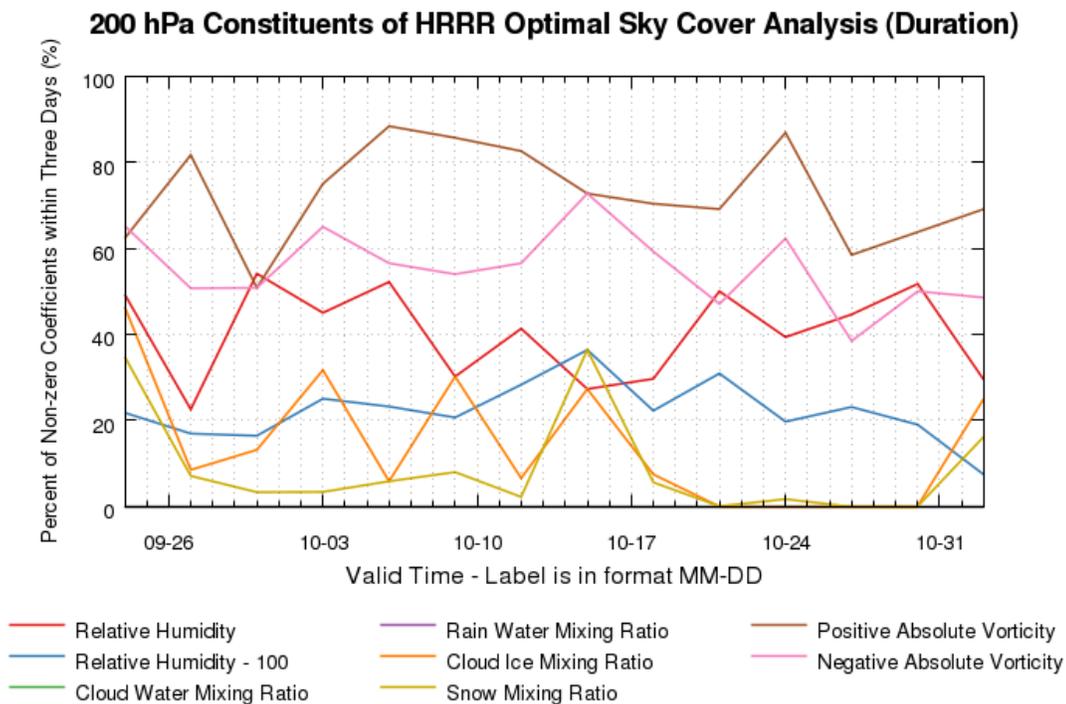


Figure 42. This figure shows the trend in the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The three-day percentage of non-zero coefficient values from individual runs, compared to the total number of solved solutions within the three-day window, is shown for the following quantities at the 200 hPa level: relative humidity, relative humidity less 100%, cloud water mixing ratio, rain water mixing ratio, cloud ice mixing ratio, snow mixing ratio, positive absolute vorticity, and negative absolute vorticity. Plotted quantities are in units of percent (%). Increments on the abscissa are one day.

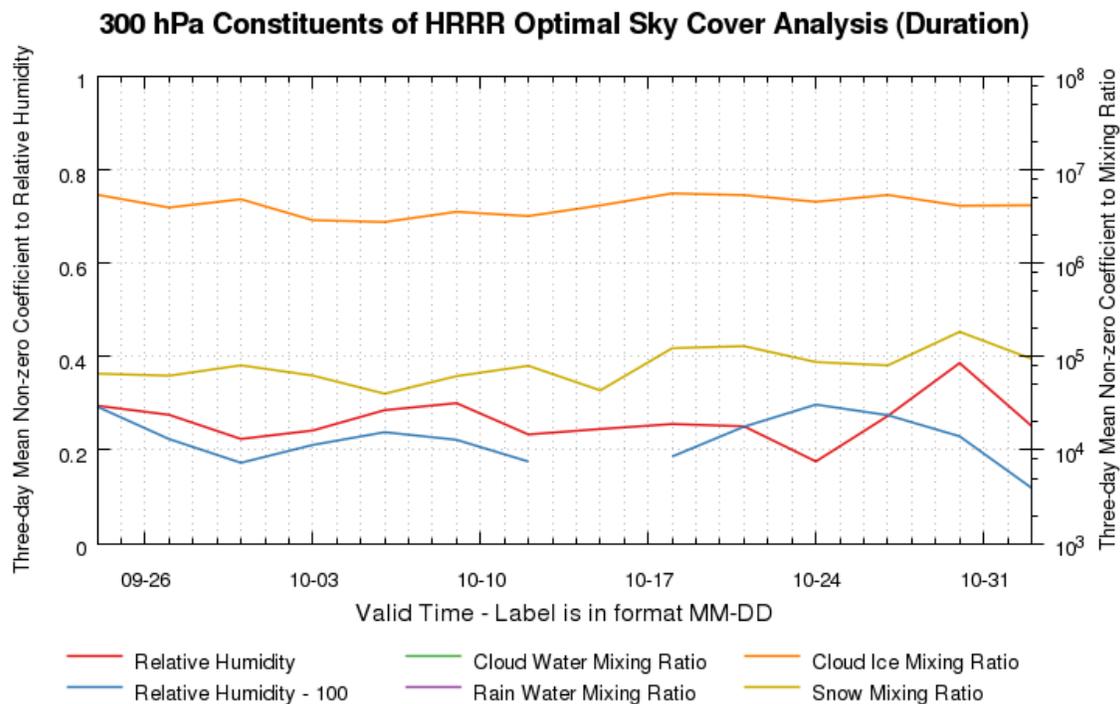


Figure 43. This figure shows the trend in the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The three-day mean non-zero coefficient value from runs where the optimizer produced a solution, is shown for the following quantities at the 300 hPa level: relative humidity, in units of %/%, relative humidity less 100%, in units of %/%, cloud water mixing ratio, in units of %•kg/kg (absent), rain water mixing ratio, in units of %•kg/kg (absent), cloud ice mixing ratio, in units of %•kg/kg, and snow mixing ratio, in units of %•kg/kg. The scale on the left for the relative humidity coefficients is linear. The scale on the right for the condensate coefficients is logarithmic. Increments on the abscissa are one day. A discontinuity in a line, or missing line on the graph, indicates that there are no non-zero coefficient values for that quantity within the open interval.

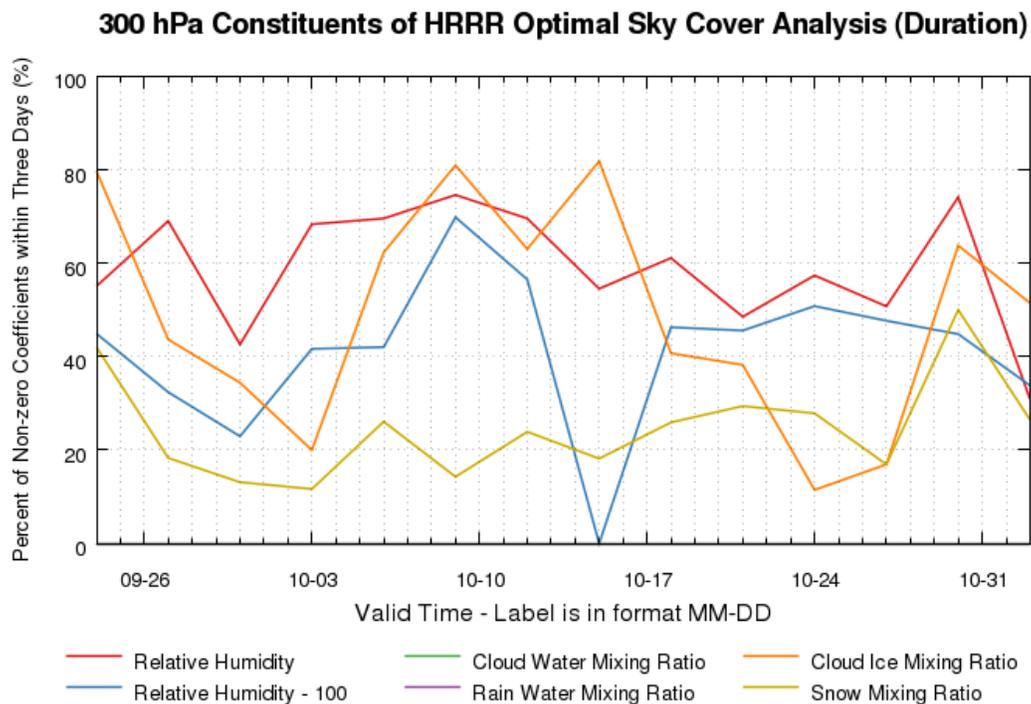


Figure 44. This figure shows the trend in the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The three-day percentage of non-zero coefficient values from individual runs, compared to the total number of solved solutions within the three-day window, is shown for the following quantities at the 300 hPa level: relative humidity, relative humidity less 100%, cloud water mixing ratio, rain water mixing ratio, cloud ice mixing ratio, and snow mixing ratio. Plotted quantities are in units of percent (%). Increments on the abscissa are one day.

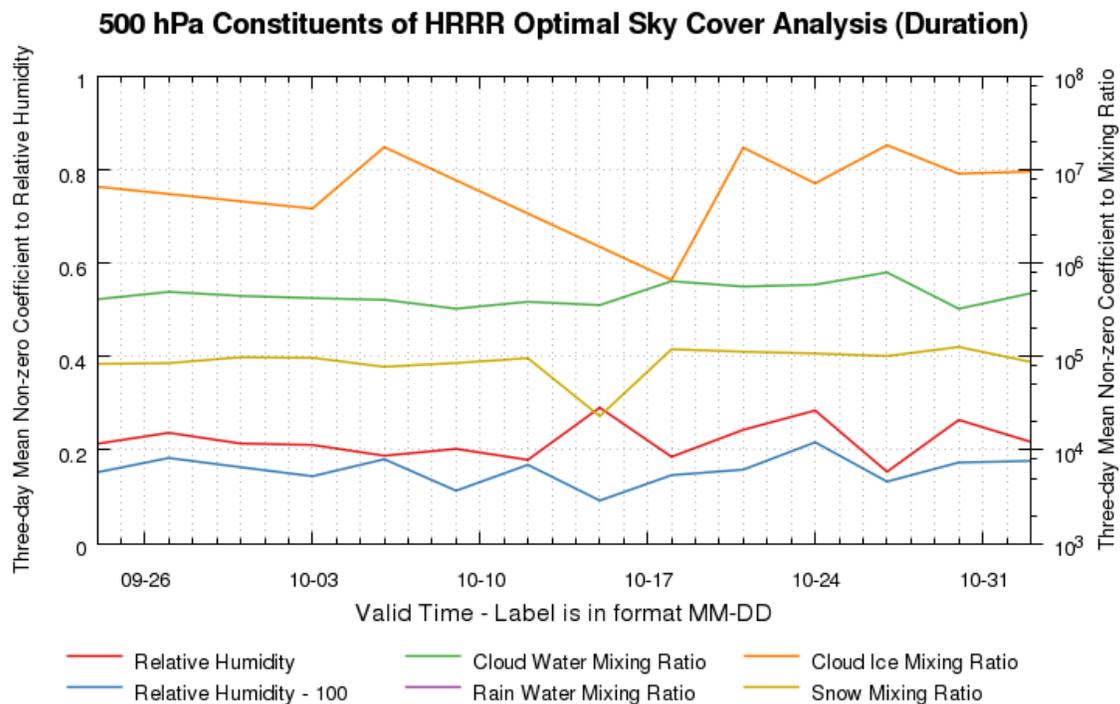


Figure 45. This figure shows the trend in the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The three-day mean non-zero coefficient value from runs where the optimizer produced a solution, is shown for the following quantities at the 500 hPa level: relative humidity, in units of %/%, relative humidity less 100%, in units of %/%, cloud water mixing ratio, in units of %•kg/kg, rain water mixing ratio, in units of %•kg/kg (absent), cloud ice mixing ratio, in units of %•kg/kg, and snow mixing ratio, in units of %•kg/kg. The scale on the left for the relative humidity coefficients is linear. The scale on the right for the condensate coefficients is logarithmic. Increments on the abscissa are one day. A discontinuity in a line, or missing line on the graph, indicates that there are no non-zero coefficient values for that quantity within the open interval.

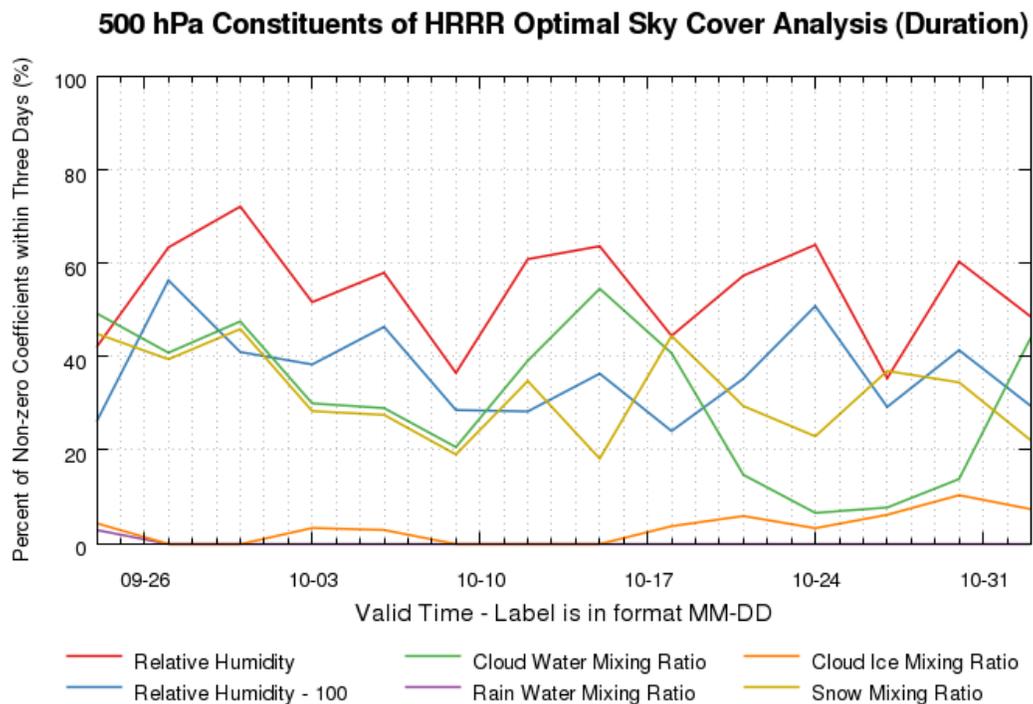


Figure 46. This figure shows the trend in the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The three-day percentage of non-zero coefficient values from individual runs, compared to the total number of solved solutions within the three-day window, is shown for the following quantities at the 500 hPa level: relative humidity, relative humidity less 100%, cloud water mixing ratio, rain water mixing ratio, cloud ice mixing ratio, and snow mixing ratio. Plotted quantities are in units of percent (%). Increments on the abscissa are one day.

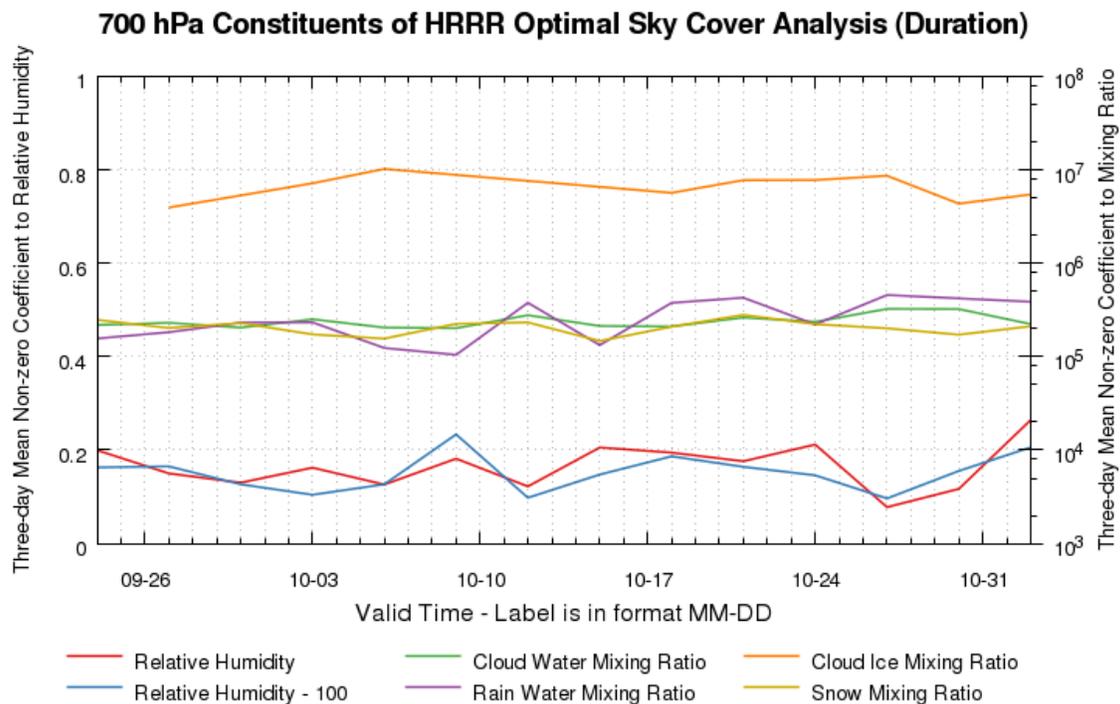


Figure 47. This figure shows the trend in the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The three-day mean non-zero coefficient value from runs where the optimizer produced a solution, is shown for the following quantities at the 700 hPa level: relative humidity, in units of %/%, relative humidity less 100%, in units of %/%, cloud water mixing ratio, in units of %•kg/kg, rain water mixing ratio, in units of %•kg/kg, cloud ice mixing ratio, in units of %•kg/kg, and snow mixing ratio, in units of %•kg/kg. The scale on the left for the relative humidity coefficients is linear. The scale on the right for the condensate coefficients is logarithmic. Increments on the abscissa are one day. A discontinuity in a line, or missing line on the graph, indicates that there are no non-zero coefficient values for that quantity within the open interval.

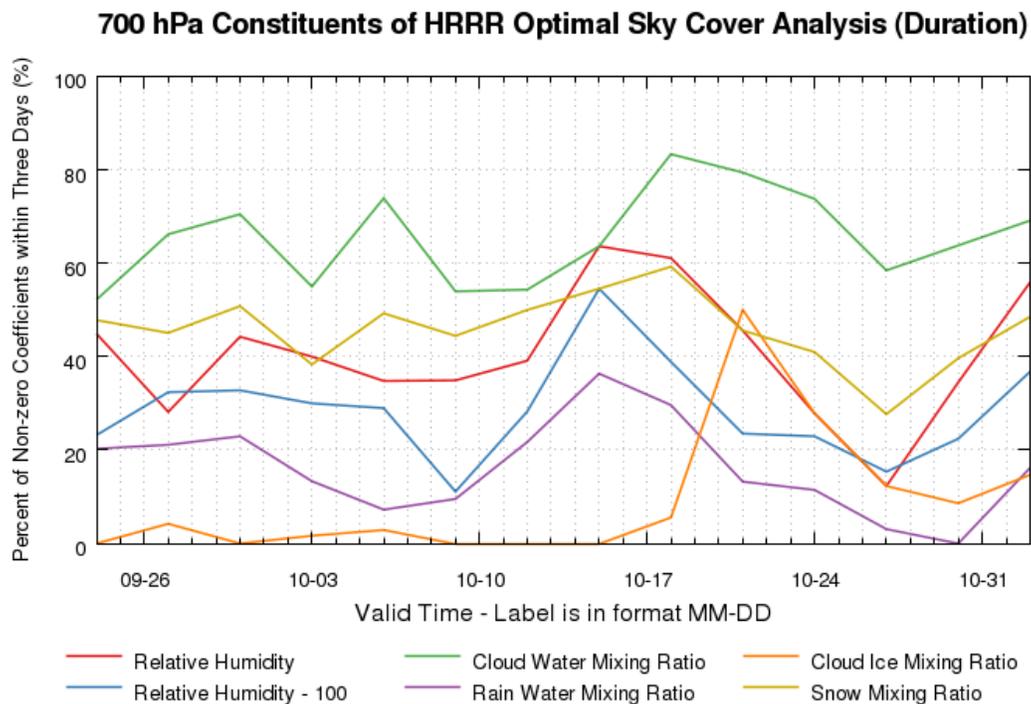


Figure 48. This figure shows the trend in the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The three-day percentage of non-zero coefficient values from individual runs, compared to the total number of solved solutions within the three-day window, is shown for the following quantities at the 700 hPa level: relative humidity, relative humidity less 100%, cloud water mixing ratio, rain water mixing ratio, cloud ice mixing ratio, and snow mixing ratio. Plotted quantities are in units of percent (%). Increments on the abscissa are one day.

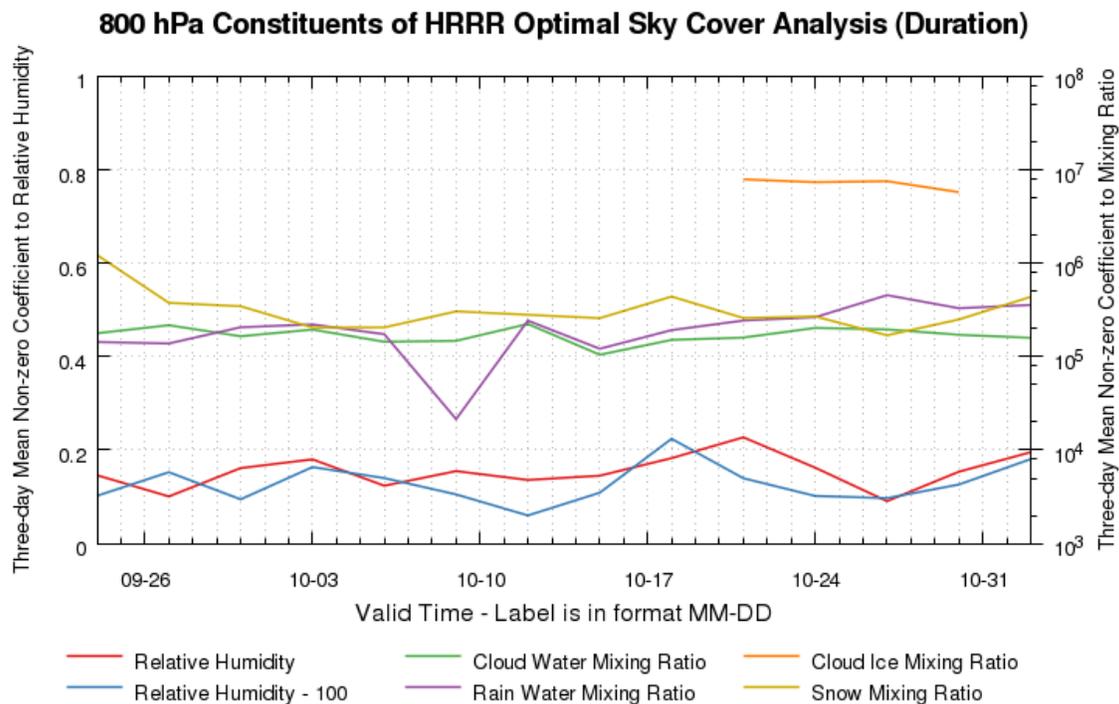


Figure 49. This figure shows the trend in the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The three-day mean non-zero coefficient value from runs where the optimizer produced a solution, is shown for the following quantities at the 800 hPa level: relative humidity, in units of %/%, relative humidity less 100%, in units of %/%, cloud water mixing ratio, in units of %•kg/kg, rain water mixing ratio, in units of %•kg/kg, cloud ice mixing ratio, in units of %•kg/kg, and snow mixing ratio, in units of %•kg/kg. The scale on the left for the relative humidity coefficients is linear. The scale on the right for the condensate coefficients is logarithmic. Increments on the abscissa are one day. A discontinuity in a line, or missing line on the graph, indicates that there are no non-zero coefficient values for that quantity within the open interval.

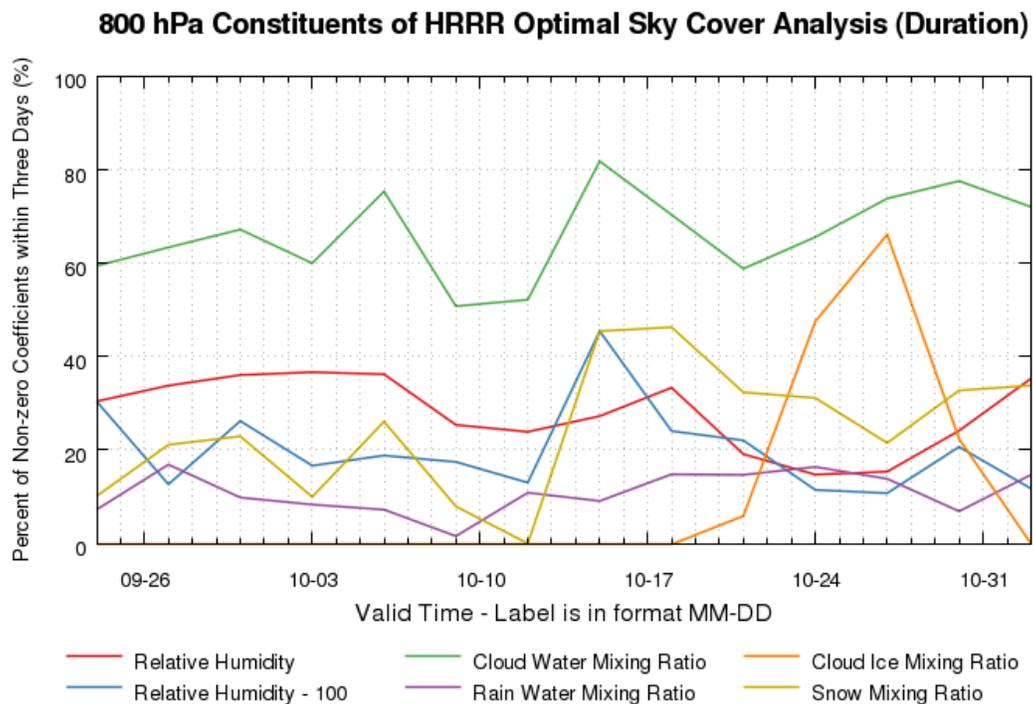


Figure 50. This figure shows the trend in the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The three-day percentage of non-zero coefficient values from individual runs, compared to the total number of solved solutions within the three-day window, is shown for the following quantities at the 800 hPa level: relative humidity, relative humidity less 100%, cloud water mixing ratio, rain water mixing ratio, cloud ice mixing ratio, and snow mixing ratio. Plotted quantities are in units of percent (%). Increments on the abscissa are one day.

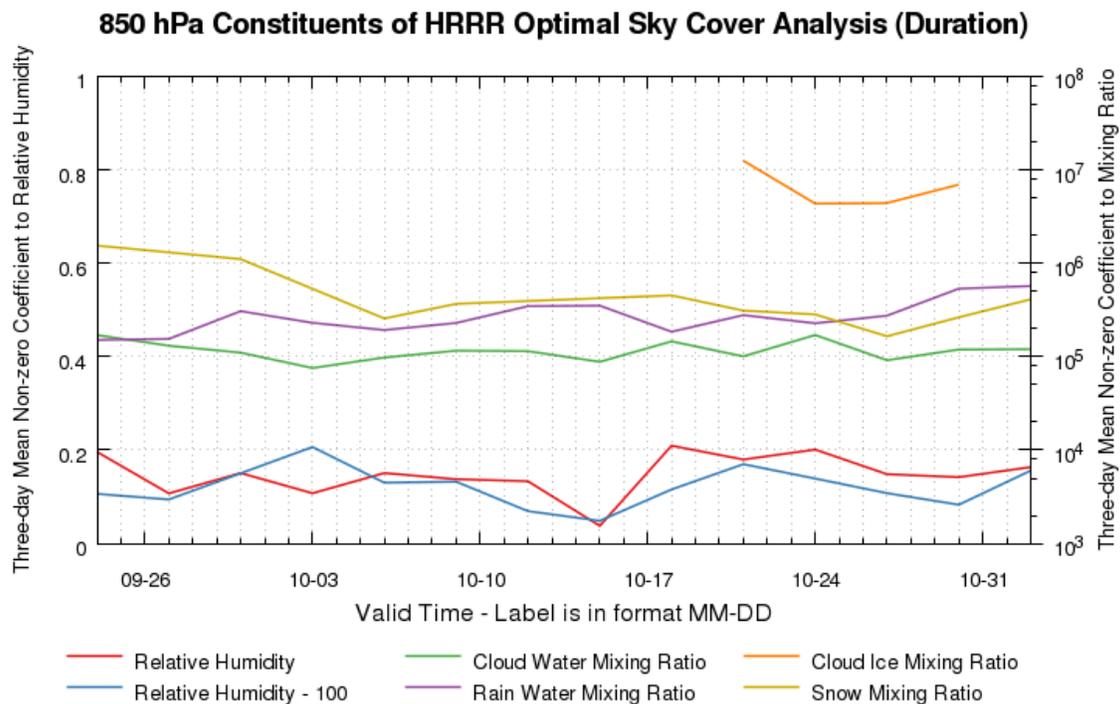


Figure 51. This figure shows the trend in the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The three-day mean non-zero coefficient value from runs where the optimizer produced a solution, is shown for the following quantities at the 850 hPa level: relative humidity, in units of %/%, relative humidity less 100%, in units of %/%, cloud water mixing ratio, in units of %•kg/kg, rain water mixing ratio, in units of %•kg/kg, cloud ice mixing ratio, in units of %•kg/kg, and snow mixing ratio, in units of %•kg/kg. The scale on the left for the relative humidity coefficients is linear. The scale on the right for the condensate coefficients is logarithmic. Increments on the abscissa are one day. A discontinuity in a line, or missing line on the graph, indicates that there are no non-zero coefficient values for that quantity within the open interval.

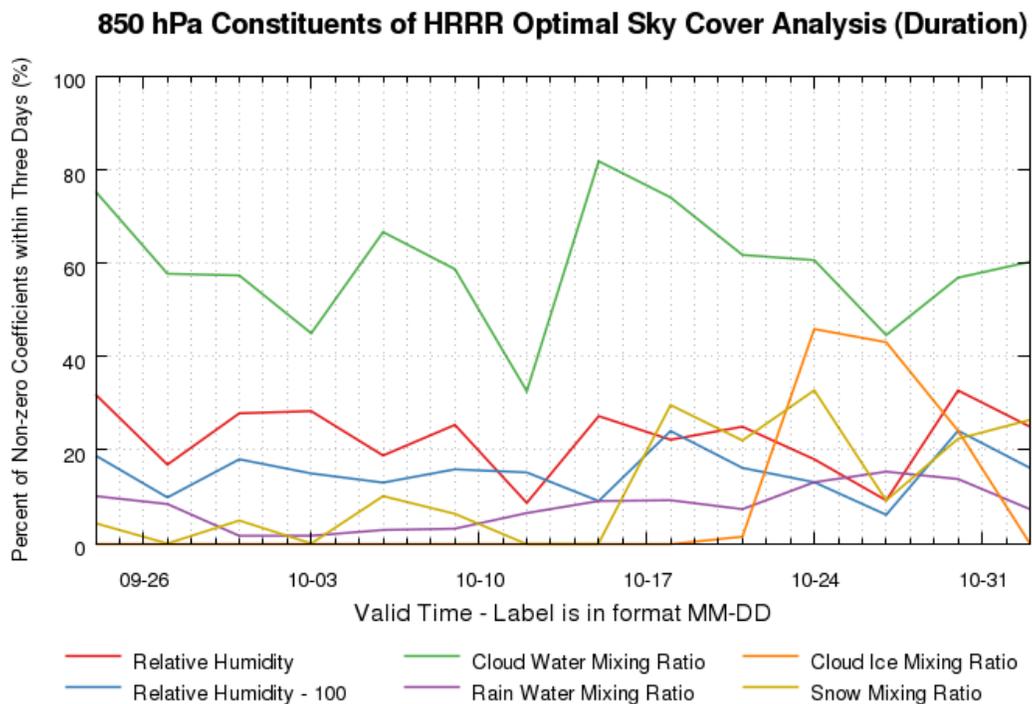


Figure 52. This figure shows the trend in the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The three-day percentage of non-zero coefficient values from individual runs, compared to the total number of solved solutions within the three-day window, is shown for the following quantities at the 850 hPa level: relative humidity, relative humidity less 100%, cloud water mixing ratio, rain water mixing ratio, cloud ice mixing ratio, and snow mixing ratio. Plotted quantities are in units of percent (%). Increments on the abscissa are one day.

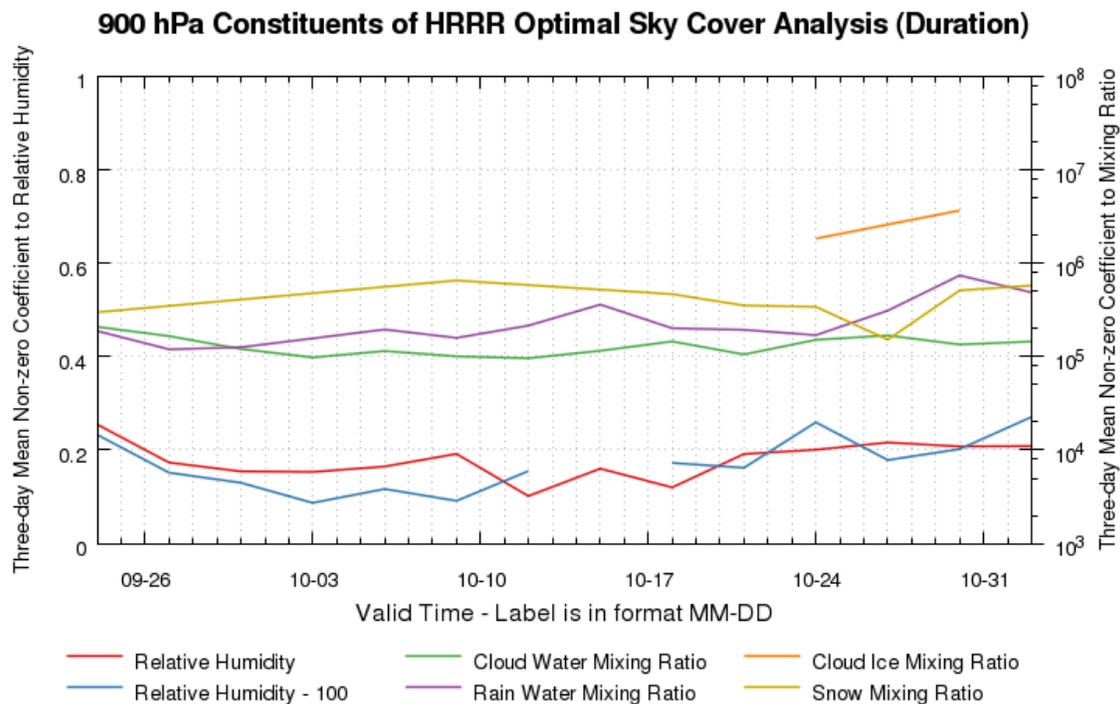


Figure 53. This figure shows the trend in the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The three-day mean non-zero coefficient value from runs where the optimizer produced a solution, is shown for the following quantities at the 900 hPa level: relative humidity, in units of %/%, relative humidity less 100%, in units of %/%, cloud water mixing ratio, in units of %•kg/kg, rain water mixing ratio, in units of %•kg/kg, cloud ice mixing ratio, in units of %•kg/kg, and snow mixing ratio, in units of %•kg/kg. The scale on the left for the relative humidity coefficients is linear. The scale on the right for the condensate coefficients is logarithmic. Increments on the abscissa are one day. A discontinuity in a line, or missing line on the graph, indicates that there are no non-zero coefficient values for that quantity within the open interval.

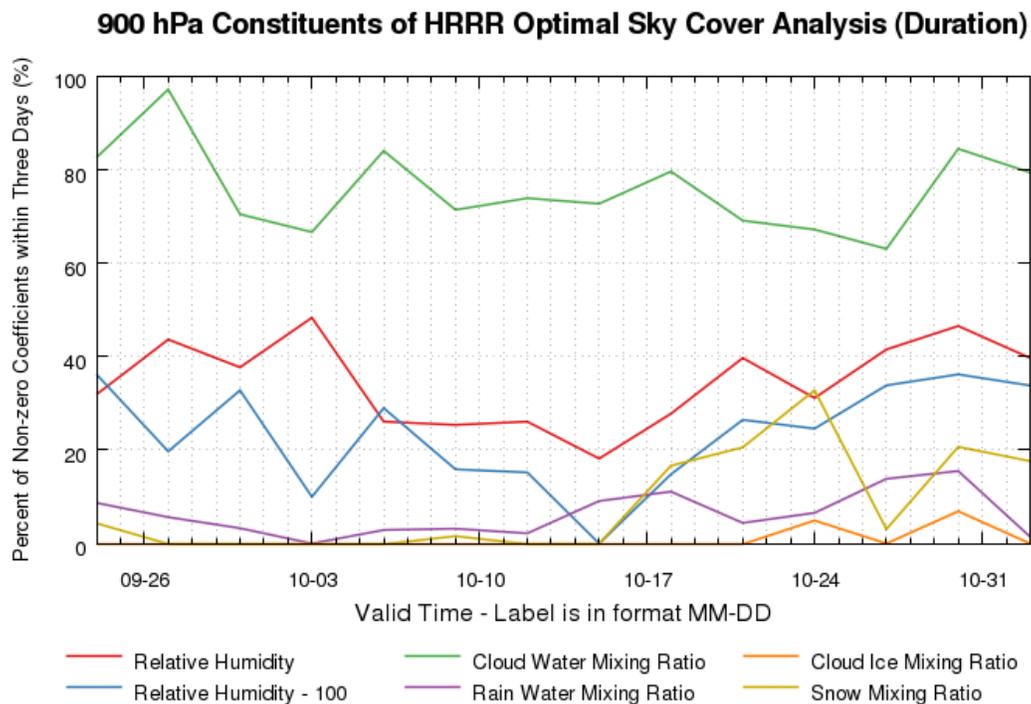


Figure 54. This figure shows the trend in the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The three-day percentage of non-zero coefficient values from individual runs, compared to the total number of solved solutions within the three-day window, is shown for the following quantities at the 900 hPa level: relative humidity, relative humidity less 100%, cloud water mixing ratio, rain water mixing ratio, cloud ice mixing ratio, and snow mixing ratio. Plotted quantities are in units of percent (%). Increments on the abscissa are one day.

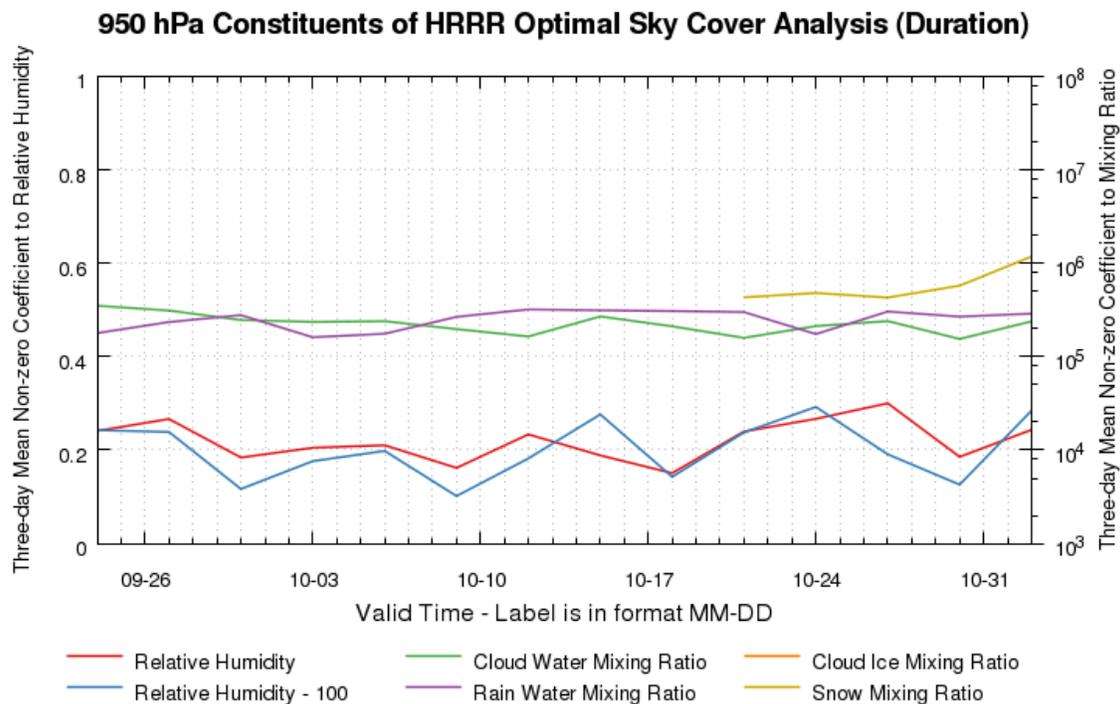


Figure 55. This figure shows the trend in the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The three-day mean non-zero coefficient value from runs where the optimizer produced a solution, is shown for the following quantities at the 950 hPa level: relative humidity, in units of %/%, relative humidity less 100%, in units of %/%, cloud water mixing ratio, in units of %•kg/kg, rain water mixing ratio, in units of %•kg/kg, cloud ice mixing ratio, in units of %•kg/kg (absent), and snow mixing ratio, in units of %•kg/kg. The scale on the left for the relative humidity coefficients is linear. The scale on the right for the condensate coefficients is logarithmic. Increments on the abscissa are one day. A discontinuity in the line, or missing line on the graph, indicates that there are no non-zero coefficient values for the quantity within the open interval.

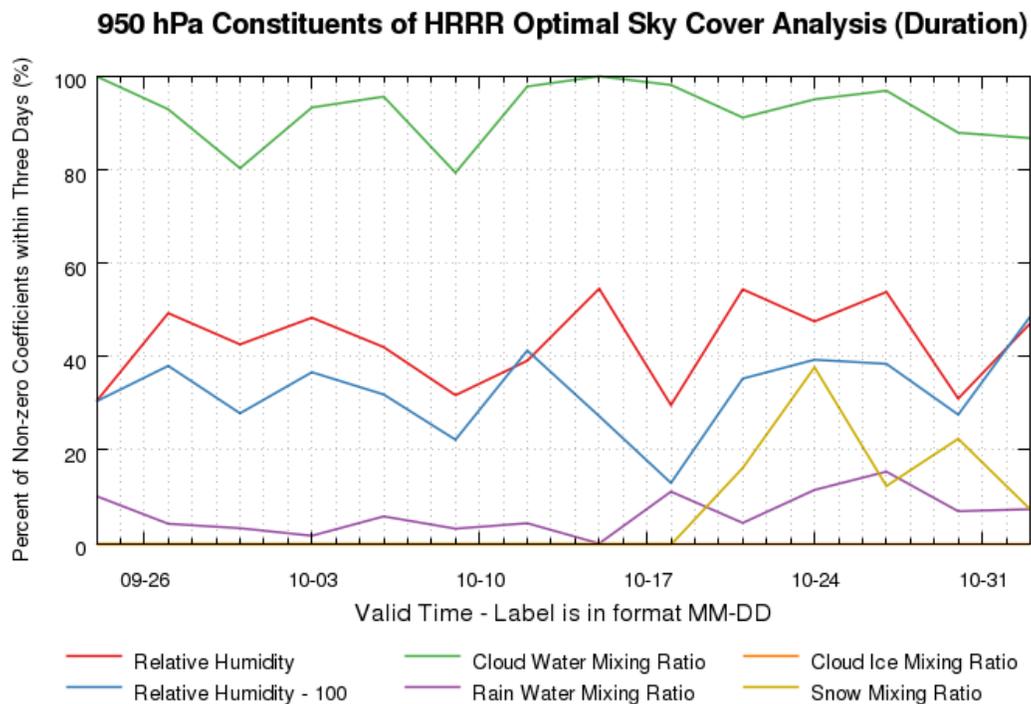


Figure 56. This figure shows the trend in the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The three-day percentage of non-zero coefficient values from individual runs, compared to the total number of solved solutions within the three-day window, is shown for the following quantities at the 950 hPa level: relative humidity, relative humidity less 100%, cloud water mixing ratio, rain water mixing ratio, cloud ice mixing ratio, and snow mixing ratio. Plotted quantities are in units of percent (%). Increments on the abscissa are one day.

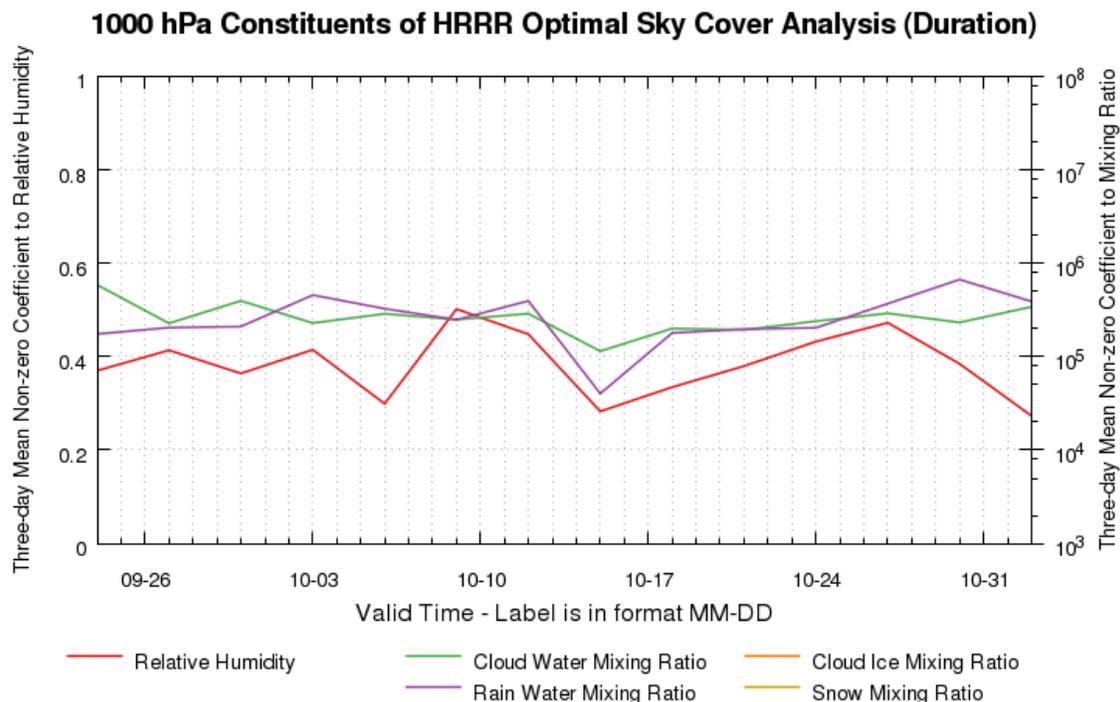


Figure 57. This figure shows the trend in the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The three-day mean non-zero coefficient value from runs where the optimizer produced a solution, is shown for the following quantities at the 1000 hPa level: relative humidity, in units of %/%, cloud water mixing ratio, in units of %•kg/kg, rain water mixing ratio, in units of %•kg/kg, cloud ice mixing ratio, in units of %•kg/kg (absent), and snow mixing ratio, in units of %•kg/kg (absent). The scale on the left for the relative humidity coefficient is linear. The scale on the right for the condensate coefficients is logarithmic. Increments on the abscissa are one day. A discontinuity in a line, or missing line on the graph, indicates that there are no non-zero coefficient values for that quantity within the open interval.

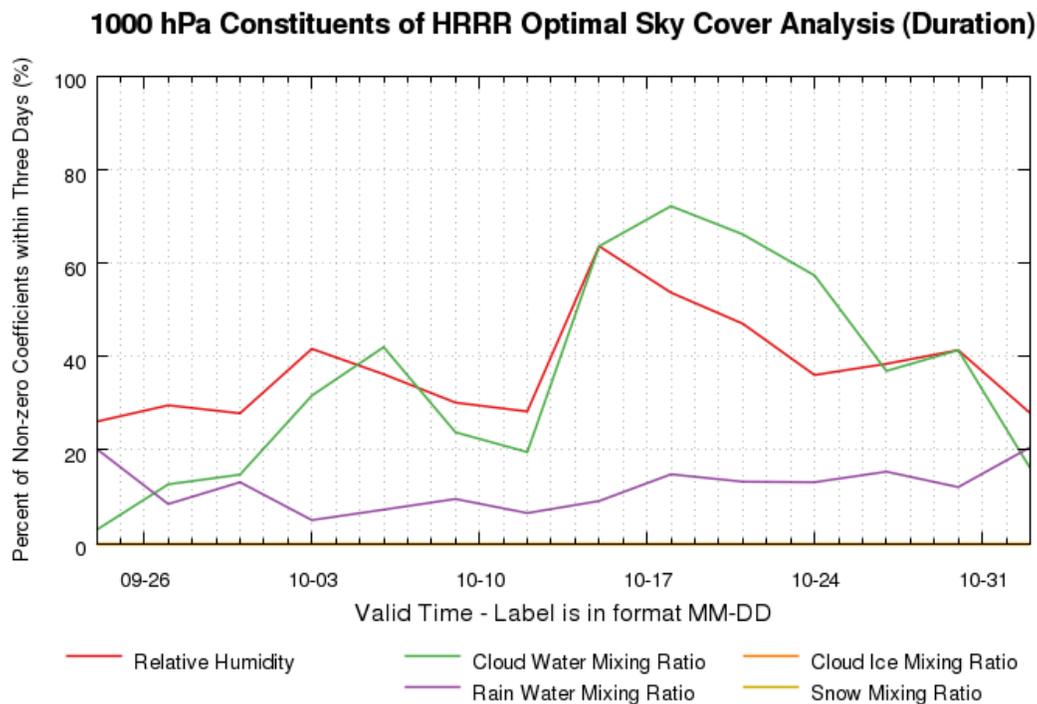


Figure 58. This figure shows the trend in the output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The three-day percentage of non-zero coefficient values from individual runs, compared to the total number of solved solutions within the three-day window, is shown for the following quantities at the 1000 hPa level: relative humidity, cloud water mixing ratio, rain water mixing ratio, cloud ice mixing ratio, and snow mixing ratio. Plotted quantities are in units of percent (%). Increments on the abscissa are one day.

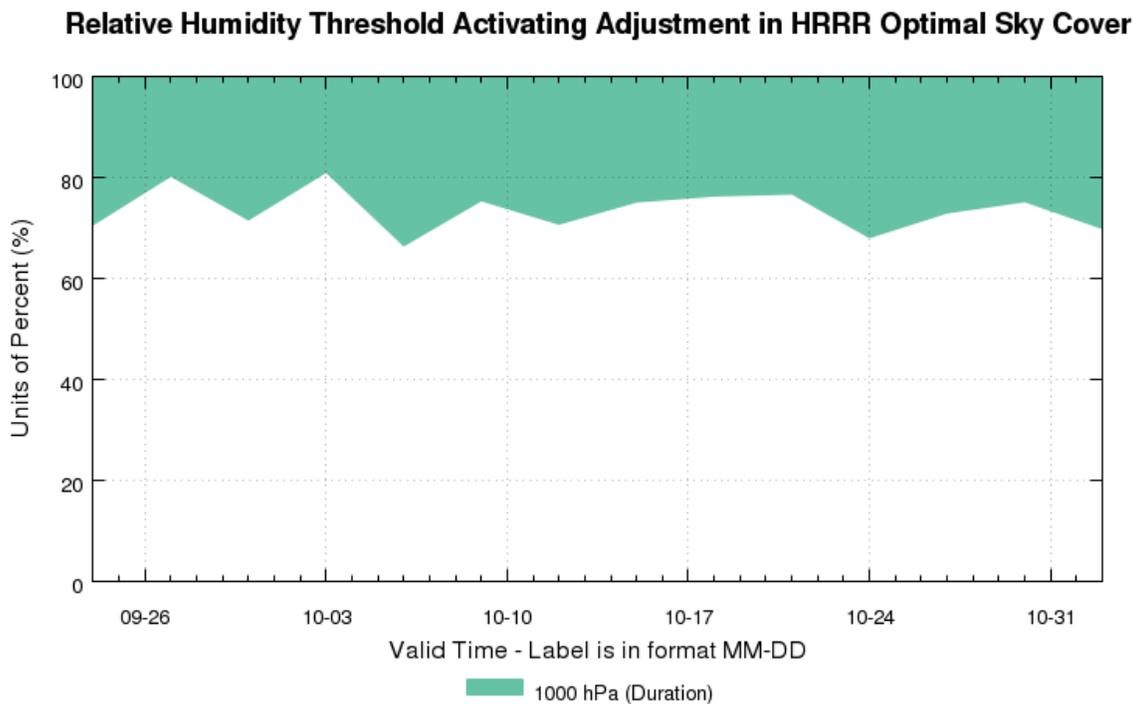


Figure 59. This figure shows the trend in the 1000 hPa relative humidity threshold output from the optimizer for a period from 21 September 2013 at 1 UTC through 1 November 2013 at 23 UTC. The three-day mean is plotted in units of percent (%). For each point on the grid, the coefficient and scalar are applied to the 1000 hPa relative humidity field value only when the value is greater than the threshold. Increments on the abscissa are one day.

Appendix A

Celestial Dome Effective Cloud Amount Logic

The satellite sky cover product is a composite of scans from two geostationary weather satellites that observe the continental United States. The primary input for the satellite sky cover product is the effective cloud amount (ECA), also known as effective cloud emissivity, output that is spatially averaged as part of the celestial dome effective cloud amount (CDECA) product. All input pixels are corrected for parallax.

One major component of the satellite sky cover product is the spatial averaging to create the celestial dome. For this, an 11 by 11 box average is used, centered on the averaged point. This product includes additional corrections to make it representative of sky cover.

One correction addresses thin high cloud above thick low cloud. In such situations, the composite ECA output is the value for the thin high cloud, producing a sharp gradient in the ECA output despite a uniform lower cloud deck. To resolve this inconsistency, the CDECA is the product of the cloud fraction and cloud probability when the:

- Cloud emissivity is less than 50%,
- Cloud fraction is greater than 95%,
- Cloud probability is greater than 95%, and
- Cloud type is opaque ice, cirrus, overlapping, or an overshooting top.

This change can occasionally lead to an overestimate of sky cover when the cloud probability and cloud fraction are both near 100% but the cloud is optically thin and there is no underneath cloud deck. The use of optical depth as a discriminator may prove insightful in addressing some of these scenarios.

Another correction addresses CDECA values of less than 100% for overcast scenes involving water cloud. This correction results in an upward increase in sky cover amount for stratocumulus clouds and other spatially homogenous clouds. In such situations, where the spatial standard deviation of the CDECA output at a satellite-observed, parallax-corrected pixel is less than or equal to 2.50, and the percent of liquid water clouds is greater than or equal to 95%, the CDECA value is set to 100%. The standard deviation threshold was chosen via inspection. For each pixel, the spatial standard deviation is calculated over the same pixel-centered 11 by 11 pixel box that is used to produce the initial CDECA value.

List of Acronyms

CDCTP	celestial dome cloud top pressure
CDECA	celestial dome effective cloud amount
CTP	cloud top pressure
ECA	effective cloud amount
GOES	Geostationary Operational Environmental Satellite(s)
HRRR	High-Resolution Rapid Refresh (model)
MAE	mean absolute error
MM-DD	date format: two-digit month, two-digit day
NDFD	National Digital Forecast Database
NWP	numerical weather prediction
NWS	National Weather Service
PDF	probability density function
POD	probability of detection
RMSE	root-mean-square error
RUC	Rapid Update Cycle (model)
WRF	Weather Research and Forecast (model)