2012-09

Toward Improving Short-Range Fog Prediction in Data-Denied Areas Using the Air Force Weather Agency Mesoscale Ensemble

Ryerson, William R.

Monterey, California. Naval Postgraduate School

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

DISSERTATION

TOWARD IMPROVING SHORT-RANGE FOG PREDICTION IN DATA-DENIED AREAS USING THE AIR FORCE WEATHER AGENCY MESOSCALE ENSEMBLE

by

William R. Ryerson

September 2012

Dissertation Supervisor: Joshua P. Hacker

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE		Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE September 2012	3. REPORT TYPE AND DATES COVERED Dissertation

4. TITLE AND SUBTITLE: Toward Improving Short-Range Fog Prediction in Data-Denied Areas Using the Air Force Weather Agency Mesoscale Ensemble	5. FUNDING NUMBERS
6. AUTHOR(S) William R. Ryerson	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number _____N/A_____.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited	12b. DISTRIBUTION CODE A

13. ABSTRACT (maximum 200 words)

This work develops and tests the viability of a new framework for producing short-range (<20 h) probabilistic fog predictions using post-processing of a 4-km, 10-member Weather Research and Forecasting (WRF) ensemble configured to closely match the Air Force Weather Agency Mesoscale Ensemble Forecast System. The raw WRF predictions produce excessive forecasts of zero cloud water, mainly caused by a negative relative humidity bias, which is largely traced to a warm overnight bias. Post-processing mitigates these systematic errors by leveraging traits of a joint parameter space in the predictions to modify individual ensemble members not predicting fog on their own. The method is generally most effective when the space is defined with a moisture parameter and a low-level stability parameter.

Cross-validation shows the method adds significant overnight skill to predictions in valley and coastal regions compared to the raw WRF forecasts, with modest skill increases after sunrise. Post-processing does not improve the highly skillful raw WRF predictions at the mountain test sites. Since the framework addresses only systematic WRF deficiencies and identifies parameter pairs with a clear, non-site-specific physical mechanism of predictive usefulness, it is transferable without the need for recalibration, and therefore does not require any observational record to employ.

14. SUBJECT TERMS Fog Prediction, Visibility Prediction, Numerical Weather Prediction, WRF, Ensemble, MEPS, Post-processing, NWP	15. NUMBER OF PAGES 249
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

TOWARD IMPROVING SHORT-RANGE FOG PREDICTION IN DATA-DENIED AREAS USING THE AIR FORCE WEATHER AGENCY MESOSCALE ENSEMBLE

William R. Ryerson
Major, United States Air Force
B.S., Pennsylvania State University, 2001
M.S., Naval Postgraduate School, 2006

Submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY IN METEOROLOGY

from the

NAVAL POSTGRADUATE SCHOOL
September 2012

Author: _____
William R. Ryerson

Approved by:

_____	_____
Joshua Hacker Professor of Meteorology Dissertation Supervisor	Michael Bell Professor of Meteorology
_____	_____
Patrick Harr Professor of Meteorology	Wendell Nuss Professor of Meteorology
_____	
Rebecca Stone Professor of Oceanography	

Approved by: _____
Wendell Nuss, Chair, Department of Meteorology

Approved by: _____
Douglas Moses, Vice Provost for Academic Affairs

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This work develops and tests the viability of a new framework for producing short-range (<20 h) probabilistic fog predictions using post-processing of a 4-km, 10-member Weather Research and Forecasting (WRF) ensemble configured to closely match the Air Force Weather Agency Mesoscale Ensemble Forecast System. The raw WRF predictions produce excessive forecasts of zero cloud water, mainly caused by a negative relative humidity bias, which is largely traced to a warm overnight bias. Post-processing mitigates these systematic errors by leveraging traits of a joint parameter space in the predictions to modify individual ensemble members not predicting fog on their own. The method is generally most effective when the space is defined with a moisture parameter and a low-level stability parameter.

Cross-validation shows the method adds significant overnight skill to predictions in valley and coastal regions compared to the raw WRF forecasts, with modest skill increases after sunrise. Post-processing does not improve the highly skillful raw WRF predictions at the mountain test sites. Since the framework addresses only systematic WRF deficiencies and identifies parameter pairs with a clear, non-site-specific physical mechanism of predictive usefulness, it is transferable without the need for recalibration, and therefore does not require any observational record to employ.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AFWA	Air Force Weather Agency
ARW	Advanced Research WRF
ASOS	Automated Surface Observing System
BC	Boundary conditions
BSS	Brier Skill Score
DOD	Department of Defense
ET	Ensemble transform
FAA	Federal Aviation Administration
GEFS	Global Ensemble Forecast System
GFS	Global Forecast System
IC	Initial conditions
LAMP	Localized Aviation MOS Program
LWC	Liquid water content
MEPS	Mesoscale Ensemble Prediction System
MM5	Fifth-Generation Mesoscale Model
MOS	Model Output Statistics
NCEP	National Centers for Environmental Prediction
NCV	National Ceiling and Visibility product
NWP	Numerical weather prediction
NWS	National Weather Service
PBL	Planetary boundary layer
POD	Probability of detection
RACE	Radiation and Aerosol Cloud Experiment

RH	Relative humidity
RPSS	Ranked Probability Skill Score
RUC	Rapid Update Cycle
VIF	Visibility-in-fog
WRF	Weather Research and Forecasting model

ACKNOWLEDGMENTS

Thanks go to my advisor, Prof. Joshua Hacker, who not only helped guide the arc of this research but also mentored me so that I might emerge a better scientist. Working with him was truly a pleasure, and a personally very rewarding experience. Thanks to the rest of my committee, including Profs. Michael Bell, Patrick Harr, Wendell Nuss, and CAPT Rebecca Stone. Their time commitment and insightful input is tremendously appreciated. Mary Jordan spent hours processing the NWP output and observations to make them more manageable and was always available for programming consultation. This research is significantly better because of her efforts. Similarly, Kurt Nielsen, Saurabh Barve, Bob Creasey, Dick Lind, and Dr. Karl Pfeiffer provided technical and meteorological assistance that greatly contributed to the final product. Thanks to the numerical models team at AFWA, particularly Steve Rugg and Evan Kuchera, who expressed a need for this research and offered generous access to all MEPS data and resources.

Obstacles along the way were overcome much easier thanks to support from my fellow Ph.D. students, including Lt Col Mark Leejoice; CDR Beth Sanabia; Majs Brian Belson, Scott Miller, Darren Murphy, Cedrick Stubblefield, Jack Evans, and Paul Homan; and LCDRs Joel Feldmeier and Cory Cherrett. On a daily basis, Andy Penny and LCDR Heather Hornick were available for always-sound advice and I am fortunate to have had them nearby from the very start.

Thanks to my parents, James and Sally Ryerson, who taught me through their actions the power of perseverance and positive attitude. My parents and my brothers, Jim and Dan, have enthusiastically supported me through all new endeavors, for which I am immeasurably grateful.

Finally, this research would not have succeeded without the unwavering support and dedication from my beautiful wife Teresa and daughter Janna. Thank you for being patient when I was not home, and for reenergizing me when I was. I love you both.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

With varying frequency, fog occurs nearly globally, and in certain locales occurs regularly enough to significantly disrupt military operations. Visibility is reduced to less than 1 km wholly or partially due to fog on an average of 53 days each year at Tyndall Air Force Base, FL, 52 days each year at Kunsan Air Base, South Korea; and 24 days each year at Kabul International Airport, Afghanistan. This does not include instances of lighter fog that do not result in visibility <1 km but can still impact operations. At any given location away from an airfield, where reliable, consistent observations do not exist, the frequency of fog will differ from that at the nearest airfield, especially in mountainous or coastal terrain. Although the body of research for fog prediction in these more remote locales pales in comparison to work done at airfields and airports (see review by Gultepe et al. 2007), the disruption to military operations can be just as significant. Weapons selection, targeting, intelligence collection, search-and-rescue operations, and low-altitude helicopter transit are all impacted by fog, yet regularly occur some distance from the nearest airfield.

A visibility >7 miles generally does not cause major disruption to most military operations, and this is the highest value Department of Defense (DOD) airfields are required to report (i.e., any visibility >6.5 miles is normally reported as 7 miles). It is also the threshold below which a DOD weather observation is required to report the cause of the restriction (e.g., fog, haze, precipitation); as a matter or nomenclature, a visibility >6.5 miles is simply referred to as "unrestricted". Numerous thresholds below 6.5 miles also have operational significance because they dictate restrictions on certain aircraft types and equipment, pilot level of experience, etc., and these restrictions can vary depending on the type of airspace or mission involved. Meaningful thresholds exist as low as ¼ mile for certain helicopter operations, but in most cases, 1 mile or ½ mile is sufficient as the lowest needed threshold for operational decision-making. Products in the Air Force Weather Agency's (AFWA) Mesoscale Ensemble Prediction Suite (MEPS) that relate to visibility provide threshold exceedance probabilities at visibilities of 5 miles, 3 miles, and 1 mile.

The goal of this research is to investigate the viability of a new framework for producing short-term ($\leq$20 h) stochastic visibility-in-fog (VIF) predictions using existing mesoscale ensemble output, suitable for use in data-denied areas away from existing airfields. To do so, the framework examines ensemble predictions from an ensemble configured to closely match MEPS, assesses two primary sources of error in the output, and explores methods to understand and mitigate the error to arrive at more skillful visibility predictions. The next chapter will introduce some background and inherent challenges of visibility prediction, including an account of previous and current techniques that set the stage for the approaches tested here. Chapter III details the data used in this research. Chapter IV closely examines the numerical weather prediction (NWP) output and characterizes two primary sources of error affecting its skill. Chapter V describes the methodology used to develop several approaches to mitigate the error, and Chapter VI presents the results of testing these approaches. Finally, Chapter VII provides a summary and recommendations, as well as suggestions for future research.

II. BACKGROUND

A. STATISTICAL PREDICTION METHODS

Statistical prediction methods have shown great promise for the prediction of various weather elements to include VIF. Perhaps the most widely-used example of this is Model Output Statistics (MOS; Glahn and Lowry 1972), which was originally developed by applying regression equations to NWP model output so the output is statistically calibrated at designated locations. Vislocky and Fritsch (1997) excluded model data altogether, applying regression on observations, nearby observations, and climatic terms to produce 0–6 h visibility forecasts that outperformed persistence. Jacobs and Maat (2005) somewhat combined these approaches by using nearby observations *and* NWP output, as predictors to produce skillful ceiling, visibility, and wind forecasts at Amsterdam's Schiphol airport. This framework was advanced by Ghiradelli and Glahn (2010), who used it at hundreds of sites in the U. S. to develop predictive equations for 17 variables as part of the Localized Aviation MOS Program (LAMP). With an eye toward improving temperature, dewpoint, and wind forecasts at non-airport instrumented sites (e.g., national parks, sports stadiums, etc.) Hilliker et al. (2010) used statistical regression to effectively calibrate forecasts from the National Digital Forecast Database, which itself is NWP model output that has been modified by National Weather Service (NWS) forecasters. Most recently, Chmielecki and Raftery (2011) performed Bayesian Model Averaging, a kind of statistical calibration that assigns weighting to each member of an ensemble of NWP models, to improve the visibility prediction skill in the northwestern U. S.

Besides regression, other statistical prediction methods have been used with success. The Federal Aviation Administration's (FAA) National Ceiling and Visibility product (NCV) uses a decision tree framework to assimilate surface and satellite observations and combine them with model data to make ceiling and visibility predictions to 12 h (Herzegh et al. 2006). Bankert and Hadjimichael (2007) also used a decision tree construct to data mine output from the Rapid Update Cycle (RUC) NWP model to produce ceiling height forecasts at New York's John F. Kennedy Airport. Marzban et al.

(2007) built a neural network from NWP output and surface observation that, when used to make ceiling and visibility forecasts at 39 U. S. airports, collectively outperformed MOS. Bremnes and Michaelides (2007) tested with good results an ensemble of neural networks, trained from surface observations only, to produce short-term visibility forecasts. Taking this statistical method further, they improved the 6-h forecasts by using the predictions from each member of the ensemble of neural networks as inputs for a subsequent neural network. Hall et al. (2010) developed a framework that searches an archive to find analogs to the real-time surface and satellite observations in order to make forecasts out to 5 h that were shown to outperform persistence.

Regardless of the set of predictors used, each of these techniques requires a robust archive of observations (to include adequate occurrences of heavy fog if this is to be a focus of the tool), in order to develop, or train, the tool. For this reason, highly statistical approaches are most useful for airfields and other locations with a long observational record; in many cases, they produce skillful, inherently calibrated predictions that outperform NWP predictions alone. But such tools become less skillful as the available observational record for the desired location is decreased, and transferring a highly calibrated technique to a new location will result in less skill due to different location-specific behavior. An example of this is the Fog Stability Index developed by Freeman and Perkins (1998), which uses a regression equation from NWP model predictions of several 2-m parameters (temperature and dewpoint) and 850-mb parameters (temperature, dewpoint, and wind speed) to predict VIF in Hungary. Later, Dejmal and Novotny (2011) found the index showed poor skill at certain Czech Republic locations, and could be outperformed simply by using near-surface dewpoint depression as a predictor instead.

An additional drawback for highly statistical methods is that their effectiveness is dependent on their inputs being relatively stable over time, meaning there are no major changes or updates to the platform from which they originate. For example, a tool that relies on MOS output as a predictor is degraded by platform changes to MOS that occurred during the training period. Likewise, after the tool has been completed, its calibration becomes suboptimal as future changes to the MOS platform are made, resulting in decreased skill.

B. PHYSICAL PREDICTION METHODS

Physical prediction methods rely only on uncalibrated NWP output, placing full confidence in the NWP model's ability to simulate the phenomenon of interest. Since visibility is not explicitly included in NWP output, it is also necessary to include a visibility parameterization to convert the output to the visibility parameter(s) of interest. In a purely physical method, the visibility parameterization uses strictly first principles for the computation, and excludes any ancillary predictors that do not have a direct physical linkage to visibility. The advantages of this utopian approach are particularly noteworthy for the unique challenges posed by military operations. As long as NWP output is available, the framework can be applied, with no requirement for observations. Also, since first principles are valid everywhere, there is similarly no need for any training or calibration of the visibility parameterization. The risk of encountering a location not well represented in a training dataset (a ubiquitous concern for statistical methods) is negated.

C. HYBRID METHODS AND "PERFECT PROG"

In practice, a purely physical approach to VIF prediction is unviable to the difficulty of a visibility parameterization that only uses first principles, which would require the summing of scattering effects on visible light from millions or billions or individual, non-uniform, suspended water droplets. Due to the complex nature of such a process, as well as the fact that most NWP models are not designed to provide the needed inputs, the visibility parameterization almost certainly must involve some statistical aspects (that is, it must be parameterized to some degree).

However, the first requirement of a physical prediction method – placing full confidence in the NWP output, and therefore leaving it uncalibrated – is feasible for some applications and is known as the perfect prog assumption. Many authors have experimented with VIF prediction using the perfect prog assumption, coupled with a simple visibility parameterization using one or two variables (i.e., liquid water content) from the NWP output. Geiszler et al. (2000) tested a 9-km resolution version of the Coupled Ocean / Atmospheric Mesoscale Prediction System model over coastal

California in this way, finding the results had little skill. Two suggested reasons given for the poor performance were a lack of aerosol information in the NWP model, and poor representation of model topography. The first of these explanations could implicate not just the NWP output, but also the visibility parameterization, because aerosol information would only improve the predictions if it was adequately processed by a more sophisticated visibility parameterization. The second of these explanations suggests a shortcoming of just the NWP model.

Qualitatively, Zhou et al. (2009) obtained better results than Geiszler et al. (2000) when applying the same simple visibility parameterization to NWP output from the 32-km horizontal resolution, 21-member Short Range Ensemble Forecast system produced by the National Centers for Environmental Predictions (NCEP). Although formal verification was not performed, the authors believed limited objective evaluations conducted by local forecasters were promising.

While still using the perfect prog assumption, another common approach to VIF prediction is to apply a more statistically-generated visibility parameterization, sometimes by data mining observational data, to the NWP output. This is the approach used for visibility predictions from MEPS, which has a visibility parameterization developed from regression on a one-year training dataset of RUC analyses at thousands of U. S. locations. The predictors used are total column precipitable water, 10-m wind speed, and 2-m relative humidity (RH) (Kuchera 2011; Kuchera 2011, personal communication). The AFWA deterministic (non-ensemble) WRF NWP model also uses this strategy, although with a different visibility parameterization that primarily relies on RH as a predictor (AFWA Model Analysis Team 2004). Zhou and Du (2010) used the perfect prog assumption on a 15-km resolution, 10-member ensemble and applied a visibility parameterization developed to make a yes/no radiation fog prediction based on liquid water content (LWC), 10-m wind speed, 2-m RH, and cloud top and base heights. In a test region in eastern China, they found the predictions were more skillful than when the visibility parameterization used LWC only. Similarly, Gultepe and Milbrandt (2007) showed that a visibility parameterization utilizing LWC, 2-m RH, 2-m temperature, and satellite data (an observational input) outperformed one using only LWC as a predictor.

Since these more complex visibility parameterizations are tuned for an entire training domain instead of for individual sites, they tend to perform well when verified over large regions. However, since the predictors are heavily mined and/or only have an indirect physical linkage to visibility, they may not perform well at individual sites or even in certain climates that are different from the mean climate of the training data. Furthermore, it is not immediately clear from these studies to what extent error in the predictions is due to deficiencies in the visibility parameterization as opposed to deficiencies in the NWP model output.

D. STRIKING THE PROPER BALANCE FOR DATA-DENIED REGIONS

Striking the proper balance between a statistical and physical approach in VIF prediction suitable for DoD operations is an overarching theme of this research. Conceptually, a physical approach (using both the perfect prog assumption and a physical-based visibility parameterization) is most advantageous because it does not require observations and is transferable to anywhere model data are available. After separately examining error from the NWP output and from the visibility parameterization, we will show that under most conditions, the introduction of statistical components is necessary to obtain skillful predictions. These additions must be done judiciously and conservatively, such that they do not result in location-specific calibration but instead serve to mitigate the impact of certain persistent deficiencies in the NWP output. Additionally, exploring the tie between the statistical components introduced in this work and the physical reasoning behind why they work helps to focus future NWP and VIF prediction research efforts. It also makes the framework more adaptable to incremental improvements in the NWP platform.

In the FAA's NCV product, Herzegh et al. (2006) interpolated between surface observations in the U. S. to help produce the initialization state, which likely improves the skill of the predictions during the first few hours. While a similar approach is feasible in many parts of the world with an adequate observation network, others have sparse networks with hundreds or thousands of kilometers between reliable surface observation sites (e.g., North Africa, parts of Central Asia), and so this strategy is not used in this

work. Satellite observations may also be used to provide an observational element (e.g., Herzegh et al. 2006, Guidard and Tzanos 2007, Gultepe et al. 2009a, Hall et al. 2010), but these techniques struggle to distinguish ground fog from low clouds, especially at night, and are not included here.

By excluding an observational element in this VIF prediction framework, we likely sacrifice potential gains in skill (relative to persistence) during the early hours of the predictions. This concept was discussed by Ghiradelli and Glahn (2010), whose LAMP paradigm is to combine observations with MOS to increase the skill of MOS most during the first few hours, and more modestly thereafter (Figure 1). Vislocky and Fritsch (1997) noted that their observation-only statistical technique outperformed MOS until 6 h, with MOS having higher skill beyond that time. Furthermore, even with a sophisticated assimilation process, statistically-derived products such as NCV usually struggle to beat persistence during the first 4–6 h, with the noted exception of the analogue techniques of Hansen (2007) and Hall et al. (2010). It is worth noting that observational inputs are not completely excluded in an NWP-only framework since that they are obviously part of the NWP model assimilation process. Indeed, the multi-agency Joint Center for Satellite Data Assimilation is a dedicated research office that examines assimilation of satellite observations into NWP models, albeit with a broad focus as opposed to focusing specifically on VIF initialization and prediction. Regardless, existing research on NWP model and data assimilation in general seeks to provide the best possible initialization field, using all available observational sources and techniques as warranted. This research seeks ways to best leverage the NWP output derived from existing mainstream assimilation processes, instead of examining the assimilation itself.

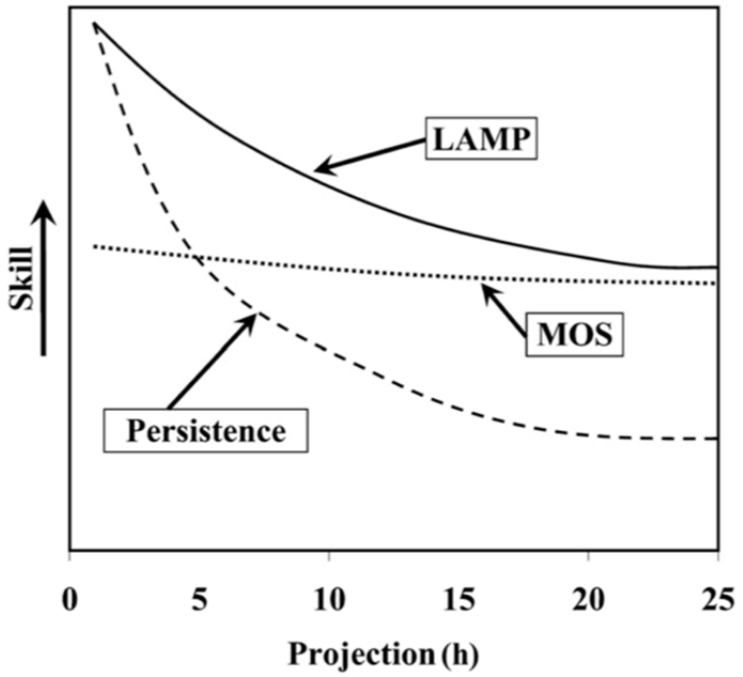

Figure 1. Notional concept of the LAMP paradigm, which combines observations with MOS to yield the most improvement over MOS during the first few hours. The improvement over MOS is more modest at later hours. (From Ghiradelli and Glahn 2010).

E. ADDITIONAL CONSIDERATIONS

Using a conceptual model of VIF prediction that includes two distinct sources of error, it is worth considering how using an ensemble system (as opposed to a single deterministic NWP model) fits into this conceptual model. Perhaps it is best to recognize that every WRF run will have error whether it is a deterministic run or a member of an ensemble, but the benefit of using an ensemble is to be able to sample at least part of that error so that it may be better understood and incorporated into a decision process by the end user. (For a general history and summary of ensemble forecast systems, see Kalnay 2003; for a real-world example of the cost-benefit of using an ensemble for ceiling and VIF prediction in the airline industry, see Keith and Leyton 2007). While the primary focus here is to identify and adapt for deficiencies in the WRF that result in prediction error in individual integrations, we perform this analysis in the context of an ensemble for several reasons. First, since each member of the ensemble varies not only in initial

conditions (IC) but also in physics suites (the ensemble setup is detailed in Chapter III), we can be more confident that consistent errors occurring in every member are likely to be attributable to a systematic WRF deficiency rather than due to a particular physics configuration or errors in the IC. Secondly, MEPS and other ensembles are already in wide use in DOD and elsewhere, and so we limit the operational value of our findings if we examine NWP VIF prediction errors without also considering and measuring the ensemble dispersion characteristics of those errors; that is, the degree to which the members tend to collectively sample the errors. By using deterministic verification techniques, we will show that the WRF output in MEPS is subject to systematic deficiencies that will negatively impact its skill in VIF prediction but can be improved with the addition of a conservative statistical component to the framework. Although the aim is not to revisit the design of the ensemble itself in this work (i.e., number of members, perturbation strategies, etc.) typical probabilistic verification practices are used to demonstrate how the skill of the MEPS is impacted by this work's findings, with the understanding that probabilistic verification measures are affected by both the errors from individual WRF members and ensemble dispersion shortfalls. With little modification, the methodology and results developed here could just as well be applied to deterministic WRF output to reduce error and improve skill, albeit without the benefit of error sampling an ensemble provides.

Furthermore, the focus on systematic WRF deficiencies rather than individual member behavior is quite different from an ensemble calibration, which Eckel and Mass (2005) suggested should be performed separately on each member. Recent history suggests MEPS members will continue to be periodically added, deleted, and modified in attempts to improve some aspect of prediction (but not necessarily always improving VIF prediction), so addressing the observed systematic deficiencies demonstrated by most or all of the members represents the most impactful, enduring contribution toward achieving our aim. Instances where individual member behavior is particularly noteworthy will be highlighted to help inform future research on NWP development, particularly with regard to planetary boundary layer and microphysics parameterizations.

Besides error from NWP predictions and from visibility parameterizations, other sources of error exist that will not be thoroughly examined in this work but warrant consideration. In their work, Geiszler et al. (2000) alluded to error incurred by using a single model grid point for verification. Known as subsubgrid-scale variability or representativeness error, this error stems from the fact that the NWP predictions represent average values in a model grid box, yet the verifying observations are taken at a single point within that box. Even for the 4-km model grid used in this research, smaller-scale fog structure exists within the grid square that will contribute to error when verification is performed against a point observation. This research will not closely investigate subgrid-scale variability, but it is briefly examined and discussed in Chapter IV to gauge its potential impact. Where examined, it was not believed to substantially affect the results.

Observation error can be defined as the measurement error of a given instrument or procedure. In an ensemble verification, Hacker et al. (2011) found that ignoring observation error had the effect of making the ensemble appear less dispersive than it is, which can in turn affect its overall skill. It is not as crucial to address observation error when performing comparative verification since it affects all techniques relatively equally over time, and it will not be considered in this work. Nevertheless, the challenges inherent in gathering VIF observations mean observation error is likely to be greater than what might be expected for verification of temperature, for example. These challenges are documented in the next chapter.

Three other previous studies helped inform the setup and approach ultimately used in this research. Bang (2006) tested deterministic VIF predictions for a heavy fog case at Incheon, South Korea using both the Weather Research and Forecasting (WRF) model and Fifth-Generation Penn State/NCAR Mesoscale Model (MM5) at various horizontal grid spacing from 54 km to 2 km. The high-resolution WRF predictions were the most skillful, lending promise to the prospects of using MEPS, which is based off of 4-km grid spacing WRF runs, for this work. They found the WRF model runs tended to underforecast fog, and dissipate it too rapidly.

Tardif (2007) examined the impact of NWP model vertical resolution on radiation fog prediction at the Paris-Charles De Gaulle airport. Using a sophisticated 1-D model

designed specifically for fog (COBEL), he found having more vertical layers near the surface improved the timing of fog onset, which tended to be delayed in the lower-resolution experiment due to the inability to create a shallow fog layer, resulting in inadequate radiative cooling (note that fog droplets have higher longwave emissivity than unsaturated air, and therefore will cool a layer more quickly when present). When increasing the resolution isn't possible, he suggested examining radiative cooling rates in the NWP model for signatures that may assist with radiation fog initiation. The lowest model level in MEPS (about 20 m above ground level) is even higher than the lowest model level in the low-resolution COBEL case (about 12.2 m above ground level), and we will show that similar behavior was observed.

Lastly, Zhou and Ferrier (2008) described a process for obtaining LWC values during radiation fog events by explicitly solving the governing equation that describes LWC as a function of turbulent exchange coefficient, droplet gravitational settling flux, condensation rate due to cooling, and height of the fog layer. Verification of the technique during an observed fog event was promising, and the authors suggest the technique could be successfully utilized to adjust the initial LWC predictions provided by NWP predictions if the NWP model is able to provide accurate predictions of the dependent variables. Our research examined the prospects for such an approach in MEPS, but as we will show, it would not provide large skill improvements due to the high number of cases in MEPS of missed fog, for which the fog depth is zero and the technique maintains zero LWC.

F. VISIBILITY PARAMETERIZATIONS

The traditional role of an NWP microphysics scheme is to predict water vapor and hydrometeor mixing ratios. In the last decade, these single-moment schemes (termed such because they predict only one parameter – the mixing ratio – for each species) have been joined by double-moment schemes, which make physics-based predictions of hydrometeor size distribution in addition to mixing ratio. In some cases, this double-moment capability is reserved only for precipitation species (Thomspon et al. 2008), but others include predictions of cloud water droplet distribution that are based on turbulence and instability parameters (Morrison et al. 2005), or cloud condensation nuclei

concentration, if available (Lim and Hong 2010). For a more complete history of how microphysics scheme capabilities have evolved, see Seifert (2009).

Many operational NWP models and ensemble systems being run at large centers, to include MEPS, have not yet assumed the additional complexity and computational expense needed to implement double-moment schemes. Instead, in the single-moment schemes in widespread use, the shape of the size distribution is held constant. To overcome this deficiency without compromising the essence of a first principles approach, the experiments in this research will use as a launching point two visibility parameterizations that rely only on the crucial variable available in the NWP output (i.e., liquid water mixing ratio), yet were developed with the benefit of field measurements.

Before describing the visibility parameterizations, note that both rely on inputs of cloud water mass concentration in units of g m^{-3}. This is different from the liquid water mixing ratio provided in most NWP output, which is in units of kg kg^{-1}. To avoid confusion, this research will always refer to cloud water in terms of the mass concentration in units of g m^{-3}, denoted by the symbol q_c. In addition, note that each parameterization provides output in terms of extinction coefficient, β_e, which is different from visibility yet will be used as the verifying parameter in this research. The reason for this choice, as well as the relationship between β_e and visibility, are explained in the next chapter.

1. Stoelinga and Warner 1999

Kunkel (1984) used in-situ measurements of 11 fog events to measure microphysical properties of droplets, and formulated a relationship between q_c and β_e used by Stoelinga and Warner (1999), hereafter SW99, as part of a case study in NWP ceiling and visibility prediction. It has been widely used in numerical weather prediction applications ranging from limited research experiments (e.g., Geiszler et al. 2000, Bang 2006, Chmielecki and Raftery 2011) to inclusion in the FAA's NCV product (Herzegh et al. 2006) and the NCEP Very Short Range Ensemble Forecast (Zhou et al. 2010), and is often referred to as the Stoelinga and Warner parameterization when used in this context:

$$\beta_e = 144.7(q_c)^{0.88}, \tag{1}$$

where β_e is in km^{-1}.

2. Gultepe 2006

More recently, Gultepe et al. (2006), hereafter G06, used field measurements from the Radiation and Aerosol Cloud Experiment (RACE) to also prepare a relationship between q_c and β_e:

$$\beta_e = 178.6(q_c)^{0.96} \tag{2}$$

More precise visibility parameterizations exist that incorporate additional variables, yet still maintain a physically-based foundation because all the inputs have a direct physical link to visibility. Gultepe et al. (2006) showed from the RACE data that incorporating both q_c and cloud droplet number concentration, N, into the parameterization provides a better fit to the observed β_e. The importance of N in VIF lies in the fact that, for a given value of q_c, many smaller droplets have a larger total cross-sectional area, and therefore a larger β_e, than fewer larger droplets (Koenig 1971, Brenguier et al. 2000, Gultepe et al. 2006). However, like cloud droplet size distribution, N is normally held constant in most current microphysics schemes, including each scheme used in MEPS (see Skamarock et al. 2008 for a summary of each scheme as well as additional references describing their details). Therefore, using the more sophisticated parameterization without skillful predictions of N has no added benefit over the G06 parameterization in equation (2). Several techniques have been proposed to estimate N when it is not given by the NWP output, to include using the airmass characteristics (Clark et al. 2008), predicted temperature (Gultepe and Isaac 2004), or predicted level of supersaturtaion combined with airmass characteristics (Bott and Trautman 2002). Since in this work we do not have verifying observations of either q_c or N, attempting to separately account for uncertainty in these variables would be highly ambiguous. Instead, we will quantitatively examine uncertainty in the single-parameter visibility parameterizations given by (1) and (2), with the impacts of N reserved for qualitative consideration.

III. DATA

A. NWP OUTPUT

To maximize the operational utility of the findings, the ensemble system used for this research is configured to match that of the AFWA MEPS as closely as possible. The details of the MEPS configuration are based on work by Hacker et al. (2011a), hereafter H11, in which several methods of producing IC and physics perturbations were examined with a goal of finding "the most skillful ensemble, with the least degree of complexity" such that it would be operationally viable given typical computational restraints. As with most operational NWP models, incremental changes have since been made to the MEPS configuration, but the basic setup exists as it did when it was closely replicated to create the runs for this research in late 2010 (Kuchera 2011, personal communication). The configuration used for the runs is described below, with further details and justification available in H11.

The ensemble consists of 10 WRF (ARW version 3.2) members with 4-km horizontal grid spacing and 42 vertical sigma levels. This high-resolution domain is nested within a larger 12-km grid spacing middle nest, which in turn is nested within a larger 36-km grid spacing outer nest. Each member obtains its ICs and lateral boundary conditions (BC) from a different member of NCEP's Global Ensemble Forecast System (GEFS, Wei et al. 2008). H11 found that this method of direct dynamical downscaling from a global NWP model to create ICs did not perform as well as when more advanced methods, such as an ensemble-transform Kalman filter, are used. However, given the low computational expense and implementation in MEPS, it is used here. For its part, GEFS is constructed from the Global Forecast System (GFS) NWP model using an ensemble transform (ET) technique (Bishop 1999) that accounts for regional differences in analysis error variance from the operational 3D-var scheme by including regional scaling of the initial perturbation (H11).

Certain properties of the lower boundary (land surface) are assigned a different value in each member based on random draws from Γ-like distributions, with distribution

parameters selected based on physical arguments and empirical data. These properties are the albedo, soil moisture availability, and roughness length, and the values assigned to each member do not change throughout the experiment. This technique was described by Eckel and Mass (2005), and led to small error reductions in lower tropospheric predictions when tested by H11 compared to when they were not used.

NWP model uncertainty can be considered distinct from IC or BC uncertainty in that it arises from, among other things, imperfect parameterizations of subgrid-scale processes (microphysics, planetary boundary layer fluxes, deep convection), radiative forcing (shortwave and longwave), and land-surface fluxes. Running a unique combination of parameterizations for each member is one way to sample this uncertainty, ultimately resulting in more skillful predictions. This approach was promoted by Eckel and Mass (2005), and H11 demonstrated its importance for near-surface predictions, stating the technique "appears critical for probabilistic prediction in the PBL (planetary boundary layer)." The specific parameterization combinations (hereafter called "physics suites") should not be selected arbitrarily because some suites that were not tuned together during their development can produce unreasonable and even unstable predictions (H11). The 10 suites used in this work are given in Table **1**. They are the same as those used in H11, although they are numbered differently, which is explained as follows. During the testing of various suites, H11 initially identified 20 that appeared to be most viable (stable, and producing reasonable predictions), later selecting the best 10 for inclusion, which are the 10 used here. However, in this work, the member number, which has no meaning aside from identification purposes, is from its number in the original 20. References for the physics options are found in Skamarock et al. (2008).

The cumulus parameterization listed in Table 1 is used on the middle- (12-km grid spacing) and outer- (36-km grid spacing) nests only; no cumulus parameterization is used for the 4-km inner nest.

The period of the study is from 21 November 2008 through 21 February 2009, with NWP runs initialized every three or four days to minimize highly-correlated cases. In all, 29 ensemble runs were performed. Each run was initialized at 0000 UTC, and the output was compiled at hourly intervals out to 20 h. Although the 0-h water vapor field

in each ensemble member is downscaled from its parent member from the global ensemble suite, solid and liquid water phases are not initialized.

Table 1. Summary of physics suite used for each member.

Member	**Microphysics**	**PBL**	**Shortwave**	**Longwave**	**Land Surface**	**Cumulus** (none on inner-most nest)
1	Kessler	YSU	Dudhia	RRTM	Thermal	KF
5	WSM6	MYJ	CAM	RRTM	Thermal	KF
7	Kessler	MYJ	Dudhia	CAM	Noah	BM
8	Lin	MYJ	CAM	CAM	Noah	Grell
10	WSM5	YSU	Dudhia	RRTM	Noah	KF
11	WSM5	MYJ	Dudhia	RRTM	Noah	Grell
15	Lin	YSU	Dudhia	CAM	RUC	BM
16	Eta	MYJ	Dudhia	RRTM	RUC	KF
17	Eta	YSU	CAM	RRTM	RUC	BM
19	Thompson	MYJ	CAM	CAM	RUC	Grell

Since cloud water is the primary field of interest in the study of fog, the first six hours of each case are evaluated with caution to account for the spin up of the field to a stable state, and these hours are not included in certain parts of the verification where noted. As previously discussed, given the NWP-only nature of this framework, skillful predictions during the first few hours are not an emphasis of this work, and so we mainly focus on the 6–20 h prediction timeframe (2200–1200 LT) representing short-term operational planning.

Figure 2 shows the domain of each of the three nests. Verification focuses on seven airfields (Figure 3) in California and Nevada representing three regions with distinct mesoscale influences: Crescent City (airport identifier KCEC, elevation 17 m) and Arcata (KACV, 66 m) represent a coastal region as both are less than 1 mile from the Pacific Ocean; Stockton (KSCK, 9 m), Modesto (KMOC, 29 m), and Merced (KMCE, 57 m) represent a valley region subject to frequent and heavy overnight radiation fog; and Emigrant Gap (KBLU, 1610 m) and Reno represent a mountainous region, with both

sites at relatively high elevations and surrounded by mountainous terrain. The NWP predictions for any given level at these seven sites are obtained by bi-linearly interpolating from the four grid points laterally surrounding each station. In most cases, NWP values from the lowest model layer or the 2-m level are of most interest. The lowest model layer (hereafter layer 1) exists at a height of 19–21 m above the model's ground level. WRF post-processing computes 2-m values of temperature and water vapor from the heat and moisture fluxes provided by the PBL scheme using the flux-profile relationship (Stull 1988).

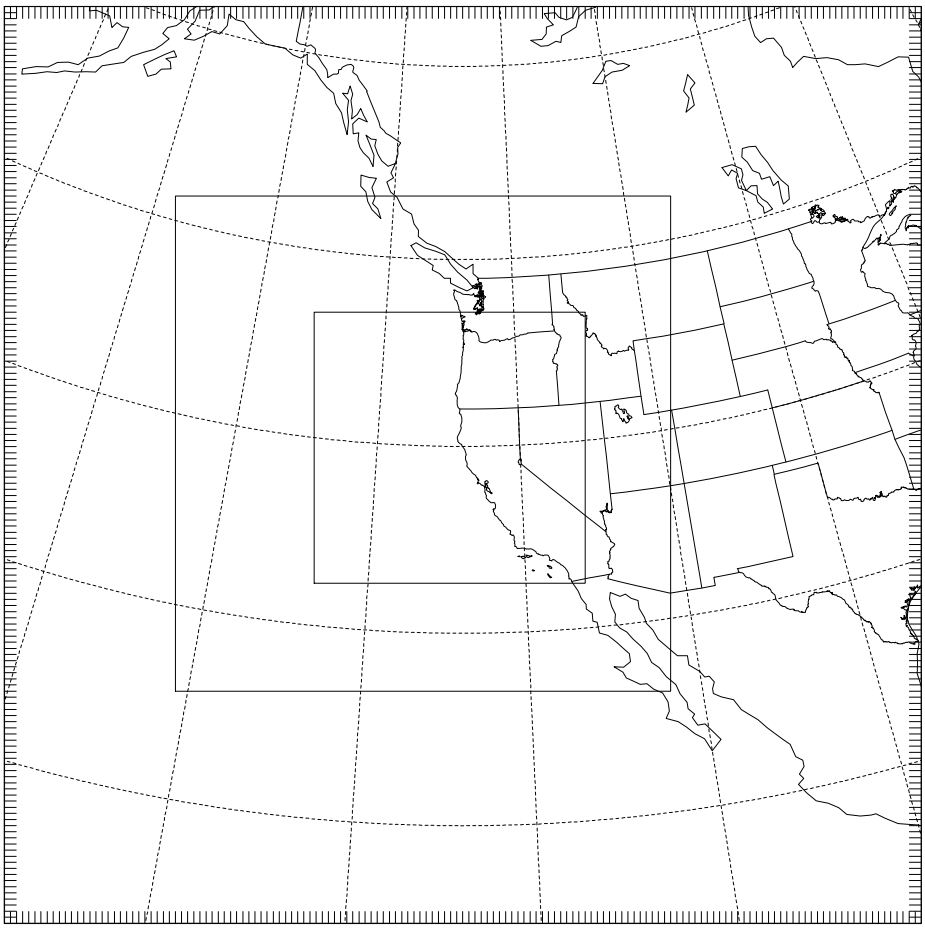

Figure 2. Domain of the three nests for WRF runs. (From Hacker et al. 2011b).

Figure 3. Location of verification sites (with elevation in meters). (Map background courtesy of Europa Technologies, Google, and INEGI 2011).

B. OBSERVATIONS

1. Physical Description of Visibility

Each of the seven airfields used for verification is instrumented with an Automated Surface Observing System (ASOS), which is maintained by NWS, FAA, and DOD. ASOS is the primary observation system in the U. S. in use at hundreds of airports and other sites (NWS 1999). Except in rare instances such as equipment malfunction or visibilities less than 0.125 mi, visibility observations are left to the ASOS's fully automated procedure, which utilizes measurements from a forward scattering sensor (Office of the Federal Coordinator for Meteorological Services and Supporting Reserarch 2005). The sensor consists of a flash lamp projector, which flashes a cone of visible light twice each second, and a detector. The detector is situated outside the lamp's projection

cone (Figure 4) so that the amount of pulsed light it receives is dependent on the collective forward scattering coefficient of the scatterers in the sample volume (National Oceanic and Atmospheric Administration/DOD/FAA/U. S. Navy 1998). Visibility is actually a function of the total extinction coefficient, β_e, but the other components of extinction (backward scattering and absorption) are comparatively negligible compared to the forward scattering. Therefore, the system assumes the measured forward scattering coefficient is also an accurate estimate of the total extinction coefficient (British Atmospheric Data Centre 2006).

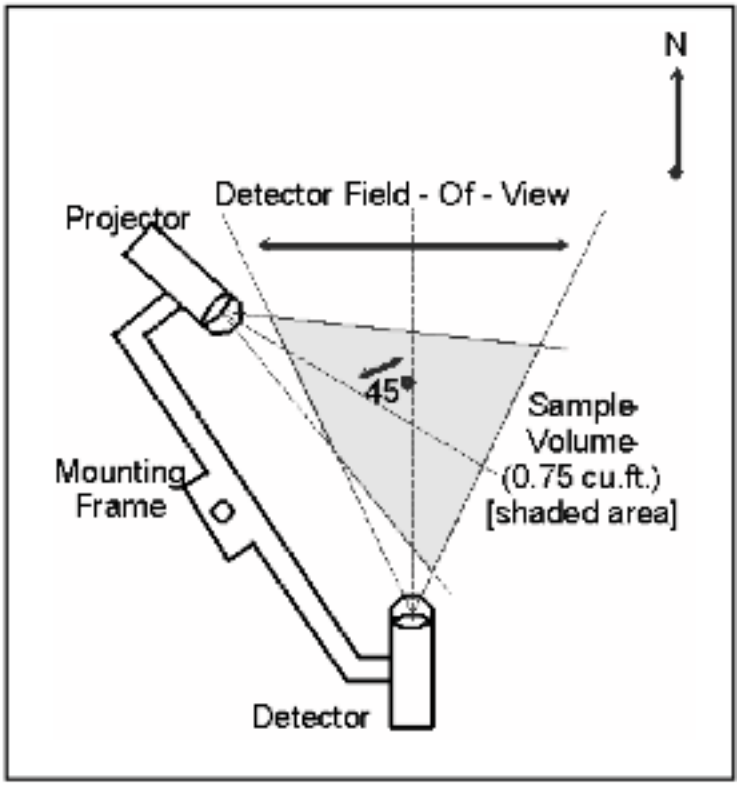

Figure 4. Top view schematic of the ASOS visibility sensor. Not shown is the integrated ambient light sensor. (From National Oceanic and Atmospheric Administration/DOD/FAA/U. S. Navy 1998).

Even with an accurate estimate of β_e, estimating the true visibility is quite complex. Consider for example the FAA definition of visibility: "The ability, as determined by atmospheric conditions, to see and identify prominent unlighted objects by day and prominent lighted objects by night" (FAA 2012). The ability to see and identify objects during the daytime is a matter of detecting the contrast, C, between the object and

its background. Middleton (1954) defines this quantity as the ratio of the brightness difference between the object and background, and the brightness of the background:

$$C = \left|(B - B')/B'\right| \tag{3}$$

where B is the brightness of the object, and B' is the brightness of the background. As viewed by an observer from a given distance, r, the apparent contrast, C_r, can be written in terms of the apparent object brightness, B_r, and the background brightness:

$$C_r = \left|(B_r - B')/B'\right| \tag{4}$$

Note that equation (4) uses the same background brightness, B', as equation (3) instead of using an "apparent" background brightness. This is because the assumption is made that the background is an infinite (flat-earth) atmosphere, and therefore the background brightness does not change regardless of r (Koschmieder 1924). The maximum reportable visibility for most ASOS stations is 10 mi, so the flat-earth, constant background brightness assumption is reasonable.

Duntley (1948) showed that the quantity $|B_r - B'|$ varies exponentially with distance as:

$$\left|(B_r - B')\right| = \left|(B_o - B')\right| \exp\left(-\int_0^r \beta_e dr\right) \tag{5}$$

By combining equations (3), (4), and (5), we can obtain an expression for the ratio of the apparent contrast at distance r to the actual contrast at distance zero. Middleton (1954) called this quantity the *contrast attenuation*:

$$\frac{C_r}{C_o} = \exp\left(-\int_0^r \beta_e dr\right) \tag{6}$$

Several less-precise assumptions are made in equation (6) to produce a visibility observation. First, the contrast attenuation does not directly indicate whether an object at distance r is visible. As mentioned earlier, the *visibility* of an object is determined by whether or not C_r is large enough to be detected by the observer. Objects with large values of C_o, such as an all-black target against a white sky, will also have larger values of C_r from any given distance than will a lighter object, even though the objects will have the same contrast attenuation. As r increases, C_r for the lighter object will eventually

become too small to be detected. The darker object, however, will remain visible until a greater distance is reached such that its value of C_r also becomes too small. Therefore, daytime visibility depends on the brightness of the object being viewed, and is greater for objects with a brightness significantly different than the background brightness (note that bright objects can also have large values of C_o if viewed against a darker background, such as an overcast sky). The fact that visibility is object-specific is not just a limitation with automated instrumentation, as a human observer viewing landmarks of various brightnesses is subject to this same complication. Nevertheless, in order to use equation (6) in an all-purpose visibility application such as ASOS, a reference value of C_o must be established. For ASOS, this reference value is 1, which can be thought of as corresponding to perfectly black reference object.

Next, the exact threshold of C_r below which an object is no longer visible will vary based on the individual and also the size of the object. Based on several laboratory and field experiments detailed in Middleton (1954) and elsewhere, values between 0.02 and 0.065 are typically used in the literature. ASOS uses a conservative value of 0.05 (Belfort Instrument 2005).

The last complicating assumption discussed here arises from the fact that β_e is only measured at the instrument and not over the entire path length. Therefore, it must be assumed the measured value is representative of the entire path.

By applying the three assumptions above, Equation (6) simplifies to

$$0.05 = \exp(-\beta_e r_t), \tag{7}$$

where r_t is the threshold distance at which the object is no longer visible. Solving for r_t results in the daytime visibility algorithm used in ASOS:

$$r_t = \frac{3}{\beta_e} \tag{8}$$

The ability to see and identify *lighted* objects, which defines nighttime visibility, involves slightly different physics than the daytime derivation. If the object has luminous intensity I_o, the illuminance, E_r, at any distance is defined by Allard's law:

$$E_r = \frac{I_o \exp(-\int_0^r \beta_e \, dr)}{r^2} \tag{9}$$

As with C_r during the daytime, there exists a critical threshold value of E_r below which the light is no longer detected. This threshold value, E_t, varies based on several factors, including the background luminance (Rasmussen et al. 1999). Using data from field testing of the first airport transmissometer, Douglas and Booker (1977) noted E_t is also affected by the distance between the observer and the source because at closer range, the glow from the source itself has a detrimental effect on the observer's ability to detect the source. Empirically, they estimated this relationship as:

$$E_t = \frac{0.052}{r}, \tag{10}$$

Replacing E_r in equation (9) with the expression for E_t in equation (10) and simplifying results in the expression

$$0.052 = \frac{I_o}{r} \exp(-\int_0^r \beta_e \, dr) \tag{11}$$

with β_e in km^{-1} and r in km.

An additional simplification is made by assuming the light source has luminous intensity, I_o, of 25 candelas (Rasmussen et al. 1999). Finally, by assuming homogeneity of β_e along the path, we may eliminate the integral as we did in the daytime derivation. The result is the ASOS nighttime visibility algorithm (Belfort Instrument 2005):

$$r_t = \frac{6.2 - \ln r_t}{\beta_e} \tag{12}$$

Unlike the daytime algorithm, the nighttime algorithm is implicit, and therefore must be solved iteratively for r_t given β_e.

Traditionally, β_e is expressed in km^{-1} and visibility, r_t, in miles. In Table 2, the ASOS daytime and nighttime equations are summarized in modified form to account for this mismatch of units.

During the verification or calibration of any fog prediction scheme in which post-processed visibility observation are used, failing to distinguish between the daytime and nighttime regimes can be a large source of error. For a given value of β_e, the daytime algorithm produces visibilities at least 20% lower than the nighttime algorithm in the visibility range of interest ($\leq$6.5 mi). The difference is larger at low visibilities, with daytime visibility barely half as large as a nighttime visibility of 1 mi (Figure 5).

Which algorithm is used depends on a separate ambient light sensor included in ASOS. The ambient light threshold determining day or night is very low (between 5 and 30 lux), such that the nighttime algorithm is normally only used when the sun is several degrees below the horizon or lower (National Oceanic and Atmospheric Administration/DOD/FAA/U. S. Navy 1998, Waynant and Ediger 2000).

Table 2. ASOS daytime and nighttime visibility algorithms. (After Belfort Instrument 2005).

Day	$r_t(miles) = \dfrac{1.862}{\beta_e(km^{-1})}$
Night	$r_t(miles) = \dfrac{5.7 - \ln r_t(miles)}{1.609 \cdot \beta_e(km^{-1})}$

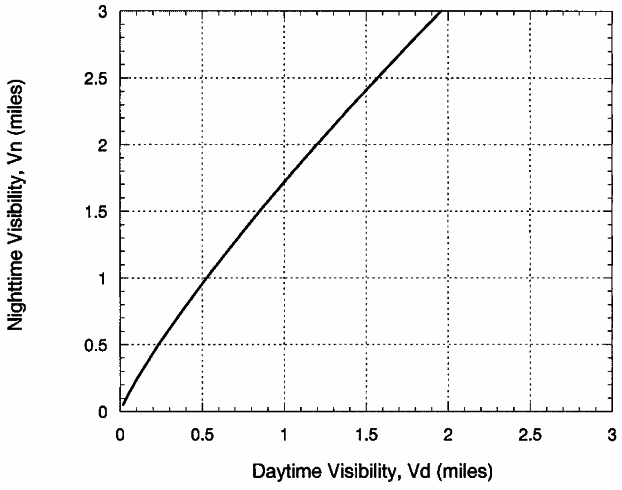

Figure 5. Comparison of results from the ASOS nighttime and daytime visibility algorithms when computed with the same extinction coefficient. (From Rasmussen 1999).

The precision of visibility reports increases as visibility decreases. Computed visibilities greater than 2.75 miles are rounded to the nearest mile, between 1.875 and 2.75 are rounded to the nearest half-mile, and between 0.125 and 1.875 are rounded to the nearest quarter-mile. If the computed visibility is below 0.125 miles, it is normally supplanted by a more precise value from a human observer, if available. Otherwise, it is simply reported by ASOS as being less than one-quarter mile (National Oceanic and Atmospheric Administration/DOD/FAA/U. S. Navy 1998).

2. Processing of Visibility Observations

It is preferable to use β_e as the verifying parameter since it is the measured quantity. When helpful for interpretation or comparison with other techniques, results will be converted to visibility using the appropriate ASOS algorithm from Table 2. While the uncertainty existing in the conversion of β_e to visibility is perhaps a significant source of error, it will not be the focus of this research. In addition to the several imperfect assumptions detailed above, producing visibility observations in practice is also subject to error from differences in the shape or color of the objects or lights being viewed, the viewing angle with respect to the horizon, and the position of the sun. Some of the assumptions made to mitigate these are necessitated by the use of automated instrumentation, and some are required even with a human observer simply due to the nature of the measurement.

Raw, one-minute β_e observational data for the seven verification sites was obtained from the National Climatic Data Center website (2011). In order condense this data into a single hourly β_e observation suitable for verification, the 10 β_e values during and prior to the top of each hour (spanning 10 min) were averaged. Other measured parameters, such as temperature, dewpoint temperature, wind direction, and current weather condition were taken directly from the official METAR observation.

The basic process used by ASOS to determine the current weather condition plays a critical role in preparing the data and is summarized in Figure 6. As with all ASOS measurements, the process is completely automated except during equipment malfunction or other extenuating circumstances (e.g., smoke in vicinity, presence of a funnel cloud,

etc.). In the overwhelming majority of cases, any reduction in reported visibility to below 7 mi as measured by the forward scattering sensor can be ascribed to precipitation (of some form), mist, fog, or haze. Precipitation is detected by the ASOS precipitation gauge and reported accordingly, regardless of the visibility. Independently, if the reported visibility is <7 mi and the dewpoint depression is <2.2 K, mist or fog is reported. The distinction between mist and fog is one of severity; fog is used if reported the visibility is <0.625 mi, while mist is used otherwise (hereafter, both will be called fog for simplicity). Note that fog and precipitation can be reported together if both conditions are met. Lastly, if the reported visibility is <7 mi but the dewpoint depression is >2.2 K, haze is reported, unless precipitation is also reported, in which case the precipitation takes precedence (National Oceanic and Atmospheric Administration/DOD/FAA/U. S. Navy 1998).

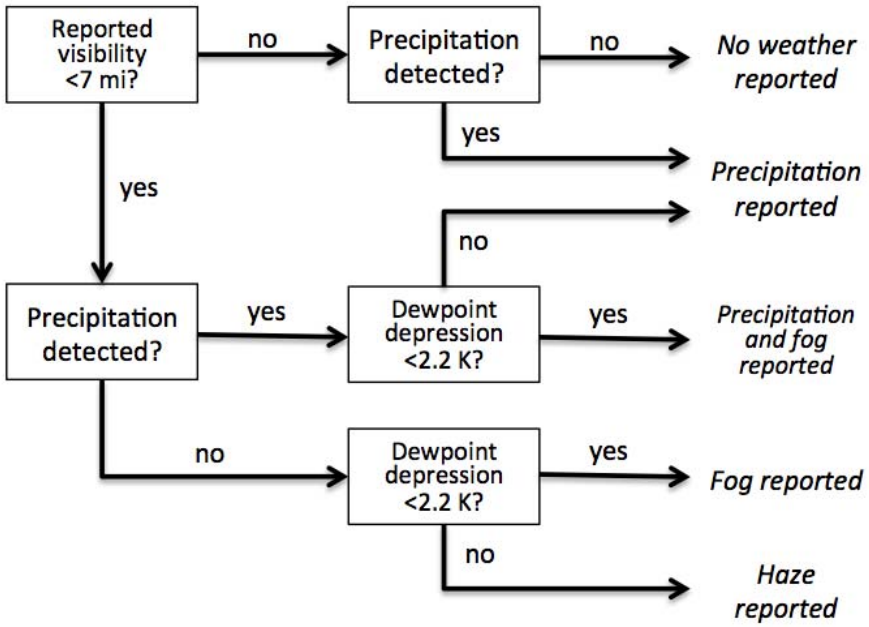

Figure 6. Summary of basic logic used by ASOS to determine present weather. Only the aspects of the logic relevant to this research are shown.

This logic makes the following the assumptions that must be deemed acceptable before using the observations as ground truth:

- Fog and haze cannot coexist
- If the reported visibility is <7 mi, the dewpoint depression is <2.2 K, and it is not precipitating, then fog must be present

Determining the presence of fog based only on the visibility and dewpoint depression may seem a crude approximation but it is consistent with a lack of distinction between fog, haze, and mist. Automated instrumentation aside, the distinction between haze and fog is quite inexact. Haze is defined as aerosol particles that "increase in size with relative humidity", but not so large that they reach their activation radii, at which point they would become mist droplets (American Meteorological Society 2012). The exact RH at which this occurs depends on the aerosol characteristics (Rogers and Yau 1989), and cannot possibly be known in every case. If the RH remains high enough, the droplets will continue to grow and eventually be classified as fog droplets. The ASOS dewpoint depression threshold of 2.2 K (which corresponds to an RH of 80–90% in most cases) is likely to be below the activation threshold of most haze particles (Rogers and Yau 1989). Referring to haze, mist, and fog, the American Meteorological Society Glossary (2012) states "there is no distinct line…between any of these categories". Given the indistinct transition between haze and fog from an observational standpoint, the ASOS logic seems reasonable. At worst, some instances of moist haze whose particles have not yet reached activation radii but *are* causing a visibility restriction will be misclassified as fog.

Once the hourly reports of temperature, dewpoint temperature, wind direction, and present weather have been combined with an hourly β_e value, additional processing is needed to isolate just the contribution of fog to the measured β_e. First, any observation with β_e <0.29 km^{-1} (approximately corresponding to daytime visibility of 6.5 mi and nighttime visibility of 8 mi), is simply classified as a no-fog case. In these cases, the actual value of β_e is not retained because 1) except for precipitation, ASOS does not report the phenomenon responsible for any reduction in visibility, and 2) it is outside the range of visibilities relevant for most DoD operations.

Next, since haze and fog cannot coexist, any observation reporting haze is also classified as a no-fog case, even if β_e>0.29 km^{-1}. In these cases, β_e is reassigned a value of 0.10 km^{-1}, an arbitrary figure that simply ensures these observations are not confused with cases of fog.

Finally, observations with β_e>0.29 km^{-1} and precipitation occurring were removed from the dataset al.together, even if fog was also reported. For a given β_e, the relative contributions of fog and precipitation are inseparable in this case.

After the filtering described above, the remaining observations are those with β_e>0.29 km^{-1} due to fog alone, thus comprising the fog cases of the verification dataset. In these cases, the β_e value was preserved.

A small percentage of the verification data did not fit into one of the above categories and required special treatment. If a nighttime observation reported a β_e value in the range 0.29-0.37 km^{-1} with no precipitation, no present weather was normally reported since this β_e range corresponds to reported visibilities $\geq$7 mi using the nighttime algorithm and subsequent rounding. In these cases, the present weather was deduced to be either haze or fog using the same dewpoint depression criteria used by ASOS.

The processing of the hourly β_e observations is summarized in Figure 7.

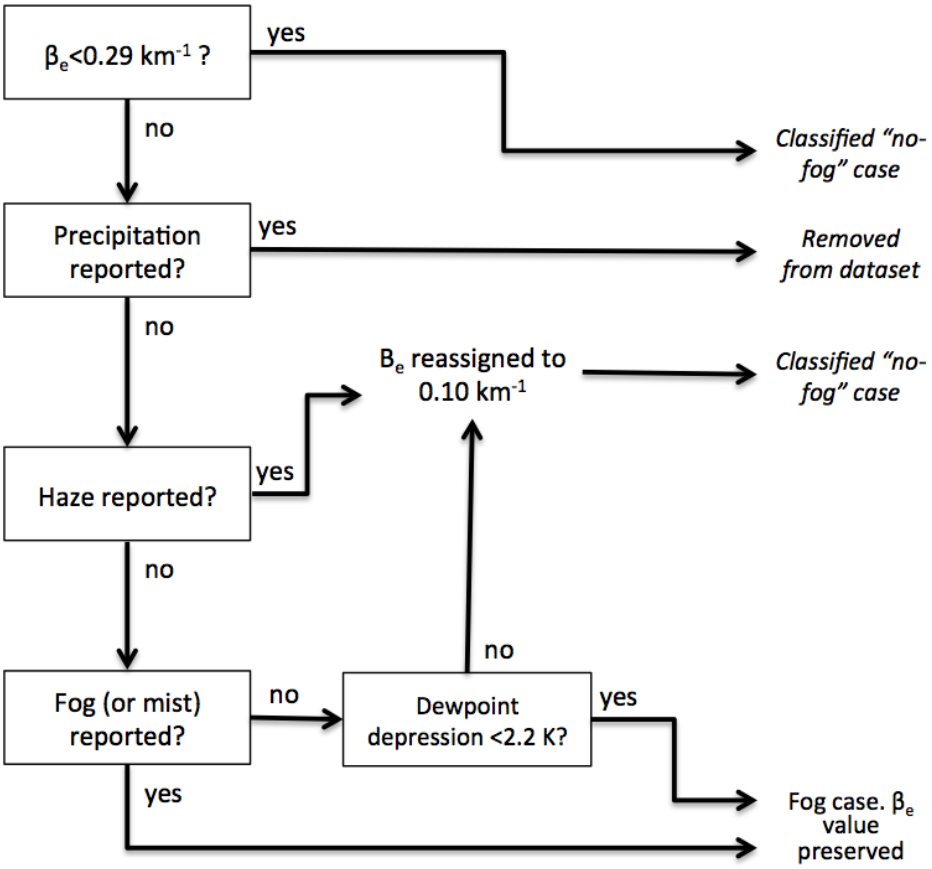

Figure 7. Summary of the processing of the hourly observations to isolate the effects of fog on the observed β_e values.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. ASSESSING VISIBILITY PREDICTIONS

A. NWP ERROR VERSUS VISIBILITY PARAMETERIZATION ERROR

1. Parametric Visibility Parameterization

To understand the relative impact of error in the NWP model predictions of q_c versus error in the visibility parameterization, a simple parametric visibility parameterization was developed to account for uncertainty in the field measurements used to formulate the SW99 and G06 visibility parameterizations. The specific goal is to roughly qualify the errors that may result from imperfect empirical relationships between β_e and LWC. Development proceeded without the raw datasets from Kunkel (1984) and G06, but was instead done by estimating characteristics of the data from the corresponding published scatter plots (Figure 8). The end result is therefore considered an approximation of the true uncertainty in the data, and is sufficient for the conclusions drawn here.

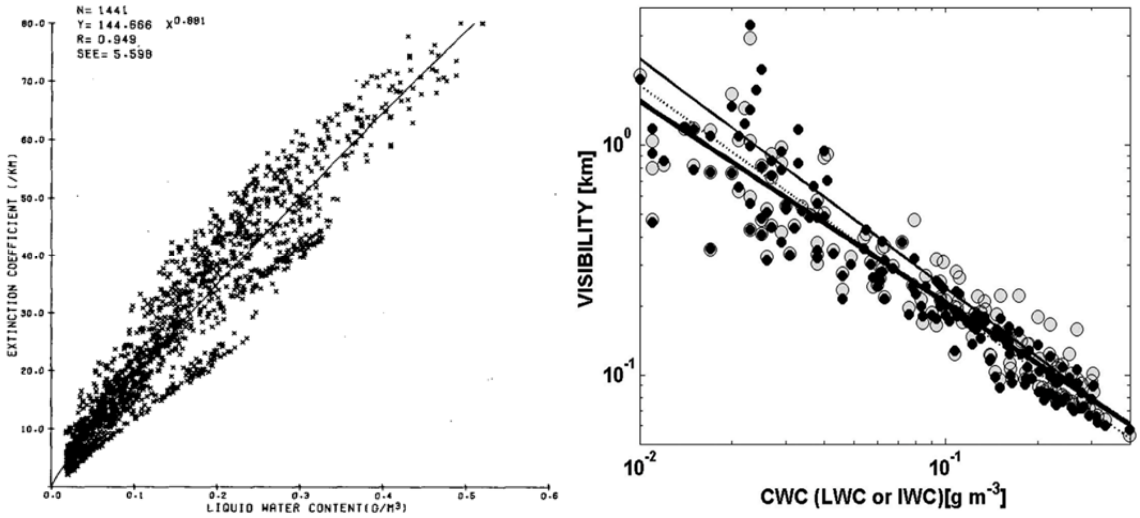

Figure 8. Scatter plots of field measurements from (left) Kunkel (1984) showing β_e vs q_c and (right) Gultepe et al. (2006) showing visibility vs q_c. The regression line shown in the left plot represents the Stoelinga and Warner (1999) visibility parameterization, and the thin dotted line in the right plot is the regression line expressing the Gultepe et al. (2006) visibility parameterization.

Kunkel (1984) and G06 both fit empirical relationships to datasets mainly limited to visibility <1 mi, leaving open the fit to smaller values of LWC and β_e. Examination of the plots in Figure 8 reveals neither dataset has any measurements when q_c is less than about 0.01 g m^{-3}, which corresponds to a daytime visibility of 0.7 mi and 0.9 mi using the SW99 and G06 visibility parameterization, respectively. This calls into the question the widespread use of the SW99 visibility parameterization during "light" fog conditions, loosely defined as fog producing visibilities in the range of 1–7 mi, which are of prime importance for DOD operations. Kunkel (1984) mentions this, noting that previous investigators (Tomasi and Tampierir 1976, Pinnick et al. 1978, and Eldridge 1971) obtained different (although not consistent) results in "observations of smaller droplets in lighter fogs". Still, the datasets in Figure 8 are used here due to various limitations in the older studies (e.g., instruments not able to measure all droplet size spectra), and in the case of the Kunkel (1984) data, its widespread use in modern NWP applications.

Uncertainty in the visibility parameterizations is represented by the spread of the data about the regression line in each scatter plot. To approximate this degree of spread, multiple points along the outer edges of the data envelope in each scatter plot, i.e., those furthest from the regression line, were transcribed to a new plot (Figure 9). Since the G06 data in Figure 8 are plotted as visibility, they are converted to β_e prior to being replotted in Figure 9 by dividing the constant -ln(0.02) by the visibility. This conversion is slightly different than the ASOS conversion given in equation (8), but is consistent with what G06 used to compute the visibilities plotted in Figure 8. A nighttime conversion is not needed, as all the G06 data was collected during daytime. The portion of the data taken in very heavy fog events with q_c >0.1 g, corresponding to daytime visibilities of <0.1 mi, is not included. The fits to the data are unphysical at greater q_c, where the lines eventually intersect.

The SW99 visibility parameterization is used to compute the mean value, $\overline{\beta_e}$, at any given q_c in the parametric visibility parameterization because it is based on a dataset that has more measurements in light fog conditions than the G06 data, and it is in widespread use. It is also used as the baseline comparison throughout this research. Both the SW99 (solid blue line) and G06 (solid black line) visibility parameterizations are

represented on the plot in Figure 9 and produce similar results in the q_c range shown. The two dashed lines define the approximate edges of the data envelope, with only a few points in either dataset falling outside this region. By definition, ~99.5% of the data should fall within three standard deviations, 3σ, of the mean value, $\overline{\beta_e}$, and so the dashed lines appear to offer a reasonable estimate for this range.

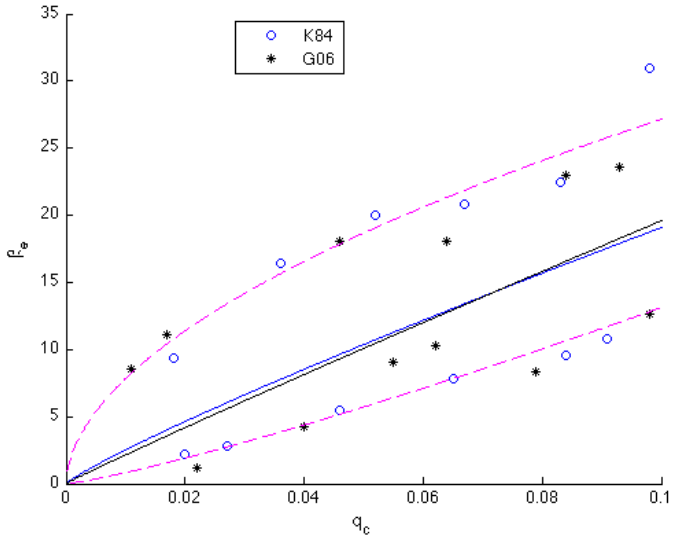

Figure 9. Plot of selected data from Kunkel (1984) and Gultepe et al. (2006). The two solid lines through the middle are regression lines for each data set, and represent the Stoelinga and Warner (1999) visibility parameterization (blue) and the Gultepe et al. (2006) visibility parameterization (black).

Examination of the Kunkel (1984) data in Figure 8 suggests the distribution of the data about the regression line at any given value of q_c is not Gaussian but heteroscedastic since it has a greater spread toward higher β_e values than it does toward lower values (the G06 data shows a similar pattern when q_c is plotted against β_e instead of visibility). This assertion is also apparent by the asymmetric shape of the data envelope about the regression lines in Figure 9. To more accurately account for the shape of this spread, the data are fitted to a log-normal distribution, where the shape of the spread of $ln(\beta_e)$ values is considered to be Gaussian about the value $ln(\overline{\beta_e})$ for any given q_c. The data in Figure 9 have been replotted in Figure 10 using $ln(\beta_e)$ as the y-axis. Symmetry of the 3σ lines about the regression lines representing $ln(\overline{\beta_e})$ supports the notion of using a log-normal

distribution for the data. Lines representing 2σ and 1σ above and below $ln(\overline{\beta_e})$ are also shown in Figure 10. The right panel in Figure 10 shows the same data using a linear y-axis once again. It is zoomed in to show only q_c values < 0.01 g m^{-3}, which corresponds to approximate daytime visibilities > 0.7 mi, and is the range of interest for this research, despite there being no observations in this range in either dataset.

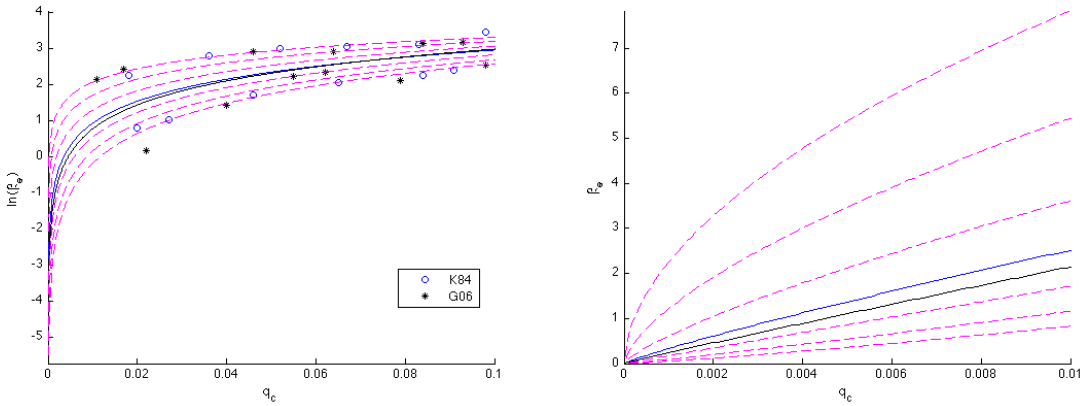

Figure 10. Left panel shows same data as in Figure 9, but plotted using $ln(\beta_e)$ as the y axis. The dashed lines represent one, two, and three standard deviations above and below the Stoelinga and Warner (1999) visibility parameterization (solid blue line). Right panel uses β_e as y-axis, and is zoomed in to show only the q_c range of interest.

The probability density function (PDF) of the log-normal distribution takes the form

$$prob\ density(\beta_e, \mu', \sigma') = \frac{1}{\beta_e \sigma' \sqrt{2\pi}} \exp\left(-\frac{(\ln \beta_e - \mu')^2}{2(\sigma')^2} \right), \tag{13}$$

where μ' and σ' are the mean and standard deviation, respectively, of $ln(\beta_e)$. The spread of the β_e probability density is greater for larger values of q_c as illustrated in Figure 11, showing the PDF of β_e for two values of q_c. The full PDF as a function of only β_e and q_c is given in Table 3, along with other key expressions used to formulate the parametric visibility parameterization. Recall that the expressions developed here only used data when q_c <0.1 g m^{-3}; the results are not valid at larger values of q_c (where σ eventually decreases and becomes unphysically negative). The precise shape of the PDF when q_c >0.1 g m^{-3} is not crucial for this research, and in that range its shape is held constant by

setting σ = 2.2 and only allowing $\overline{\beta_e}$ to change. Once an individual PDF is constructed for each member, an ensemble PDF for the entire suite of members is formed by adding together the individual PDFs, and normalizing by dividing by 10, the total number of members in the ensemble suite.

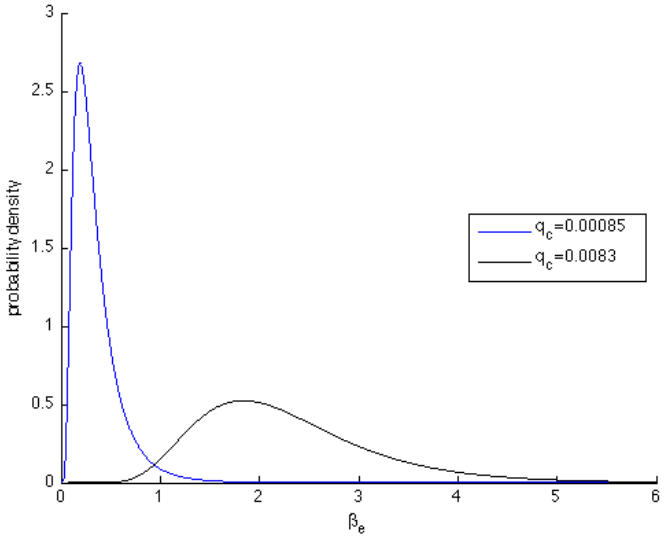

Figure 11. Parametric PDF of β_e values for q_c values of 0.00085 g m^{-3} (blue line), and 0.0083 g m^{-3} (black line). $\overline{\beta_e}$ for these PDFs is 0.29 km^{-1} and 2.1 km^{-1}, corresponding to approximate daytime visibilities of 6.5 mi and 0.875 mi, respectively.

An example of the result of this process is illustrated in Figure 12, which is from the ensemble prediction at KSCK at 29 January 2012 1800 UTC. In this forecast, five of the members have predicted non-zero q_c, and their corresponding PDFs of β_e are shown with solid blue lines. Four of the members predicted a very heavy fog event with β_e >15 km^{-3}, while one member predicted a lighter event. Five members predicted zero values of q_c and therefore have no PDF drawn. The resulting ensemble PDF from this forecast is shown with a dotted black line. The probability of exceedance for any given β_e threshold predicted by the ensemble is obtained by integrating the ensemble PDF for the desired interval. In Figure 12, the predicted probability for β_e >2.1 km^{-3} (corresponding to an approximate daytime visibility of 0.875 mi) is 0.4012, essentially because four of the ten members have their PDFs almost entirely above this threshold, while the member predicting lighter fog has only a small portion of its PDF above the threshold. As another

example, the predicted probability of β_e >0.29 km^{-3} (corresponding to an approximate daytime visibility of 6.5 mi) is 0.4929 because all five members have nearly their entire PDFs above this threshold.

Table 3. Summary of key expressions related to parametric visibility parameterization. Except for μ' and $\overline{\beta_e}$, expressions are valid only for q_c $\leq$0.1 g m^{-3}.

μ'	$\mu' = 0.88\ln(q_c) + 4.975$
σ'	$\sigma' = -0.11\ln(q_c) - 0.1437$
PDF(β_e, q_c)	$prob\ density = \frac{1}{-\beta_e\sqrt{2\pi}(0.11\ln(q_c)+0.1437)}\exp\left(-\frac{(\ln\beta_e - 0.88\ln q_c - 4.975)^2}{2(-0.11\ln q_c - 0.1437)^2}\right)$
$\overline{\beta_e}(q_c)$ (same as SW99 visibility parameterization)	$\beta_e = 144.7q_c^{0.88}$
$\overline{\beta_e}(q_c)$ + 1$\sigma(q_c)$	$\beta_e = 125.3q_c^{0.77}$
$\overline{\beta_e}(q_c)$ - 1$\sigma(q_c)$	$\beta_e = 163.4q_c^{0.99}$
$\overline{\beta_e}(q_c)$ + 2$\sigma(q_c)$	$\beta_e = 108.6q_c^{0.65}$
$\overline{\beta_e}(q_c)$ - 2$\sigma(q_c)$	$\beta_e = 184.2q^{1.10}$
$\overline{\beta_e}(q_c)$ + 3$\sigma(q_c)$	$\beta_e = 255.7q_c^{0.54}$
$\overline{\beta_e}(q_c)$ - 3$\sigma(q_c)$	$\beta_e = 207.9q^{1.20}$

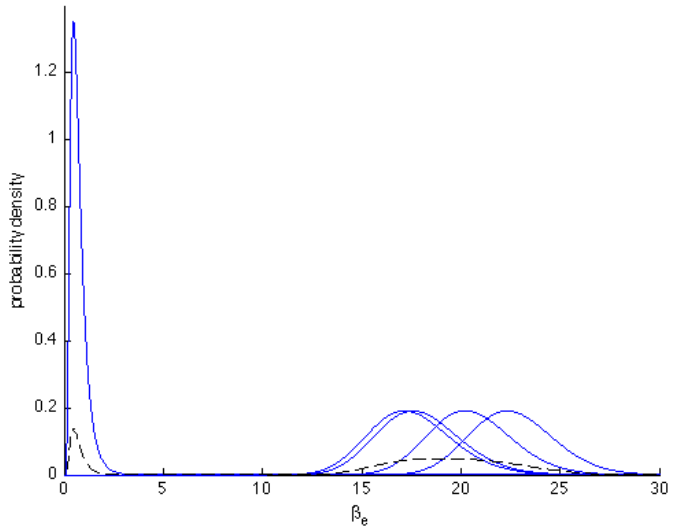

Figure 12. PDFs for ensemble prediction of β_e at KSCK for 29 January 2012 1800 UTC based on each members' q_c forecast. Five members predicted non-zero q_c, and their corresponding PDFs are plotted with solid blue lines. The ensemble PDF for the entire suite of members is plotted with a dashed black line.

2. Skill Scores

As a baseline performance metric, the Brier Skill Score (BSS) of the ensemble predictions is computed at four β_e thresholds corresponding to daytime visibilities of approximately[1] 6.5, 4.5, 2.75, and 0.875 mi. The BSS is obtained by comparing the Brier Score of the forecasts to the Brier score of a reference forecast, which for this research is persistence.

The persistence forecast is defined as the condition observed at the initialization time of the forecast preserved unchanged through the remainder of the forecast run. As noted previously, observations reporting an elevated β_e due to precipitation were removed from the dataset. However, when precipitation was occurring at the initialization time of an NWP run, it is necessary to categorize the observation as either above or below the β_e threshold of interest so the persistence forecast can be defined (even though the 00-h observation itself is still excluded from the results). In these cases, the persistence forecast was categorized as meeting the β_e criteria if the 00-h observation had a dewpoint

[1] These thresholds are approximate due to uncertainty in the relationship between β_e and visibility. The SW99 visibility parameterization is used to estimate the proper β_e thresholds for the visibilities.

depression <2.2 K (following the logic used by ASOS) *and* the observed β_e was above the threshold of interest. If either of these conditions were not met, the persistence forecast was categorized as not meeting the β_e criteria.

Following Wilks (1995), the Brier Score can be decomposed into reliability, resolution, and uncertainty, and these are also shown. A Ranked Probability Skill Score (RPSS), which is similar to BSS except it combines the performance at all four thresholds into a single metric, is also computed. Each of the relevant metrics is described in Table 4.

Except for RPSS, verifying metrics for all sites combined are provided in Figure 13. In order to assess the relative impact of NWP model error versus visibility parameterization error on the final predictions, two sets of results are shown on each plot: the results using just the deterministic SW99 visibility parameterization (solid blue lines), and the results using the parametric visibility parameterization (dashed black lines). The same metrics are provided separately for the coastal, valley, and mountain regions in Figures 14, 15, and 16, respectively. The RPSS for all regions combined and each individual region are shown in Figure 17.

As a broad summary of Figures 13–17, the NWP predictions show increasing skill with forecast hour compared to persistence, with the most skill in the mountain region and the least skill in the valley region. A close examination of these results follows in subsequent sections; for now, note that in nearly every plot in Figures 13–17, the results when the SW99 visibility parameterization was used are indistinguishable from when the parametric visibility parameterization was used.

The lack of visibility parameterization uncertainty at the four tested thresholds is evident in virtually every metric and region. The first-order error in β_e prediction from the ensemble is from the NWP predictions of q_c, and the conversion of q_c to β_e plays a negligible role. This does not mean visibility parameterization error is absent, only that it is unimportant given the magnitude and nature of the q_c predictions from the NWP model. The following section will examine the q_c prediction errors, and reveal why this is the case.

Table 4. Description of metrics used to assess stochastic predictions from the ensemble.

Metric	Formula	Description	Best Score	Worst Score
Reliability	$\frac{1}{M}\sum_{i=1}^{I} N_i((p_e')_i - \bar{o}_i)^2$	Measures how well a given forecast probability matches the observed frequency of occurrence	0	1
Resolution	$\frac{1}{M}\sum_{i=1}^{I} N_i(\bar{o}_i - \bar{o})^2$	Measures degree to which ensemble, through its probability forecasts, can parse data into subsamples having frequency of occurrence different from overall climatological frequency	Uncertainty score	0 (frequency of occurrence in every subsample = overall climatological frequency)
Uncertainty	$\bar{o}(1-\bar{o})$	Does not depend on forecast, only on climatological frequency; indicates level of difficulty in obtaining resolution	N/A – but scores may range from 0 (event occurs 0% or 100% of time, so no resolution possible) to 0.25 (event occurs 50% of time, maximizing potential resolution score)	
Brier Score	Reliability – Resolution + Uncertainty	Combines reliability and resolution to summarize overall ensemble accuracy	0	1
Brier Skill Score (relative to persistence)	$1 - \frac{Brier\ Score}{Brier\ Score_{persistence}}$	Measures overall stochastic skill of ensemble at particular threshold. Value of 0 indicates forecast is no better or worse than persistence forecast.	1	-∞
Ranked Probability Skill Score (relative to persistence)	$1 - \frac{\sum_{k=1}^{T} Brier\ Score_k}{\sum_{k=1}^{T}\left(Brier\ Score_{persistence}\right)_k}$	Combines multiple thresholds to indicate overall stochastic skill of ensemble. Value of 0 indicates forecast is no better or worse than persistence forecast.	1	-∞

M = number of forecast/observation pairs
I = number of probability bins (11)
N = number of data pairs in bin i
p_e' = center of forecast probability bin (0.025, 0.1, 0.2, 0, … 0.7, 0.8, 0.975) for bin i
$\bar{o}_i$ = observed relative frequency for bin i
$\bar{o}$ = climatological frequency (total occurrences / total forecasts)
T = number of event thresholds

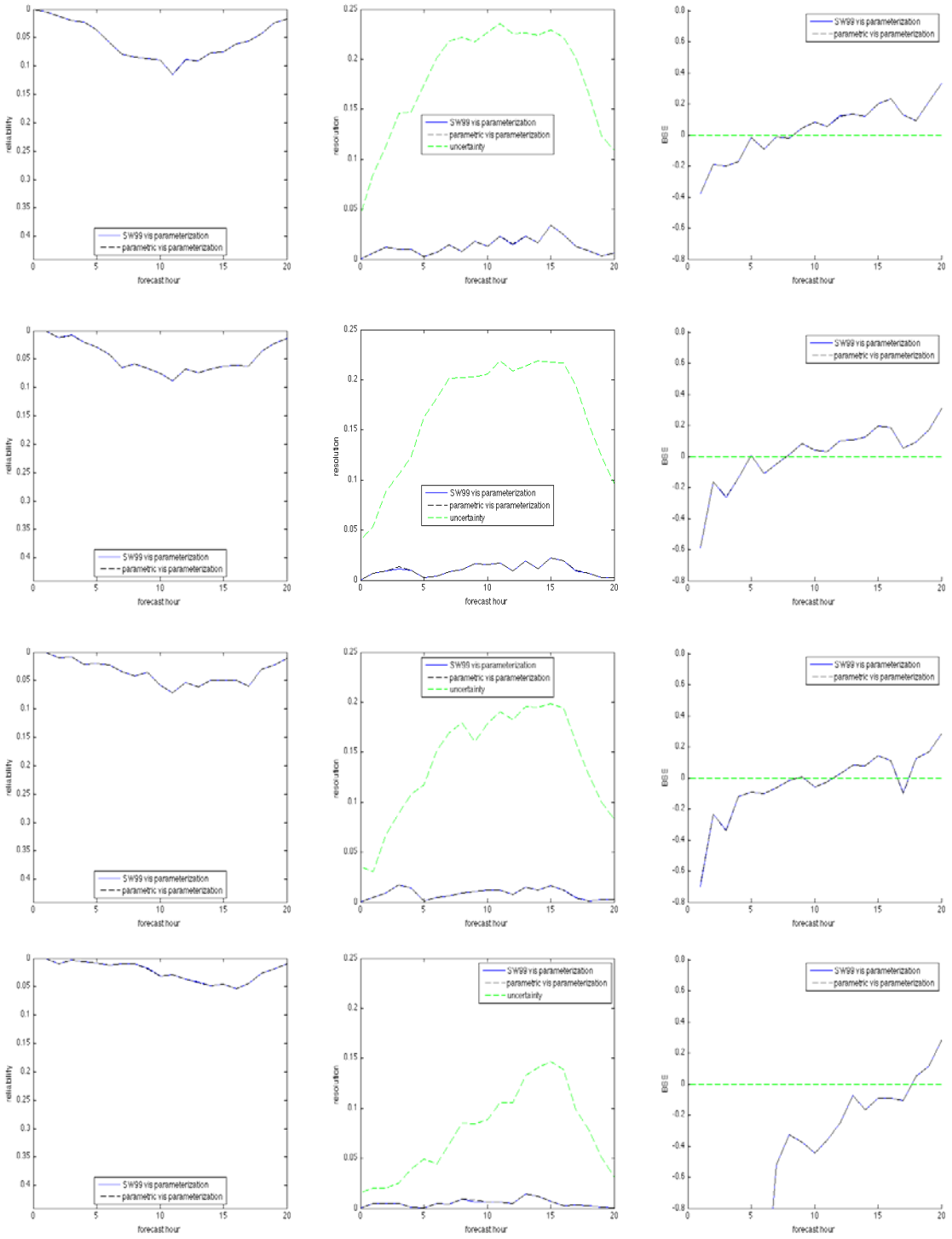

Figure 13. Ensemble reliability (left column), resolution (center column), and Brier Skill Score (right column) at four different β_e thresholds: 0.29 km^{-1} (top row), 0.41 km^{-1} (center row), 0.68 km^{-1} (third row), and 2.10 km^{-1} (bottom row). Forecast uncertainty is also shown on the resolution plots.

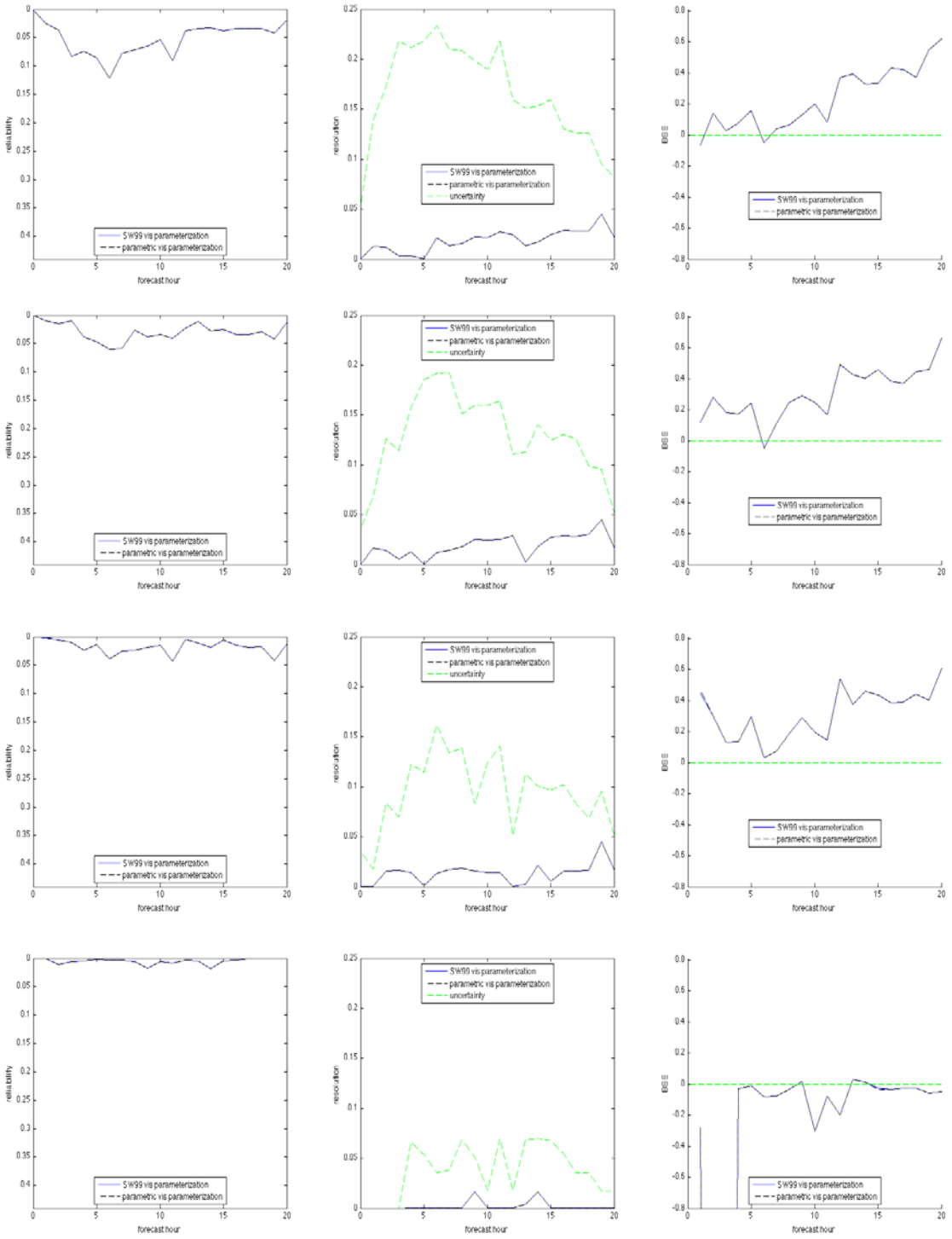

Figure 14. Same as in Figure 13, but only for the coastal sites.

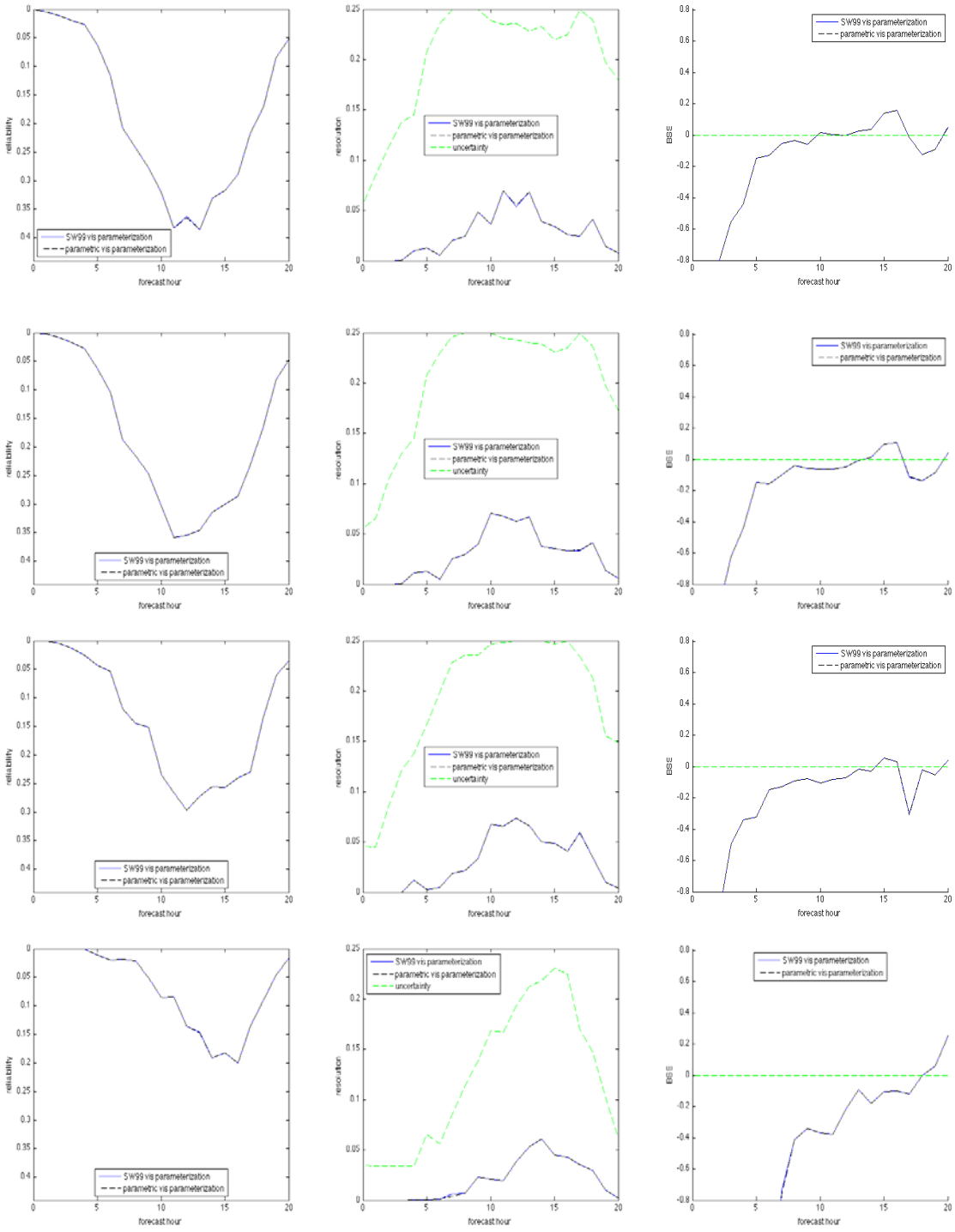

Figure 15. Same as in Figure 13, but only for the valley sites.

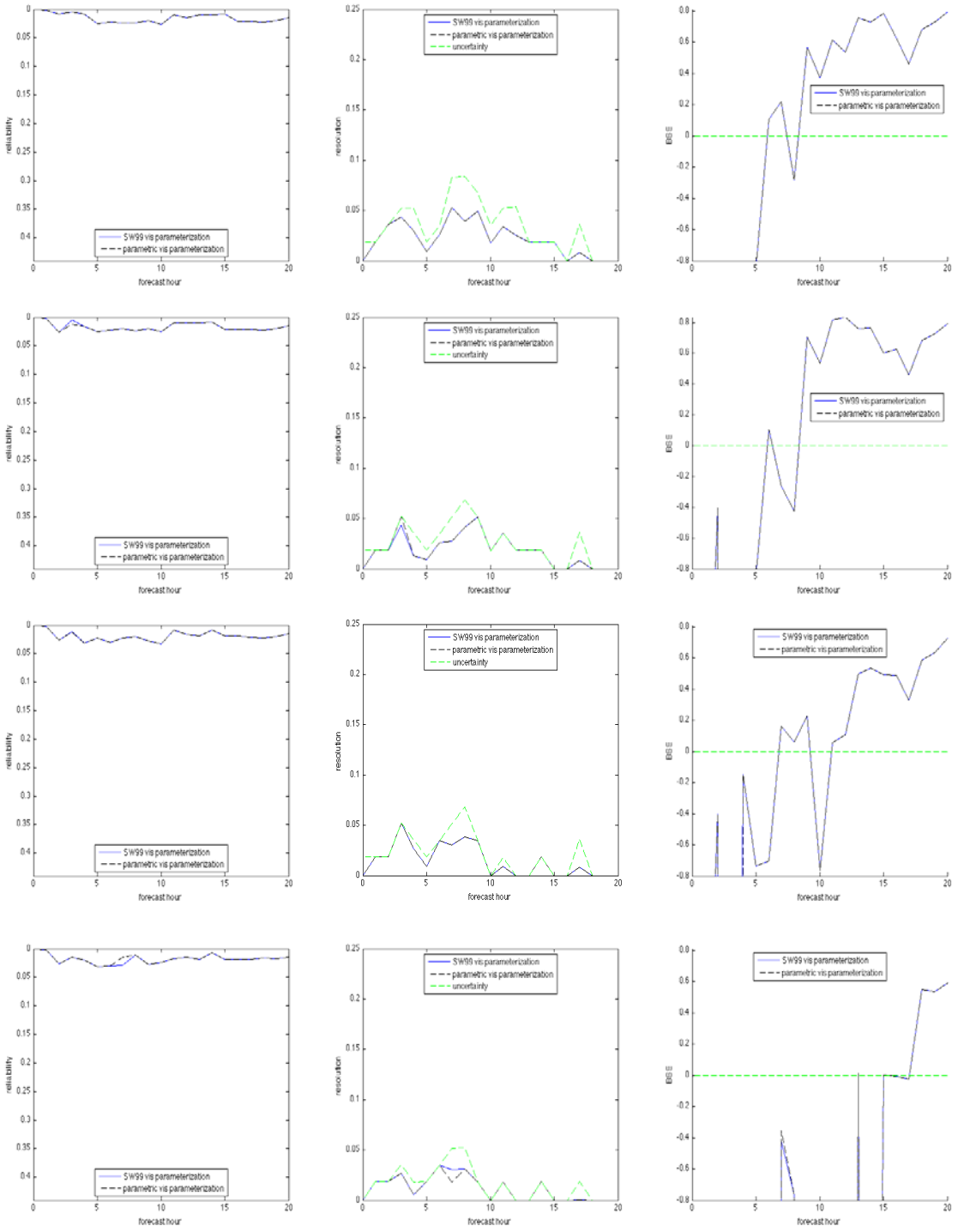

Figure 16. Same as in Figure 13, but only for the mountain sites.

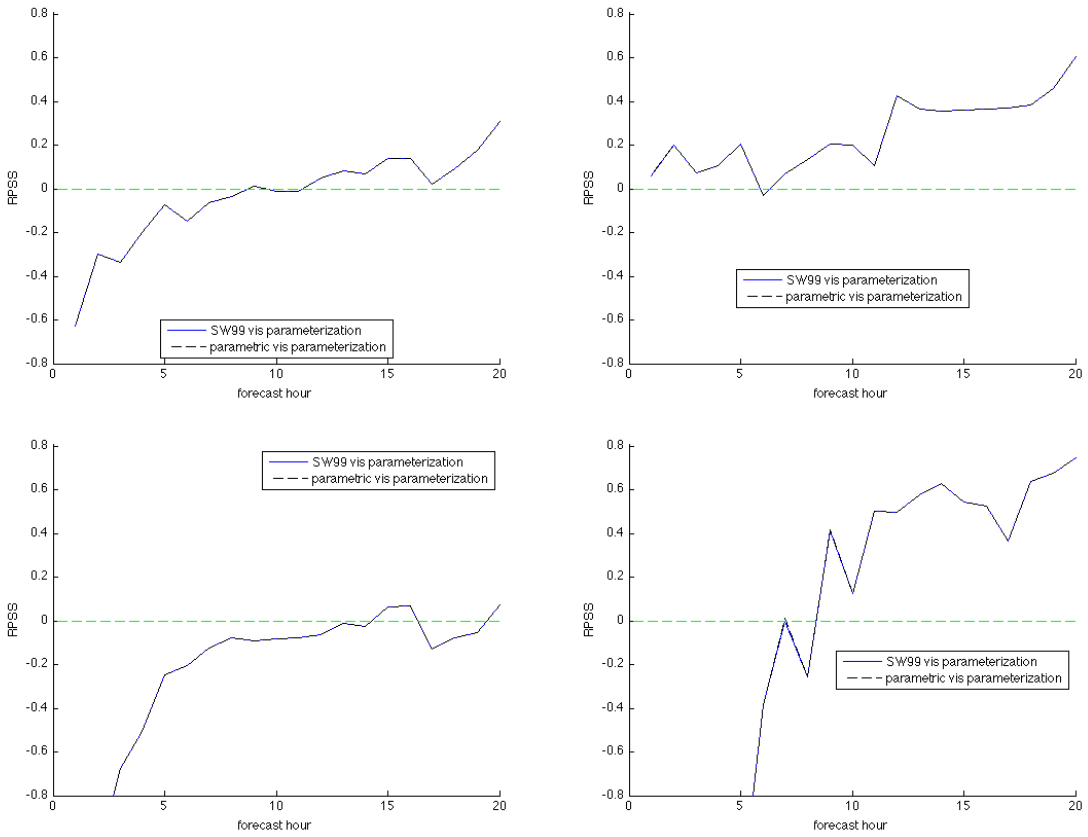

Figure 17. Ranked Probability Skill Score for all regions (top left), coastal region (top right), valley region (bottom left), and mountain region (bottom right).

3. Bimodal Nature of NWP Cloud Water Prediction Error

In this section, we will begin to examine the characteristics of the NWP q_c predictions error to better understand why it, and not visibility parameterization error, is dominant. The histograms in Figure 18 show the distribution of each members' q_c predictions (dark blue bars, top x-axis labels) for forecast hours 7-20 combined, overlaid with the distribution of observed β_e (light green bars, bottom x-axis labels). The bins for q_c and β_e are aligned based on their relationship via the SW99 visibility parameterization. For reference, the corresponding daytime visibility thresholds used in the BSS and RPSS calculations (6.5, 4.5, 2.75, and 0.875 mi) are indicated on the plot with vertical pink dotted lines. The first (leftmost) bin for q_c forecasts represents q_c values equal to zero, while the second bin represents non-zero values less than 8.5 x 10^{-4} g m^{-3}. These two bins are combined into a single bin for the observed β_e distribution because there are no

zero values of β_e. The first six hours of each case are excluded from the histograms to minimize the impact of NWP model spin up in the results.

The q_c predictions from each member show a bimodal signal, with values tending to indicate unrestricted visibility (bins 1 and 2, with q_c values $<8.5 \times 10^{-4}$ g m^{-3}) or heavy fog (bins 9 and 10, with q_c values $> 8.3 \times 10^{-3}$ g m^{-3}), with very few forecasts in the light fog range (bins 3 through 8). To a lesser extent, the observed β_e distribution is also grouped toward the outermost bins, but has a higher frequency of occurrence in the light fog range than do the predictions. For most of the members, the deficit in light fog predictions is coupled with a surplus in zero-q_c forecasts. The exceptions are members 16 and 17, whose forecasts of unrestricted visibility are split more evenly between zero q_c forecasts (bin 1) and very small, non-zero q_c forecasts (bin 2).

The behavior of these two members, which are the only members using the Ferrier microphysics scheme, is examined more closely by subdividing the q_c forecasts in bin 2 from Figure 18 into 12 sub-bins (Figure 19). This histogram shows that nearly all these q_c predictions are only slightly greater than zero, and are not near the threshold for light fog. Of the 772 q_c predictions from member 16 plotted in Figure 19, 767 of them have a q_c value $\leq 1.68 \times 10^{-9}$ g m^{-3}. Using the parametric visibility parameterization, the probability of these producing a β_e in the light fog range is $<2 \times 10^{-9}$. The results for member 17 are similar. Later, we will examine whether these small, non-zero q_c forecasts are a skillful indicator of fog if given a bias correction. For now, we may conclude that uncertainty in the visibility parameterization is insufficient to deduce a chance of light fog from these small, non-zero q_c predictions.

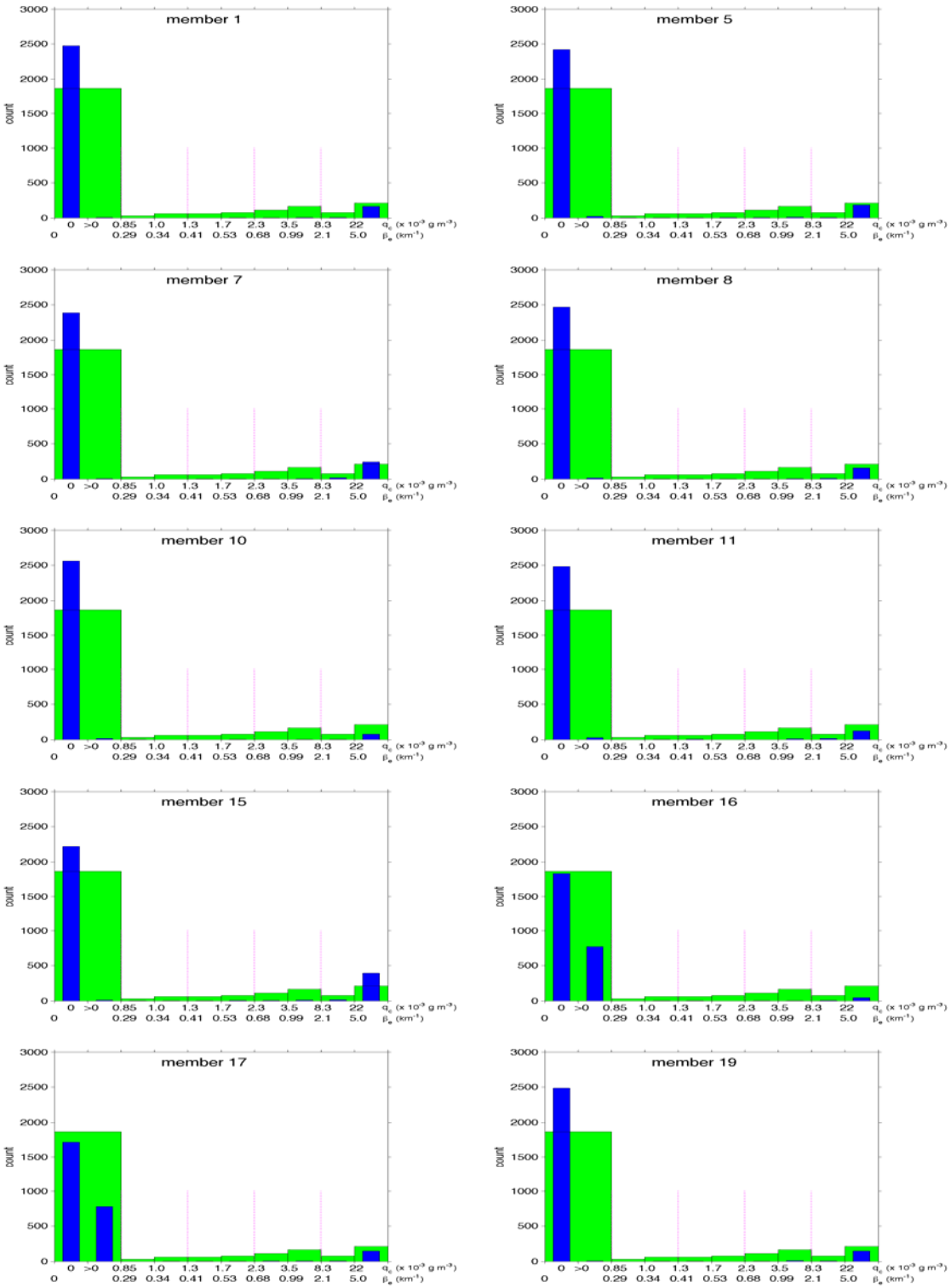

Figure 18. Histogram of distribution of NWP q_c predictions (blue bars), and β_e observations (green bars). Vertical pink dotted lines indicate approximate daytime visibility thresholds of 6.5, 4.5, 2.75, and 0.875 mi. The two leftmost q_c bins are combined into a single β_e bin. The first six hours of each case are excluded.

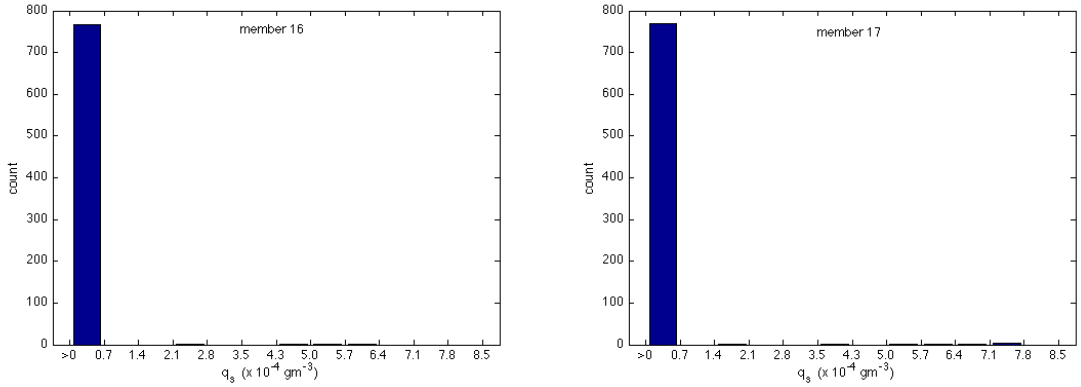

Figure 19. Distribution of q_c predictions from members 16 and 17, showing only the predictions from bin 2 in Figure 18.

Uncertainty in the visibility parameterization is also insufficient to deduce a chance of light fog from the vast majority of the heavy fog predictions, where most of the predictions reside in bin 10 in Figure 18. A q_c prediction of 0.022 g m^{-3} (the boundary between bins 9 and 10) has only a 0.0009 probability of translating to a β_e value in bin 8 or below, thus becoming a light fog prediction. Furthermore, most of the q_c predictions in bin 10 have values well above 0.022 g m^{-3}; the median q_c value for forecasts in bin 10 from member 1 is a full order of magnitude greater at 0.22 g m^{-3}, corresponding to a $\overline{\beta_e}$ of about 0.05 mi. The other members have similar median values.

To further illustrate this point, Figure 20 shows a scatter plot of observed β_e vs NWP-predicted β_e. All non-zero q_c predictions from all members and all sites are shown, with the first six hours of each case excluded. Each NWP prediction is plotted as a blue segment, which represents the range $\overline{\beta_e} \pm 3\sigma$ using the parametric visibility parameterization. The shaded pink interval indicates the range of β_e values corresponding to light fog conditions, or bins 3-8 in Figure 18. Observations of β_e that were reassigned a value of 0.10 km^{-1} during pre-processing (according to Figure 7) have been added to a small random number between -0.05 and 0.05 to prevent these cases from being plotted directly on top of each other, which conceals their incidence.

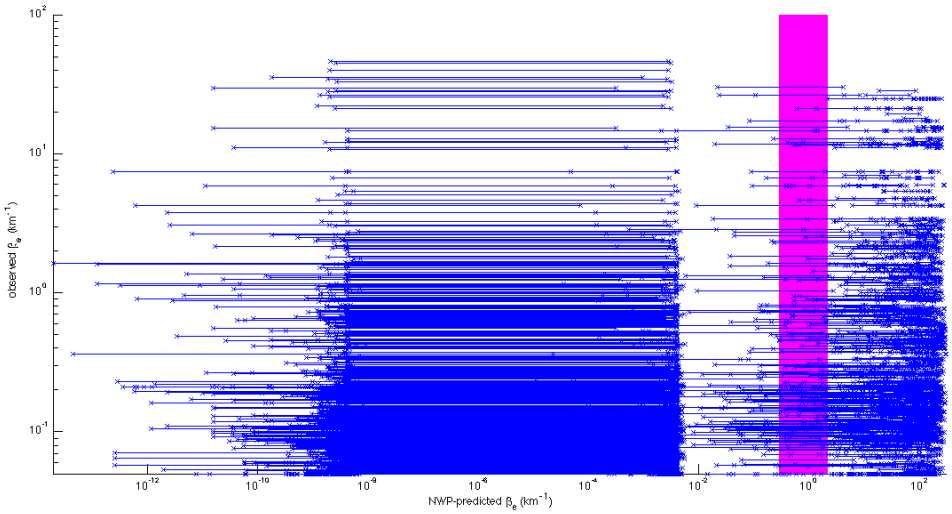

Figure 20. Scatter plot of observed versus NWP-predicted β_e for all members. Each prediction is plotted as a blue segment, which indicates the range $\overline{\beta_e} \pm 3\sigma$ from the parametric visibility parameterization. The pink box indicates the approximate range of β_e values corresponding to light fog.

Accounting for visibility parameterization uncertainty with the parametric visibility parameterization developed in this work has little effect on the BSSs at the β_e thresholds of interest because of the highly bimodal distribution of the q_c predictions from the NWP model. The bimodal nature of the data is evident in Figure 20. The abundance of small, non-zero q_c predictions (mainly from members 16 and 17) is shown to translate to very small β_e range mainly between 10^{-9} to 10^{-2} km^{-1} and below the threshold for light fog. Similarly, the large majority of heavy fog predictions have a plotted range entirely above the light fog threshold. Among all the observations, the climatological frequency of light fog is 0.196. Yet, if we include the zero-q_c predictions (which have $\sigma = 0$ and therefore a zero probability of translating to light fog), the incidence of all predictions having a plotted range that involves the light fog interval is only 0.013. If we limit the range to $\overline{\beta_e} \pm 1\sigma$ (not shown), which is essentially just the portion of the PDF with enough probability density to appreciably affect the final stochastic predictions, the incidence of all predictions involving the light fog interval is reduced to 0.006.

With this in mind, the remainder of this chapter will more closely scrutinize the q_c prediction error and other aspects of NWP model error contributing or related to the q_c error. This will allow a better understanding of the error, paving the way to develop strategies to mitigate it in Chapter V.

B. ANALYSIS OF NWP PREDICTION ERROR

1. Cloud Water

The bimodal nature of the NWP q_c predictions does not necessarily mean that, as an ensemble suite, they are unskillful at predicting the probability of exceedance at the thresholds of interest. It is a fundamental advantage of an ensemble, as opposed to a single deterministic NWP prediction, that skill is achievable if the relative number of members above and below the threshold can change with some degree of correlation to the verifying observation, even if every member has a poor prediction individually. Reexamining Figures 13–17, we see this to be the case in certain situations. The RPSS results show that for all sites collectively, skill gradually increases with forecast hour, outperforming persistence (i.e., RPSS >0) beyond 9 h. The inability to beat persistence early in the runs is consistent with the performance characteristics of many fog prediction frameworks, including NCV (Herzegh et al. 2006), and is not surprising for a model-only framework that must undergo spin up of its uninitialized q_c field. Note that the skill of persistence has a diurnal trend (not shown) that starts as a perfect forecast (0 h), decreases overnight (2~15 h) as the incidence of fog increases, then improves after sunrise near the end of the runs (16~20 h) as the incidence of fog decreases. The improving skill of the NWP predictions during the overnight hours is therefore assisted by the accompanying drop in skill of persistence, with mixed results after sunrise that are examined more closely in subsequent sections. The following two sub-sections will individually examine the resolution and reliability of the NWP q_c forecasts.

a. Resolution

Bearing in mind that RPSS and BSS are affected by the accuracy of the NWP predictions *and* the accuracy of the persistence forecasts, it is useful to isolate just the performance of the NWP predictions to better understand how the NWP model

performs. In particular, the resolution term of the BSS indicates the degree to which the ensemble distinguishes cases when the threshold is met (an event) from cases when it is not (a non-event), without regard to the accuracy of the predicted probability of occurrence. For example, if an ensemble made up of 10 highly bimodal members consistently has four members above the verifying threshold during non-events, and five members above the threshold during events, it would have a high resolution despite the fact that the predicted probabilities (0.4 and 0.5, respectively) are not particularly accurate. The ability to obtain resolution depends on the observed climatological frequency of occurrence, with resolution most easily obtainable when the event occurs half the time, and becoming progressively more difficult to obtain as the climatological frequency approaches 0 or 1. This ease with which resolution may be obtained is termed the forecast *uncertainty*, which quantitatively is the maximum possible resolution. So it is the difference between the uncertainty and the resolution that provides the best overall indication of the ensemble's ability to distinguish events from non-events (with smaller differences indicating more ability).

Examining the cases for all sites (Figure 13), the first few forecast hours are marked by a rapid increase in uncertainty caused bythe increasing incidence of observed fog with the loss of daytime heating. (Forecast hour 0 corresponds to 1600 LT, with each run ending at 1200 LT the following day). This increase is not met with a corresponding increase in resolution until about 6 h, after which point the resolution slowly increases throughout the overnight hours. After 15 h, the resolution decreases, but this coincides with a rapid decrease in the uncertainty (associated with a decrease in fog incidence due to daytime heating) such that the difference between uncertainty and resolution actually continues to decrease. Specifically, the ensemble does the poorest job of distinguishing events from non-events near midnight, then shows a consistently increasing ability to do so throughout the early morning, dawn, and late morning hours.

This upward trend is an encouraging sign for using the ensemble as the underpinning of a fog prediction framework for the traditionally challenging period during and after sunrise, but the difference between uncertainty and resolution remains quite large at all hours with room for potential improvement using a post-processing

technique. At a minimum, it should be a fundamental goal with the addition of any post-processing technique to not inadvertently destroy forecast resolution that already exists.

b. Reliability

For a strictly statistical calibration that entails a bias correction to the final predicted probabilities from the ensemble, the resolution will not change, and so the amount of resolution initially present is of prime importance for the success of the final calibrated product. Ensemble reliability, which indicates the conditional bias of the probability predictions (i.e., conditioned on the predicted probability bin), is of less consequence in this case aside from simply informing the bias correction to be applied.

For our purpose of pursuing an adaptable, worldwide-transferable VIF prediction framework rather than a location-specific calibration, the reliability is of prime importance since we cannot simply maximize it with a statistical correction. Instead, our approach to addressing conditional biases must be to first understand why they exist and whether they are likely due to a systematic deficiency in the NWP model. Examining the reliability for all sites shows near-perfect reliability at initialization, which is attributed to the 0.0471 observed frequency of fog at this late afternoon hour closely matching the predicted probability from the ensemble, which is 0 in every case due to the lack of q_c initialization. As the incidence of fog increases during the afternoon and evening hours (evident by the increasing uncertainty), reliability worsens. For the verification at the lowest β_e threshold (top row in Figure 13), the worsening reliability continues until 11 h, which corresponds to the period of highest fog incidence (0.3802). After this period, the reliability improves while the incidence of fog decreases. The reliability changes and changes in fog incidence appear to be highly correlated in the verification at all β_e thresholds.

The reliability results suggest the ensemble probabilistic forecasts have a negative q_c bias throughout the runs. To conceptually illustrate this point, consider the extreme example of an ensemble that always predicts 0 probability of an event occurring. The ensemble will be quite reliable when the true incidence of occurrence is low, but becomes less reliable as the incidence of occurrence increases. Without precise

observations of q_c against which we can verify, it is difficult to exactly quantify such a bias, but we can deduce from the distributions in Figure 18 that a negative bias exists in every member for all post-spin up forecast hours collectively. To confirm this bias at individual forecast hours, the next sub-section presents a deterministic verification on each member at the lowest of the four β_e thresholds used in the stochastic verification (β_e = 0.29 km^{-1}, or a daytime visibility of 6.5 mi).

c. Deterministic Member Verification

As before, the q_c predictions from each member were converted to β_e using the SW99 visibility parameterization. The metrics used in the deterministic verification are summarized in Table 5 (some of these metrics are presented elsewhere, but their descriptions are included in the table for convenience). Results for all sites are shown in Figure 21.

At this relatively low threshold, the small q_c bias ratios are present in all members at nearly all hours. The negative q_c bias also manifests in the probabilities of detection, which generally remain below 0.2 for most members. The bias ratios are predictably small early in the runs, then show very slight improvement with forecast hour. We know that the observed incidence of fog is increasing between 0–11 h, so the steady or slightly improving biases during this interval indicates the members are actively producing fog in the runs. Pre-sunrise forecast hours 10–15 are characterized by a high incidence of observed fog (between 0.33 and 0.39 – not shown), yet the bias ratios continue to improve while the false alarm ratios and probabilities of detection also improve. This matches well with the period of increasing ensemble resolution (Figure 13), and reinforces the fact that the ensemble is able to distinguish fog events from non-events to some extent despite the significant negative q_c bias of all its members at this threshold. The final few hours of the runs are characterized by more erratic results associated with daytime heating and a lower incidence of observed fog, although nine of the 10 members still have a bias ratio <1. Eight of the members maintain a bias ratio <0.5 at all forecast hours. The persistent negative q_c bias is also evident in the probabilities of detection, which generally remain below 0.2 for most members.

Table 5. Description of metrics used to assess deterministic predictions from each ensemble member. A "yes" forecast or observation means it is above the verification threshold. False positive rate is included in the table but is not used until later figures.

Metric	Formula	Description	Best Score	Worst Score
Bias Ratio	$\frac{(total\ "yes"\ forecasts)}{(total\ "yes"\ observations)}$	Reveals whether predictions, on average, are too ambitious or too conservative in forecasting event.	1	Overforecast: $+\infty$ Underforecast: 0
False Alarm Ratio	$\frac{(incorrect\ "yes"\ forecasts)}{(total\ "yes"\ observations)}$	Answers question "when event is forecast, at what rate does is occur?"	0	1
Probability of Detection (each member)	$\frac{(correct\ "yes"\ forecasts)}{(total\ "yes"\ observations)}$	Answers question "when event occurs, at what rate was it forecast?"	1	0
False Positive Rate (also called False Alarm Rate)	$\frac{(incorrect\ "yes"\ forecasts)}{(total\ "no"\ observations)}$	Answers question "when event does not occur, at what rate was it incorrectly forecast to occur?"	0	1

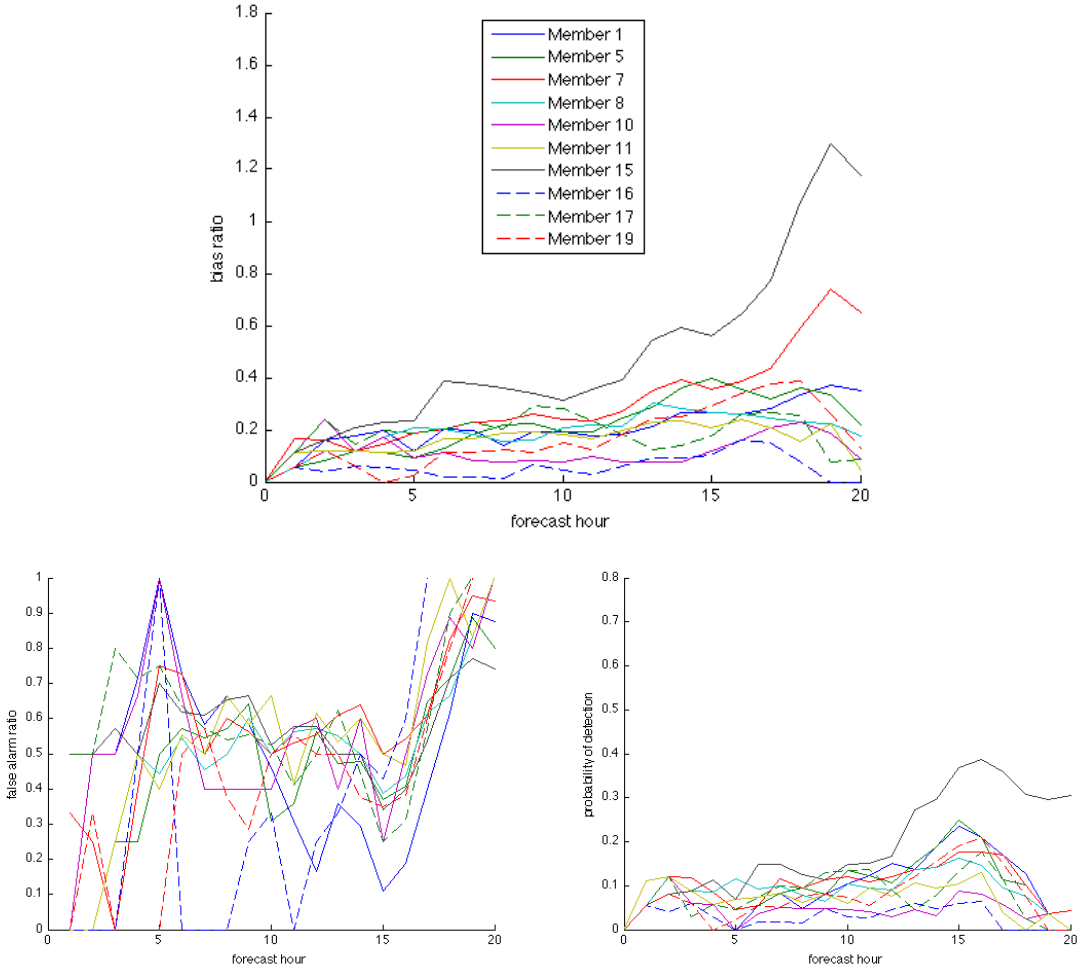

Figure 21. Results from deterministic verification of each ensemble member in all regions using a verification threshold of $\beta_e = 0.29$ g m^{-3}: (top) bias ratio, (bottom left) false alarm ratio, (bottom right) probability of detection.

d. Regional Results

Until now, we have only examined the observations and predictions of all sites collectively, but the data from individual regions is useful because different regions have different physical processes controlling visibility (e.g., radiation fog in the valley region, radiation and advection fog in the coastal region, etc.). A better understanding of the regional results also helps formulate potential approaches to improve the forecasts in later chapters. Figures 22–24 show the post-spinup distribution of the NWP model q_c predictions and β_e observations for the coastal, valley, and mountain regions,

respectively. The bimodal distribution of the q_c predictions is evident in each region, though it is not as pronounced in the coastal region, which is distinguished by the small number of fog predictions of any severity by any member. Despite an obvious negative q_c bias in the coastal region, the weaker bimodality in the prediction distribution compared to the other regions accurately reflects the unique observations distribution, which is not bimodal.

The distributions in the valley region are similar to the overall data, with the bimodal predictions displaying a surplus of no-fog forecasts (bins 1 and 2), and mostly lacking predictions in the light fog range. Unlike the other regions, light fog is common in this region, occurring in 32% of all observations. The frequency of predictions of the heaviest fog events in bin 10 generally matches the observed frequency of these events.

The mountain region is characterized by only 27 observed fog events, and a frequency of no-fog predictions that generally agrees with the observed frequency of no fog. The predictions are also bimodal, with virtually all predictions for fog in the rightmost bin.

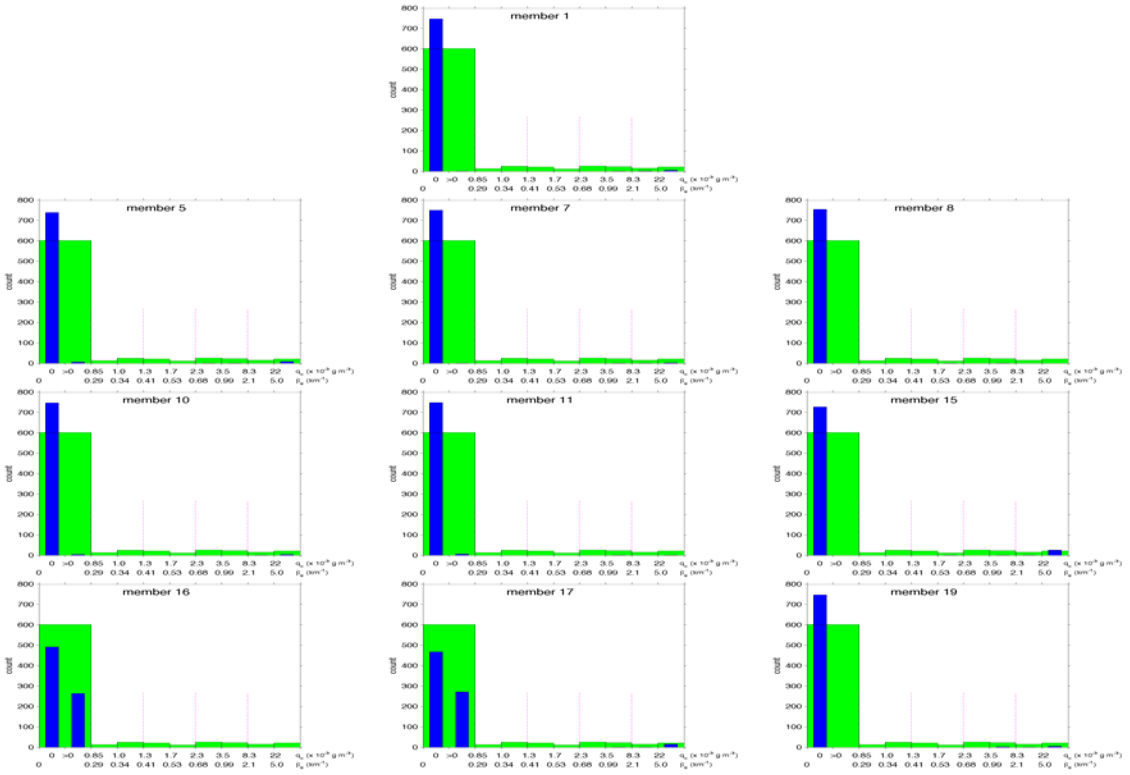

Figure 22. Same as in Figure 18, but only for the coastal sites.

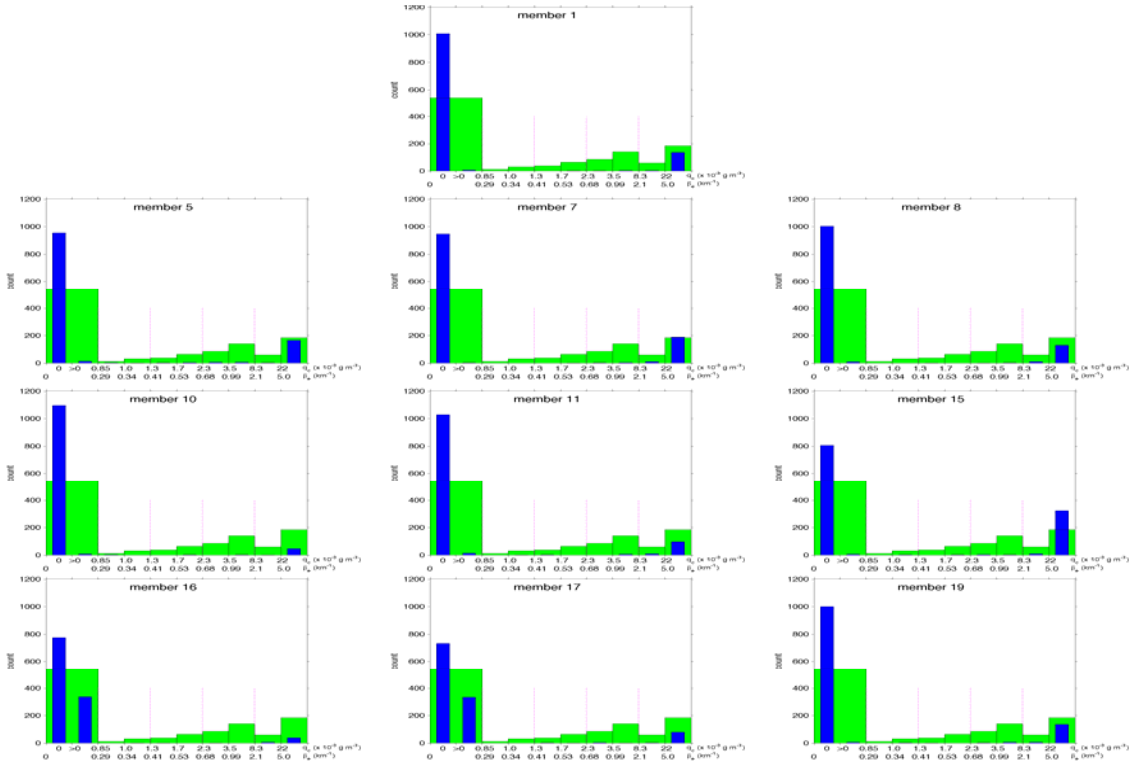

Figure 23. Same as in Figure 18, but only for the valley sites.

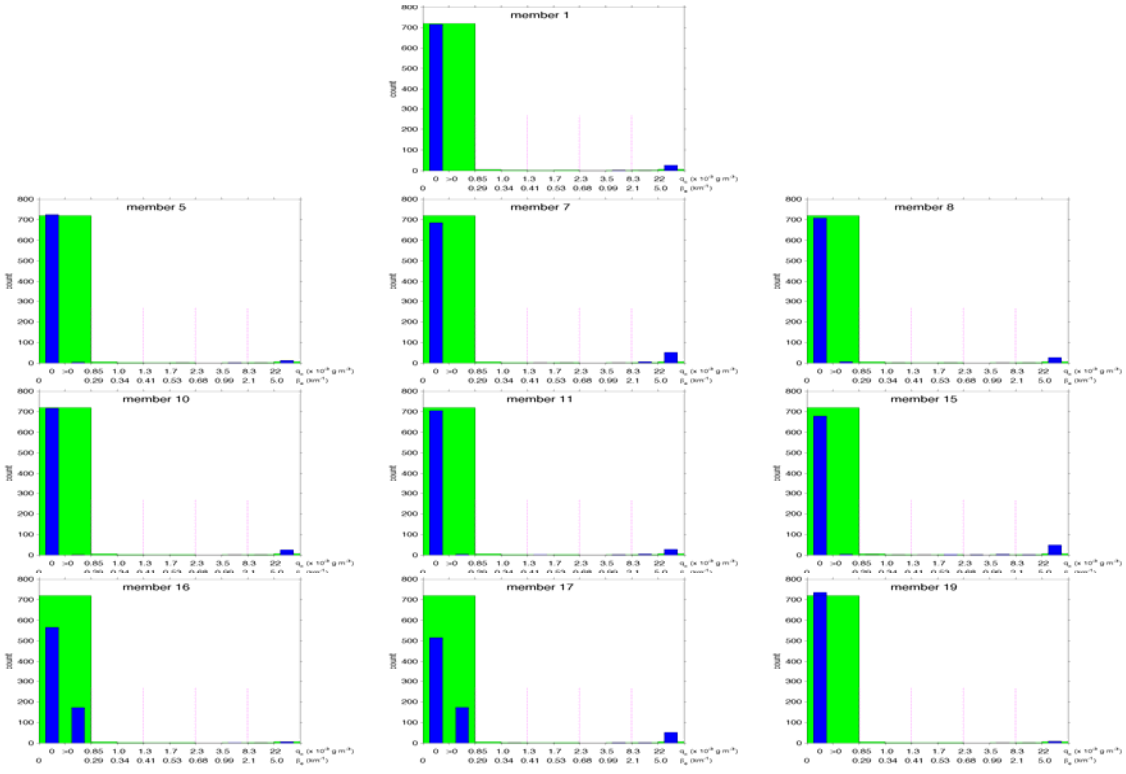

Figure 24. Same as in Figure 18, but only for the valley sites.

As with the full dataset, a deterministic verification of each member at the lowest of the four β_e thresholds was performed for each region, and these results are displayed in Figure 25. Predictions in the coastal region show small bias ratios by all members at all hours at this threshold, a trait also reflected in the reliability at this threshold (Figure 14), which appears well-correlated to the uncertainty during the first 15 h of the runs. The bias ratios are the lowest of any region, but the ensemble still displays consistent resolution and positive skill after the spin up period. While the reliability and resolution remain fairly steady during daytime heating, the uncertainty decreases from 16-20 h, causing the BSS to increase to 0.6 by 20 h. During these hours, there are no false alarms (the false alarm ratio is quite erratic at earlier hours due to the small number of predicted events) by any member, and only members 5 and 15 have any fog predictions at all as evidenced by their non-zero probabilities of detection (POD). This illustrates how the influence of just a few members can impact resolution and ensemble skill if they can occasionally distinguish an event, regardless of overall ensemble bias.

In the valley region from 0–17 h, all members have slowly increasing bias ratios that generally do not exceed 0.5. The low bias ratios combined with the high incidence of observed fog results in the poorest ensemble reliability of all the regions. Ensemble resolution is slightly higher than in the other regions during the overnight hours, but is relatively small in relation to the uncertainty. As in the other regions, the BSS at this threshold gradually increases overnight as the resolution increases, but here it only briefly exceeds 0 from 13–16 h and the ensemble is otherwise outperformed by persistence. A skill decrease after sunrise matches corresponding decreases in probabilities of detection while false alarm ratios increase to >0.8 for most members. Unlike in the coastal region, the negative q_c bias and modest resolution is not enough to provide sustained skill in a region where the observed frequency of fog is much higher.

The low incidence of observed fog events in the mountain region makes the deterministic verification data at any single hour rather volatile. Bias ratios are higher than in the other regions, with single-member averages from 0.3 (member 16) to 2.3 (member 7) across all post-spin up hours. The average bias ratio from all members at all post-spinup hours is 1.3, indicating a slightly positive q_c bias at this threshold. The ensemble is shown to have resolution nearly equal to uncertainty for most hours, indicating events are distinguished by the predictions far better than they are in other regions. The BSS at this threshold shows mostly increasing skill from 5–8 h, followed by a score between 0.4 and 0.8 for the remainder of the runs.

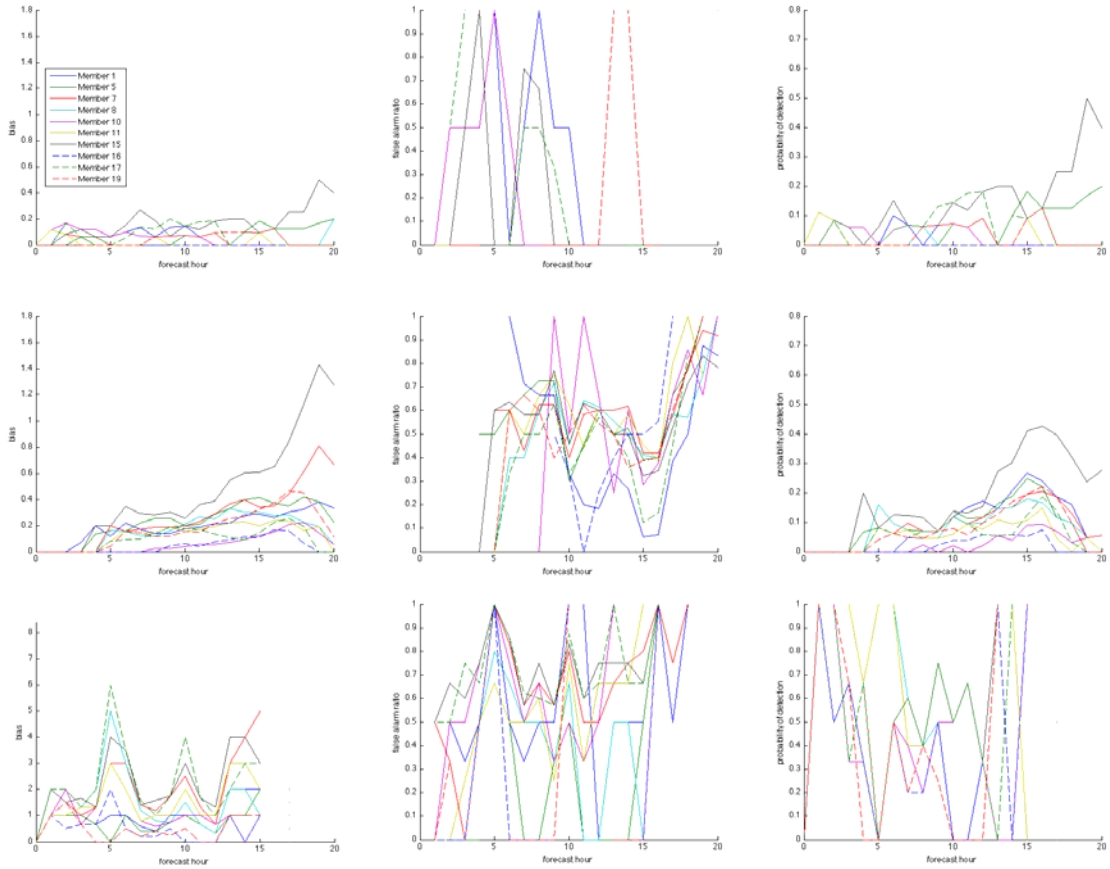

Figure 25. Results of deterministic verification at β_e threshold of 0.29 g m^{-3} in the coastal region (top row), valley region (center row), and mountain region (bottom row). Metrics shown are bias ratio (left column), false alarm ratio (center column), and probability of detection (right column).

The first few hours after sunrise are traditionally a period of difficulty for radiation fog forecasting. This period is often characterized by fog dissipation, with the rate of dissipation dependent on the depth and heating rate of the fog layer, as well as changes in the turbulent vertical moisture flux. Predictions in the valley region in particular exhibit indications of these challenges with a sudden decline in RPSS and BSS at most thresholds shortly after sunrise. To more closely examine how well the NWP predictions handle radiation fog dissipation during this period in the valley region, instances when the members correctly predicted fog at 14 h (1–2 h prior to sunrise) were tracked through the dissipation process over the subsequent 6 h (Figure 26). The lowest of the four β_e thresholds was used as the fog/no fog delineator. Of the 53 cases of observed fog at 14 h, each member correctly verified between 2 (member 10) and 16

(member 15) of them at that hour, so each plot represents only a small fraction of the total observed fog cases. Tracking this specific subset of data in this way eliminates the impact of fog that forms after sunrise, which is uncommon but does occur both in the observations and predictions, and is arguably not radiation fog. As the cases are tracked forward in time, the number that maintained observed fog at each hour is plotted with a black line, and the number that maintained fog in the predictions is represented by the shaded area, which is divided into correct fog predictions (i.e., "hits", indicated with red shading), and false alarms (blue shading). Note that the number of hits cannot exceed the number of observed cases, so any predictions of fog above the black line must necessarily be false alarms. Conversely, it is possible to have a false alarm area below the black observations line if the member prematurely dissipates some cases yet incorrectly prolongs others.

The plots show that, on occasions when fog is correctly present in the NWP model prior to sunrise, the dissipation biases vary by member. Three of the members (1, 7, and 15) tend to dissipate the fog cases too slowly, creating an abundance of false alarms by 20 h. In contrast, two members (16 and 17) are shown to dissipate their cases rather quickly after sunrise, with the remaining five members showing little bias in dissipation rate for this subset of the data.

With so few cases, it is impossible to draw definitive conclusions about any systematic NWP deficiency regarding the post-sunrise dissipation rate. These limited results do not suggest a clear systematic error exists. Bias ratios in Figure 25 show mixed trends after sunrise in this region depending on the member. The increasing false alarm rates and decreasing probabilities of detection during the post-sunrise hours are mostly due to volatility from a small and declining sample size. The occasional cases of observed and/or predicted fog formation during the period generally do not verify well but do appreciably affect the metrics due to the small sample size. The post-sunrise declines in RPSS and BSS are further affected by an increasingly accurate persistence forecast (which is for no fog in 94% of the cases in this region) as the number of fog cases declines.

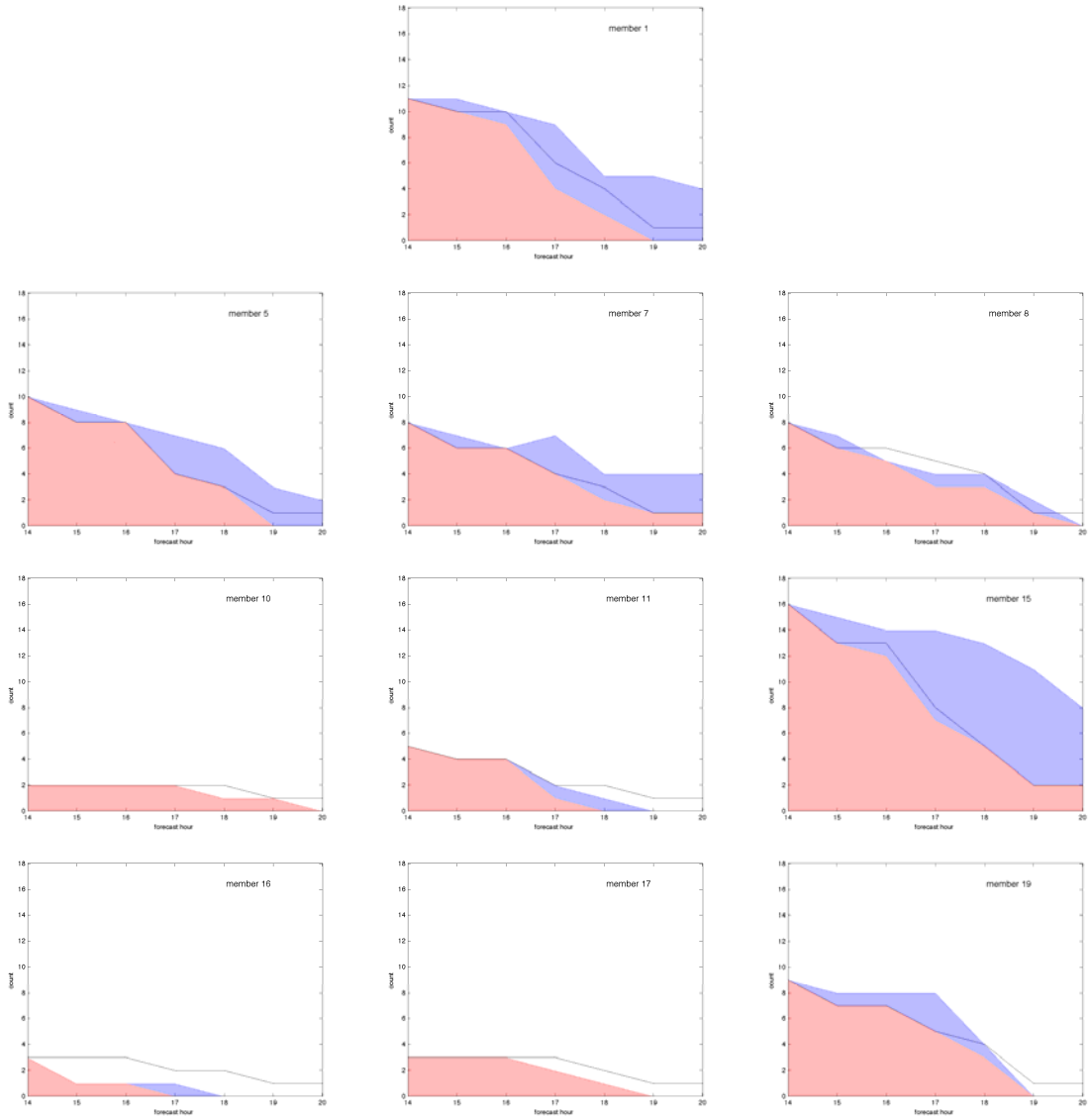

Figure 26. Observed cases of fog (black line) and predicted cases of fog (total shaded area) for each member in the valley region. The plots only include cases when the model correctly predicted fog at 14 h. The shaded region is divided into hits (red) and false alarms (blue).

The more obvious systematic deficiency remains the negative q_c bias in this region, typified by the fact that 33 of the 53 observed fog cases at 14 h were not predicted by *any* member. These results partially agree with those of Bang (2006), whose WRF runs tended to underforecast radiation fog, but also dissipate it too rapidly in a heavy fog case study at Incheon, South Korea. Here, post-sunrise dissipation rates are

inconclusive and are not shown for the coastal or mountain regions due to the limited number of fog predictions in the latter (fewer even than the valley region), and the limited number of observed cases in the former.

For much of the verification of NWP cloud water predictions discussed thus far, we have focused on the lowest of the four β_e thresholds, approximately corresponding to the important delineator between unrestricted visibility and light fog. But with the bimodal nature of the NWP predictions (92.66% of predictions above the lowest β_e threshold are also above the highest of the four β_e thresholds), it is fitting to also examine their relative ability to predict just the heavy fog events corresponding to a daytime visibility < 0.875 mi. The BSSs at this highest β_e threshold (Figure 16) are generally lower than at other thresholds, but are also subject to volatility given the fewer number of heavy fog cases. To provide context to the skill scores, Figure 27 compares the false alarm ratios and PODs at the lowest and highest β_e thresholds for each member. The data from all post-spin up hours has been combined for the plots.

The skill apparent in predicting the lowest β_e threshold (corresponding to daytime visibility < 6.5 mi) is lacking in predictions of the highest β_e threshold (corresponding to daytime visibility < 0.875 mi). At the lowest β_e threshold, we saw that predictions in the coastal region had the largest negative q_c bias of any region, but maintained sufficient resolution to produce skillful forecasts after 7 h. The same is not true for verification at the highest β_e threshold, which shows the predictions are unskillful at most hours due to virtually no resolution. Of the eight members that predicted heavy fog at least once, all have a false alarm ratio >0.88. Of 36 total instances of observed heavy fog in this region, only two members verified any of them, accounting for only 4 total hits.

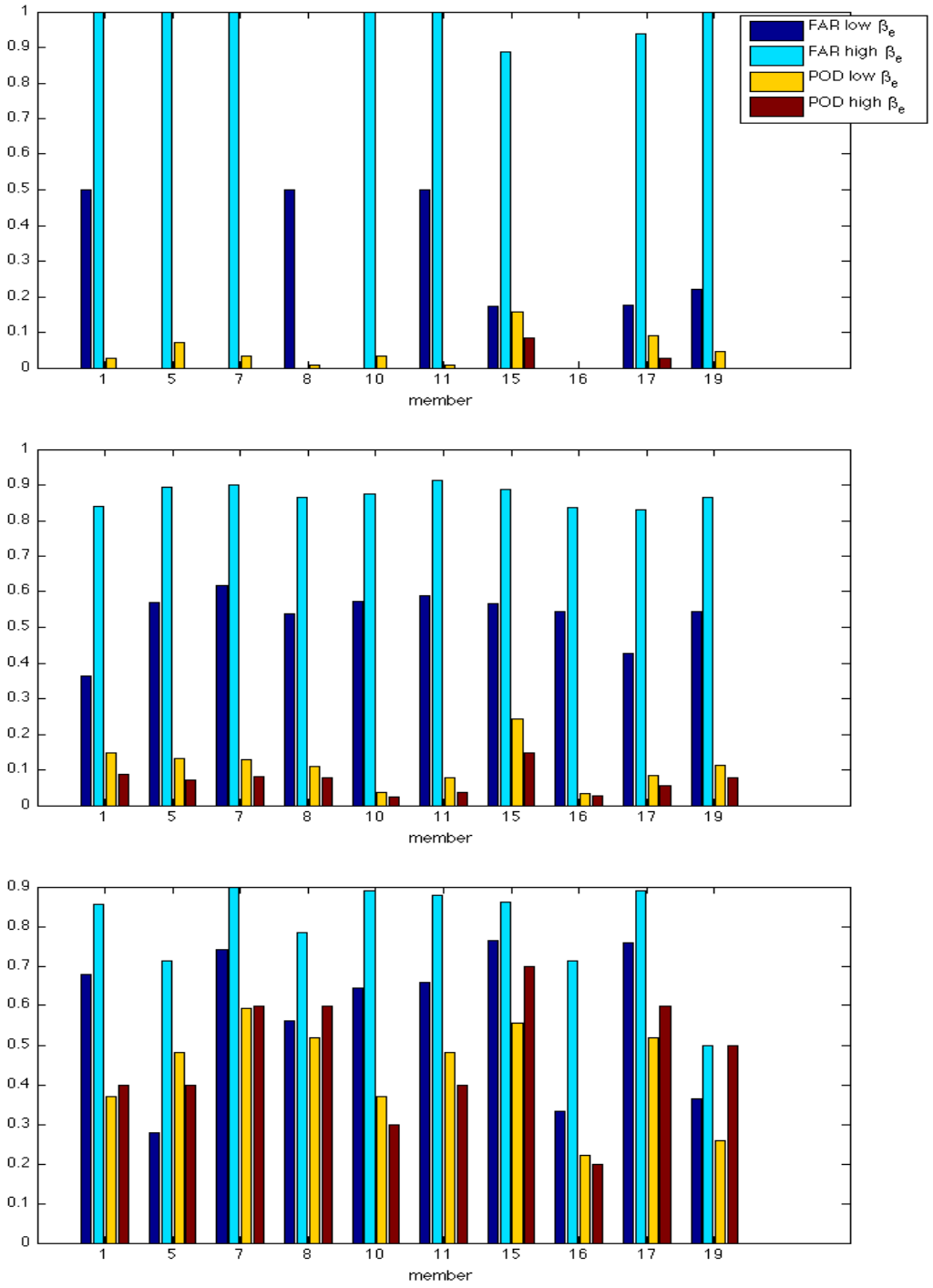

Figure 27. Comparison of false alarm ratio and probability of detection at the low β_e threshold (0.29 km^{-1}) and high β_e threshold (2.1 km^{-1}) for the coastal (top), valley (center), and mountain (bottom) regions. The data includes forecast hours 7-20.

Predictions in the valley region have a BSS <0 at most hours at the higher β_e threshold partly due to significantly higher false alarm ratios and lower PODs. Despite a decreasing resolution, the BSS shows an abrupt increase at the end of the runs. The improvement results from fewer false alarms as the members simply predict heavy fog at a lower rate, thereby improving the reliability. By 19–20 h, the fewer false alarms is enough to improve the reliability such that the NWP predictions beat persistence, which has a false alarm ratio of 1 from 16–20 h (not shown).

The mountain region had only 10 observed heavy fog events, causing the BSS at the higher β_e threshold to be an especially volatile and incomplete picture of NWP model performance. When all post-spin up hours are combined, false alarm ratios for most members are <40% higher than at the lower β_e threshold (a smaller increase than in the other regions), and the probabilities of detection are comparable or higher. These results are promising, but also not entirely surprising since the observed fog distribution is most bimodal in this region (i.e., the bimodal predictions have already shown skill at predicting fog, and most fog events are heavy fog events). More cases of heavy fog are needed to draw clearer conclusions about the NWP predictive skill for heavy fog in the mountains.

With the possible exception of the mountain region, the poor scores at the highest β_e threshold serve to emphasize that the ensemble's skill in predicting the existence of fog is better than its skill in specifically predicting heavy fog. In general, the BSSs in each region get progressively worse for greater β_e thresholds, with the largest decrease occurring between the third and fourth thresholds (corresponding to daytime visibilities of 2.75 mi and 0.875 mi, respectively). However, even at the third β_e threshold (corresponding to a daytime visibility of 2.75 mi), the scores show non-trivial positive skill in the coastal and mountain regions, suggesting the predictions are useful for more than just delineating between fog and no fog in some situations.

e. Summary

To summarize the key findings drawn from examination of the NWP q_c predictions, the skill of the ensemble suite in predicting fog increases throughout the run,

and is highest in the mountain region and lowest in the valley region, where it generally does not demonstrate skill. The ensemble is more skillful at predicting fog events than it is at specifically predicting heavy fog events.

Variations in IC and physics suites among the members are shown to produce variations in the prediction distributions, but every member exhibits highly bimodal predictions in all regions. This results in very few q_c predictions in the light fog range, despite a significant observed incidence of light fog in the coastal and valley regions. This is suggestive of a deficiency in the underlying NWP model physics rather than initial condition error since the observed fog climatology and the NWP model climatologies simply do not match. Possible sources of the deficiency include an inaccuracy in the amount of supersaturation needed for the condensation of fog droplets, error in the predicted moisture or temperature fields themselves, or a model layer that is simply too high above the ground to adequately resolve some fog events. These hypotheses will be examined in subsequent sections, but since the behavior is observed in every region by every member regardless of the physics suite used, we may reasonably conclude the deficiency is systematic.

In the coastal and valley regions, the negative q_c biases and lack of predictions corresponding to light fog are accompanied by a surplus of predictions for zero or near-zero q_c. This results in q_c bias ratios <0.5 at the light fog threshold for every member at nearly all hours. The implications of this negative q_c bias on the overall stochastic predictions is illustrated in Figure 28, which shows the distribution of ensemble mean q_c predictions for all post-spin up cases of observed fog. Of 795 total observed fog events in all regions, nearly 500 of them (62%) have an ensemble mean q_c prediction of zero, which is only possible if every member predicts zero q_c. If we also include cases when the ensemble mean q_c is below the threshold to be considered fog (q_c $<8.5 \times 10^{-4}$ g m^{-3}, or a predicted daytime visibility >6.5 mi), which often happens when one or two members have a very small but non-zero q_c prediction while the remaining members predict zero q_c, we have accounted for 96% of all observed fog cases. This systematic deficiency of producing bimodal predictions and therefore too many zero q_c predictions in the coastal and valley regions is believed to significantly reduce ensemble

skill, and addressing it represents the most impactful avenue of research to improve WRF-based VIF prediction without location-specific calibration.

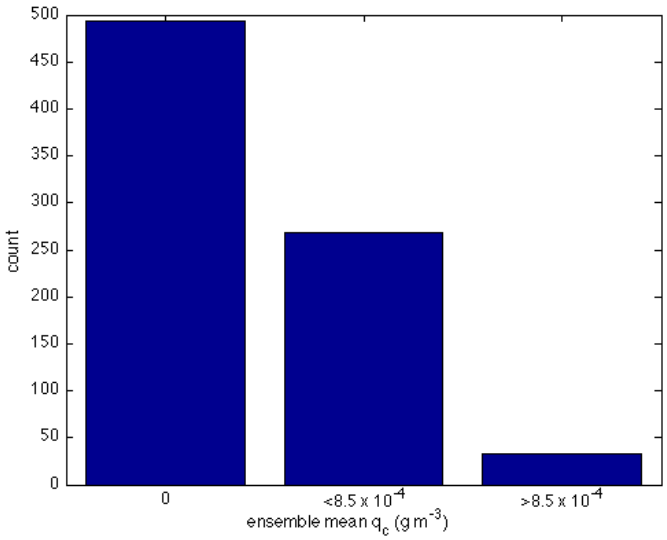

Figure 28. Distribution of ensemble mean q_c for all cases of observed fog in all regions. The first six hours of each case are excluded.

As shown in Figure 20, the bimodal tendency of the members' q_c forecasts also means many q_c predictions are too high, significantly overforecasting the severity of fog. Investigating these might also reveal a strategy to mitigate this deficiency, but this research will not pursue this avenue for two reasons. First, instances of a member overforecasting q_c values happen with less frequency than when q_c is underforecast, as evidenced by the prediction distribution histograms in Figure 18. Therefore, addressing the deficiency that causes the underforecasting of q_c (specifically, predictions of zero q_c $<88.5 \times 10^{-4}$ g m^{-3}) is believed to have more potential to positively impact predictive skill simply because it is more common.

Second, the individual member forecasts are not as important to overall skill as is the stochastic prediction from the entire ensemble suite, and rarely do the majority of the members predict heavy fog at the same time. Among all instances when heavy fog is predicted by at least one member, it is predicted by two or fewer members in 52% of the cases, and five or fewer members in 86% of the cases. This results in significant ensemble dispersion, which tempers the overall impact of erroneously high q_c

predictions from individual members. In contrast, in the 62% of observed fog cases where all members predict zero q_c, there is no ensemble dispersion. It is believed that adding dispersion to these cases provides the best chance to increase the resolution and reliability of the ensemble, while altering the high q_c prediction cases (and effectively reducing ensemble dispersion) runs a greater risk of negatively impacting the ensemble resolution that already exists.

Naturally, the challenge in adding resolution to the ensemble by statistically adjusting low q_c predictions from the members is knowing whether fog is likely in the absence of predicted non-negligible q_c. In principle, this strategy has several advantages. First and foremost, it attempts to address prediction errors that show clear evidence of being the result of a systematic NWP deficiency, and so it seems an appropriate place for the judicious introduction of a statistical element. Second, it only engages a specific and well-defined aspect of the predictions, allowing individual members producing cloud water on their own to do so unabated and still affect the predictive PDF. In this way, the approach is intentionally restrained, and ensures the framework remains largely physical-based when the NWP model predicts fog. Third, it offers the potential for improvement not just in reliability, but also in resolution since each individual member and case will be affected differently, unlike in an ensemble bias correction. Finally, it is only possible to make the adjustment in one direction (increasing q_c), which reduces complexity and simplifies tuning of the technique if it is found to destroy existing resolution.

Although the NWP q_c predictions are highly bimodal in all regions, the nature of the prediction error in the mountain region is unique in that it does not exhibit a surplus of zero or near-zero q_c predictions. It is proposed this is mostly due to a unique and highly bimodal observation distribution as opposed to any unique behavior of the NWP model. Regardless, since the overall q_c bias is near-neutral or positive for most members and the predictions are shown to produce the highest RPSS of any region beyond 10 h, attaining additional skill in this region is not a driving force behind the development and refinement of the techniques described in this work. Instead, the techniques are developed with the goal of increasing skill in the coastal and valley

regions while minimizing collateral impacts in the mountain region. To specifically seek skill improvements in the mountain region, it is suggested a more comprehensive approach is needed that involves more than just the cases of low q_c predictions.

The remainder of this chapter will examine the low-level thermodynamic properties of the NWP predictions to further uncover the source of systematic deficiency causing the excessive zero or near-zero q_c predictions. Techniques to address the deficiency are examined in Chapter V.

2. Layer 1 Relative Humidity

Given the critical role of RH in fog dynamics, we next examine the RH predictions from the NWP model to determine the role it plays in the systematic lack of q_c predictions. NWP output at layer 1 refers to the lowest NWP model layer on which full integrations are performed in the NWP model, and is where the q_c predictions examined thus far are produced. The layer is 19-21 m above the model ground level. Later, we will examine predictions from the 2-m level that are produced by WRF post-processing.

The predicted and observed RH distributions are presented in Figures 29-35. The data are presented for each site rather than each region to show the amount of variation among the sites within each region. Except for a few aspects of the data discussed below, the results show very little intra-region variability relevant to the conclusions of this thesis. This supports the notion that the NWP deficiencies identified in the layer 1 RH predictions are systematic since they are evident at multiple sites within a region. The remainder of the thermodynamic variables examined later in this work also show minimal intra-region variability pertinent to the conclusions made. For brevity, their results will be shown for each region rather than each site.

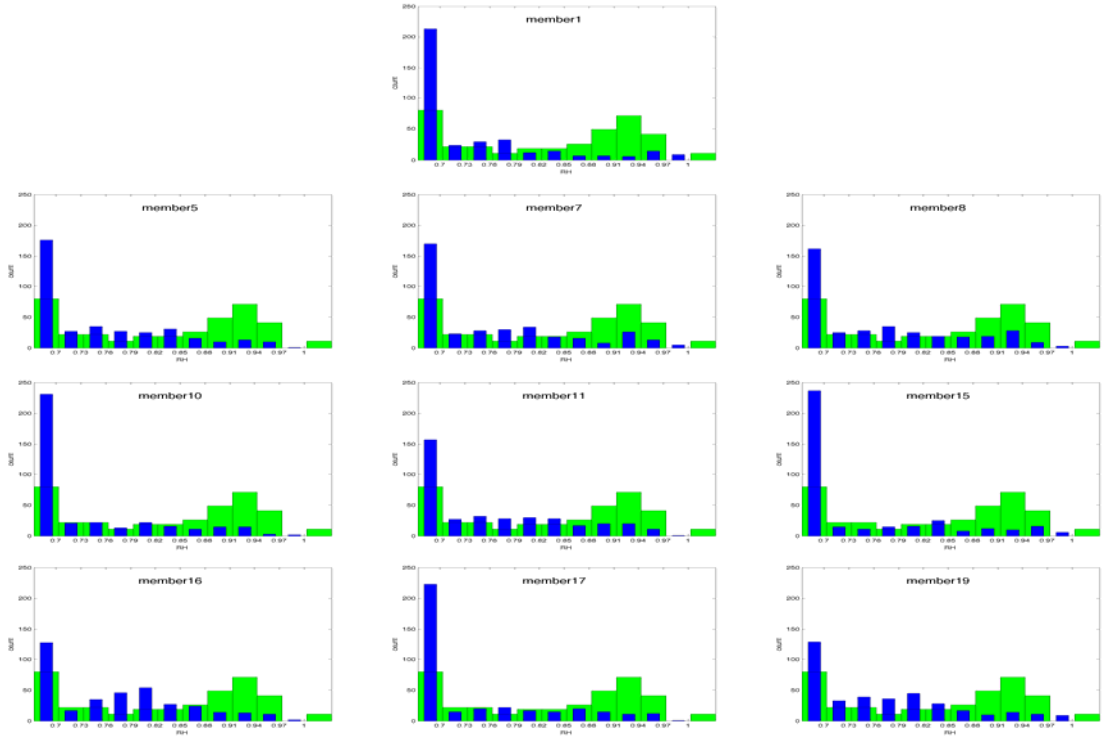

Figure 29. Distribution of NWP layer 1 relative humidity predictions (blue bars), and KCEC observations (green bars). The first six hours of each case are excluded.

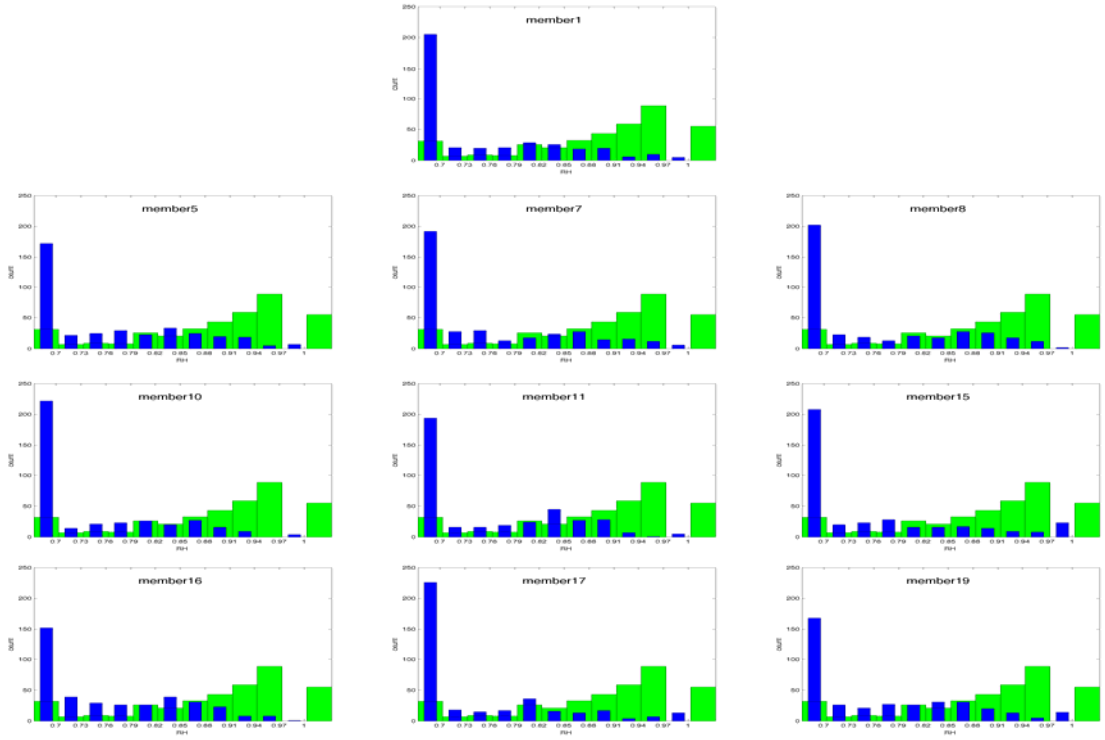

Figure 30. Same as Figure 29, but for KACV.

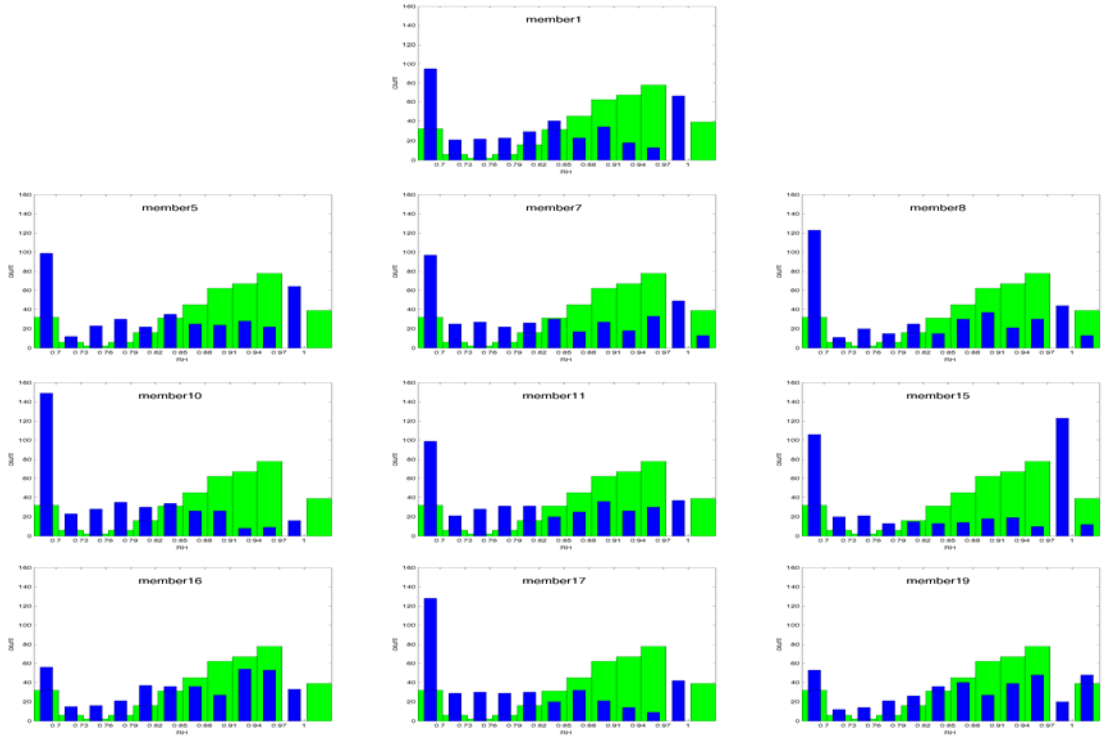

Figure 31. Same as Figure 29, but for KSCK.

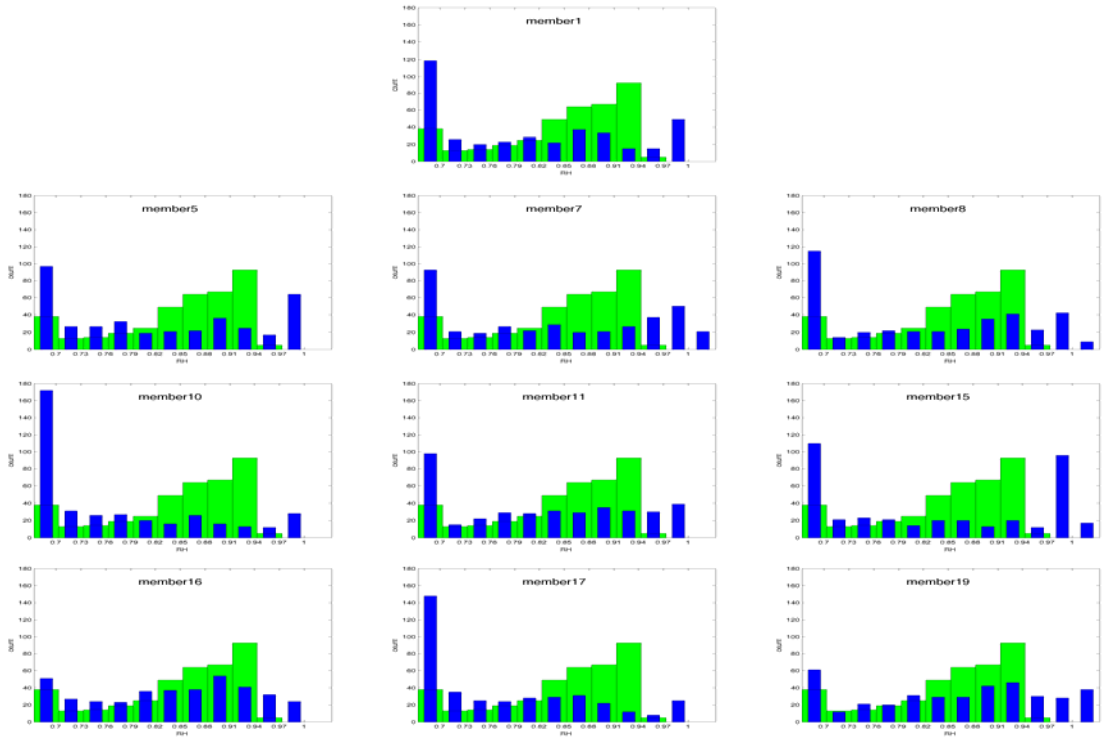

Figure 32. Same as Figure 29, but for KMOD.

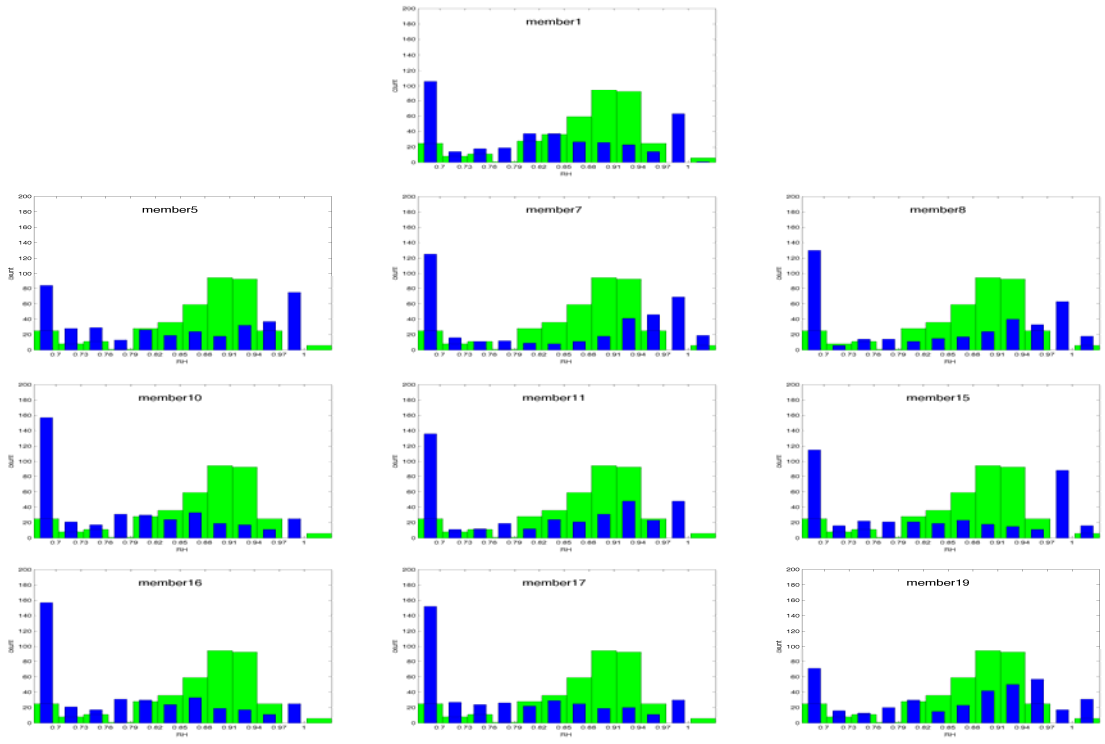

Figure 33. Same as Figure 29, but for KMCE.

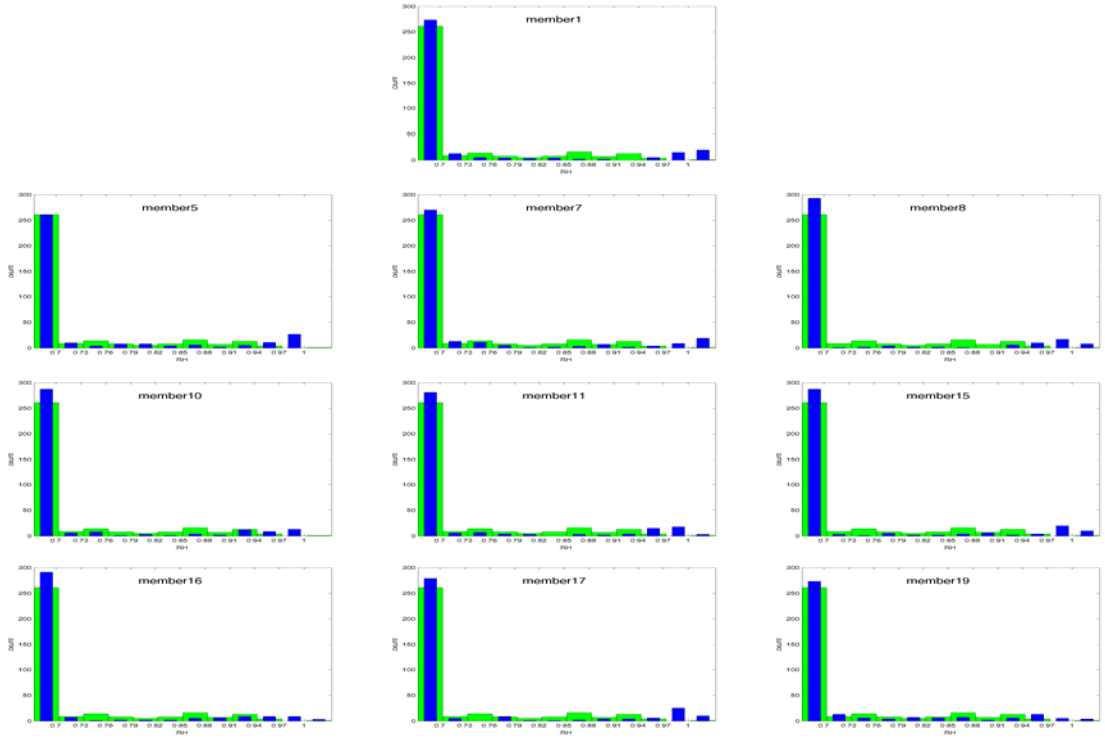

Figure 34. Same as Figure 29, but for KBLU.

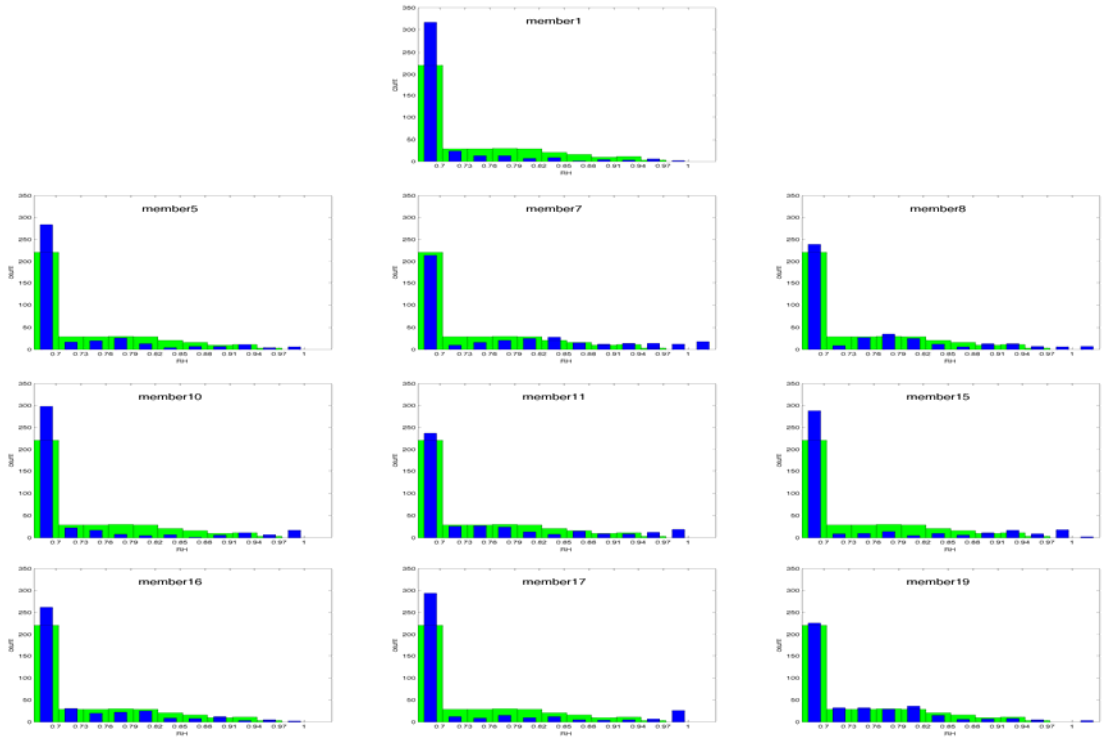

Figure 35. Same as Figure 29, but for KRNO.

The coastal sites (KCEC and KACV) are shown to have a negative RH bias by every member, with a surplus of predictions with RH <0.7 and insufficient predictions with RH >0.85. The observations exhibit a local maximum in the distribution for RH values of 0.88–0.97 that does not exist in the prediction distributions. The character of the distributions is similar in the valley region, although the extent to which the predicted distributions underestimate the local maximum in observed high RH values (which in the valley region is between 0.82 and 0.97) is less and varies substantially by member. The observed RH is shown to reach saturation quite often at KSCK, but for reasons not clear, there are few instances of observed RH reaching saturation at KMCE, and no instances of this at KMOD. Minor ASOS temperature and/or dewpoint instrument error is suggested, as the stations have similar observed frequencies of fog (0.57, 0.49, and 0.54 at KSCK, KMOD, and KMCE, respectively) and heavy fog (0.23, 0.21, and 0.19) during the study, which would not be expected if only some of the sites were reaching an RH of 1 while others were not. Additionally, RH values of 0.97–0.99 were never observed at any site, which is believed to be due to a rounding routine employed by ASOS.

More significantly, the members are shown to have substantial differences in their incidence of saturated or supersaturated (RH $\geq$1) predictions. Several members (e.g., members 8 and 15) produced predictions at or above complete saturation fairly regularly, while others (members 5 and 10) never produced saturation at any site. Additionally, some members (e.g., member 15) show a high incidence of near-saturation predictions (RH $\geq$0.97 but <1) compared to others.

Some variation in RH predictions among the members is expected and indeed desired, as the ensemble is intended to sample the uncertainty of the prediction. However, these profound differences near the limit of saturation, coupled with the fact that saturation was never predicted by any member in the coastal region, raise questions about the reconciliation of saturation and cloud water by each member's respective microphysics scheme.

To more closely examine the relationship between RH and q_c within each member, Figure 36 shows each member's entire RH distribution for all sites, with instances corresponding to predicted q_c >8.5 x 10^{-4} g m^{-3} (the lowest threshold for fog)

indicated in light blue. For comparison the observed data are also shown, with observed fog events plotted in light blue. Clearly, each member is able to predict non-trivial levels of q_c when RH is below saturation, making the fact that some of the members never actually reach saturation largely irrelevant from a fog prediction perspective. One of these is member 10, which predicted fog with RH as low as 0.80. However, the plots show the members are not likely to predict fog until the RH is at least 0.93, and much higher than this in some members. This does not agree with the observed data, which shows fog being more likely than not at an RH of only 0.88, and being observed with RH as low as 0.81 (lower than the lowest predicted RH coinciding with predicted fog in nine of the members). As discussed in Chapter III, fog is included in observations rather liberally by the ASOS algorithm, likely involving some instances of moist haze whose particles have not yet reached activation radii. At issue is the point at which these moist haze particles are considered cloud water by individual microphysics schemes, and the data in Figure 36 suggests each scheme uses a more restrictive criterion than does ASOS. The criterion in the microphysics schemes may be more physically sound, but in practical terms, it likely results in the members missing some visibility restrictions due to moist haze, which the schemes do not consider. Absent a modification in the ASOS fog identification algorithm, it is likely the microphysics schemes used in this research will miss many instances of observed fog when RH is 0.81–0.93 *if* their RH predictions are accurate in these cases. The extent of the impact will be tested in Chapter V by using predicted RH values as a proxy for fog.

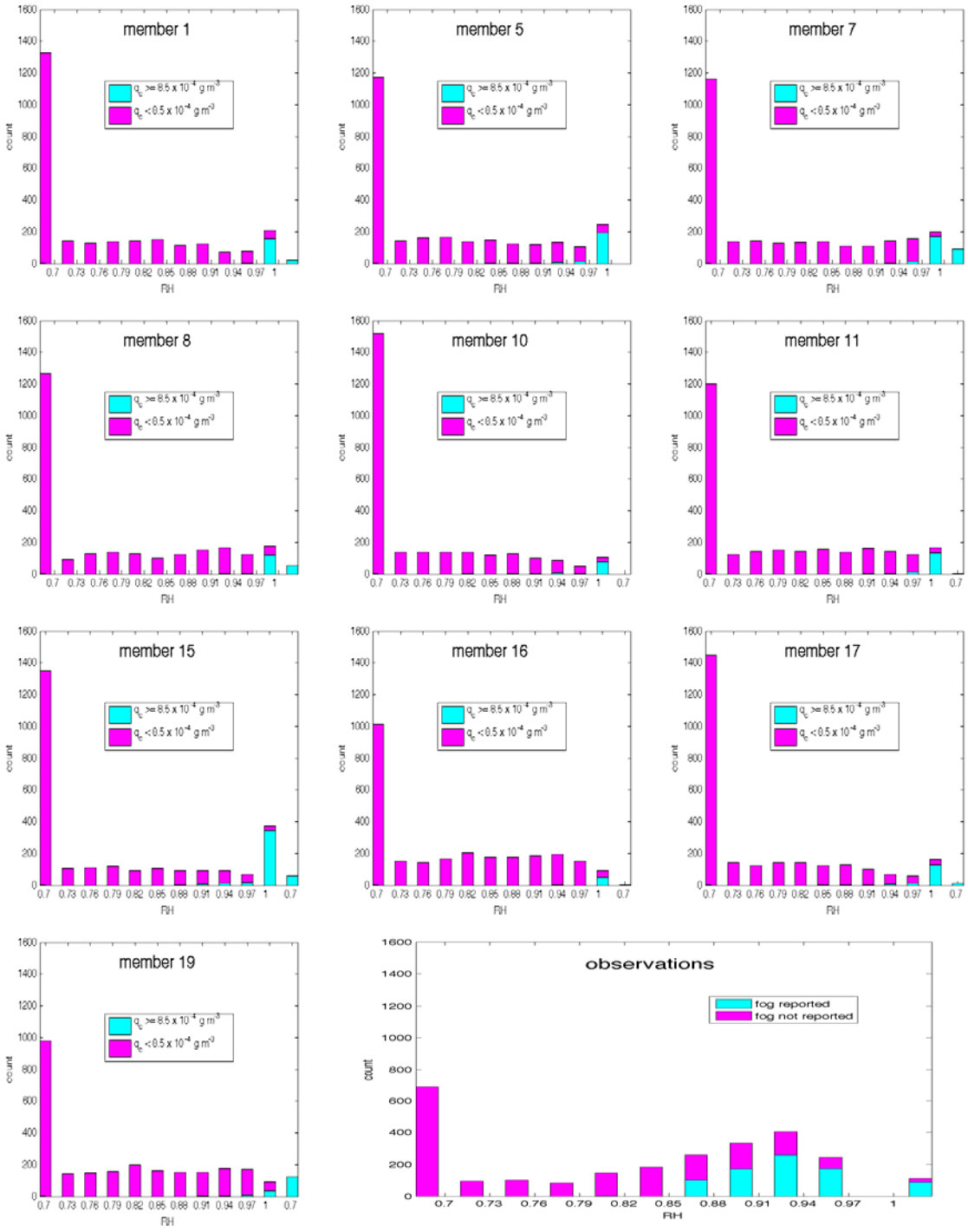

Figure 36. Distribution of layer 1 relative humidity predictions from each member at all sites. Predictions coinciding with $q_c \leq 8.5 \times 10^{-4}$ g m^{-3} (the lowest threshold for fog) are plotted in light blue. The observed relative humidity distribution is also included, with instances coinciding with observed fog plotted in light blue. The first six hours of each case are excluded.

Aside from the inconsistencies at near-saturation, the larger discrepancies between the distributions of RH predictions and observations suggest more fundamental NWP prediction errors. The negative bias of nearly all members in the coastal and valley regions (as well as the near-neutral bias in the mountain region) is reflected in the verification rank histograms of layer 1 RH (Figure 37) for all post-spin up data, which show that the observed RH is higher than the predictions of all members at a rate that exceeds 0.6 in the coastal region, and exceeds 0.35 in the valley region. These rates are inflated to some extent due to the reluctance of some members to reach saturation (something the observations do with some regularity), but they still provide strong indication that the deficiencies of each member's RH forecasts also significantly hinder the quality of the ensemble stochastic predictions. Additionally, the frequency of observations falling above *or* below all member predictions is excessive in the valley and mountain regions. This indicates the ensemble is underdispersive, or that the uncertainty in the prediction is not adequately sampled by the ensemble members, even after correcting for bias. In the coastal region, the dispersion characteristics are difficult to determine due to the strong negative bias overwhelming the signal.

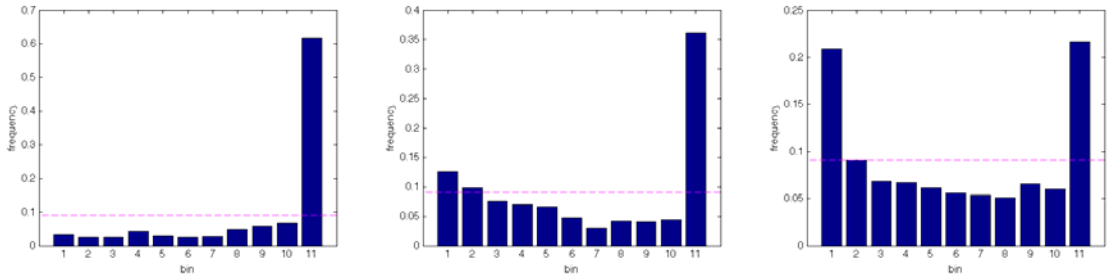

Figure 37. Verification rank histograms of layer 1 relative humidity for the coastal region (left), valley region (center), and mountain region (right). The first six hours of each case are excluded.

In the left column of Figure 38, the layer 1 RH bias and error variance for all cases are shown for the coastal region (top two panels), valley region (center two panels), and mountain region (bottom to panels) for each member as a function of forecast hour. These two metrics function as a decomposition of the total mean squared error of the predictions into a bias component (or the mean error at each hour for the given member), and the mean square of the remaining error after the bias has been subtracted from the

member's predictions. Decomposing the error into these components shows the potential effectiveness of a bias correction to the data (i.e., predictions with low error variance offer more promise for an effective correction, especially if the observed bias is relatively consistent among the members and forecast hours).

Not surprisingly, RH biases in the coastal region are between -0.10 and -0.25 for each member throughout most of the run, with all members improving to a small negative bias from 17–20 h. Even after this bias is subtracted from the predictions, the error is significant, with error variances for the majority of members ranging from 0.02 to 0.04. (Taking the square root of the error variance yields the bias-corrected standard deviation of the RH error, σ, which in this case is between 0.14 and 0.20.) The biases in the valley region are smaller in magnitude and more consistent throughout all forecast hours, ranging from about -0.15 to 0 for most members.

The negative biases that decrease in magnitude after sunrise are consistent with the NWP model layer 1 not adequately capturing a low-level inversion, whether due to the model layer being too high and/or inadequate cooling at the layer itself (consistent with the findings of Tardif 2007, whose model layer 1 was only half as high at 10 m above the model ground level). This scenario might be expected in some radiation fog events, and perhaps during some advection fog. The bias improves after sunrise as the boundary layer is heated and mixed, destroying any low-level inversions. We will show later that this likely contributes to at least part of the negative bias in layer 1 RH.

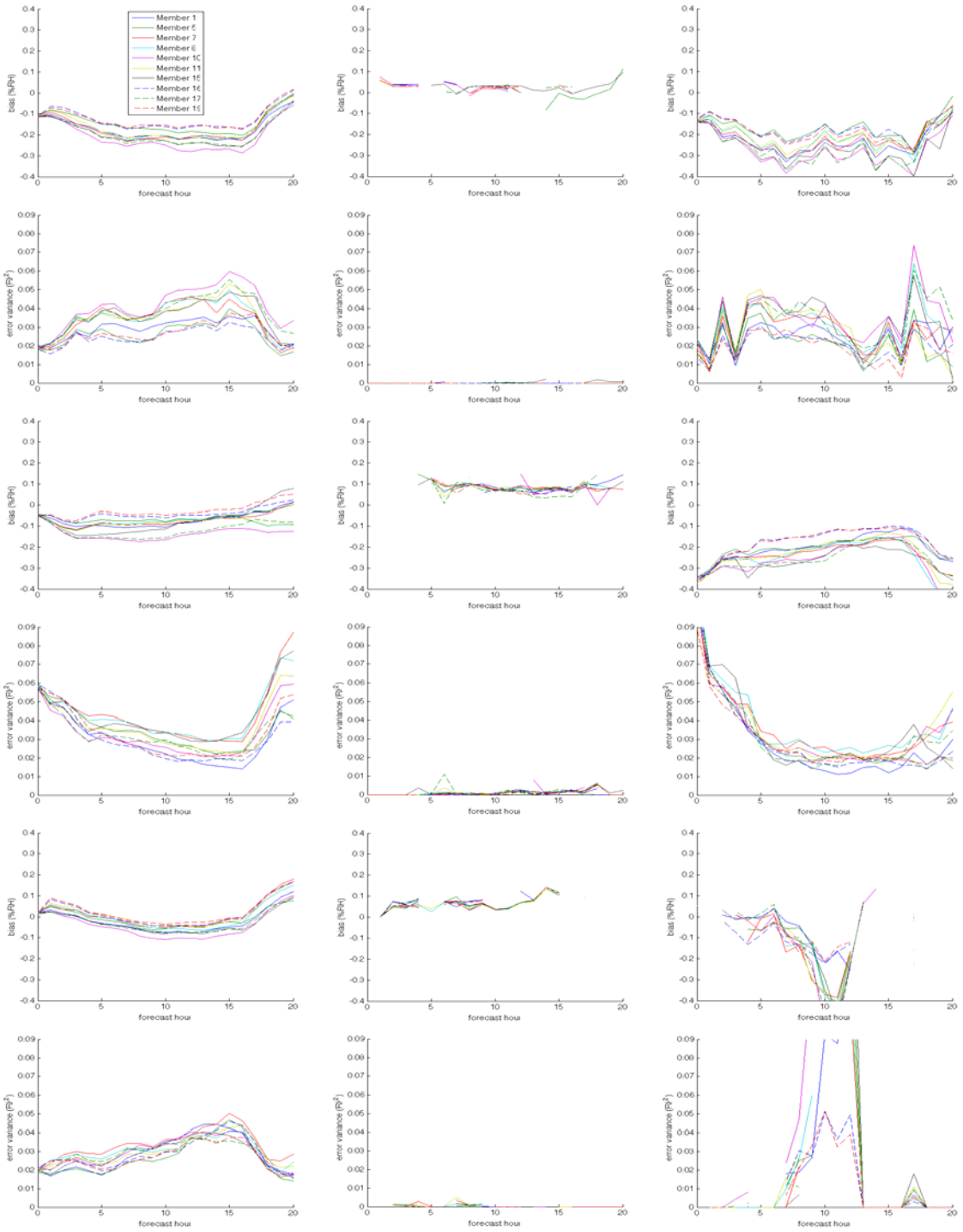

Figure 38. Layer 1 relative humidity bias and error variance of each member for coastal (top two rows), valley (center two rows), and mountain (bottom two rows) regions. The left column shows all data, the center column includes only fog hits (fog observed and predicted) and the right column includes only fog missed opportunities (fog observed and not predicted).

The error variances in the coastal region gradually increase overnight before dropping during the post-sunrise hours. This trend closely matches the observed incidence of fog (i.e., the error variance is higher when the incidence of fog is highest), raising doubts about whether the layer 1 RH predictions alone offer adequate predictive skill to inform adjustments of low q_c predictions in this region. The inconsistent biases could necessitate the additional complexities of a q_c adjustment strategy that is time-dependent, something that is preferably avoided due to the risks of using a much smaller dataset at any given hour as the basis for mitigating the impacts of an NWP systematic deficiency. Still, the mountain region also shows somewhat inconsistent biases with error variances only slightly smaller than in the coastal region, yet the q_c predictions are the most skillful presumably because the bias is near-neutral. Although layer 1 RH predictions in the coastal region are not ideally suited for our purpose, they certainly cannot be excluded as an option to help inform q_c adjustments.

In the valley region, error variances are shown to be comparatively lower during the nighttime before increasing after sunrise. Since the overnight hours are also characterized by a fairly consistent bias, the prospect of leveraging available RH predictive skill to inform q_c adjustments is higher than in the coastal region, excluding the post-sunrise period.

Since we are limiting our statistical approach to upward adjustments of zero q_c predictions, it is useful to compare the biases and error variances of instances when the members correctly predicted fog (i.e., the hits, shown in Figure 38 center column) to instances when fog was observed but not predicted ("missed opportunities", shown in Figure 38 right column). The interpretation here is different than for the overall data in the sense that we do not have the option of correcting for the missed opportunity biases since we do not know a prediction is a missed opportunity until after the fact; indeed, identifying low q_c predictions likely to be missed opportunities is precisely our primary objective. Instead, viewing the parsed biases and error variances in this way potentially provides insight into why the NWP model sometime predicts observed fog events and at other times misses them.

For the hit cases, the data shows very small error variances because the predicted layer 1 RH must be very close to 1 in order to predict q_c. The biases in each region are slightly >0, reinforcing our earlier results that RH in the model must be closer to 1 (or slightly >1) to produce q_c than what is required in the observations. However, attempting to account for this discrepancy by using a slightly lower RH as a proxy for fog is unlikely to have great effect since the biases for the missed opportunities are far from neutral, ranging from -0.35 and -0.2 for most members. Furthermore, the magnitude of these biases is larger than for the overall data, especially in the valley region, which means even with a bias correction to the RH predictions (informed by the overall RH bias), they might have limited usefulness in skillfully reducing fog missed opportunities.

Another important inference we may draw from the parsed biases is the degree to which error in the layer 1 RH predictions is linked to error in the q_c predictions. For example, if the biases are similar for fog hits and missed opportunities, it suggests layer 1 RH errors are independent of excessive zero q_c predictions, and therefore could not be traced as a cause of the q_c prediction deficiency. In the coastal and valley regions, this is clearly not the case, indicating there is a high correlation between RH errors and q_c errors. Given our physical understanding of fog and the critical role of RH in its dynamics, we may reasonably conclude that prediction error in layer 1 RH plays a role in the systematic NWP deficiency that ultimately manifests as excessive zero q_c predictions in the coastal and valley regions.

Our next step is to continue to trace the error backward through the predictions of the fundamental elements of RH to better understand the source of the q_c error. Specifically, layer 1 temperature and layer 1 water vapor are examined next. Before proceeding to the analysis of water vapor prediction errors, two brief observations are made regarding the layer 1 RH data that is not central to this work but noteworthy nonetheless. Unlike the q_c field, the RH field is initialized in each member with ICs from a member of GEFS. However, the initialization in this dataset provided layer 1 RH ICs that were too low by an average of 0.10 in the coastal region, and 0.05 in the valley region (Figure 38, left column). After a few hours, the effect of this IC bias is likely smaller than the effect of the systematic NWP deficiency evidenced by the mismatched

model climatology and observed climatology of layer 1 RH, but further examination is needed to better grasp the full impacts of IC bias.

Secondly, there is some evidence of significant spin up fluctuations in the layer 1 RH field in all regions when the members are initialized in moist conditions. This can be seen by the oscillations of error variances during the first few hours of the missed opportunities cases (Figure 38, right column). Additionally, the mountain region during these hours has a bias of near zero during missed opportunities (the only time in any region this is observed), indicating the layer 1 RH values are accurate (i.e., near saturation), but there is no cloud water in the predictions. Whether this is due to spin up of the q_c field or a case of moist haze being identified as fog by ASOS cannot by known without further investigation.

3. Layer 1 Temperature

Systematic NWP error causing RH predictions to be too low could be due to temperature predictions that are too warm, moisture predictions that are too low, or a combination of both. Distributions of predicted and observed layer 1 temperature for each region are shown in Figures 39–41. In the coastal region, the NWP model climatology from every member is shifted several degrees warmer than the observed climatology, resulting in a clear warm bias. Seven of the members had no predictions <276 K, yet the observed climatological incidence of temperatures below this threshold is 0.2019. The same deficiency is present in the valley region, although it appears to be less severe in most of the members. The distributions of predictions in the mountain region do not show a clear warm bias. The mountain region is also unique for its bimodal distribution of observations, a feature also reflected in the prediction distributions of most members.

The verification rank histograms for the layer 1 temperature (Figure 42) show that the stochastic predictions from the entire ensemble suite also have a clear warm bias in both the coastal and valley regions. The bias in the coastal region is the most severe, with over 70% of the observation verifying below every member's prediction. A minor warm bias is evident in the mountain region. The ensemble is shown to be underdispersive in each region.

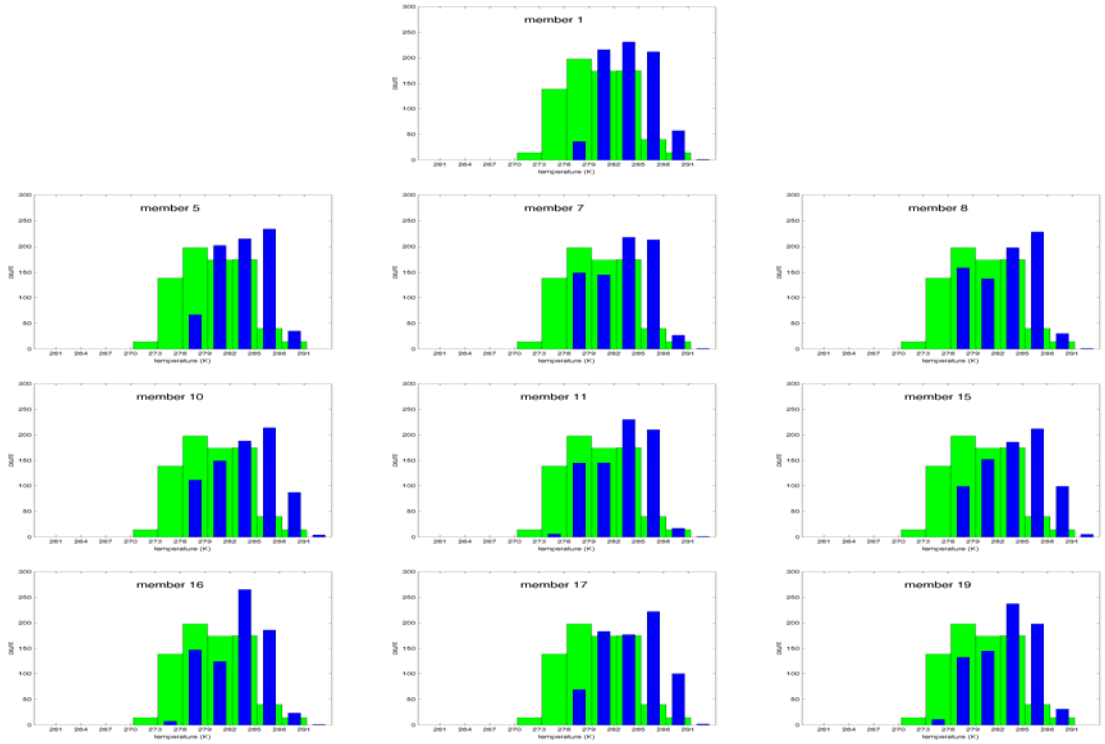

Figure 39. Histogram of distribution of NWP layer 1 temperature predictions (blue bars), and observations (green bars) for coastal region. The first six hours of each case are excluded.

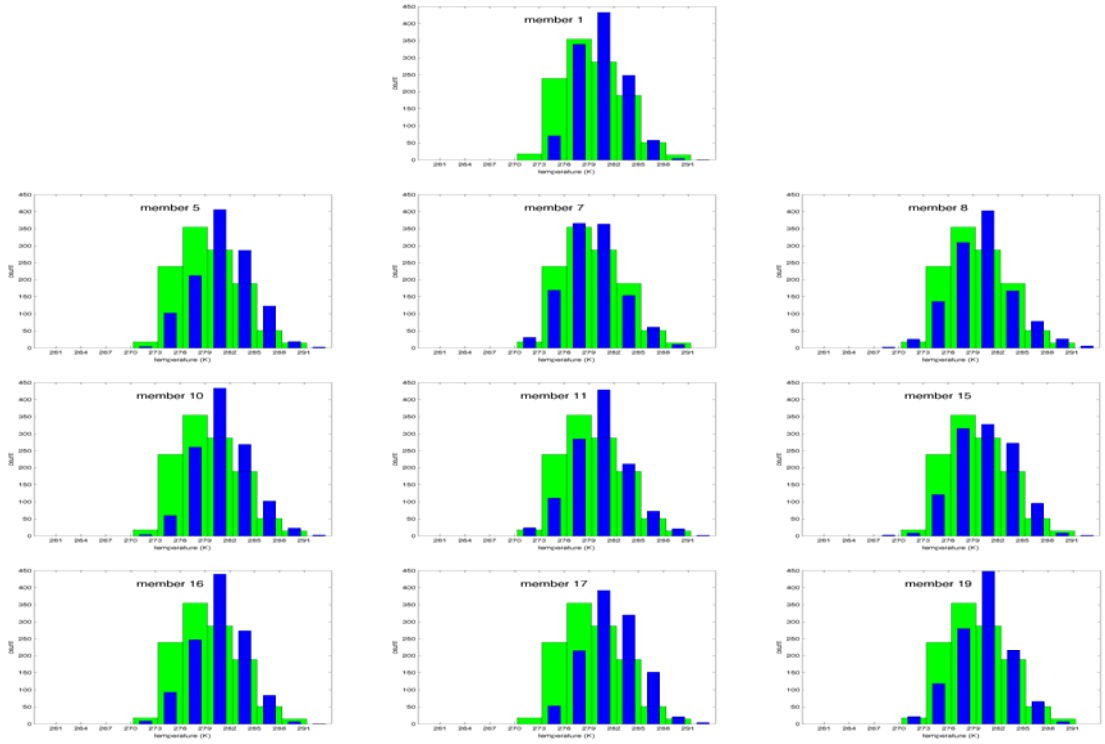

Figure 40. Same as in Figure 39, but only for the valley sites.

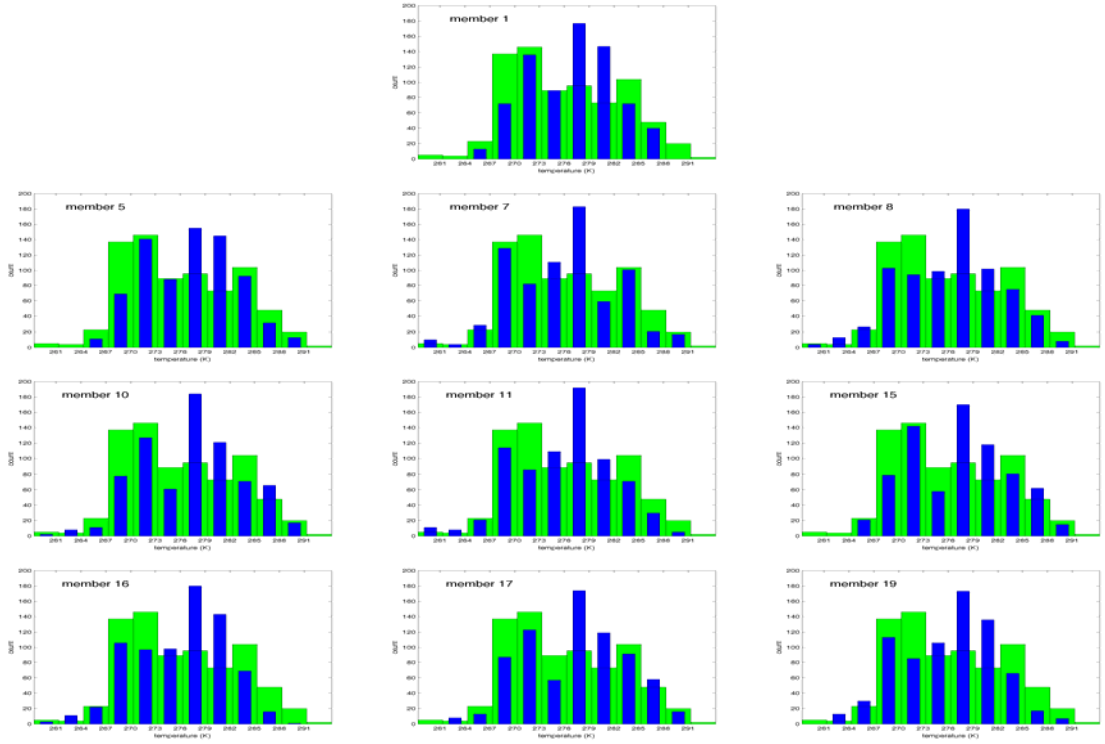

Figure 41. Same as in Figure 39, but only for the mountain sites.

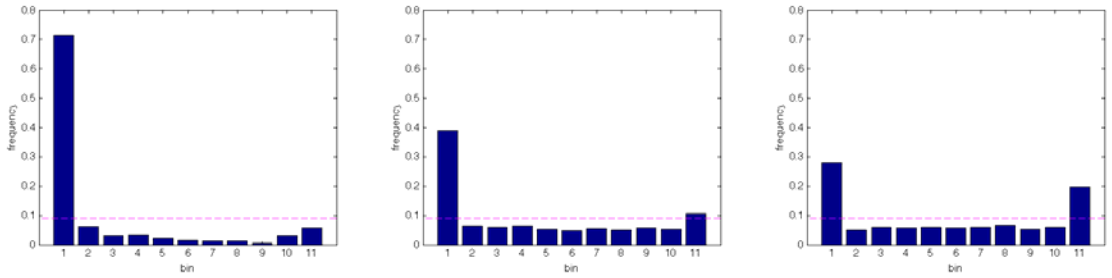

Figure 42. Verification rank histograms of layer 1 temperature for the coastal region (left), valley region (center), and mountain region (right). The first six hours of each case are excluded.

Figure 43 confirms that, when all the data is included (left column), the coastal region exhibits the largest warm biases, which gradually increase during the night and reach nearly 5 K by all members just prior to sunrise. The pattern is the same with less magnitude in the valley region, with the warm biases reaching 2–3 K for most members before returning to near-neutral after sunrise. The nature of the error variances, however, is quite different between the two regions. At the coastal sites, the error variances reach nearly 20 K^2 pre-sunrise, then decrease to about 5 K^2 during the late morning. In contrast, error variances in the valley region are relatively low overnight, then increase by 5–15 K^2 after sunrise. This pattern closely follows those of the layer 1 RH error variances in each respective region, suggesting the temperature prediction errors are at least partially responsible for the layer 1 RH errors.

To compare these results more closely in the context of diurnal temperature changes, Figure 44 shows the mean temperature change of observations (green) and predictions (blue) during the interval 7–15 h (2300-0700 LT), and again from 15–20 h (0700-1200 LT) for all cases. Although it is mean temperature changes that are shown, the line for the predictions does not start at zero but has been displaced upward above the line for the observations so that the mean bias of the predictions is also portrayed throughout the plots. The thin dashed lines represent ± 1σ of the temperature changes (not the biases) for each of the two intervals. The plots show that both regions exhibit mean observed diurnal temperature changes of several degrees, but the valley region predictions have the diurnal changes more accurately forecast. Of particular note is the mean cooling rate of the predictions in the valley region, which is in close agreement

with observations. This is a different result than that achieved by Tardif (2007), who found delayed fog onset due specifically to inadequate cooling.

In contrast, the coastal region predictions have a total temperature range that averages <1 K across the entire post-spin up period (7–20 h), suggesting a general deficiency in the handling of boundary layer temperature forcings. The difference between the two regions is especially evident during the interval 15–20 h, when the coastal region predictions show mean warming of only 0.8° C.

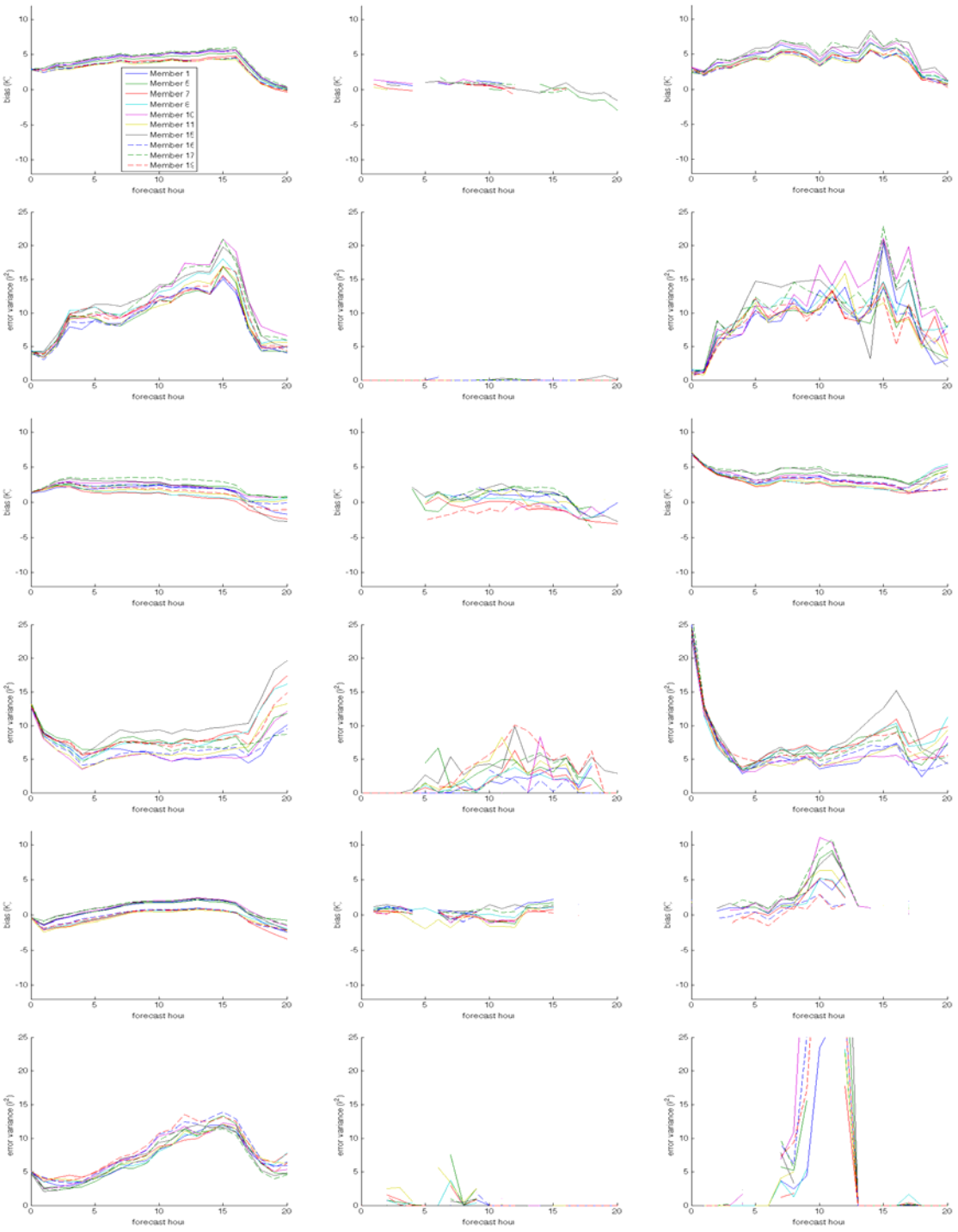

Figure 43. Layer 1 temperature bias and error variance for each member for coastal (top two rows), valley (center two rows), and mountain (bottom two rows) regions. The left column shows all data, the center column includes only fog hits, and the right column includes only fog missed opportunities.

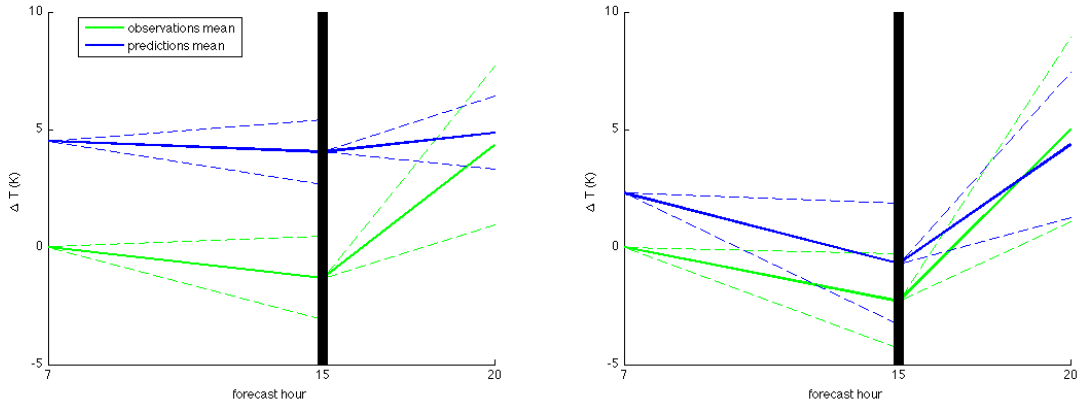

Figure 44. Layer 1 mean temperature change for observations (solid green line) and predictions (solid blue line) from 7–15 h, and again from 15–20 h in the coastal region (left) and valley region (right). The line for the mean prediction change is offset above the line for the mean observations change so that the mean bias of the predictions is also portrayed throughout the plot. The dotted lines represent ± 1σ of the temperature change within each interval.

It is proposed the difference in overnight error variances between the two regions is attributable to a more consistent nighttime boundary layer structure in the valley region, which is subject to large-scale radiative cooling and weak drainage flow on the majority of nights, as opposed to a mix of less-consistence radiative cooling and advection complicated by larger low-level temperature gradients inherent to the coastal region. While the valley region structure is seemingly more predictable for the WRF members than the coastal boundary layer, the warm bias in both regions suggests the NWP model does not fully resolve the coldest air near the surface. Perhaps the coastal region predictions are also sensitive to IC bias, which is shown to average 3–4 K warm in all members. After sunrise, the decreasing error variances in the coastal region are due to observed warming that is more consistent in timing and amplitude, whereas warming in the valley region has more day-to-day variation not resolved by the predictions.

The biggest reason for greater variation in warming rates in the valley region may be the greater tendency for fog to linger well into the late morning, with most cases absent in the predictions; at 20 h, the incidence of observed fog is 0.2338 in the valley region, and only 0.0893 in the coastal region. These post-sunrise trends are consistent

with the fog BSSs in each region, which generally increase after sunrise in the coastal region, and decrease in the valley region.

Examining the layer 1 temperature biases and error variances for fog hits (Figure 43, center column) and fog missed opportunities (Figure 43, right column) shows that the members in both the coastal and valley regions have virtually no temperature bias when fog is correctly predicted. However, predictions resulting in fog missed opportunities are characterized by a warm bias of at least 3 K at most hours in both regions. This disparity is additional evidence that the observed temperature deficiencies are linked to q_c prediction deficiencies (via RH prediction deficiencies).

For most members, the layer 1 temperature biases in the coastal region are larger by <1 K larger during missed opportunities compared to the biases for all the data. This aspect of the predictions makes layer 1 temperature a good candidate for a bias correction in this region. However, the large overnight error variances are a drawback, as are the abrupt change in biases after sunrise. Even with a successful bias correction, the full impact on improving the skillfulness of the RH and q_c predictions is also dependent on the nature of the water vapor predictions, which are examined in the next section.

The layer 1 temperature predictions are perhaps slightly less suitable for a bias correction in the valley region given the larger overnight biases by 0.5–1.5 K during missed opportunities compared to the biases for all the data (the differences become larger after sunrise). However, the reasonably consistent nature of the biases as a function of forecast hour, and the low error variances relative to the coastal region are positive characteristics of the predictions that might be leveraged to inform q_c adjustments using methodology other than a bias correction. Whether this is the case is explored in subsequent chapters.

4. Layer 1 Water Vapor

The systematic warm bias in the NWP predictions has been shown to play a role in the negative RH bias, but moisture predictions may also contribute to the low RH predictions. Distributions of layer 1 water vapor mixing ratio, q_v, predictions are in generally close agreement with the observed distribution in each region (Figures 45–47), with only minor discrepancies apparent in individual members. Unlike the layer 1 RH

and temperature predictions, no systematic NWP model deficiency affecting the model climatology of q_v predictions is immediately apparent.

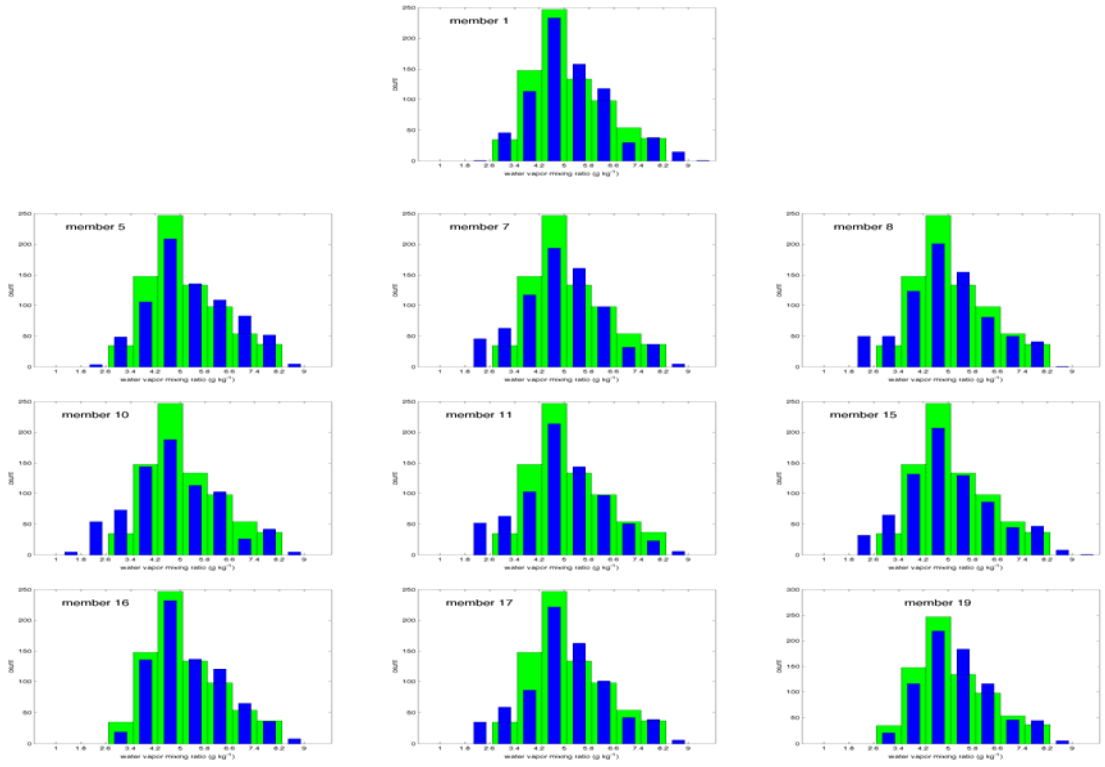

Figure 45. Histogram of distribution of NWP layer 1 q_v predictions (blue bars), and observations (green bars) for coastal region. The first six hours of each case are excluded.

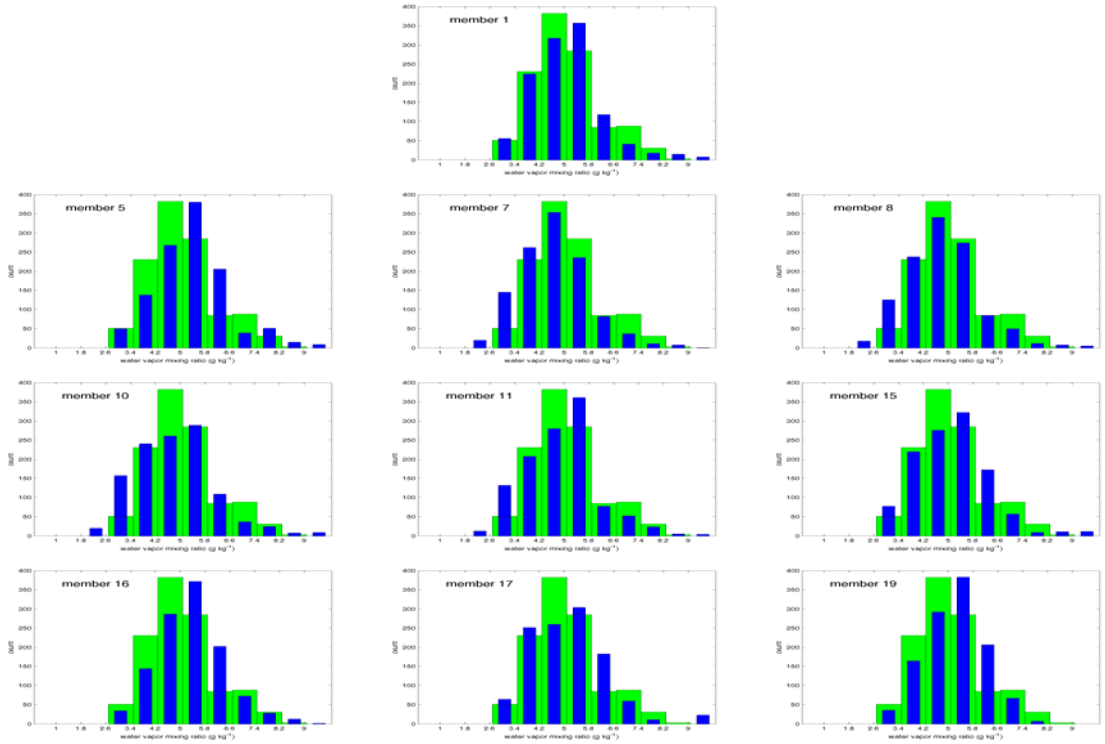

Figure 46. Same as in Figure 45, but only for the valley sites.

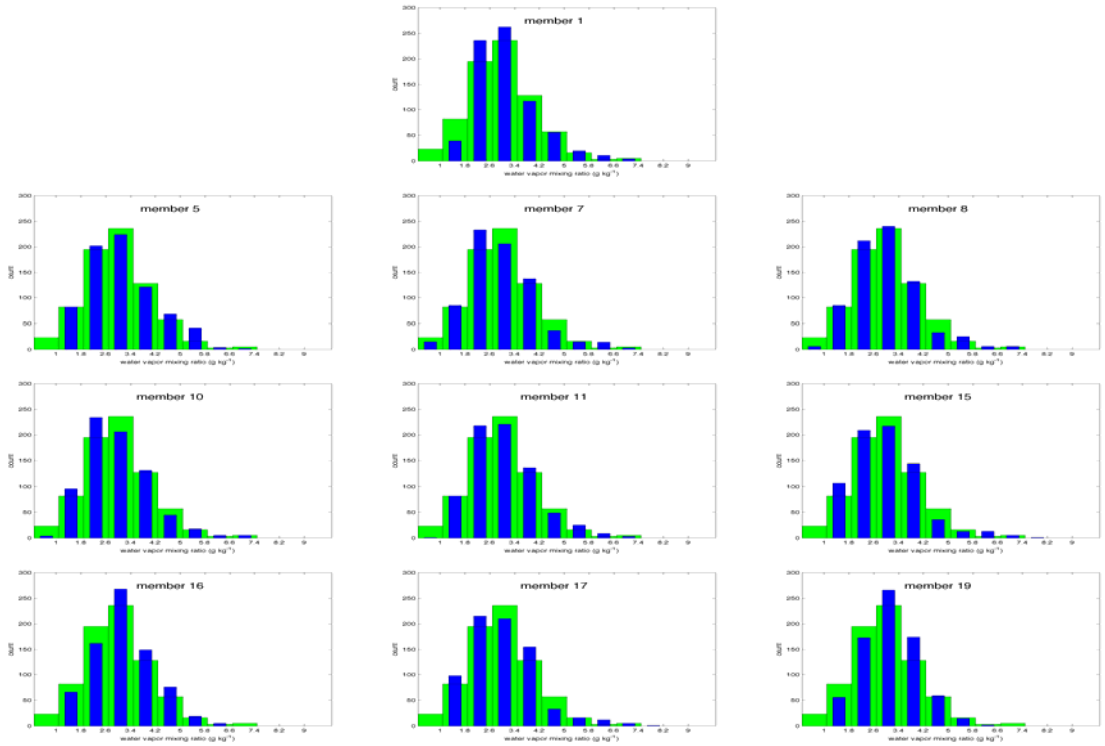

Figure 47. Same as in Figure 45, but only for the mountain sites.

The verification rank histograms of layer 1 q_v (Figure 48) show that the ensemble suite perhaps has a slight moist bias in each region, further suggesting that q_v bias is not a primary cause of the NWP model RH bias. Furthermore, the magnitude of the bias implied by Figure 48 is less than that implied by the rank histograms of layer 1 RH and temperature predictions, indicating there is comparatively little bias in the q_v predictions. The stochastic predictions in each region are clearly underdispersive, particularly in the valley and mountain regions.

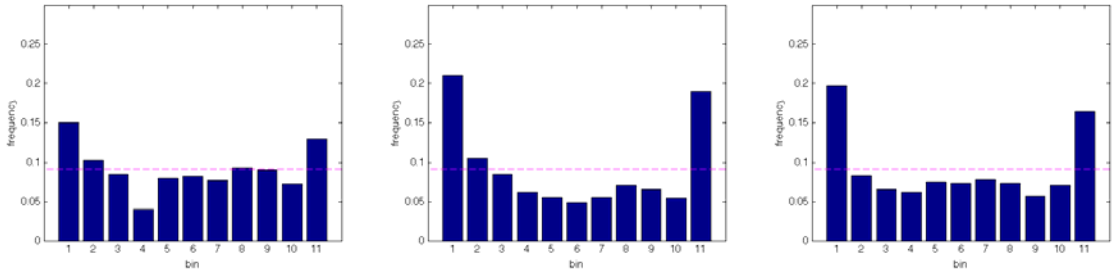

Figure 48. Verification rank histograms of layer 1 q_v for the coastal region (left), valley region (center), and mountain region (right). The first six hours of each case are excluded.

In Figure 49, the member biases for all cases (left column) are shown to be near-zero throughout the overnight hours in each region. Aside from a spin up period in the valley region, error variances are 0.5–1.0 $g^2\ kg^{-2}$ (translating to σ of 0.7–1.0 g kg^{-1}) for all members in all regions through most forecast hours. To compare the relative impact on RH of this error variance versus layer 1 temperature error variance, consider that at 1000 hPa with a temperature of 278 K and an RH of 0.9, a decrease in q_v of 0.85 g kg^{-1} (or about 1σ in the data) results in an RH of 0.74, which it the same effect as a temperature increase of 2.7 K (which when squared translates to an error variance of 7.3° K^2). The relative effect varies substantially at different RH, but as a first-order estimate, we may conclude the q_v predictions and temperature predictions have comparable error variances in the valley region in regard to their effect on RH during the overnight hours, with the temperature predictions having larger error variances (and likely less predictive skill) after sunrise. In the coastal region, the temperature predictions have greater error variances during the nighttime, and similar error variances as the q_v predictions after sunrise.

Since the biases are near-neutral during the nighttime, the q_v predictions offer less immediate opportunity to leverage for an bias adjustment technique during this period. The change in bias after sunrise in the coastal and valley regions is also not well-suited for a correction due to its time dependence, but it is still worth examining to better understand the behavior of the NWP model. In Figure 50, the q_v changes for these two regions are plotted in the same format as Figure 44. On average, both regions show an observed q_v decrease overnight (7–15 h), followed by an increase after sunrise (15–20 h). Despite the deficiencies observed in the coastal region in capturing diurnal temperature trends, the NWP model appears to model the diurnal q_v trends relatively accurately in this region. The plots suggest the small positive bias overnight evolves into a negative bias by the end of the runs due generally to insufficient moistening of the boundary layer after sunrise. This characteristic is more pronounced in the valley region, where the average rate of observed moistening is higher but the average rate of predicted moistening is lower than in the coastal region.

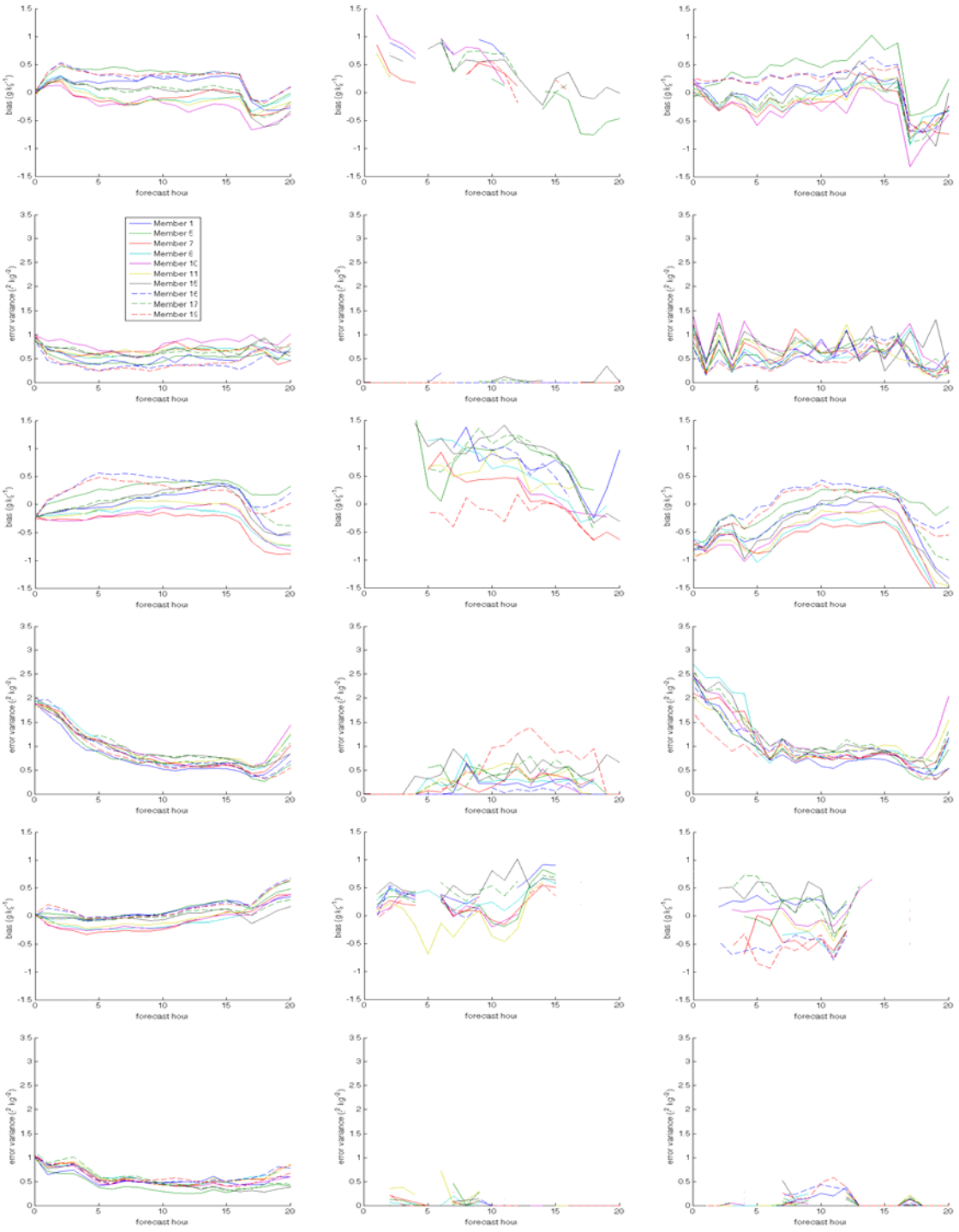

Figure 49. Layer 1 q_v bias and error variance of each member for coastal (top two rows), valley (center two rows), and mountain (bottom two rows) regions. The left column shows all data, the center column includes only fog hits, and the right column includes only fog missed opportunities.

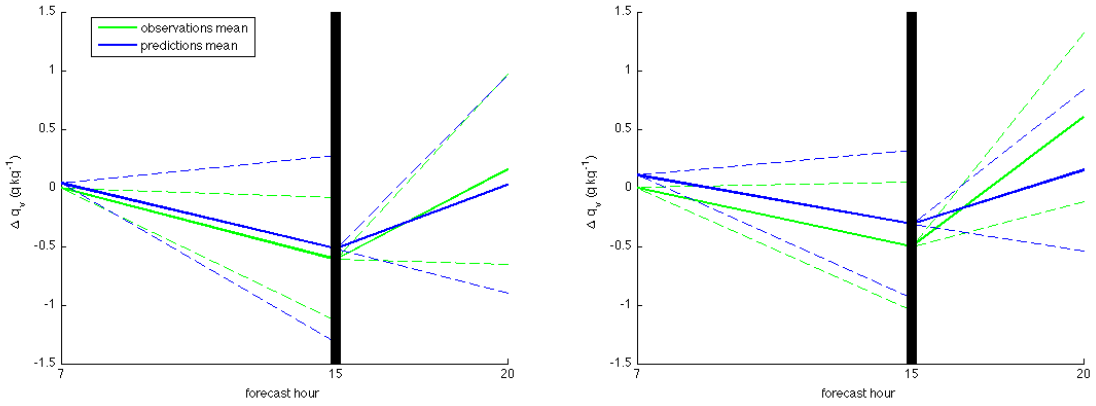

Figure 50. Layer 1 mean q_v change for observations (solid green line) and predictions (solid blue line) from 7–15 h and from 15–20 h in the coastal region (left) and valley region (right). The line for the mean predictions change is offset above the line for the mean observations change so that the mean bias of the predictions is also portrayed throughout the plot. The dotted lines represent ± 1σ of the q_v change within each interval.

Dai et al. (1999) and Dai et al. (2002) proposed several influences on diurnal q_v trends, including evapotranspiration, synoptic scale vertical motion, precipitation, and convective vertical mixing. Any of these might have varying influence on any given day, with evapotranspiration perhaps playing the largest overall role due to its tendency to increase with insolation and peak around noon (therefore being consistent with post-sunrise moistening), and the abundance of water sources in both regions (moist soil, vegetation canopy, bodies of water, etc.). If this is the case, the evapotranspiration dynamics (or the representation of water sources) in the NWP model may have important errors in both the coastal and valley regions, but the larger warm biases in the coastal region predictions may counteract this shortcoming (since evapotranspiration rate increases with temperature). Whether this or other factors are important will not be exhausted here. Recall that layer 1 RH biases in both regions have an upward trend after sunrise, indicating that the decreasing temperature biases, not the downward-trending q_v biases, are the dominant influence during this period. Still, further analysis is warranted, especially since the negative post-sunrise q_v bias is larger during fog missed opportunities (Figure 49, right column).

During the post-spin up overnight hours (7–15 h), q_v biases during missed opportunities are near-neutral in all regions, suggesting the warm temperature biases are the primary systematic deficiency leading to the negative RH bias and excessive zero q_c predictions.

During fog hits, q_v biases are slightly positive. This accounts for the positive RH bias exhibited during fog hits (Figure 38 center column) since temperature biases were shown to be near-neutral bias during fog hits (Figure 43 center column),

While the coast and mountain regions exhibit near-neutral biases of their q_v fields at 0 h, the initialization biases in the valley region average about -0.2 g kg^{-1} for all the data, and are 3–4 times higher when fog is observed at the initialization hour (there are no fog hits at 0 h, so the missed opportunities data represents all observed fog cases). When combined with an approximately 7 K warm bias at 0 h when fog is observed, the valley region appears to undergo an especially ponderous spin up of both the q_v and temperature fields when fog is present at the initialization hour. The magnitude of these initialization errors during 0 h fog events raises questions about the extent to which they affect the predictions throughout the run, even though the biases level off and the error variances decrease rapidly during the spin up period. At a minimum, it indicates the initialization process needs further attention if either of these fields are to be used in moist conditions without the benefit of a generous spin up period.

In summary, the layer 1 q_v predictions demonstrate minimal biases, and are not primarily responsible for the negative RH biases at any post-spin up hour. This is not to say the q_v predictions are highly accurate, as they still contain significant error. However, the error variances are comparable to or lower than those of the layer 1 temperature predictions in regard to their impact on RH. As an ensemble, the q_v predictions are underdispersive. With the possible exception of the post-sunrise period, which is characterized by insufficient moistening of the boundary layer in the valley region with relatively minor impact on RH, the NWP model exhibits no obvious systematic deficiencies regarding its q_v predictions. We may therefore reasonably conclude the first-order NWP model systematic deficiency responsible for excessive zero q_c predictions is a negative RH bias attributable to a warm temperature bias.

It is worth considering whether, in order to best assess the NWP general moisture verification, it is better to verify the entire water budget field, $q_v + q_c$, rather than each field separately as we have done, especially since we know the q_c field is significantly underforecast by the NWP model. A simple scale analysis of these components reveals that even in heavy fog with $q_c = 0.05$ g m^{-3} (corresponding to a daytime visibility of about 0.2 miles), the liquid water mixing ratio is only about 0.042 g kg^{-1}, significantly less than both the typical magnitude of the q_v biases (0.5 g kg^{-1}) and the bias-corrected σ of the error (0.7 g kg^{-1}). We may conclude q_v is a reasonable estimate of the total moisture content, and the results of the q_v verification alone are sufficient to assess the NWP model's general verification of moisture.

5. 2-Meter Temperature

In the next three sections, we break from our strategy of tracing sources of the q_c error backward through the predictions, and instead examine the impact of the WRF post-processing routine used to produce predictions at 2-m and whether it might be leveraged to increase q_c predictive skill. Slightly different metrics are used that are more suited to this task.

WRF post-processing derives the 2-m predictions of temperature and q_v from the layer 1 predictions by employing a flux-profile relationship (Stull 1988), where fluxes of heat, moisture, and momentum are provided by the PBL scheme used in the member. q_c is not included among the variables predicted, but temperature and q_v at 2 m above model ground level are, and these are examined next (along with 2-m RH). As these sub-layer 1 predictions are strictly post-processed after the WRF has completed its integrations, there is no feedback mechanism for them to affect the layer 1 predictions. Therefore, they cannot be a source of the q_c error.

Distributions of 2-m temperature predictions in the coastal region (Figure 51) show a systematic warm bias similar to that observed in the layer 1 predictions. As in layer 1, the model climatologies of every member in this region are distinctly offset to the warm side of the observed climatologies. Additionally, compared to the layer 1

predictions, the 2-m predictions in several of the members (members 1 and 16[2]) appear to be less dispersive, with a very high incidence of predictions in the range 282-285 K. The smaller dispersion is confirmed by computing the average variance of each member's predictions, which is 11.7 K^2 at layer 1, but only 7.3 K^2 at 2 m. Both of these are less than the observed temperature variance of 14.5 K^2.

The 2-m temperature distributions in the valley region (Figure 52) also appear to maintain the systematic warm bias observed at layer 1 in this region, although the bias is perhaps not as large at 2 m. The shape of the prediction distributions appears slightly less underdispersive for several of the members compared to the distributions at layer 1, but otherwise no obvious differences are evident.

Prediction distributions in the mountain region (Figure 53) have the largest variety among the members, with most members' distributions appearing to be centered a few degrees colder than the layer 1 predictions.

The 2-m temperature verification rank histograms (Figure 54) show that the stochastic predictions in the coastal region suffer from a warm bias comparable to that of the layer 1 predictions. The stochastic predictions remain underdispersive in the coastal region. In the valley region, the warm in 2-m temeprature is smaller than that of the layer 1 predictions, and the stochastic predictions are less underdispersive than the layer-1 stochastic predictions. Predictions in the mountain region are characterized by a cold bias (in contrast to a slight warm bias at layer 1). Underdispersion is slightly improved compared to layer 1.

[2] In a WRF model update notice dated 21 December 2011, primary model developers at the University Corporation for Atmospheric Research (UCAR) reported a bug affecting 2-m temperature predictions when the RUC land surface model is used in conjunction with the YSU PBL scheme. Members 15 and 17 are configured with these two schemes, and their results indeed deviate from the rest of the member predictions in certain aspects of the verification. Although a new version of WRF was released by UCAR with the bug resolved, it was too late in this work to reproduce the NWP model runs, and therefore the 2-m verification results presented in this section include output from the affected members even though their results are largely excluded from the discussion. Verification results of 2-m q_v and 2-m RH from these two members are also erratic at times, and so these are likewise excluded from discussion despite inclusion in the figures. During development and testing of the q_c adjustment techniques proposed later in this work, the members were largely excluded when 2-m predictions were involved, with exceptions to this rule noted in those chapters.

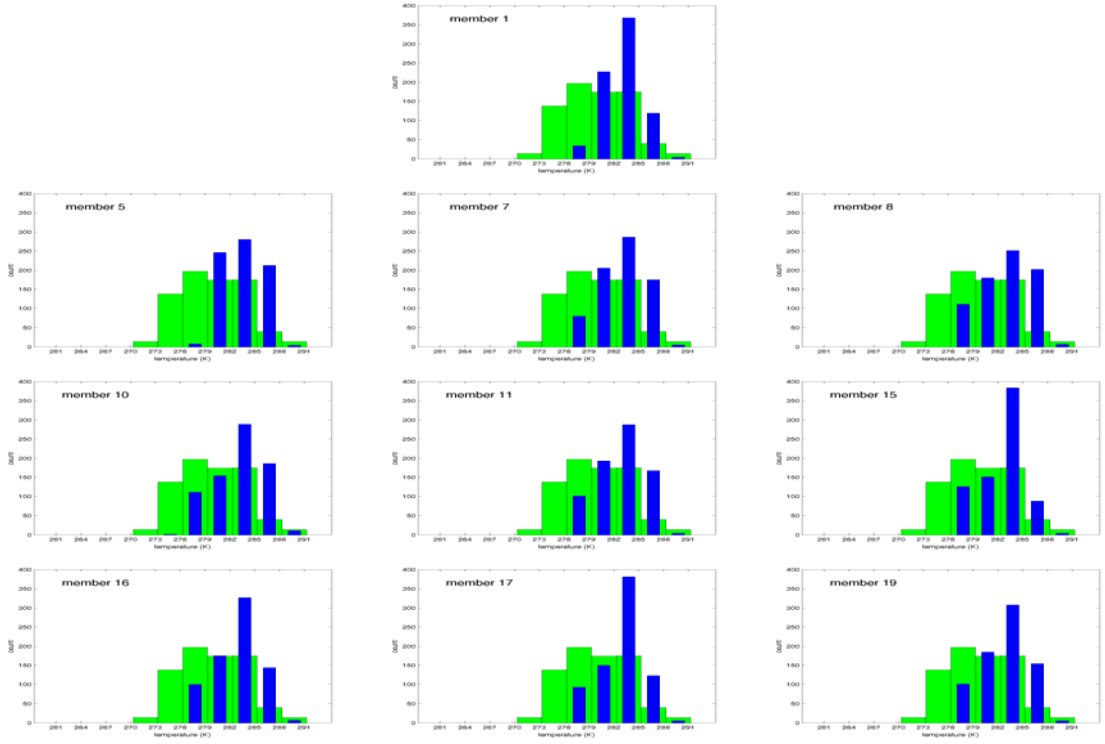

Figure 51. Histogram of distribution of NWP 2-m temperature predictions (blue bars), and observations (green bars) for coastal region. The first six hours of each case are excluded.

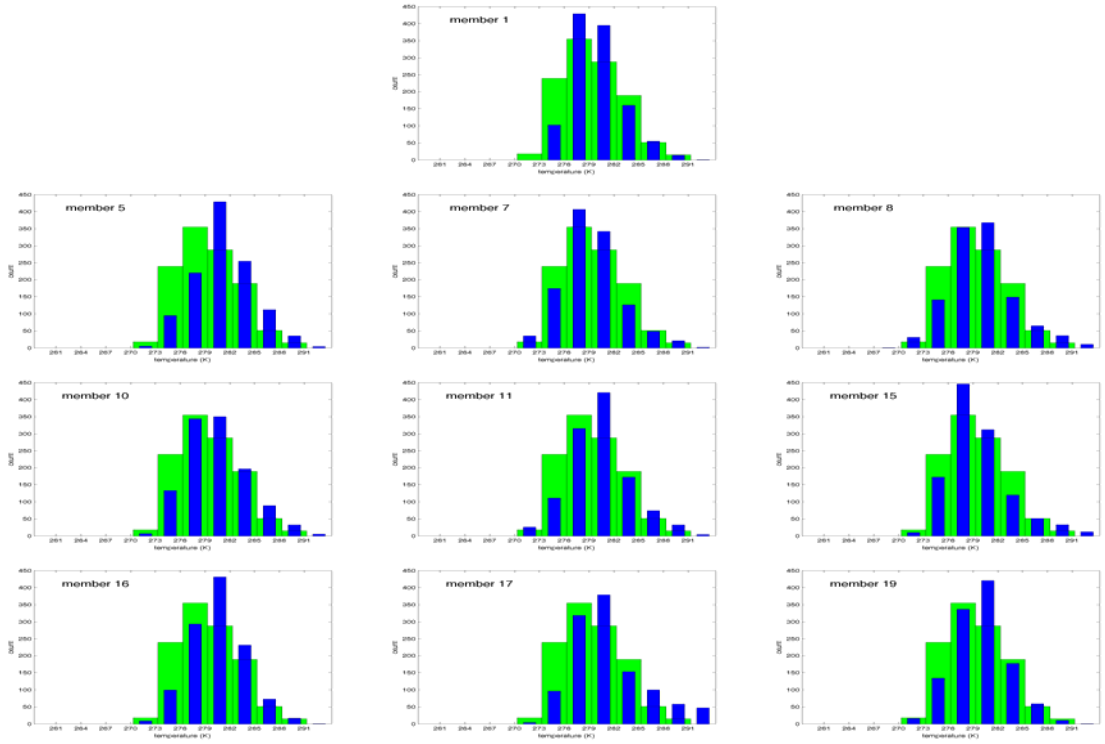

Figure 52. Same as in Figure 51, but only for the valley sites.

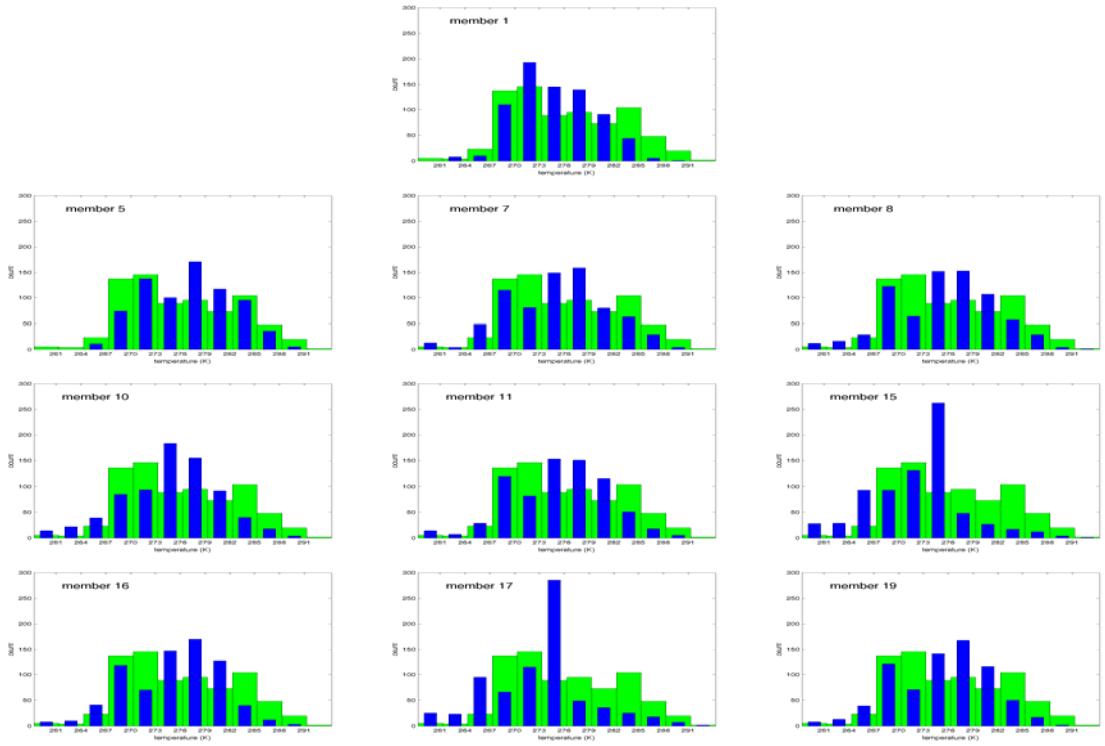

Figure 53. Same as in Figure 51, but only for the mountain sites.

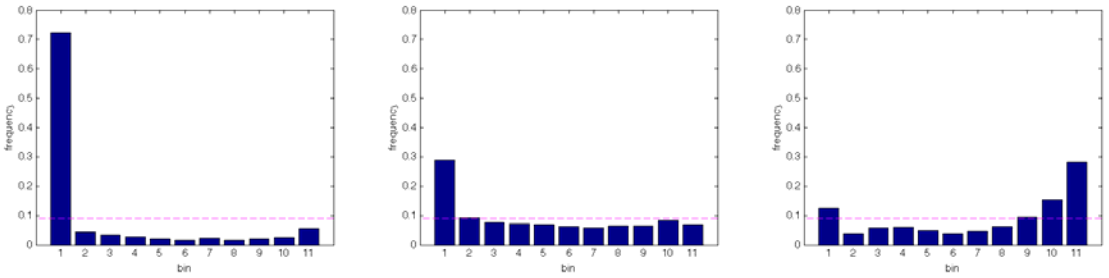

Figure 54. Verification rank histograms of 2-m temperature for the coastal region (left), valley region (center), and mountain region (right). The first six hours of each case are excluded.

The more accurate dispersion of the 2-m predictions in the valley region does not necessarily indicate less error in the predictions, but it does signify that for any given forecast, the members more thoroughly sample the uncertainty and are thereby less likely to be clustered together either above or below the verifying temperature. Quantifying the reasons for the increased dispersion – which likely include differences in each member's land surface scheme, PBL scheme, and land surface properties such as soil moisture—is not an emphasis of this work. But should these 2-m predictions be found useful to inform a q_c adjustment technique, the greater dispersion is an added benefit that translates to better dispersion in the results of the technique. Notably, the varied model physics and land surface parameters did not result in increased dispersion in the coastal region, and only slightly increased dispersion in the mountain region. This suggests the physics variations among the members in these regions are not sufficiently aggressive to sample the full physics uncertainty, or that significant sources of unsampled uncertainty exist elsewhere in the NWP model (e.g., sea surface temperature in the coastal region).

Biases and error variances as a function of hour are shown for each region in Figure 55. The two columns in the figure represent results from all data (left column), and fog missed opportunities (right column). Verification during fog hits is excluded here because our emphasis is no longer on tracing the source of the q_c prediction deficiency, but instead to simply assess the potential to use the 2-m data to inform our q_c adjustment technique. Since the fog hits would not be affected by this technique and the 2-m predictions have no effect on layer 1 predictions, the results during fog hits are not relevant.

The 2-m temperature predictions in the coastal region have biases similar to the biases at layer 1 at all hours, including the decreasing bias after sunrise indicative of insufficient warming in the predictions. At 2 m, the biases for the fog missed opportunities average ~1 K warmer than the biases for all the data, which is nearly similar to the bias differences at layer 1. However, note that biases at 2-m are more consistent among the members (especially during fog missed opportunities), which is significant since any potential bias correction would not be member-specific. Additionally, error variances are lower for the 2-m predictions, an indication of higher predictive skill than the layer 1 predictions.

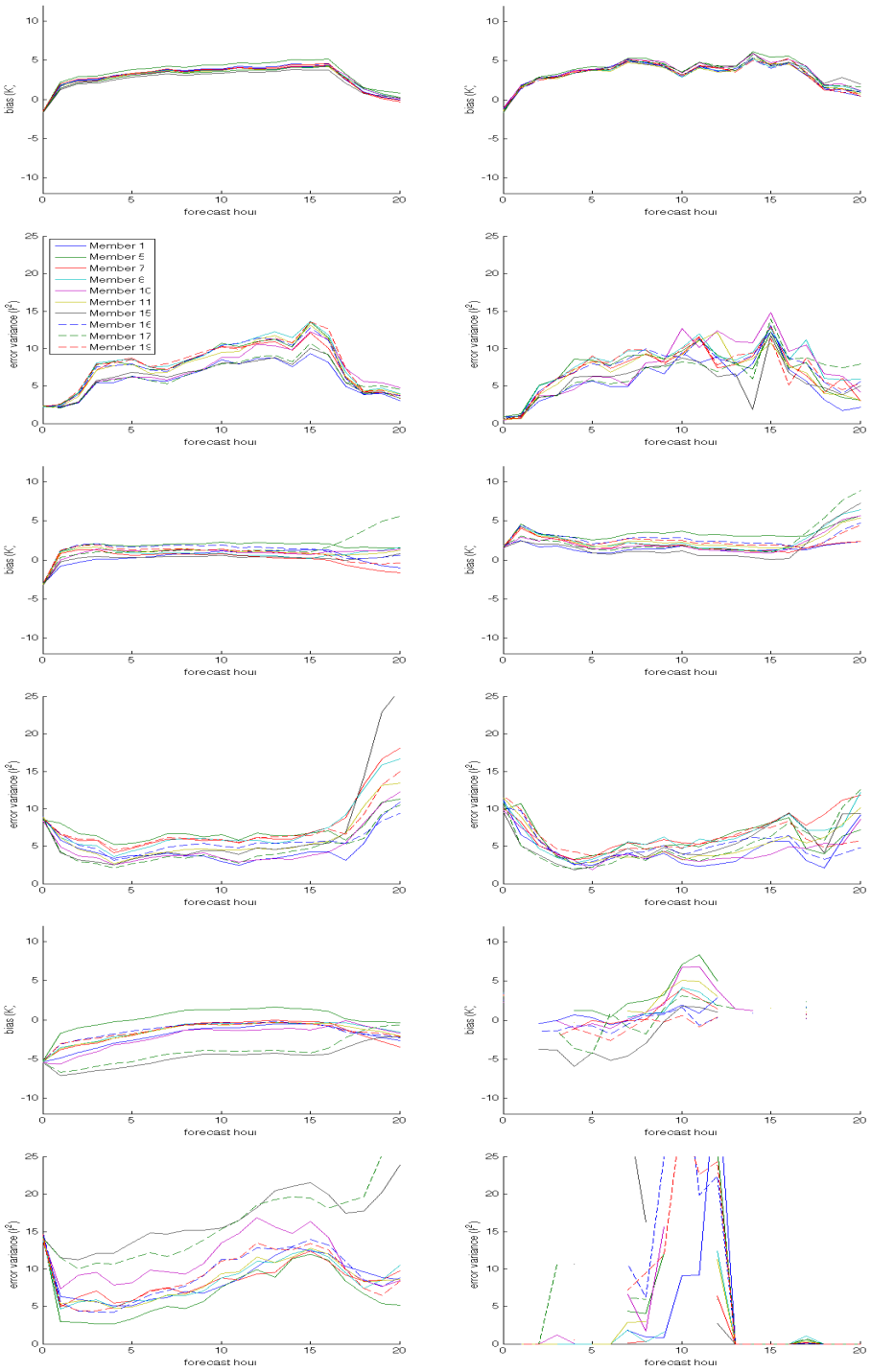

Figure 55. 2-m temperature bias and error variance of each member for coastal (top two rows), valley (center two rows), and mountain (bottom two rows) regions. The left column shows all data, and the right column includes only fog missed opportunities.

At 2-m, the warm biases in the valley region are reduced to <1 K in most members. In contrast to the coastal region predictions, diurnal changes in the 2-m temperature biases are practically eliminated in the valley region predictions at 2 m, indicating the post-sunrise heating in the predictions matches the magnitude of the observed heating. The error variances of the 2-m predictions in this region are also generally lower than at layer 1 until sunrise, after which they are comparable at both levels. In all, the 2-m temperature predictions appear to be more skillful and (based on the verification rank histogram) more accurately dispersed.

Additionally, the valley region 2-m predictions do not show the very warm-biased initialization and large error variances of up to 25 K^2 during spin up that were indicated in the layer 1 predictions when fog is present at initialization. However, overnight warm biases of 2–3 K are still present during fog missed opportunities and become worse after sunrise, raising questions about the prospects of an effective bias correction.

In the mountain region, a near-neutral bias was observed in the layer 1 temperature predictions. However, at 2 m, a cold bias exists for nearly all members at all hours after being initialized with a 5 K cold bias. Error variances at 2-m are comparable to those at layer 1 for most members. The inter-member variability of error variances is larger for the 2-m predictions, a characteristic only observed in this region, suggesting that certain physics suites (such as those used by members 5 and 7) perform significantly better in this region than others (those used by members 10 and 16). In contrast to the other regions, the 2-m temperature predictions do not appear to offer better predictive skill than the layer 1 predictions, and indeed appear less skillful.

Similar biases in the coastal region layer 1 and 2-m temperature predictions suggests that, if the systematic warm bias of the NWP model is caused by unresolved inversions below layer 1, the WRF post-processing does not adequately reveal them in this region. As the region is characterized by a mix of radiation and advection fog, it is unlikely a 3–5 K cold bias in the layer 1 predictions can be explained solely by shallow inversions not at least partly revealed during post-processing for the 2-m temperature predictions. More likely, there is a systematic warm bias at layer 1 itself, worse during the nighttime, that causes a systematic warm bias in the 2-m predictions as well.

In contrast, the layer 1 warm bias in the valley region is reduced by the post-processing of 2-m temperatures. If we assume the post-processing is at least somewhat skillful at detecting temperature trends in the first few meters above the ground (which the low error variances suggest is the case), then the bias improvement at the 2-m level indicates unresolved inversions below layer 1 are a contributing factor to the systematic layer 1 warm bias.

6. 2-Meter Water Vapor

2-m q_v predictions are examined next to evaluate their potential to be used to inform q_c adjustments. Distributions of each member's 2-m q_v predictions in each region are shown in Figures 56–58. In the coastal region, the predictions exhibit a moist bias in every member in contrast to the near-neutral biases in the layer 1 q_v predictions in this region. The valley region distributions show no noticeable systematic bias, and in fact each member's distribution has only minor differences from its distribution of layer 1 q_v predictions. Predictions in the mountain region, where each member had a near-neutral bias at layer 1, exhibit more variability among the members than the other regions. Some members (e.g., members 7 and 8) maintain a similar distribution at both levels and a near-neutral bias, while others (members 16 and 19) show model climatologies with a moist bias at 2-m that was not present at level 1. None of the members have a noticeably drier distribution at 2-m than at layer 1.

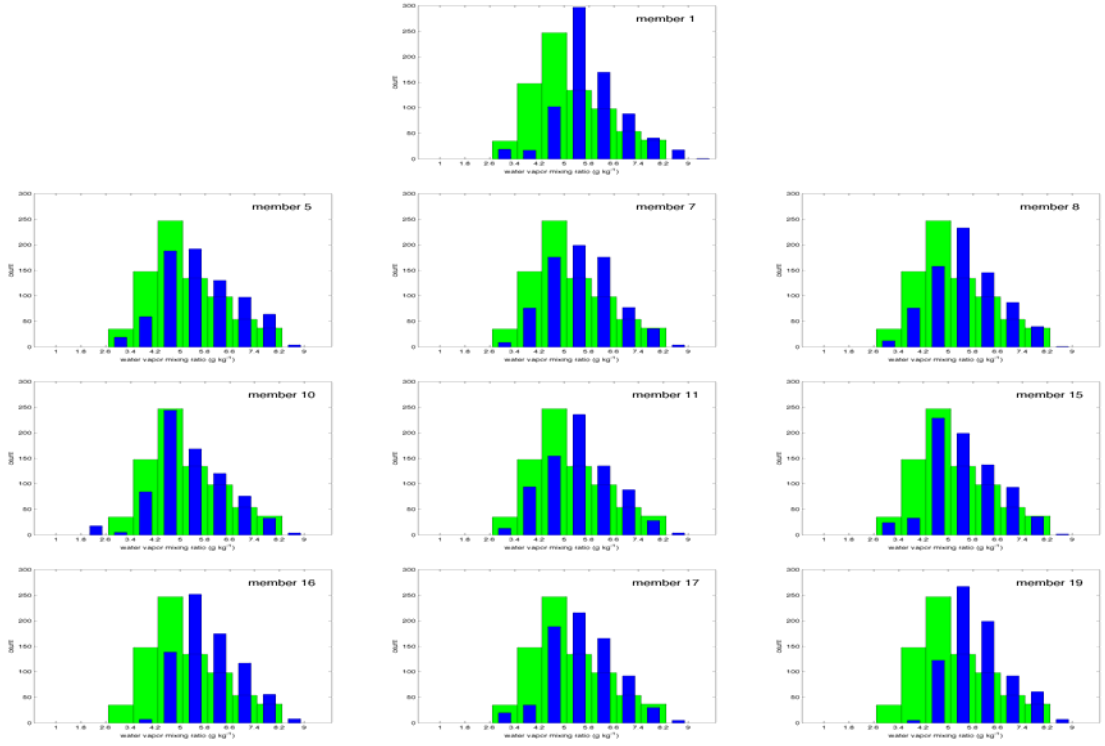

Figure 56. Histogram of distribution of NWP 2-m q_v predictions (blue bars), and observations (green bars) for coastal region. The first six hours of each case are excluded.

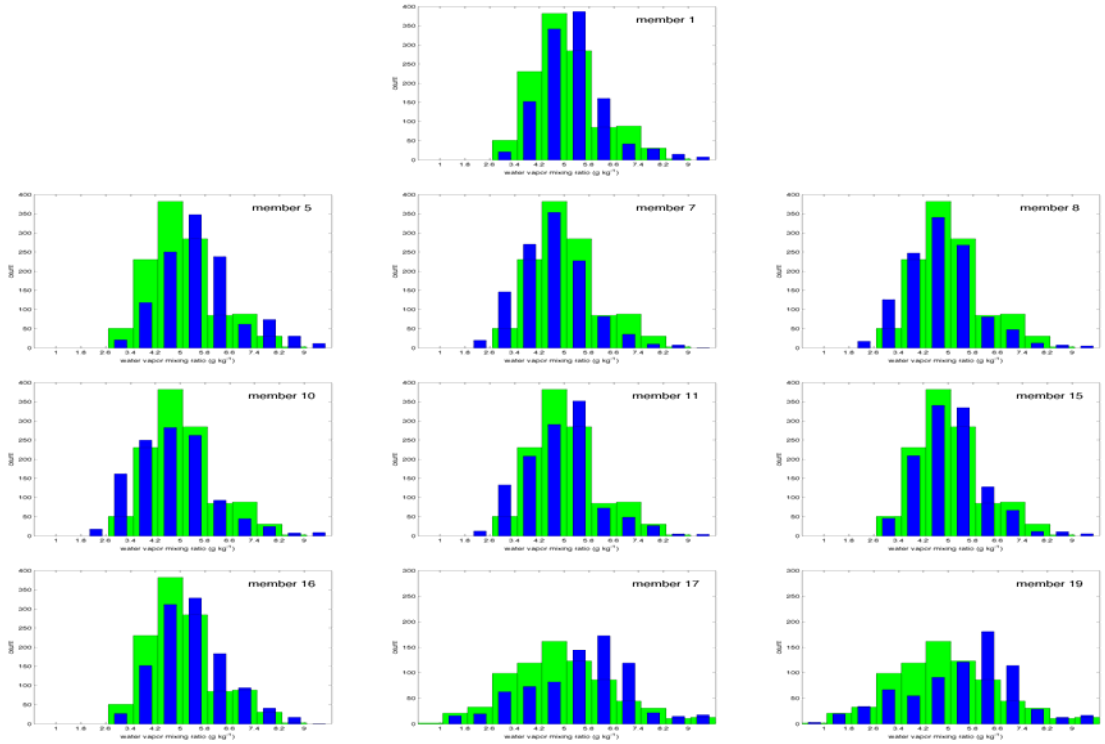

Figure 57. Same as in Figure 56, but only for the valley sites.

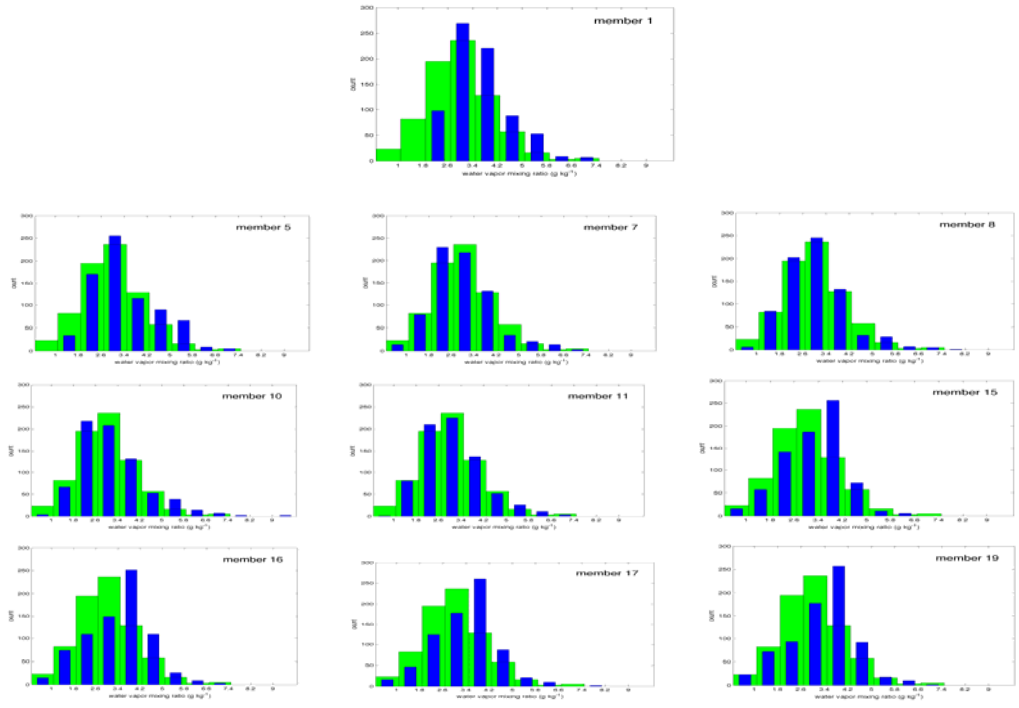

Figure 58. Same as in Figure 56, but only for the moutnain sites.

Stochastic biases of layer 1 q_v were shown to be minimal in each region, but Figure 59 indicates the ensemble is distinctly too moist in each region at 2-m. The moist bias is largest in the coastal region, and smallest in the valley region. The verification rank histograms also indicate the valley and mountain region stochastic predictions are less underdispersive than the layer 1 q_v predictions. As with the 2-m temperature predictions, the dispersion of the 2-m q_v predictions is likely aided by the multi-physics and multi-land surface properties (including soil moisture) used in the ensemble. In the case of the moisture field, the effect is evident not only in the valley region, but especially in the mountain region, which shows accurate dispersion (i.e., after correcting for bias, the uncertainty in the prediction is fully-sampled by the members). The dispersion condition at 2-m in the coastal region is difficult to determine in Figure 59 due to the large moist bias, but the layer 1 q_v predictions were more dispersive in this region than in the others.

The biases and error variances of the 2-m q_v predictions (Figure 60) show that the moist bias in the coastal region is present in every member during the overnight hours, in contrast to the near-neutral overnight biases in most members in layer 1. However, compared to the biases in layer 1, the 2-m biases are more consistent throughout the forecast period for each individual member. Even after sunrise, when the biases decrease in both layers due to insufficient moistening of the boundary layer, the decrease is not as large in the 2-m predictions. Between 10–16 h, the 2-m biases during fog missed opportunities in the coastal region are roughly the same as the biases for all the data. After sunrise the bias decreases are larger during missed opportunities, indicating the NWP model particularly struggles to moisten the boundary layer during this period when fog is present. This characteristic of the predictions was also observed at layer 1.

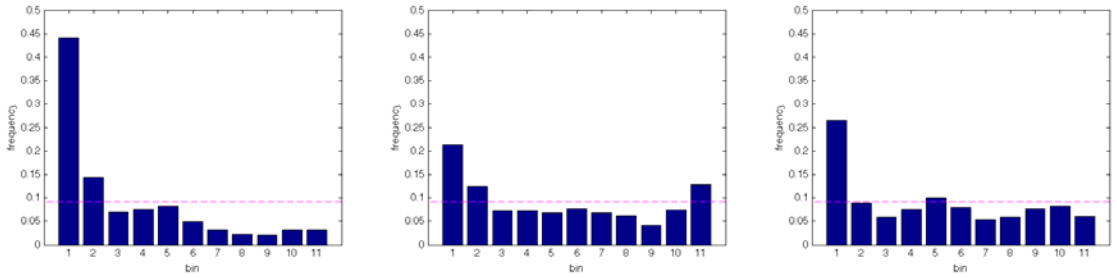

Figure 59. Verification rank histograms of 2-m q_v for the coastal region (left), valley region (center), and mountain region (right). The first six hours of each case are excluded.

Error variances are generally lower at 2-m than at layer 1, with less inter-member variability, indicating none of the physics suites is particularly better or worse at predicting moisture changes once they are corrected for bias. Overall, the layer 1 q_v predictions are more reliable than the 2-m predictions due to their near-neutral biases. However, with an appropriate bias correction, the 2-m predictions might actually be more useful to inform a q_c adjustment due to their lower error variances and the consistent nature of the 2-m biases.

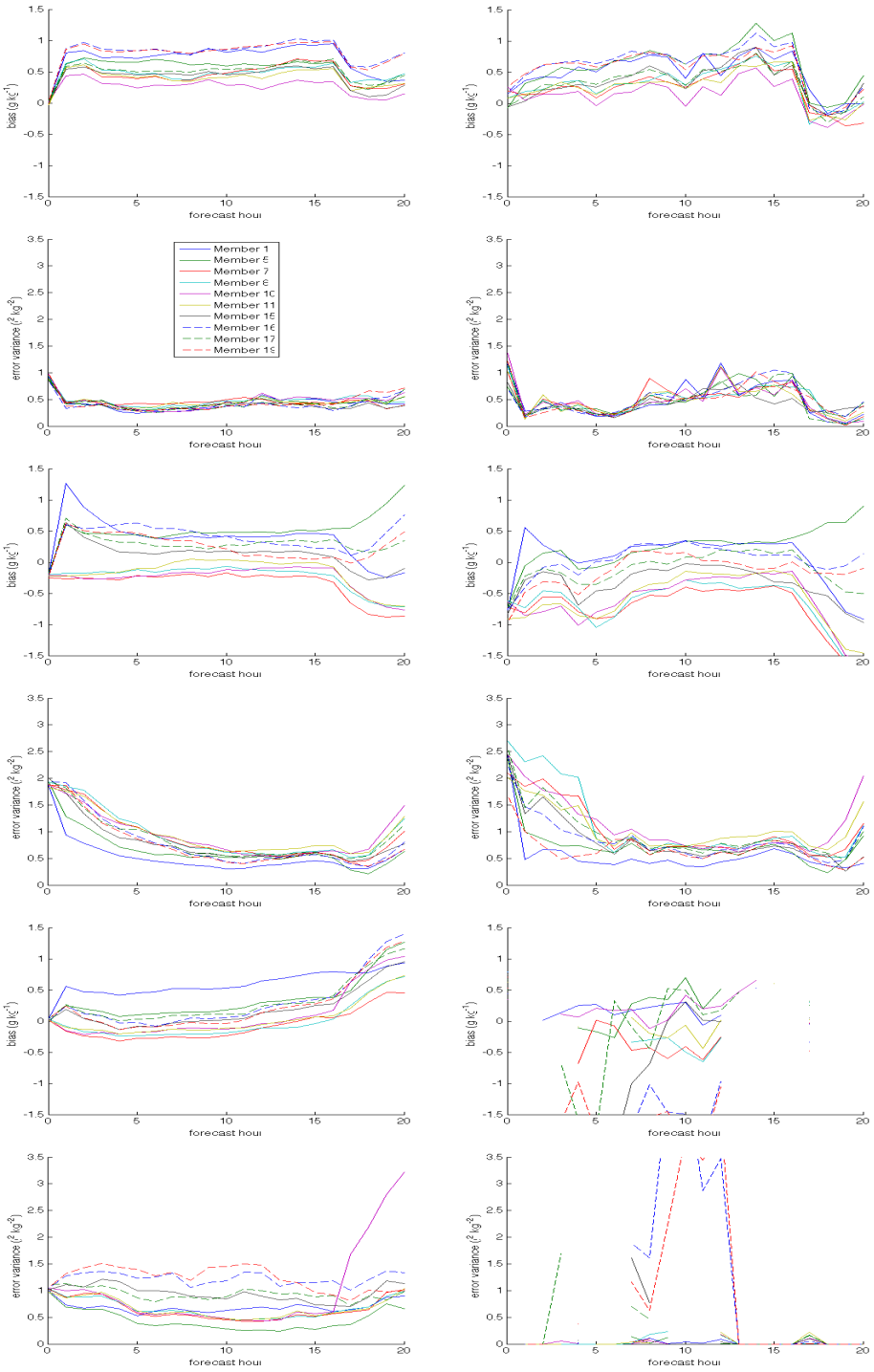

Figure 60. 2-m q_v bias and error variance of each member for coastal (top two rows), valley (center two rows), and mountain (bottom two rows) regions. The left column shows all data, and the right column includes only fog missed opportunities.

When all the data is included, overnight biases in the valley region vary from -0.2 to 0.5 g kg^{-1} for the members, with a slightly moist bias on average. Between 10–15 h, the biases during fog missed opportunities are 0.1–0.2 g kg^{-1} lower than the biases for all the data, indicating a bias correction on the 2-m predictions in this region would still leave a slight dry bias during fog missed opportunities. This was also the case with the layer 1 q_v predictions. With saturated air at 1000 mb and a temperature of 278 K, a q_v bias of -0.2 g kg^{-1} results in a predicted RH of 0.964, so the impact of this deficiency is relatively minimal, especially considering the bias for some members would be even smaller in magnitude than this. The error variances for the 2-m q_v predictions are comparable or slightly less than at layer 1, and the dry bias that averages -0.7 g kg^{-1} at initialization when fog is present is still evident in the 2-m predictions.

Overall, the 2-m q_v predictions do not appear markedly more useful than the layer 1 predictions as far as being leveraged to inform a q_c adjustment in the valley region, with perhaps the largest advantage relating to the increased dispersion of the ensemble suite at 2-m. Incidentally, the greater dispersion is likely due in part to the wider variety of biases among the members at 2-m, which is not normally a desirable way to achieve dispersion (because it does not represent sampling of the true uncertainty in the prediction) and would actually be eliminated during a traditional member-specific bias correction. Since no member-specific procedure will be pursued here, it remains to be seen whether the variety of uncorrected biases among the members negates the added benefit of slightly lower error variances at 2 m. This question will be explored in subsequent chapters.

2-m q_v dispersion in the mountain region also appears to benefit from a wide variety of biases among the members. Greater dispersion notwithstanding, the predictions appear to offer little added value over the layer 1 predictions, with moist and inconsistent biases at 2 m, and error variances that are larger than at layer 1.

7. 2-Meter Relative Humidity

The final predicted 2-m variable we will examine for its potential to be leveraged to improve the q_c predictions is RH. Of particular interest are the error characteristics of the 2-m RH predictions compared to the layer 1 RH predictions, as either could

potentially be used as a proxy for fog to adjust the zero q_c forecasts. 2-m RH is computed using the 2-m predictions of temperature and q_v from the WRF post-processing procedure. Unlike in layer 1, there is no microphysics scheme used at this level, and so all members are able to supersaturate without bounds (similarity theory treats temperature and moisture profiles as independent variables, making no concessions for saturation). To prevent this from skewing the biases and error variances presented here, 2-m RH predictions exceeding 1 were reassigned a value of 1 prior to plotting in Figure 65.

The distributions of coastal RH predictions (Figure 61) show a systematic negative bias, which is also evident in the stochastic predictions as shown in the verification rank histogram (Figure 64) for this region. Since the 2-m q_v predictions have a moist bias in this region, we can conclude that the 2-m warm bias is the dominating deficiency leading to the negative 2-m RH bias. Figure 65 shows that the negative RH biases and error variances are smaller at 2-m than they are at layer 1 (Figure 38). Biases during fog missed opportunities average 0.049 lower than the total bias, which is not ideal but is less than the 0.081 average discrepancy in the layer 1 predictions. Overall, the 2-m RH predictions appear slightly better suited than the layer 1 RH predictions to help identify zero q_c predictions likely to be fog missed opportunities.

In the valley region, the distribution of 2-m RH predictions (Figure 62) has similar characteristics as the layer 1 RH predictions. The predictions from most members remain bimodal, with excessive predictions of RH <0.7, and insufficient predictions of RH from 0.82–0.94, which account for 65.7% of the observations (many of which include fog). However, to varying degrees, the members also have excessive predictions with an RH >0.94, which was not present in the layer 1 RH distributions. The biases of individual members is smaller than at layer 1, with some members having a positive bias (Figure 65). The average bias of all members at all post spin-up hours is -0.032, which is reflected as a small negative stochastic bias in the verification rank histogram (Figure 64). With the absence of any microphysics schemes, the inconsistent distributions among the members near saturation is no longer evident as it was at layer 1.

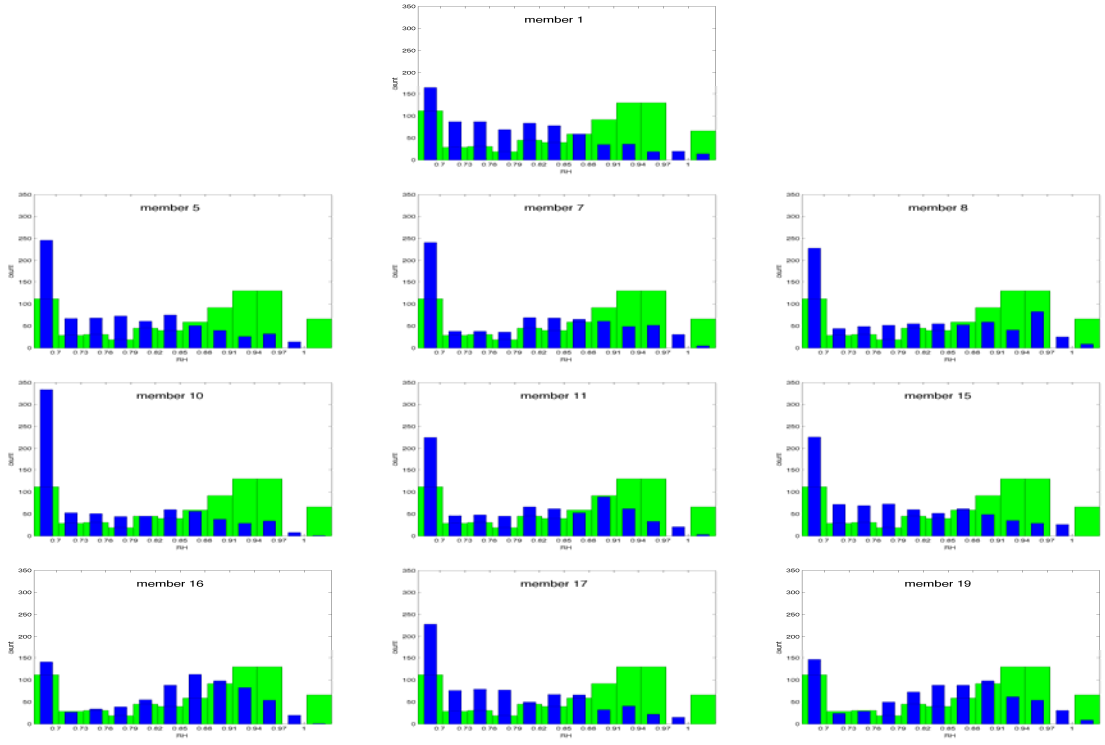

Figure 61. Histogram of distribution of NWP 2-m RH predictions (blue bars), and observations (green bars) for coastal region. The first six hours of each case are excluded.

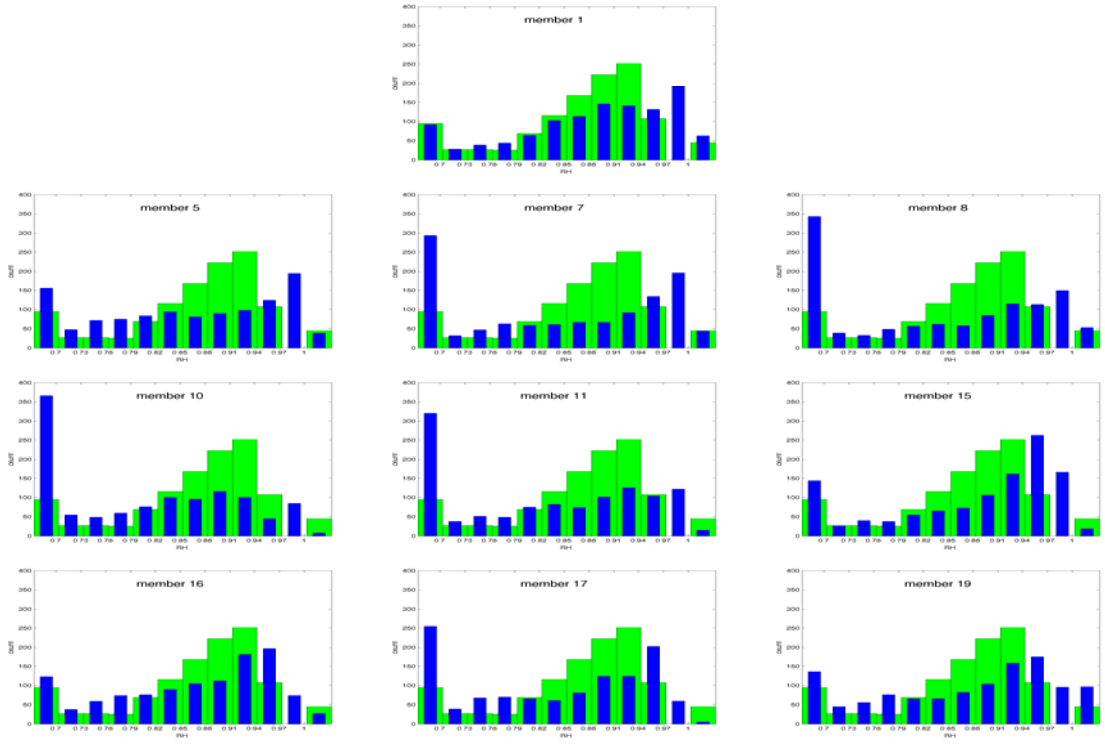

Figure 62. Same as in Figure 61, but only for the valley sites.

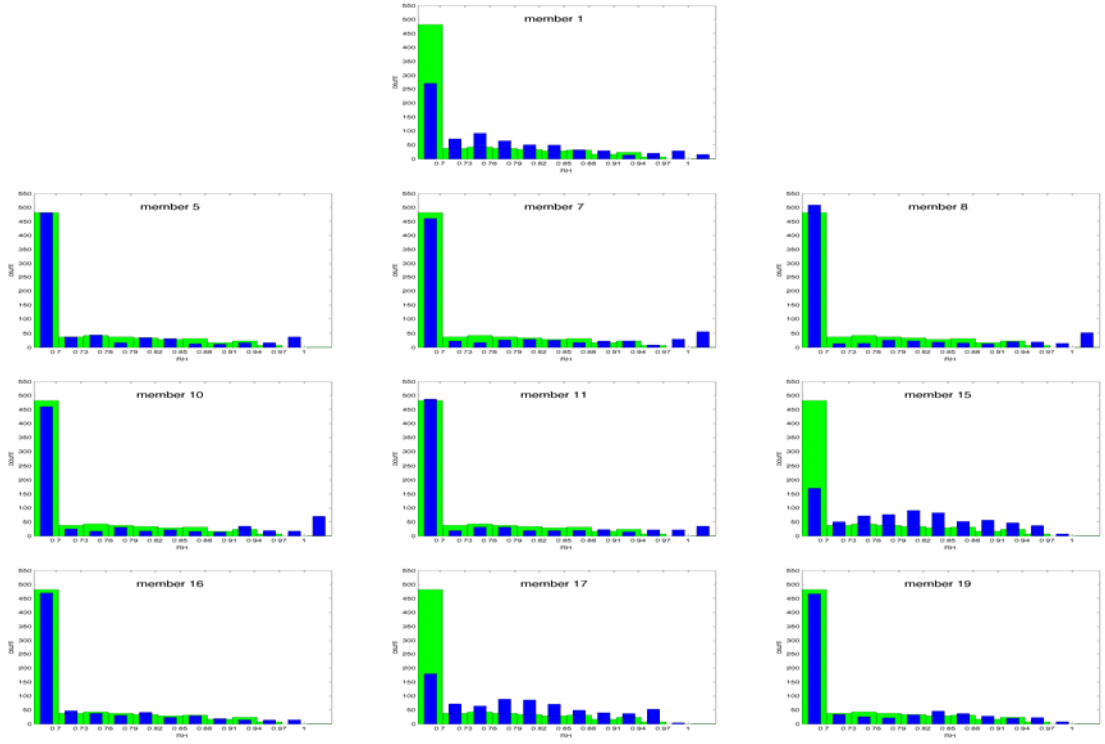

Figure 63. Same as in Figure 61, but only for the valley sites.

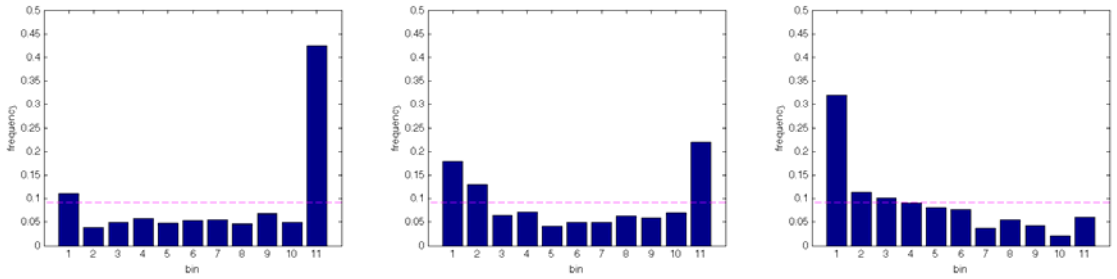

Figure 64. Verification rank histograms of 2-m RH for the coastal region (left), valley region (center), and mountain region (right). The first six hours of each case are excluded.

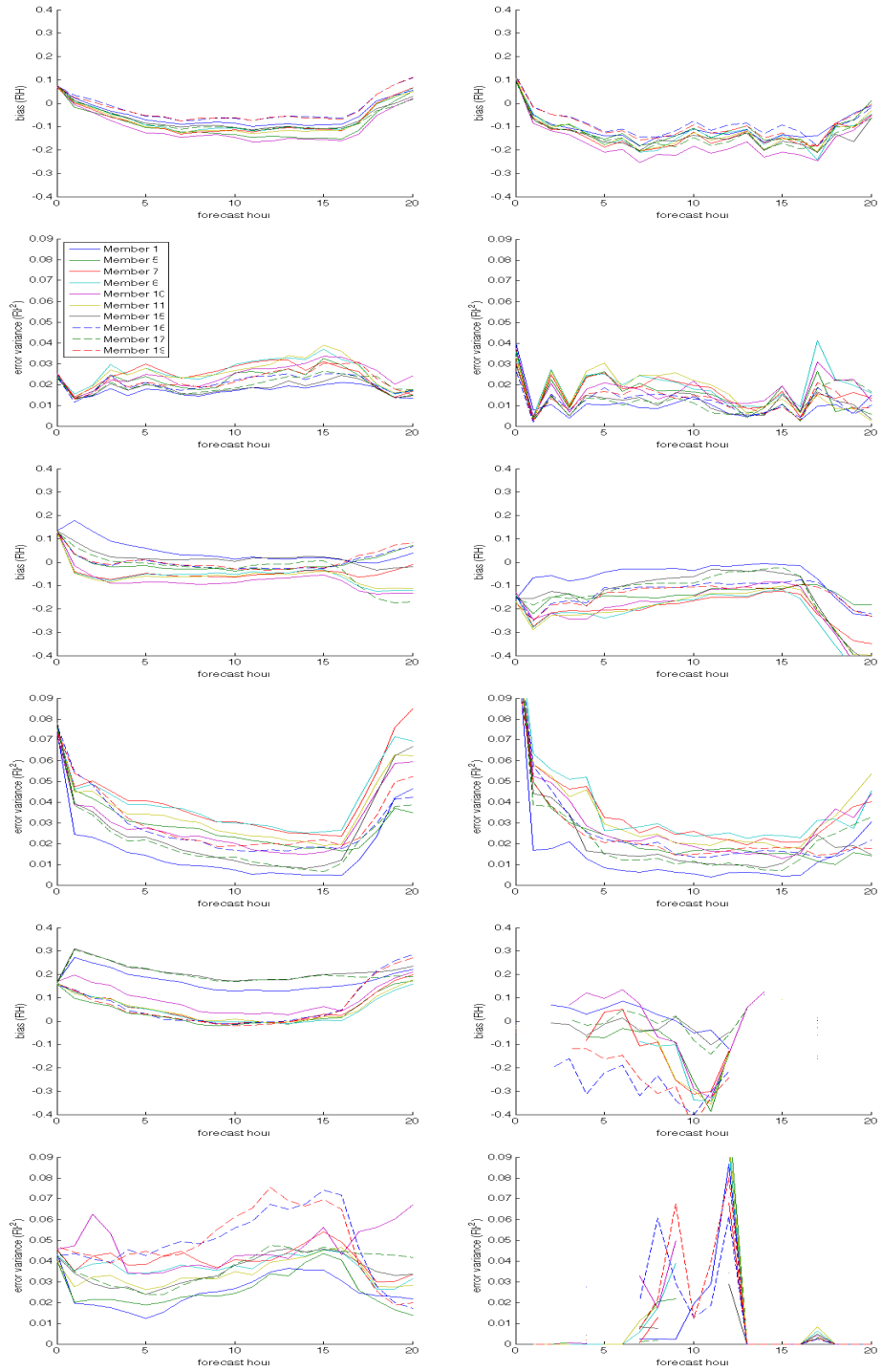

Figure 65. 2-m RH bias and error variance of each member for coastal (top two rows), valley (center two rows), and mountain (bottom two rows) regions. The left column shows all data, and the right column includes only fog missed opportunities.

The near-neutral biases by most members is explained by offsetting warm temperature and moist q_v biases at 2-m, which makes the RH bias somewhat tenuous since it would be made worse by correcting for only one of the biases in either one of the components of RH. Furthermore, the 2-m temperature and 2-m q_v biases were shown to be substantially greater during fog missed opportunities, raising doubts about the value of a bias correction in either variable in this region. Not surprisingly, the 2-m RH biases during fog missed opportunities are lower by 0.1–0.3 than the biases for all data, similar to the discrepancy seen in the layer 1 RH predictions. The lower error variances and slightly better dispersion of the 2-m RH predictions suggests they are perhaps more useful as is than the layer 1 RH predictions to inform a q_c adjustment technique, but the shortcomings in RH predictions at both levels makes it unlikely that using RH alone as a proxy for fog could be as successful as it might be in the coastal region.

2-m RH predictions in the mountain region are characterized by a positive bias attributed to the moist q_v bias in the 2-m predictions. Error variances are generally larger than in the layer 1 RH predictions, and vary substantially by member consistent with the 2-m temperature and 2-m q_v predictions in this region.

V. NWP POST-PROCESSING

In the previous chapter, we examined the performance and error characteristics of NWP model q_c predictions, as well as predictions of the primary thermodynamic variables at the layer 1 and 2-m levels. This revealed that a layer 1 negative RH bias is largely responsible for the lack of q_c predictions in the coastal and valley regions, which is mostly due to a layer 1 warm bias that is strongest overnight. It also revealed that certain aspects of the predictions have no obvious systematic error and are relatively accurate, such as the 2-m q_v predictions in the valley region.

In this chapter, we develop several potential approaches to leverage the most useful aspects of the predictions to skillfully predict the probability of fog when the NWP model does not do so on its own (thereby mitigating the primary NWP model deficiency of insufficient fog predictions). The most basic of these approaches are informed by the most obvious error characteristics revealed in Chapter IV; for example, applying a temperature bias correction. In addition, predictions of some of these variables will be shown to have less obvious predictive usefulness for fog, and these are pursued and explained as well. The viability of each approach is tested using a form of "leave one out" cross-validation.

All nine of the NWP post-processing approaches developed and presented in this chapter are aimed at making skillful upward adjustments to zero or near-zero q_c predictions. Since our goal is to mitigate the impact of NWP systematic deficiencies rather than perform a member-specific calibration, the techniques are not tailored for individual members. Regardless of the β_e threshold being used for verification, the subset of predictions subject to post-processing does not change; it is only those with a q_c prediction below the *lowest* β_e threshold (0.29 km^{-1}).

Furthermore, to reduce complexity, the probabilisitic post-processing techniques described here are designed to directly provide a stochastic β_e prediction rather than an adjusted q_c prediction, a strategy that considers the combined effects of NWP prediction error and visibility parameterization error, but also renders them indistinguishable. Although the end result is largely the same, estimating the errors separately would better

facilitate an understanding of them, as well as help to develop future strategies for managing them. Relatively little is known about visibility parameterization uncertainty in light fog, and so it is left for future research to more fully explore this specific source of error. Doing so might entail NWP output that includes additional variables vital to the relationship between q_c and β_e (such as N, droplet size distribution, etc.), a more refined parametric visibility parameterization designed for light fog conditions, and/or observations of q_c against which to verify.

An ideal VIF post-processing technique intended for operational use is able to be applied indiscriminately in any of the three region categories (i.e., across an entire NWP model domain), bypassing the need to pre-define region categories within the NWP model domain, which can be time-consuming and rather arbitrary (e.g., the geographical transition from a valley region to a mountain region is typically gradual, with further research needed to understand the nature of the NWP model error in these transition zones). Therefore, each technique is first developed with optimization for an *all regions domain*; that is, with no region specificity. Subsequently, most of the techniques are re-optimized for three additional domains made up of individual regions or region combinations, which leverages the unique severity of the systematic NWP error in each domain and/or the aspects of the predictions with the most predictive skill (e.g., error variances of the 2-m temperature and q_v predictions are lower than at layer 1 in the coastal region, but significantly higher than at layer 1 in the mountain region). In addition to the all regions domain, the three additional domains for which optimization is performed are a coastal-only domain, a valley-only domain, and a combined valley/mountain domain. This work does not develop a post-processing technique for a mountain-only domain since VIF prediction skill from the NWP model is already comparatively high and is not likely to be aided by upward q_c adjustments alone. A combined coastal/mountain domain is also excluded due to the relatively fewer locales where these two regions exist absent some semblance of an intervening valley region.

The domain-specific optimizations are intended for applications such as small NWP model domains with little geographical variation, or point forecasts for which the domain category can be appropriately defined. Significant consideration and discussion

are given toward maintaining merely domain-specific optimization (with the intent that the techniques are transferable to other like domains) as opposed to approaching a site-specific optimization.

The techniques described in this chapter are presented in order of increasing sophistication, culminating in the use of joint parameter space of the NWP output to adjust the low q_c predictions, which is generally shown to be most effective and to which we devote the majority of the discussion. The techniques presented and tested before it are intended to document the viability of a variety of post-processing strategies, as well as serve as foundational building blocks for the joint parameter space techniques.

Following a description of the post-processing techniques (which are summarized in Table 6) is an explanation of the cross-validation method used to test them. Chapter VI discusses the testing results.

Table 6. Summary of post-processing techniques tested, with symbols used in later figures. All the techniques are first developed and tested without regional specificity, and some are then refined for specific regions or region combinations, which are listed.

Symbol	Name	Description	Optimization Domains
	Cntrl	Unaltered NWP predictions	N/A
	SCW	Small, non-zero cloud water values	All regions
	RH_D	RH threshold, deterministic	All regions, coast, valley, valley/mountain
	BiasRH_D	RH threshold with 2-m temperature bias correction, deterministic	All regions, coast, valley, valley/mountain
	RH_P	RH, probabilistic	All regions, coast, valley, valley/mountain
	BiasRH_P	RH with 2-m temperature bias correction, probabilistic	All regions, coast, valley, valley/mountain
	JP_B	Joint parameter space, best overall	All regions, coast, valley, valley/mountain
	JP_LB	Joint parameter space, large bins	All regions
	JP_SB	Joint parameter space, small bins	All regions
	JP_U	Joint parameter space, best universal	All regions, coast, valley/mountain
Line Type Used in Results to Denote Domain Optimization			
		All regions domain	
		Individual coast or valley domain	
		Combined valley/mountain domain	

A. POST-PROCESSING TECHNIQUES

1. Small, Non-Zero Cloud Water Values

SCW tests whether small, non-zero q_c predictions that are below the lowest verification threshold of 8.5×10^{-4} g m^{-3} represent a skillful fog indicator, or whether they are unskillful noise that should be treated as a zero q_c forecast and therefore be subject to post-processing in the remaining experiments. Assessment of the NWP predictions in Chapter IV revealed a surplus of zero q_c predictions compared to observations, but also a

significant incidence of these small, non-zero q_c predictions from members 16 and 17 (Figures 18 and 19). *SCW* is performed by deterministically adjusting these forecasts upward beyond each of the four verification thresholds. The adjustment is made to any member whose q_c prediction falls in the range $0 < q_c < 8.5 \times 10^{-4}$ g m^{-3}, although the vast majority of the affected predictions are from members 16 and 17. Rarely are more than two members affected by *SWC* at any given hour, so when the technique is invoked, the upward adjustment to the ensemble probabilistic forecast is a 10–20% increment in the probability of event forecast in almost all cases.

The results of *SCW*, which are fully presented in the next chapter, will suggest that the small, non-zero q_c forecasts are not random events but they also do not add appreciable skill improvement at any verification threshold. For this latter reason, these predictions were treated as zero q_c forecasts in the remaining post-processing experiments, and were subject to upward adjustments accordingly.

2. RH Threshold, Deterministic

RH_D tests the prospect of using an RH prediction threshold as a proxy for fog. In this technique, each zero q_c forecast is deterministically adjusted upward beyond the fog verification threshold if the member's RH forecast exceeds a fixed value. 2-m RH predictions are used instead of layer 1 RH prediction due to their lower error variances in the coastal and valley regions, total biases that better match the missed opportunity biases in the coastal region, and larger dispersion in the valley region. The 2-m RH predictions were found to have larger error variances and biases than the layer 1 predictions in the mountain region.

The optimal RH thresholds are determined by using the receiver operating characteristics (ROC curve) shown in Figure 66. The plots show the false positive rate and POD achieved by using various RH thresholds as a proxy for fog at the lowest β_e threshold (ROC curves and optimal thresholds are similar at the three other β_e thresholds used for verification, and are not shown). The plots were generated using only instances

when the member did not predict fog[3], and also excludes the first six hours of each case. The optimal RH threshold is one with a low false positive rate and a high POD such that it is furthest from the diagonal green line toward the upper-left corner of the plot. These are annotated in each plot with a large red marker.

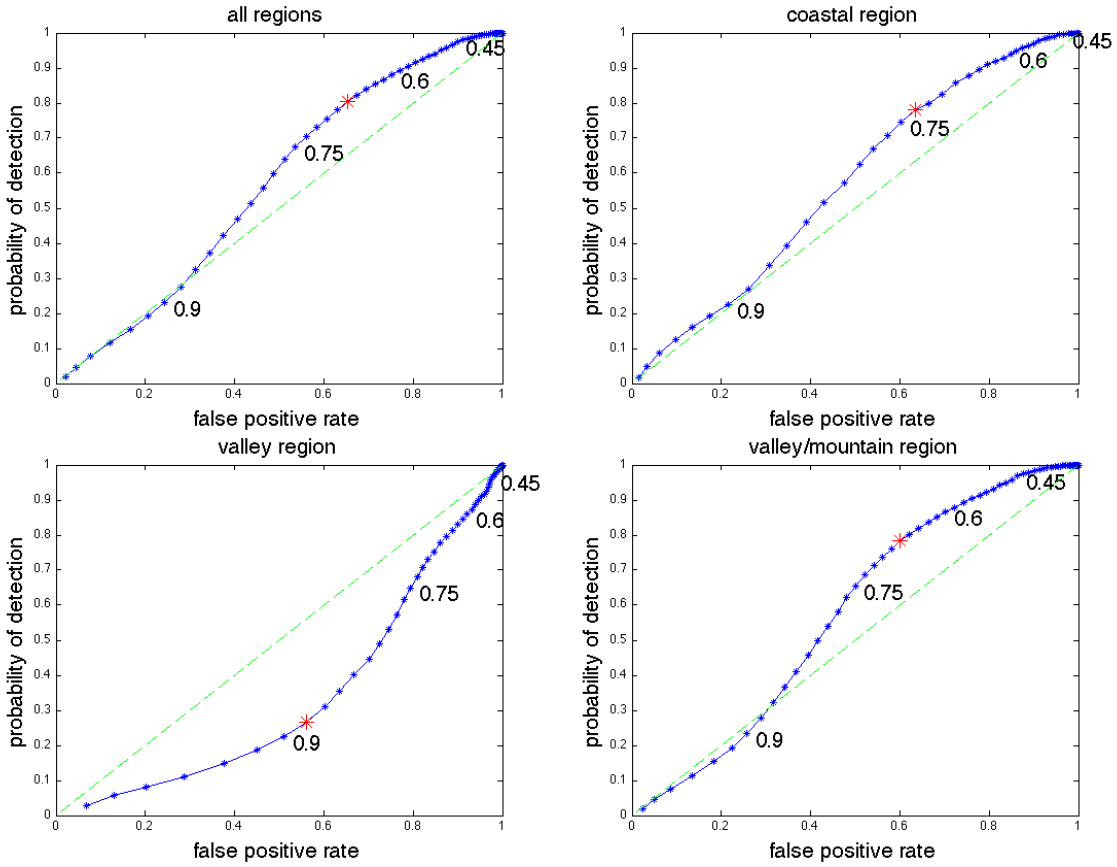

Figure 66. Receiver operation characteristics (ROC curve) for various 2-m RH prediction thresholds as a classifier for observed fog in each of the four domains. The optimal threshold is indicated with a large red marker. The data only includes cases when the member did not predict fog. The first six hours of each case are excluded.

The optimal threshold in the coastal domain is shown to be 0.735, significantly lower than saturation due partly to the negative RH bias exhibited in this region.

The data in the valley region indicates nearly all thresholds produce results to the lower-right of the green line, where they are less accurate as a fog classifier than random

[3] Members 15 are 17 are excluded from this technique's development and testing, as they are with all techniques involving 2-m predictions.

guessing because the false positive rate exceeds the POD. However, if we instead consider these unskillful thresholds as a classifiers for no fog—such that predictions below the thresholds are deterministic adjusted to a fog prediction—the false positive rate (the rate at which the event is predicted among the times it does not occur) and the POD (the rate at which the event is predicted among the times is does occur) are reversed. Graphically, this has the effect of the plotted thresholds being reflected about the green line, resulting in some measure of accuracy for these thresholds as fog classifiers (or more appropriately, as fog reverse classifiers).

If we consider all the RH thresholds plotted in the valley region ROC to be fog reverse classifiers, and therefore reflect the plotted points about the green line, an RH threshold of 0.885 would be furthest from the green line toward the upper-left of the plot and thus provide the most skill. Physically, this is counterintuitive, as it means RH predictions *below* 0.885 are more likely to be observed fog cases than predictions above this threshold. In this case, the reason for the results being reversed has to do with a warm bias that appears to preferentially exist when conditions are more favorable for fog, thus yielding erroneously low RH values during many fog cases. This unique characteristic of the 2-m RH classifier in the valley region will appear in later experiments and be explored further, but for *RH_D,* the threshold of 0.885 is applied as a reverse classifier for fog, and the results are tested accordingly.

Note that the 2-m RH thresholds in the all regions domain and valley/mountain domain are not reverse classifiers, but are significantly lower (0.675 in both domains) than in the single-region domains. When the unique characteristics of the valley region classifier profile are combined with the more conventional profile (i.e., higher predicted RH correlated to observed fog) from another region or regions, the optimal RH threshold ends up being lowered to the point that it simply undercuts the majority of valley region predictions corresponding to observed fog with predicted RH <0.885. But it also groups these predictions with the RH predictions >0.885, an abundance of which correspond to observed no fog but which will be classified as fog by the post-processing. This does not signify a great deal of promise for obtaining skill improvement in the valley region using the simple technique *RH_D* across a combined domain.

3. RH Threshold with 2-m Temperature Bias Correction, Deterministic

BiasRH_D is identical to *RH_D* except that a correction of the 2-m temperature sample bias is applied to the predictions prior to computation of the optimal RH threshold. The sample bias is computed in each domain using only the cases when members did not predict fog (since this is the subset of data subject to post-processing), which differs slightly from the overall bias (Table 7). During testing of *BiasRH_D*, the bias itself is subject to leave one out cross-validation, and therefore changes slightly as the developmental data sample is changed. This process is explained further later in this chapter. The sample bias is computed with no member-specificity or time-dependency; the correction addresses the average member sample bias in the domain during the interval 7-20 h.

Table 7. Summary of the average 2-m temperature prediction bias (K) in each domain among all members for the period 7-20 h. The bias used to perform a bias correction in *BiasRH_D* varies slightly from the overall bias because it is computed using only instances when fog was not predicted by the member.

Domain	Overall Bias	Bias used for *BiasRH_D*
All Regions	+1.11	+1.23
Coastal Region	+3.19	+3.23
Valley Region	+0.93	+1.15
Valley/Mountain Region	+0.29	+0.35

Since correcting for the bias lowers the 2-m temperature in each domain, the first-order effect is to increase the optimal RH threshold used as the fog classifier compared to *RH_D*. More significantly, correcting the bias has a non-linear effect on 2-m RH that is a function of the temperature (the correction will cause a larger RH increase at lower temperature), and it is the impact of these non-linear interactions that is examined in *BiasRH_D*. 2-m q_v biases could be corrected in addition to or instead of 2-m temperature biases, but here we limit the correction to one variable to better evaluate the impact. Temperature biases are selected for correction instead of q_v biases because the previous chapter revealed the negative RH bias at this level is primarily caused by a warm bias, while the q_v bias is slightly positive in each region.

The ROC curves in Figure 67 show that the bias correction has the expected effect of raising the optimal RH threshold in each domain, particularly in the coastal region, where the largest bias correction was applied and the optimal RH threshold increased by nearly 0.17. The threshold shows little change in the valley/mountain domain, where the negative biases of the mountain sites largely offset the positive biases in the valley sites, leading to a modest bias correction of only -0.35 K.

Aside from an increase in the threshold in three of the four domains, there are no significant changes in the false positive rate or POD achieved in each domain at the optimal threshold, and the overall shape of the curves is virtually identical to those in *RH-_D*. Full verification results are presented in Chapter VI, but the similarity in ROC curves in *RH_D* and *BiasRH_D* suggests the non-linear relationship between temperature and RH is of minimal consequence to the RH error. Applying a homogenous bias correction to the entire domain may have little affect on overall skill.

As in *RH_D*, the ROC curves and optimal RH_thresholds for verification at the three higher β_e thresholds (not shown) are similar to those shown in Figure 67.

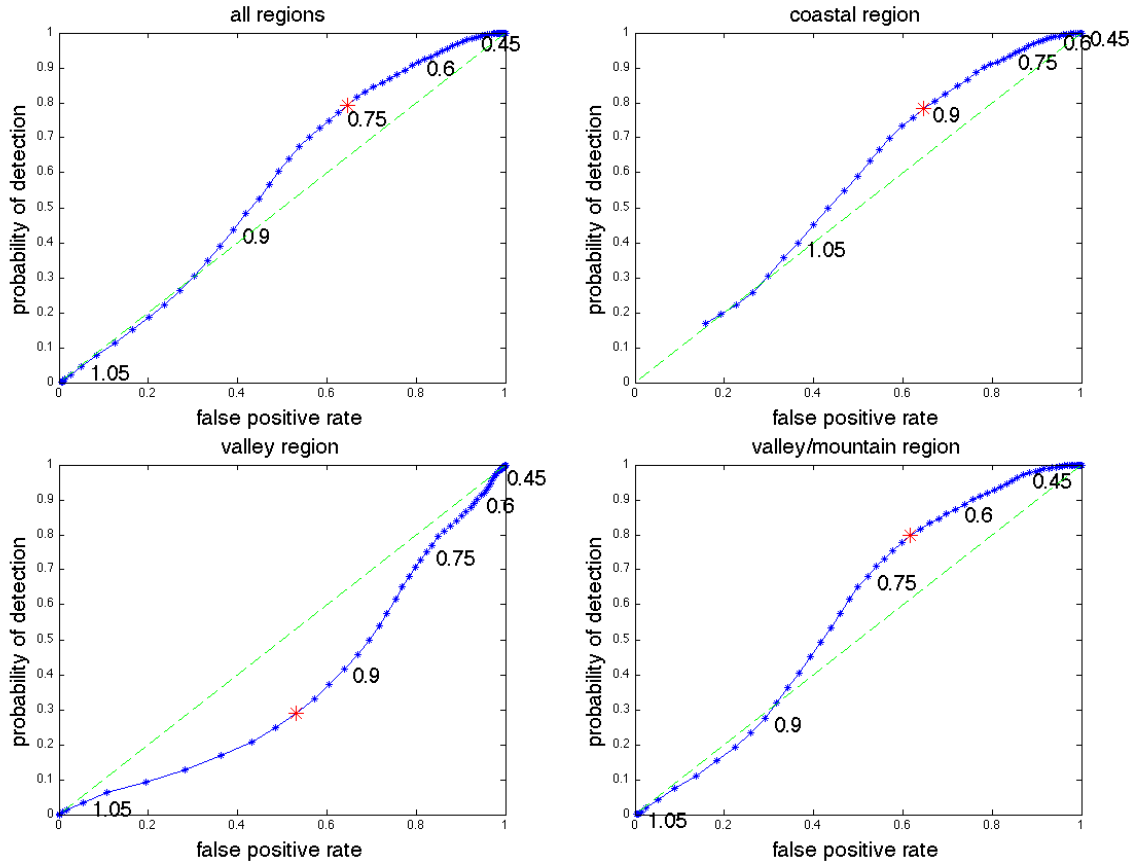

Figure 67. Same as in Figure 66, but after a 2-m temperature bias correction has been applied in each domain.

4. RH, Probabilistic

RH_P examines the impact of using a probabilisitic as opposed to deterministic framework for the post-processing of each member. By nature of using an ensemble in this work, all the VIF forecasts already provide some measure of stochastic information. However, RH_P further develops the framework of RH_D by producing a probability of exceedance of each β_e verification threshold, rather than using a fixed 2-m RH threshold to arrive at a deterministic β_e exceedance prediction.

The procedure for producing the probability of β_e threshold exceedance is described using the data plotted in Figure 68. For each of the four domains, the figure shows the total distribution of 2-m RH predictions when fog was not predicted, with the light blue portion of the distribution representing predictions coinciding with observed fog (using the lowest β_e threshold), or the missed opportunities. The purple portion of the

distribution therefore represents instances when fog was neither predicted or observed, also called the correct rejections. Using the data from all regions as an example (top left panel), we see that when the members' 2-m RH predictions fall in the bin 0.30–0.315, the ratio of observed fog cases (missed opportunities) to total plotted cases (missed opportunities plus correct rejections) is 4:63, or an incidence of 0.063. This is lower than in the 0.90–0.915 RH bin, where the ratio is 221:742 for an incidence of 0.298. However, simply using the ratio in each fixed bin as our post-processed probability of exceedance becomes problematic when the number of cases in the bin is small. Consider that the 63 total cases in the 0.30–0.315 bin could represent as little as 8 h of data from one day (one prediction per hour from eight members), a rather small dataset to evaluate a meaningful pattern to leverage in post-processing. In contrast, large bins might have too many cases, which can conceal meaningful patterns in the data that would emerge if the bin were smaller.

This issue is addressed by using flexible bin sizes, such that each bin has the same number of cases. This is achieved by defining the limits of the bin for any given RH prediction as one that captures a fixed number of nearest RH predictions. In *RH_P*, this number is set to one-twelfth of all the data in the domain, which means each bin contains 1660 predictions (out of nearly 20,000 total predictions) in the all regions domain. The probability of β_e exceedance for the member is then found by using the incidence of observed fog among the 1660 cases in the bin. The corresponding predicted probability for any given RH prediction using this procedure is plotted with a black line in Figure 68.

The range of the bins using this method can vary widely, but this trait serves to equally balance across the entire prediction space the competing interests of overfitting the data (by making the bins too small) and surrendering predictive resolution (by making the bins too large). Updating our previous example, an RH prediction of 0.3 uses as its bin predictions ranging from 0.1500 to 0.4495, which is a large range compared to other portions of the prediction space but buffers against the uncertainty that would otherwise exist in the procedure since there are very few cases with RH predictions this low. The incidence of fog in this bin is 0.0946, and Figure 68 shows that the output probabilities change very little near the tails of the distribution where data are scarce.

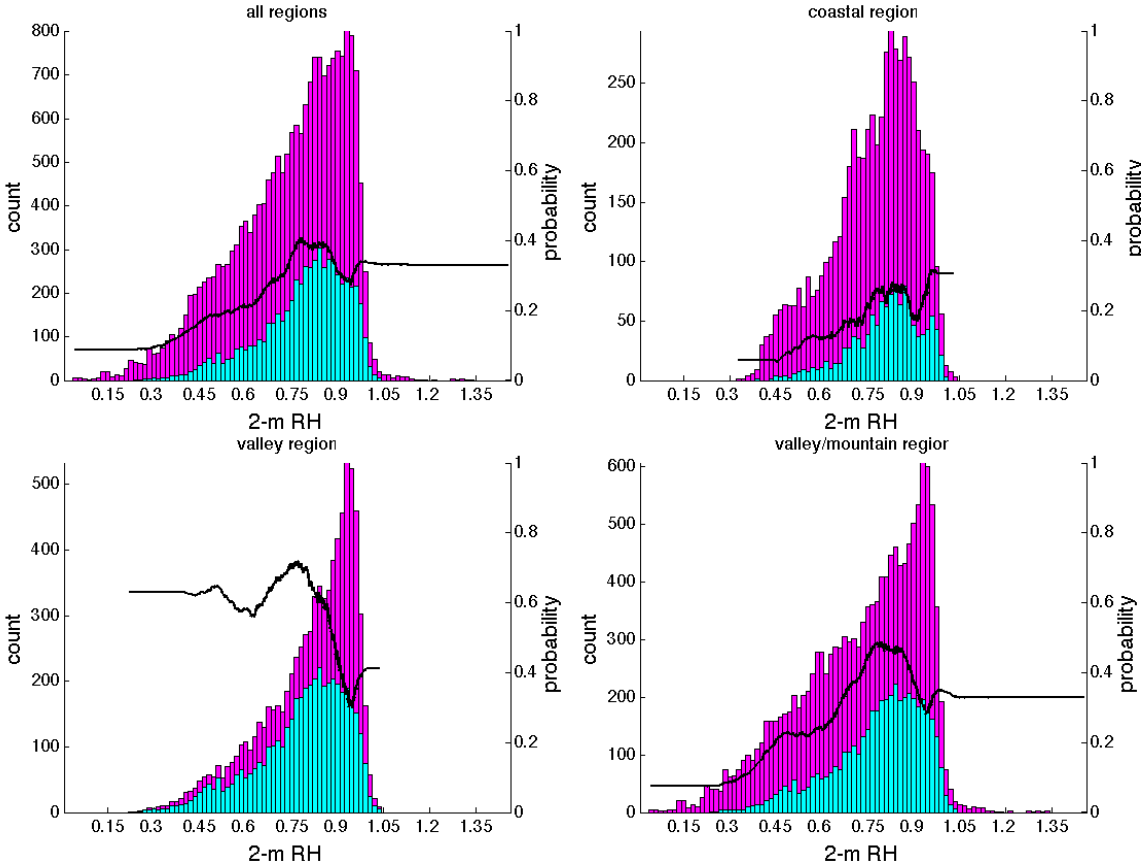

Figure 68. Distribution of 2-m RH predictions for each of the four domains when fog was not predicted by the member ($q_c \leq 8.5 \times 10^{-4}$ g m^{-3}). The light blue portion of each distribution represents predictions corresponding to observed fog ($\beta_e > 0.29$ km^{-1}). The black line represents the predicted probability of fog based on the post-processing procedure. The first six hours of each case are excluded.

In contrast, an RH prediction of 0.9 benefits from high data density during post-processing, and so the bin is accordingly smaller, ranging from 0.8833 to 0.9167 with a fog incidence of 0.310. Since there is more data at these values, the output probabilities are permitted greater sensitivity to small changes in the RH predictions, which allows them to leverage patterns in the data that might otherwise be diluted with larger bins. An example of this is in the valley domain (bottom left panel), where the decreasing incidence of fog with increasing RH in the range 0.75–0.95 is evident and is consistent with the unique reverse classifier found in *RH_D*.

As bin size (i.e., the number of cases included in each bin) increases, we can expect each bin to produce output probabilities closer to the climatological incidence of the entire data, which by definition will destroy resolution in the final predictions, but increases our likelihood of reliability improvements since the degree of overfitting the data will be reduced. Decreasing bin sizes aims at greater resolution, but risks overfitting and instead reducing both reliability and resolution. Until cross-validation is performed, it is impossible to know if overfitting has occurred (the reliability of the training data is always perfect). A thorough optimization of bin sizes is not performed in this work, and without it, we use the one-twelfth size parameter in most experiments as a fairly conservative value after limited testing. We will briefly examine the sensitivity of skill improvement to bin size in later experiments. Even in the coastal region, which is the smallest domain, one-twelfth of the data translates to a bin size of 499 predictions, which equates to an average of 62 predictions per member, or at least five separate days of data (recall that each case is spaced three or four days apart to further reduce correlation among the cases). For this work, we give high priority to pursuing an incremental skill increase with a framework that can transfer to other regions, an objective that must necessarily place emphasis on suppressing overfitting within reason. Further experimentation with a larger dataset is warranted to determine if smaller bins are advisable in the interest of more aggressively pursuing resolution gains.

The impact of larger bin sizes is evident in the probability output profiles of *RH-_P* as shown in Figure 68, which have a relatively limited range. For example, output probabilities in the coastal region range from 0.056 to 0.316, while the climatological incidence of fog for all the data in the domain is 0.197. This suggests large resolution improvements are unlikely in this case. However, since the post-processing is only applied to members without fog already in their prediction, additional resolution in the final ensemble VIF prediction can still be achieved if the stochastic probabilities are preferentially increased for the observed fog cases, even if only by a few percent.

Simple logistical regression of the RH predictions against the observed incidence of fog is not pursued because Figure 68 shows the relationship between these variables is highly non-sigmoidal, or does not resemble a monotonic “S” shape prescribed by

logistical fitting if the relationship between RH and q_c were linear. Nonlinear regression might be used to describe the relationship, but this is a premature and perhaps inappropriate (without a larger dataset) course. Alternatively, the nonlinear and physically unclear relationship implied by the data might be clarified by the inclusion of an additional predictor, which is the course chosen and developed in future experiments.

In the meantime, a simpler process is used to describe the curves in *RH_P* that more easily and quickly facilitates cross-validation. Using the data in Figure 68 as an example, once the incidence of fog within each customized bin has been computed at every prediction value, the RH range between predictions is populated by linearly interpolating the incidence values from the two adjacent predictions. This allows the framework to provide a probability of β_e exceedance for any given 2-m RH prediction within the total RH range of the plot, which for the all regions domain is 0.042–1.454. The process of formally fitting a non-linear expression to the data, which is not required to employ any of the frameworks presented here, is left for future work.

Once the post-processed probability is computed for each member that did not predict fog on its own, the probabilities are combined with the predictions from the members that did predict fog (and therefore were not post-processed), and all the probabilities are normalized as described in Chapter IV.A.1 to arrive at a final probability of exceedance prediction.

The data used to develop the post-processing output for the probability of exceedance at the three other β_e thresholds are shown in Figure 69 for each domain. Generally, the forecast probabilities decrease at higher β_e thresholds, but the shape of the profiles are similar. One notable exception is in the coastal region, which has a distinct absence of heavy fog events (i.e., at the highest β_e threshold of 2.10 km^{-1}) at higher RH predictions.

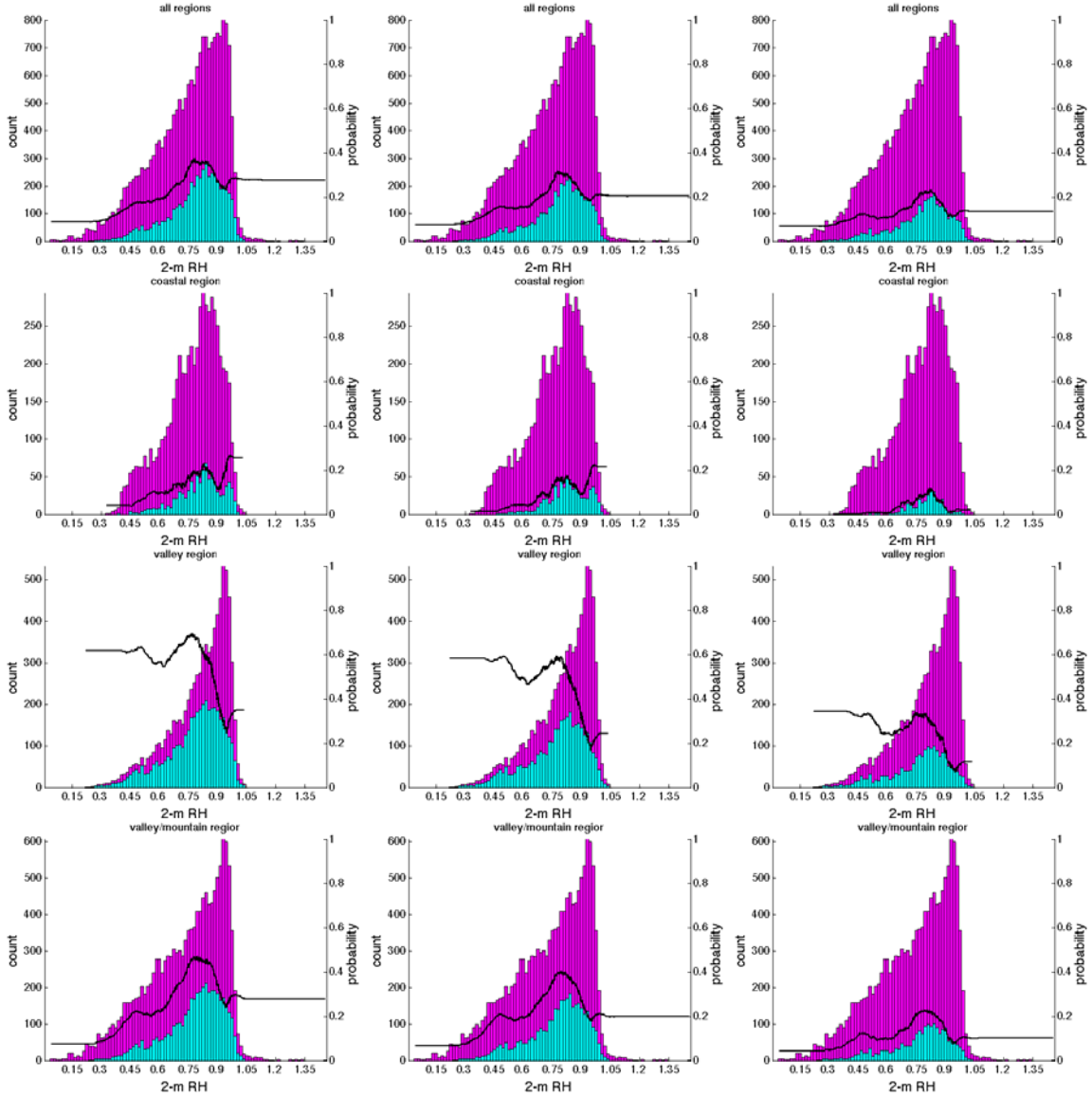

Figure 69. Same as in Figure 68, but with the light blue portion of the distributions and the output probability of exceedance corresponding to each of the three other β_e thresholds: 0.41 km^{-1} (left column), 0.68 km^{-1} (center column), 2.10 km^{-1} (right column). The rows correspond to each of the four domains: all regions (top row), coastal region (second row), valley region (third row), and valley/mountain region (bottom row). The first six hours of each case are excluded.

5. RH with 2-m Temperature Bias Correction, Probabilistic

BiasRH_P applies the same post-processing procedure as *RH_P*, but after the 2-m temperature bias correction has been applied according to Table 7. The distribution of the predictions following the bias correction, as well as the probability output used in the post-processing, is shown in Figure 70 for each β_e threshold. As expected, the probability output profile of each plot is shifted toward higher RH values compared to the pre-bias corrected data in Figures 68 and 69. The largest shift is in the coastal region domain, which was subject to the greatest bias correction. More subtly, the bias correction has the effect of increasing the variance of the overall distribution, indicating that it has a larger impact on higher RH predictions than it does on lower RH prediction. This can only be because the higher RH predictions coincide with lower temperature predictions. Whether or not the performance of the post-processing is improved by the bias correction is determined by whether it affected the observed fog cases differently than the observed no-fog cases, and this is not immediately apparent, but is addressed by testing *BiasRH_P* with cross-validation.

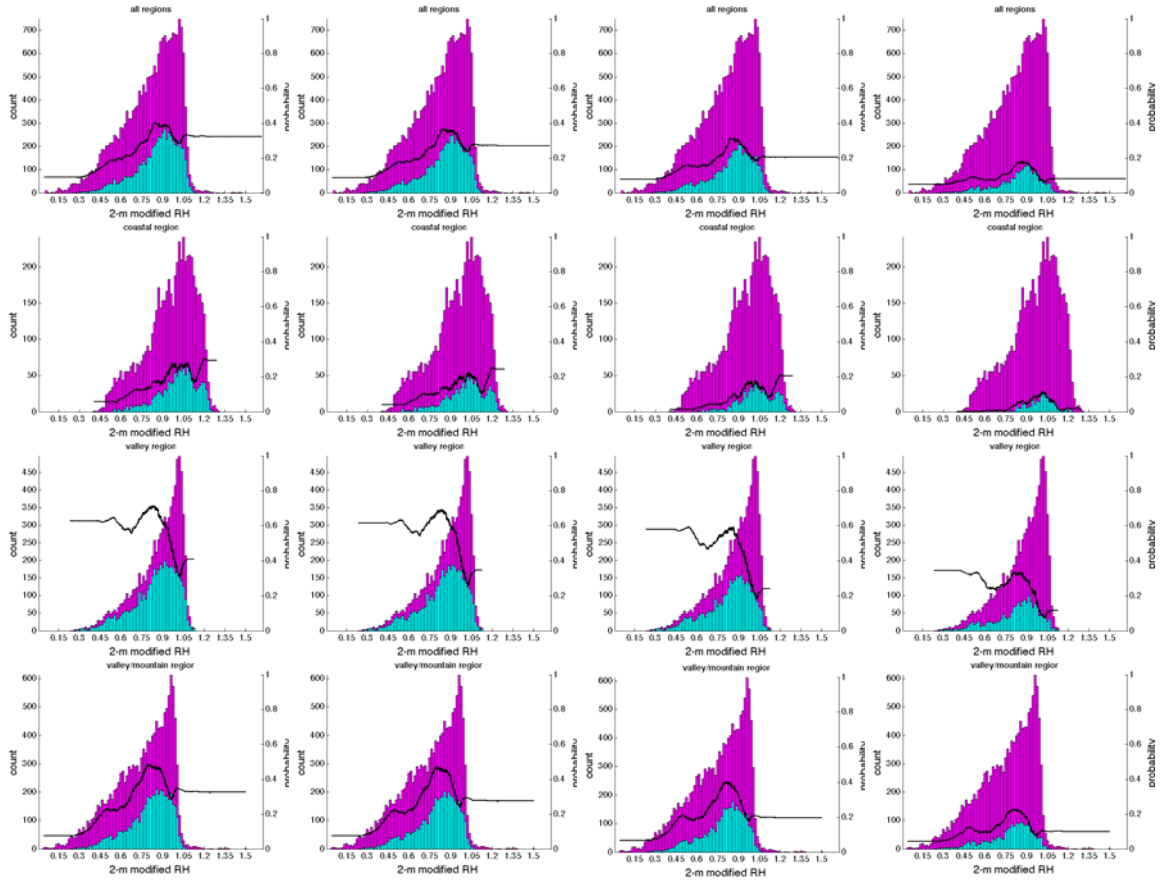

Figure 70. Same as in Figure 68 and Figure 69, but after a 2-m temperature bias correction has been made in the predictions. The columns correspond to each of the four β_e thresholds, increasing from left to right. The rows correspond to each of the four domains: all regions (top row), coastal region (second row), valley region (third row), and valley/mountain region (bottom row). The first six hours of each case are excluded.

6. Joint Parameter Space, Best Overall

a. Description

The rather limited range of probability forecasts prescribed by the profiles of *RH_P* and *BiasRH_P* suggest 2-m RH predictions alone have somewhat limited predictive usefulness for fog. With the joint parameter space techniques developed in the following sections, we examine other NWP model parameters for their fog predictive usefulness, while also expanding the interrogation of predictors to two dimensions.

The framework of *RH_P* and *BiasRH_P*, in which the incidence of fog is measured within a flexible bin at every prediction value, is extended to joint parameters space in *JP_B*. A tangible example of the advantage of joint parameter space is illustrated in the left panel of Figure 71, which takes the same data as in the coastal region plot of Figure 68 and extends it to two dimensions. Here, predictions corresponding to observed fog (using the lowest β_e threshold of 0.29 km^{-1} in this example) are plotted in red, and those corresponding to observed no fog are plotted in blue. If we examine only the distribution in the x direction, the plot adequately conveys the same pattern shown in Figure 68, with a somewhat increasing but rather erratic incidence of observed fog as predicted RH increases. However, by including a second parameter on the plot in Figure 71, we see a large portion of the observed no-fog cases with high RH predictions can be distinguished from the observed fog cases by nature of their lower 2-m vapor pressure predictions. Incidentally, the reason for this is not due to any substantial change in NWP model error; at high RH, fog is simply observed less often at lower temperatures (and therefore lower vapor pressures) in the coastal region. This example illustrates why two predictors are advantageous, and this particular characteristic of coastal region fog will prove to have significant predictive usefulness and will be examined in later experiments.

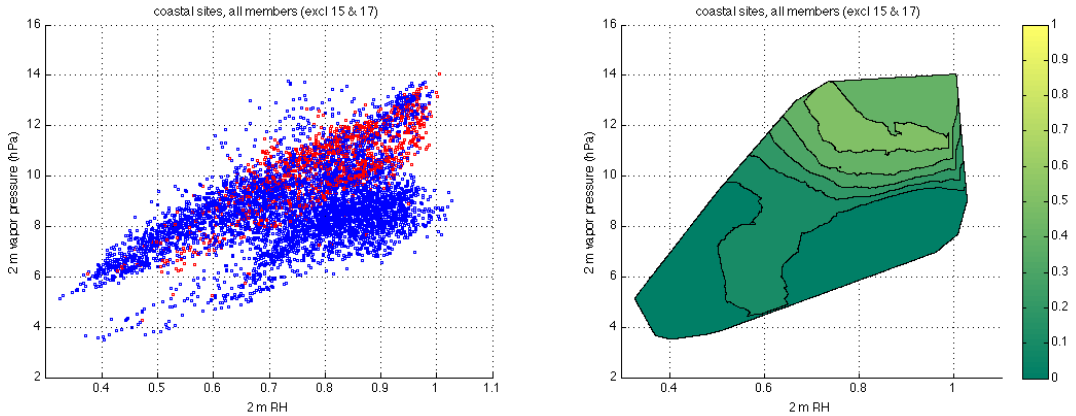

Figure 71. Scatter plot of fog missed opportunities (red) and fog correct rejections (blue) within a joint parameter space using 2-m RH predictions and 2-m vapor pressure predictions as the parameter pair. The right panel shows the forecast probability map derived from the plotted data. The first six hours of each case are excluded.

In order to arrive at a forecast probability of fog using the joint parameter space, the concept of flexible bins used in *RH_P* and *BiasRH_P* is also extended to two dimensions. The bin for any given prediction plotted in the 2-D space consists of a circle centered at the prediction, and of sufficient radius to capture one-twelfth of the total data. Since the two axes of the joint parameter space plot will normally have different scaling, a correction is applied to the circle axes based on the ratio of the total range of the data for each parameter. The effect of this correction is to keep the bins relatively circular as they would appear on the plot, as opposed to becoming highly elliptical in some cases.

Once the bin is established, the post-processed probability forecast is based on the incidence of fog within the bin. The right panel of Figure 71 shows the probability forecasts of the prediction space after contouring has been applied. As with the probabilistic single parameter experiments, no attempt is made to fit the multivariate relationship to an expression via multiple regression. For a linear relationship to exist between β_e and the predictors, u_1 and u_2, the predicted probabilities of β_e threshold exceedance (as plotted in Figure 71) do not need to also be linear along u_1 and u_2, but should be monotonic along u_1 (at all values of u_2) and u_2 (at all values of u_1), ideally taking on a two-dimensional sigmoid shape (Figure 72) in some orientation. In Figure 71 and every other joint parameter experiment presented in this work, the probability forecasts are non-monotonic in both axes directions, indicating the relationship between the predictors and β_e is non-linear. This suggests a multiple nonlinear regression technique is needed to properly fit the data to an expression.

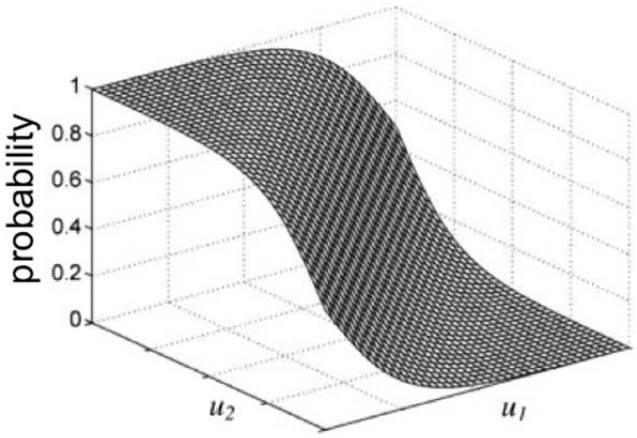

Figure 72. Notational illustration of a two-dimensional sigmoid plotted in the joint parameter space u_1, u_2, with probability plotted in a third dimension rather than contoured as in other plots. (After Yang 2009).

A multiple nonlinear regression technique is prematurely complex for this stage of the framework development, and ultimately unnecessary for implementation. Instead, the exceedance probabilities in the portions of the joint space between each prediction are two-dimensionally interpolated using a Delauney triangulation scheme (Delauney 1934), which for irregularly spaced (i.e., non-gridded) data is preferable to bilinear interpolation. In Delauney triangulation, the joint parameter space is broken into small triangles with vertices located at the data points (Figure 73). Conceptually, for any given triangle, each of its three vertices can be raised to a height corresponding to its probability of exceedance value, and any given point within the triangle then also has a height and corresponding value. Using this method, all portions of the joint space have a defined probability of β_e exceedance value that is be applied to any new predictions of u_1 and u_2 from the NWP model. Further details on Delauney triangulation are found in Barber et al. (1996).

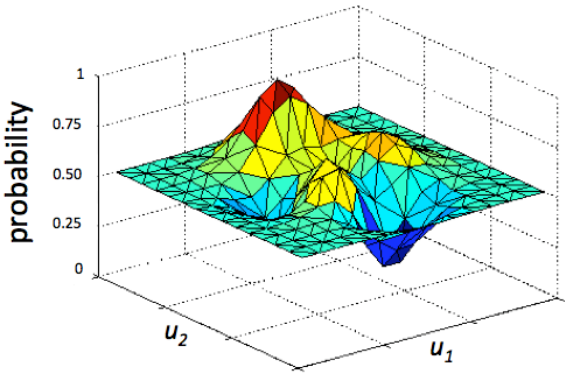

Figure 73. Notational illustration of Dulauney triangulation in the two-dimensional joint parameter space u_1, u_2. In this example the probability forecast values plotted on a third axis to aid in the conceptual visualization of the interpolation scheme. (After The Mathworks, Inc. 2009).

Using a strict interpolation strategy such as Delauney triangulation for every data point in the joint parameter space might at first seem to risk drastic overfitting of the data. However, recall that the dependent variable being fit is the probability of β_e exceedance based on the observed incidence within a bin containing hundreds of nearby predictions. Therefore, the probability of β_e exceedance changes very little over short distances in the joint space. The degree of data overfitting is controlled by the bin size.

Clearly, using joint parameters achieves some measure of additional separation between the observed fog cases and no-fog cases, allowing output probabilities to range from 0.549 (at high 2-m RH predictions and high 2-m water vapor predictions) to 0.002 (at low 2-m water vapor predictions). This range is significantly larger than the range obtained with single predictors in *RH_P* and *BiasRH_P*, but the results are also quite different. In those experiments, the lowest probability values were found at low RH prediction values, while the 2-D space of Figure 71 shows that probabilities are just as low (in fact slightly lower) during high RH predictions if the 2-m water vapor prediction is low. Furthermore, Figure 71 suggests that 2-m water vapor predictions are a better predictor of fog than the 2-m RH predictions (when fog is not already predicted by the member), as the probabilities have more variation in the y-direction than in the x-direction in Figure 71.

Examining the highest and lowest output probabilities on any single-parameter or joint-parameter plot provides some indication of how well the observed fog cases are spatially separated from the no-fog cases, which in turn provides an indication of likely improvement in predictive resolution. But more thoroughly, the likely predictive resolution can be assessed by computing the *variance* of the output probabilities themselves; that is, the mean squared difference between the output probability at each point and the climatological fog incidence of the entire plotted data. This is analogous to the resolution measurement of a stochastic prediction, where very low or very high probabilities are preferable to probabilities near the climatological incidence. Since the reliability of any plot of this kind is inherently perfect for the training data, and the bin size is standardized (which effectively equalizes the potential impact of data overfitting for any given parameter pair), we are able to use the variance of the plot as a rather powerful quantitative assessment tool for evaluating the merit of numerous parameter pairs and revealing patterns of NWP model behavior prior to performing a full cross-validation. There is no presumption that, in real-world use, the reliability of the parameter space will be perfect and the resolution can be exactly measured by the plot variance; indeed the degree to which these assertions break down depends on the degree of overfitting of the training data, which will be examined during cross-validation. For now, we use these assumptions to assist with selecting the most promising parameter pairs prior to cross-validation, while emphasizing the fact that standardizing the bin sizes makes this simplification reasonably valid.

Since the variance is computed using the mean squared difference at each plotted point, it is naturally weighted toward portions of the plot where the data density is highest (and where future predictions are most likely to exist). In addition to variance of the plot, subjective evaluation is also required to establish a physical mechanism by which the parameters achieve their predictive usefulness (and furthermore, the likely transferability of the procedure to other locales), and this is clearer in some cases than in others.

Compelling arguments can be made for evaluating a wide variety of basic and derived parameters, especially if a location-specific statistical calibration is the aim

(e.g., Bankert and Hadjimichael 2007, Marzban et al. 2007). In total, nearly 1000 joint parameter pairs were initially evaluated in each domain in order to select the most promising parameter pairs for full cross-validation. In the interest of facilitating an interpretation of the results within the context of the systematic NWP model errors detailed earlier in this work, we mainly limit the parameter candidates to temperature and moisture variables at the layer 1 and 2-m levels, as well as parameters that are easily derived from them, such as RH, virtual temperature, and vapor pressure depression (i.e., the difference between the saturation vapor pressure and the vapor pressure). In addition, the variable *deficits*, which are defined as the 2-m prediction values minus the layer 1 prediction values, are evaluated as parameter candidates as part of a parameter pair.

Some of the NWP model deficiencies examined in Chapter IV exhibit a time dependence, which degrades the effectiveness of a simple bias correction unless it too is time dependent. This might be alleviated by including forecast hour or time of day as a parameter in the joint parameter space techniques, but as an option to address time-dependent deficiencies we instead include the time rate of change of each parameter as its own parameter candidate. As an example, the post-sunrise hours might be characterized by increasing predicted temperature or decreasing predicted RH, and so the two distinct presentations of 2-m RH biases (for instance) might be effectively parsed by including the time rate of change of one of these parameters (instead of the parameter itself) in the parameter pair. From the standpoint of maximizing transferability of the technique, this approach is believed preferable for addressing time-dependent biases because it is based on output from the NWP model itself. In contrast, using time of day as a predictor may not transfer well to other latitudes or seasons since any diurnal cycles (to include sunrise and sunset) could vary by several hours. For any given parameter, its time rate of change is computed by subtracting the prediction from the previous hour to obtain the 1-h change as predicted by the NWP model.

Predictions of 850-hPa wind direction were also evaluated as a parameter candidate based on the rather primitive proposition that they provide some information on airmass type, and therefore the droplet number concentration, N. Results mostly rejected this premise, but one notable finding is presented in Appendix A.

The complete list of parameter candidates initially considered for the joint parameter space experiments are listed in Table 8. Including the time rates of change of each parameter, there are 946 possible joint parameter combinations. For more than half of these, the plot variance is computed in each domain at the lowest β_e threshold (0.29 km^{-1}), with the remainder of the parameter pairs able to be logically ruled out as viable options due to the poor predictive usefulness of one of the parameters in the pair. Once the plot variances were evaluated at the lowest β_e threshold, the 20 parameter pairs producing the highest plot variance in each domain had their plot variances computed at the remaining three β_e thresholds. Any other parameter pairs subjectively determined to be promising or interesting also had their plot variances computed at the remaining three β_e thresholds[4].

Some of the parameters in Table 8 might appear redundant, such as temperature and saturation vapor pressure, but they produced a plot variance that differed by up to 7% (when paired with the same parameter), enough to potentially make appreciable resolution differences in the final predictive skill. While saturation vapor pressure is a function of only the temperature, the relationship is exponential, indicating that differences in scaling of otherwise similar parameters could be an important factor in parsing observed fog from observed no fog in the joint space.

[4] The initial evaluation of the joint parameter pairs was performed with the output from members 15 and 17 included. Once the bug regarding the 2-m predictions in these members was discovered, all the plots discussed in this work affected by the bug (i.e. if either of their variables involved a 2-m prediction) were reevaluated with the two members removed. For joint parameter pairs that do not involve the 2-m level, no change was made.

Table 8. Predicted parameters considered for use in a parameter pair to define a joint parameter space. In addition, the one-hour time rate of change of each parameter is also considered as its own parameter. The cloud water mass concentration predictions tested for use in a parameter pair only include values $\leq 8.5 \times 10^{-4}$ g m^{-3} since anything larger than this is not subject to post-processing and therefore is not in the training data.

Parameters	
Layer 1 temperature	2-m virtual temperature
Layer 1 water vapor mixing ratio	2-m RH
Layer 1 virtual temperature	2-m vapor pressure depression
Layer 1 RH	Temperature deficit
Layer 1 vapor pressure depression	Saturation vapor pressure deficit
Layer 1 saturation vapor pressure	Vapor pressure deficit
Layer 1 vapor pressure	Virtual temperature deficit
2-m temperature	RH deficit
2-m water vapor mixing ratio	Vapor pressure depression deficit
2-m saturation vapor pressure	850-hPa wind direction
2-m vapor pressure	Cloud water mass concentration, q_c

No attempt was made to apply any bias correction to the predictions prior to plotting and evaluating them in joint space. Applying a bias correction to a plotted parameter itself would serve to uniformly shift the data along its axis, having no effect on the skill of the post-processing procedure. Applying a bias correction to a component of a parameter such that there was a non-linear effect (e.g., correcting temperature prior to plotting RH as we did in *BiasRH_D* and *BiasRH_P*) would affect the results, but previous experiments showed a relatively minor impact in most cases. If we correct for 2-m temperature bias prior to producing the joint parameter plot in Figure 71 for the coastal region, the plot variance changes by a negligible 0.11%. Presumably, the impact could be larger depending on the parameters involved and the magnitude of the biases, but this will not be examined in this work.

The discussion here will primarily focus on the parameter pairs that performed well across all four β_e thresholds, with the most emphasis on the lowest

threshold (i.e., the ability to predict any fog of any severity is given higher priority than the ability to predict only the heavy fog cases). Other evaluation approaches are possible, including using one parameter pair to predict fog, and another to predict the fog severity (i.e., the conditional probability of heavy fog), which we believe is a promising future path. However, here we aim to establish the single best parameter pair to be used across all severities of fog.

Experiments *JP_B* and *JP_U* present the most promising parameter pairs for each domain that were subject to full cross-validation. For *JP_B*, we examine and cross-validate the single parameter pair in each domain producing the largest sum of plot variances at each of the four β_e thresholds. For any given parameter pair in a domain, this sum is inherently weighted toward the lower β_e thresholds because the plot variances have more variability among parameter pairs at the lower β_e thresholds. For some domains, it is reasonable to believe that the predictive usefulness of the parameter pairs in *JP_B* are closely based on a rather localized aspect of the climatology. If this is the case, even cross-validation might not fully expose this shortcoming because each site within a region is subject to similar climatology. Later, *JP_U* will take a more critical view and examine parameter pair options with more transferability.

b. Coastal Optimization

The coastal and valley domains are examined first so that we are better able to later interpret the results in the combined domains. The parameter pair producing the largest sum of plot variances at each of the four β_e thresholds in the coastal domain is the time rate of change of virtual temperature paired with 2-m vapor pressure (Figure 74). The plots show that distinguishing heavy fog events in the joint parameter space is less successful than distinguishing *any* fog event, as the probabilities (and plot resolution) are significantly lower at the higher β_e thresholds.

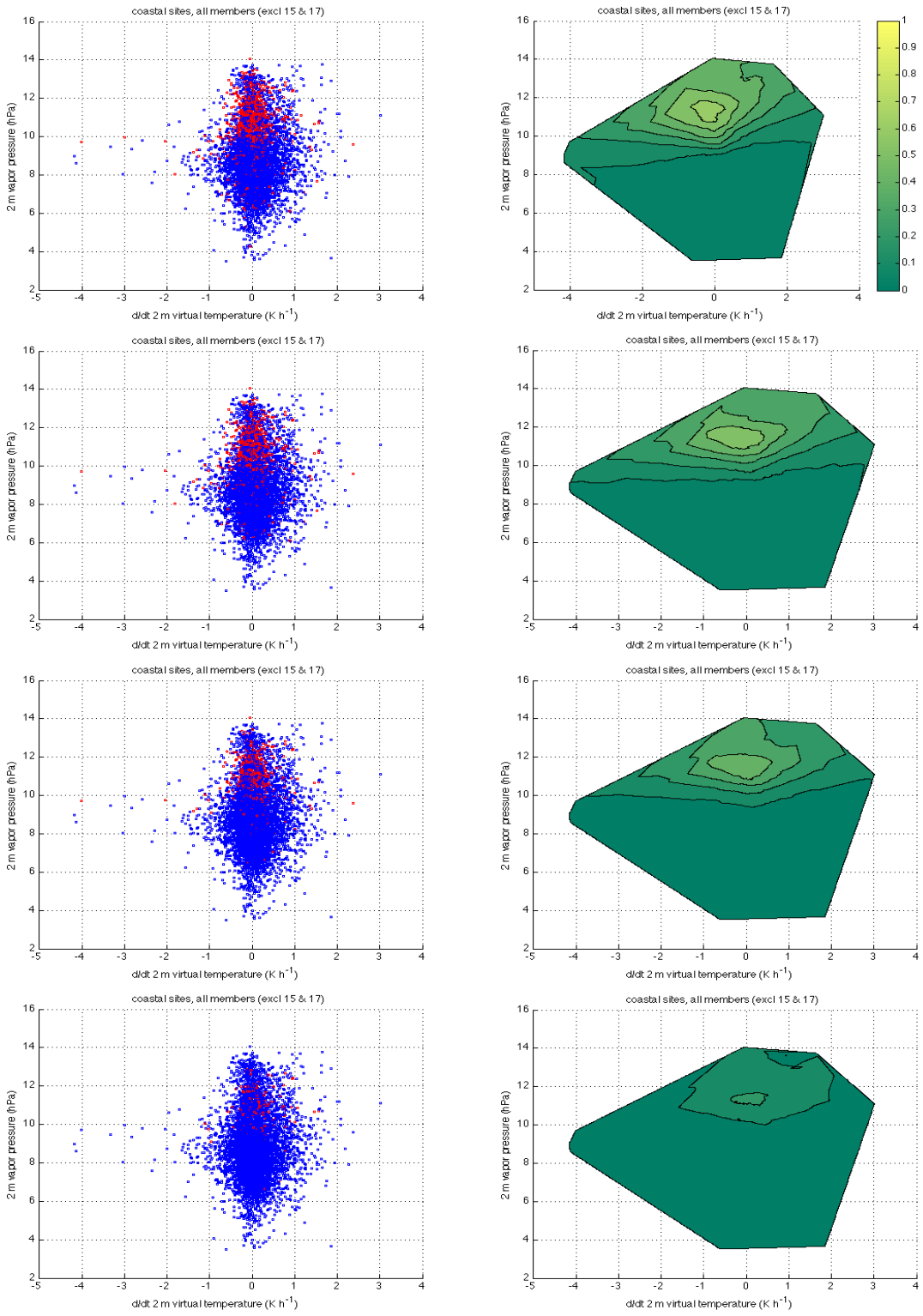

Figure 74. Same as in Figure 71, but for d/dt 2-m virtual temperature vs 2-m vapor pressure. The rows correspond to each of the four β_e thresholds, increasing from top to bottom.

As previously discussed and shown in Figure 71, the 2-m vapor pressure predictions exhibit high predictive usefulness in this region, despite having a significant moist bias. Regardless, error variances of 2-m q_v were shown to be quite low in this region, and so the data in Figure 74 conveys that fog simply has a low incidence when the 2-m vapor pressure (predicted or observed) is low. In large part, the mechanism behind this connection is proposed to be related to marine boundary layer stability. During the overnight hours, the vapor pressure in this region is closely correlated to the temperature, and at low temperatures, upward heat flux from the sea surface maintains a weakly turbulent boundary layer that favors low stratus clouds as opposed to fog. In contrast, at higher vapor pressures and temperatures, the boundary layer is stable and fog is more easily formed.

In fact, if we ignore bias, the 2-m vapor pressure predictions are a better predictor of observed temperature than the 2-m temperature predictions themselves during the overnight hours. This is illustrated in Figure 75, which shows the mean error variance across all members of the 2-m saturation vapor pressure predictions (solid blue line) are higher than the error variance when the 2-m vapor pressure predictions are verified against the 2-m saturation vapor predictions (dashed red line). It is believed this is why 2-m vapor pressure, as opposed to saturation vapor pressure or temperature, better accounts for the stability condition above the sea surface. The overnight bias between the vapor pressure predictions and saturation vapor pressure observations is <0.2 hPa (not shown). Therefore, Figure 74 suggests the probability of fog abruptly increases when the observed saturation vapor pressure exceeds roughly 10 hPa, which translates to a temperature of about 286 K. According to buoy data at the Trinidad pier situated between KCEC and KACV, the water temperature during the period of study ranged from 282-285 K (National Data Buoy Center 2012), just below this critical air temperature threshold and supporting the notion that the air-water temperature difference and resulting marine boundary layer stability plays a role in the fog predictive usefulness of the 2-m vapor pressure predictions.

A vapor pressure prediction >10 hPa does not guarantee fog, but simply makes it more likely (the maximum probability output is 0.653 for the lowest β_e threshold

of 0.29 km^{-1}). Examination of the synoptic pattern during the period of study reveals that elevated vapor pressures most often occur during the 1–2 days leading up to a frontal passage associated with an offshore low pressure system (not shown). During this scenario, southwesterly (onshore) flow is often present, which not only raises the vapor pressure but also increases the probability of offshore fog being advected inland. Although the output probabilities of Figure 74 show less variation in the direction of 2-m virtual temperature changes, the dependence of fog on this parameter is believed to be tied to the diurnal cycle. When the vapor pressure is high enough, the plot shows fog is most likely if the predicted 2-m virtual temperature change is zero or slightly negative, which occurs in the model for more frequently during the overnight hours than during the day (not shown). Increases in the 2-m virtual temperature predictions are consistently present after sunrise, when the incidence of fog is lower.

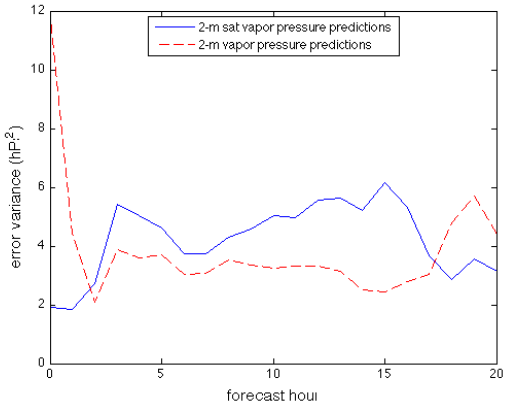

Figure 75. Mean error variance (across all members) of two NWP model variables when verified against the observed saturation vapor pressure in the coastal region: the 2-m saturation vapor pressure (solid blue) and the 2-m vapor pressure (dashed red).

This particular parameter pair, while clearly offering the potential for high resolution in the test region, might be significantly less useful in a coastal locale with a different water temperature, or even in the test locale but during a different season. If so, this flaw might not be revealed even with cross-validation since the testing sites do not change, and water temperatures change by only a few degrees over the course of the study period. One potential preventative measure for this might be to adjust the technique to account for the local water temperature. In a later experiment, we will take

another approach by examining a different parameter pair with predictive usefulness believed to be less-specific to the local climatology of the test sites.

c. Valley Optimization

In the valley domain, the parameter pair producing the largest sum of plot variances at each of the four β_e thresholds is the saturation vapor pressure deficit paired with layer 1 vapor pressure depression (Figure 76—note that the y-axis has been inverted such that smaller vapor pressure depressions, which generally correspond to higher RH, are toward the top of the plot). Saturation vapor pressure deficit appears to possess more predictive usefulness, with negative values (i.e., the 2-m prediction is less than the layer 1 prediction) associated with high fog probabilities. As saturation vapor pressure depends only on temperature, this region of the plot corresponds to predicted low-level temperature inversions, which in this region are typically produced by overnight radiational cooling of the ground and are a requisite condition for radiation fog. To a certain extent, leveraging the temperature deficit predictions helps mitigate the impact of volatility in the temperature, q_v, and RH predictions, which were shown to have inconsistent biases during fog missed opportunities. Regardless of these biases at each level of the NWP model, the predictions of temperature deficit appear to be a viable predictive indicator of fog, with a large portion of the space producing fog probabilities exceeding 0.8 at the lowest β_e threshold.

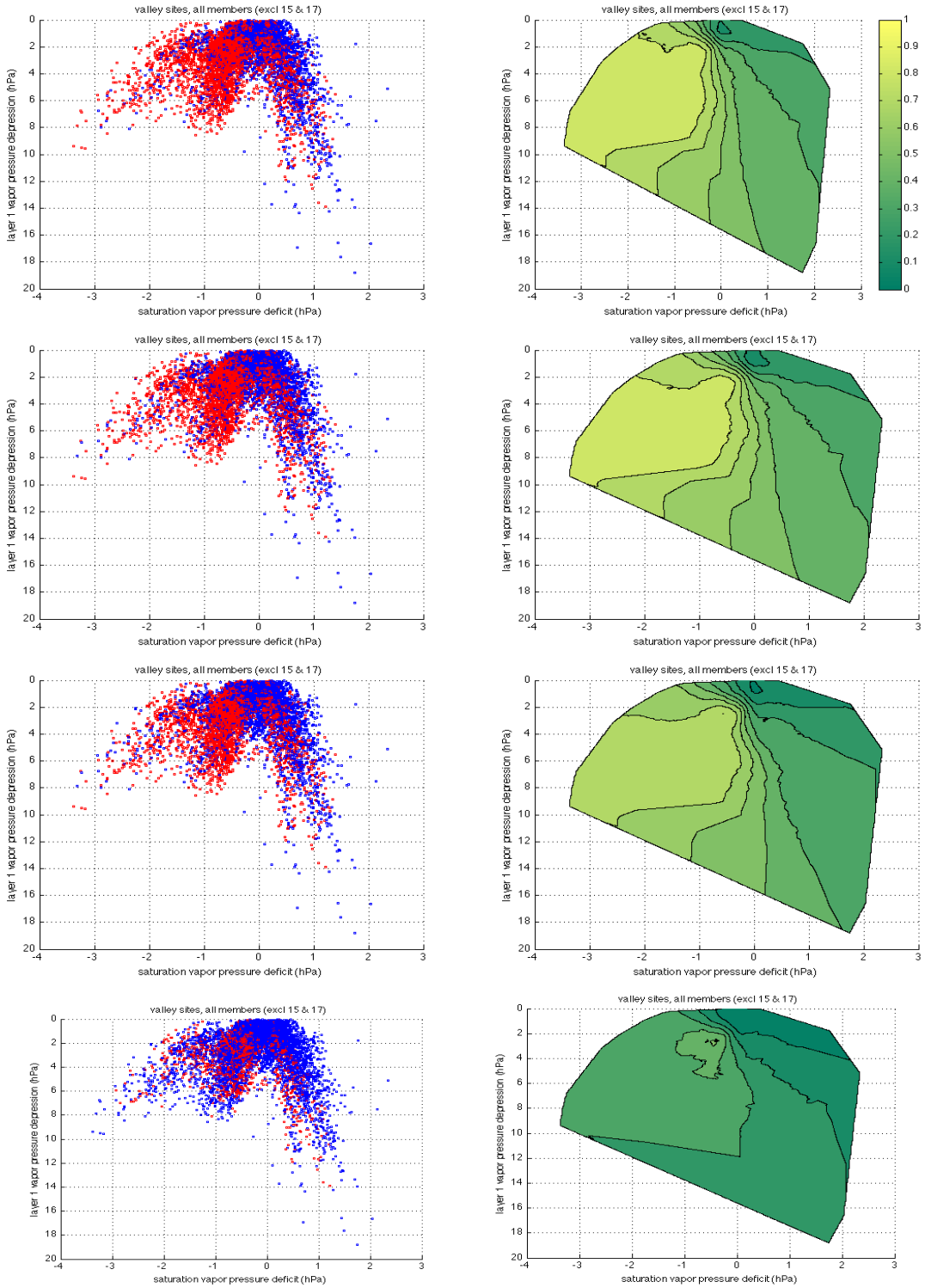

Figure 76. Same as in Figure 71, but for the valley region. The parameters are saturation vapor pressure deficit and layer 1 vapor pressure depression. The rows correspond to each of the four β_e thresholds, increasing from top to bottom.

Even when an inversion is predicted, the data show fog is less likely when the layer 1 vapor pressure deficit is very small, which corresponds to high RH. A similar trend was observed in the 2-m RH post-processing data earlier in this chapter, where high RH values were associated with lower fog probabilities. The reason for this connection is traced to the temperature initialization of the model and subsequent cooling rates during the early evening prior to fog development. Figure 77 plots the mean observed and predicted saturation vapor pressure for the valley sites on days when overnight or morning fog would eventually be observed (left panel), and on days without fog. The plots do not include cases when fog was predicted. The vapor pressure is also plotted for context, although it does not appear to play a crucial role. The fog days are characterized by more rapid cooling, which continues until sunrise near 16 h. This is consistent with a conventional radiation fog scenario, which is often supported by minimal cloud coverage and light winds that aid in the cooling rate.

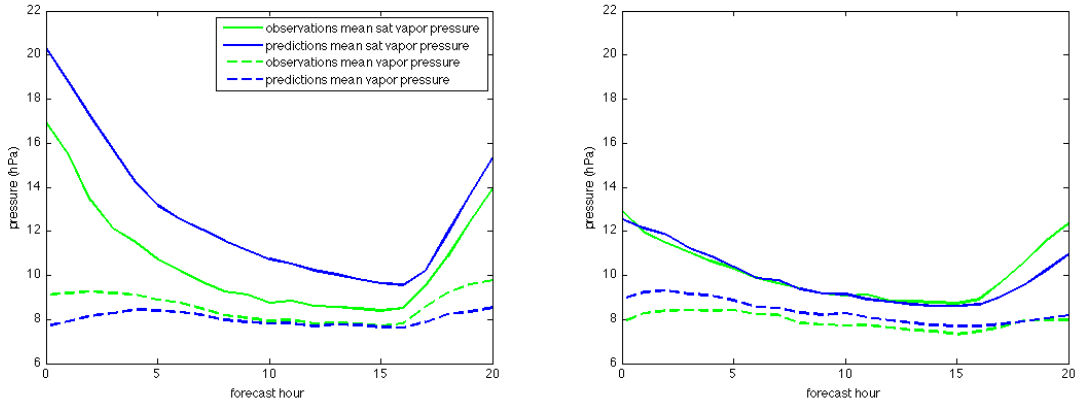

Figure 77. Mean observed and predicted saturation vapor pressure and vapor pressure at the valley region sites for (left) days when fog occurred and was not predicted between 10–17 h, and (right) days when fog did not occur and was not predicted between 10–17 h.

On average, the cooling rate predictions are accurate, but saturation vapor pressure is initialized too high by about 3 hPa (or about 2-3 K), and maintains this bias throughout the night, resulting in erroneously low RH predictions. In contrast, the cases without fog have lower afternoon temperatures and smaller cooling rates throughout the nighttime, oftentimes due to cloud cover and/or higher wind speeds. In these cases, the

NWP model predictions have minimal temperature biases during initialization and throughout the nighttime, and RH biases are much lower in these cases. Furthermore, in cases when the model correctly predicts a fog day (which the average member does for 26% of the fog days), the initialization bias is slightly <0, and is followed by a relatively unbiased cooling rate (not shown).

Clearly, initialization bias is associated with the missed fog events and warrants further examination in future studies. Notably, on the correctly predicted fog days (for which the initialization bias is slightly negative), the observed temperature at initialization averages 1.7 K lower than on days when fog is missed by the members. This suggests the initialization error is more likely or more severe on warmer days, which are also aided by clear skies and light winds and may explain why it preferentially affects the NWP model on nights with fog.

Observed RH during the nighttime shows little difference between the fog and no-fog cases plotted in Figure 77. Furthermore, the predicted RH values during the no-fog cases are reasonably accurate with just small positive biases stemming from slightly positive vapor pressure biases. So although the warm initialization error and warm biases during the fog cases result in larger RH biases, the deficiency seems to serve as an unconventional but effective predictor for fog when paired with saturation vapor pressure deficit in the joint parameter space. Since observed RH values show only minor difference between the fog and no-fog cases, correcting the initialization deficiency and RH bias might actually reduce the predictability of radiation fog absent a suitable replacement that similarly leverages a thermodynamical indicator.

These results offer a subtle contrast to the low-level cooling rates suggested by Tardif (2007) for use as a radiation fog predictor. As it were, cooling rates produced post-processing plots with variances about 30% lower than those in Figure 76, and even then only when paired with saturation vapor pressure deficit in the joint parameter space. Even so, Figure 77 suggests cooling rates could be a valuable alternative for identifying radiation fog likelihood, perhaps more so if post-processed in a way that allows the response in fog probability to lag the indicator (e.g., a high cooling rates result in high fog probabilities at a later forecast hour). No such capability is tested

here, and the individual performance characteristics of the NWP model used will certainly inform the results (particularly regarding something as specific as initialization error, which could be unique to the downscaling process or assimilation system used). Nevertheless, saturation vapor pressure deficits are conceptually tied to cooling rates, and for these WRF runs they are found to offer the most promising predictive skill in the valley region when paired with layer 1 predictions of vapor pressure depression.

d. Valley/Mountain Optimization

As we detailed in Chapter IV, q_c predictions in the mountain region are already more skillful than the other regions and do not contain a strong negative bias. Therefore, making upward adjustments to the q_c predictions alone is not believed to offer the same potential for skill improvement, and the post-processing framework developed in this work is not well-suited for the region. When combined with other regions to simulate operational realities, the parameter pairs with the most predictive usefulness are those where the mountain region predictions exist in a different sector of the space than the rest of the data, and can therefore be assigned appropriately low probabilities (since fog has the lowest incidence in this region). This is beneficial for the other regions involved as well, as their probabilities are not erroneously lowered by excessive influence from the mountain region predictions.

Different valley and mountain behavior leads to the parameter pair with the largest sum of plot variances at each of the four β_e thresholds in a combined valley/mountain domain (Figure 78), which utilizes predictions of virtual temperature deficit paired with layer 1 vapor pressure to distinguish the fog cases from the no-fog cases. The predictive usefulness of this former variable is not surprising, as it serves to identify inversions similar to how the saturation vapor pressure deficit was utilized in the valley region. In fact, saturation vapor pressure deficit could be substituted into this combined region plot, and still produce the second-highest variance of all the parameter combinations tested. The difference is nearly negligible.

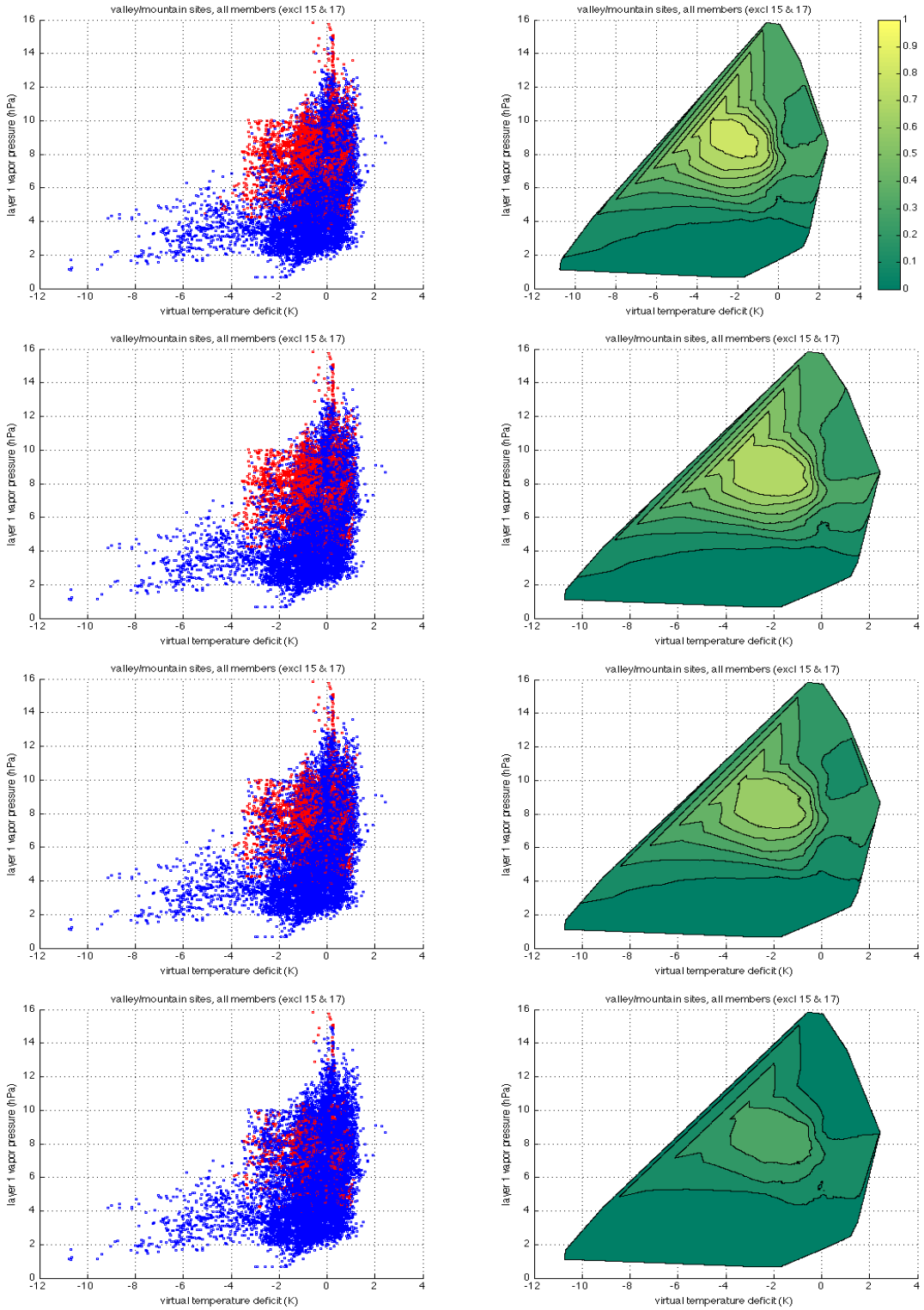

Figure 78. Same as in Figure 71, but for the valley/mountain domain. The parameters are virtual temperature deficit and layer 1 vapor pressure. The rows correspond to each of the four β_e thresholds, increasing from top to bottom.

Layer 1 vapor pressure acts to effectively separate many of the mountain data from the valley data. The majority of mountain vapor pressures <6 hPa due to dryer conditions at higher elevation. (Note that vapor pressure is a function only of dew point temperature, and is not *directly* impacted by pressure changes associated with changes in elevation. However, it is likely to be lower at lower pressure by nature of the cooler temperatures and lower dew point temperatures typical of a high-elevation environment). These low vapor pressure predictions translate to the lowest probability outputs of anywhere in the joint parameter space, a large portion of which is associated with probabilities <0.1.

In contrast, the vapor pressure predictions in the valley region rarely drop below 5 hPa, and are therefore mostly affected by the upper portion of the space where the virtual temperature deficit plays a primary role. Note that although the area of highest probabilities associated with temperature inversions is smaller than what was achieved in the valley-only region (Figure 76), the probabilities at the lowest β_e (0.29 km^{-1}) in the combined domain still exceed 0.8 near the center of the space, indicating the presence of the mountain data does not appear to drastically impede the predictive usefulness of these features. Fog is relatively rare in the valley region at observed vapor pressures <6hPa (not shown), and the low probabilities in this portion of the plot are not necessarily incompatible with valley region predictions. The limited data in the uppermost portions of the space with vapor pressure predictions >12 hPa are mostly associated with a few cases of warm frontal passage, all of which occur in the valley region and some of which occur with fog.

Using vapor pressure as a mechanism to separate data from each region is done at the expense of being able to use vapor pressure depression as a parameter in the pairs to refine the valley fog probabilities as was done in the valley-only domain. The results section will formally quantify the impact of this tradeoff on the valley region VIF skill, as well as detail the impact (detrimental or otherwise) the post-processing has on VIF skill in the mountain region.

e. All Regions Optimization

A single pair of predictors viable for all regions would be most desirable from an operational standpoint since it could conceivably be applied across a large model domain without the need to pre-define region categories. For this *JP_B* experiment optimized for the all regions domain, cross-validation will evaluate whether combining all the data is feasible for a simplified framework.

The joint parameter pair producing the largest sum of plot variances at each of the four β_e thresholds (Figure 79) is the same as for the valley/mountain domain. This is logical considering the high predictive usefulness of 2-m vapor pressure predictions revealed in the coastal domain, and the fact that both the coastal and valley domains are shown to have their highest fog probabilities within a similar range of predicted vapor pressures. Specifically, the highest fog probabilities in the coastal domain are between 10–12 hPa, slightly higher than the 9–10 hPa values corresponding to the maximum probabilities in the valley/mountain parameter space. The addition of the coastal prediction data draws the area of highest fog probabilities to slightly higher predicted vapor pressures compared to Figure 78. The values of these highest probabilities is between 0.7 and 0.8, which is higher than the maximum probabilities in the coastal domain (0.6–0.7) and lower than those in the valley/mountain domain (0.8–0.9).

The coastal region has different sensitivity to predicted radiation inversions from the valley region, but the nature of the pattern is the same. Fog is favored during negative virtual temperature deficits. The coastal data contains a large number of no-fog observations during predictions of low vapor pressure (6–9 hPa) and a statically unstable lower boundary layer (virtual temperature deficits of 0–2), a relatively common scenario in this region even during the nighttime. This has lowered the probabilities in this portion of the space, which also contains a limited number of fog observations in the valley region mostly associated with dissipating heavy radiation fog

that has lingered into the late morning hours. These valley fog events not associated with a predicted inversion are not very well resolved in any joint parameter space, but the lowering of probabilities in this space caused by the coastal data may limit any potential VIF skill increases in the valley region during these hours.

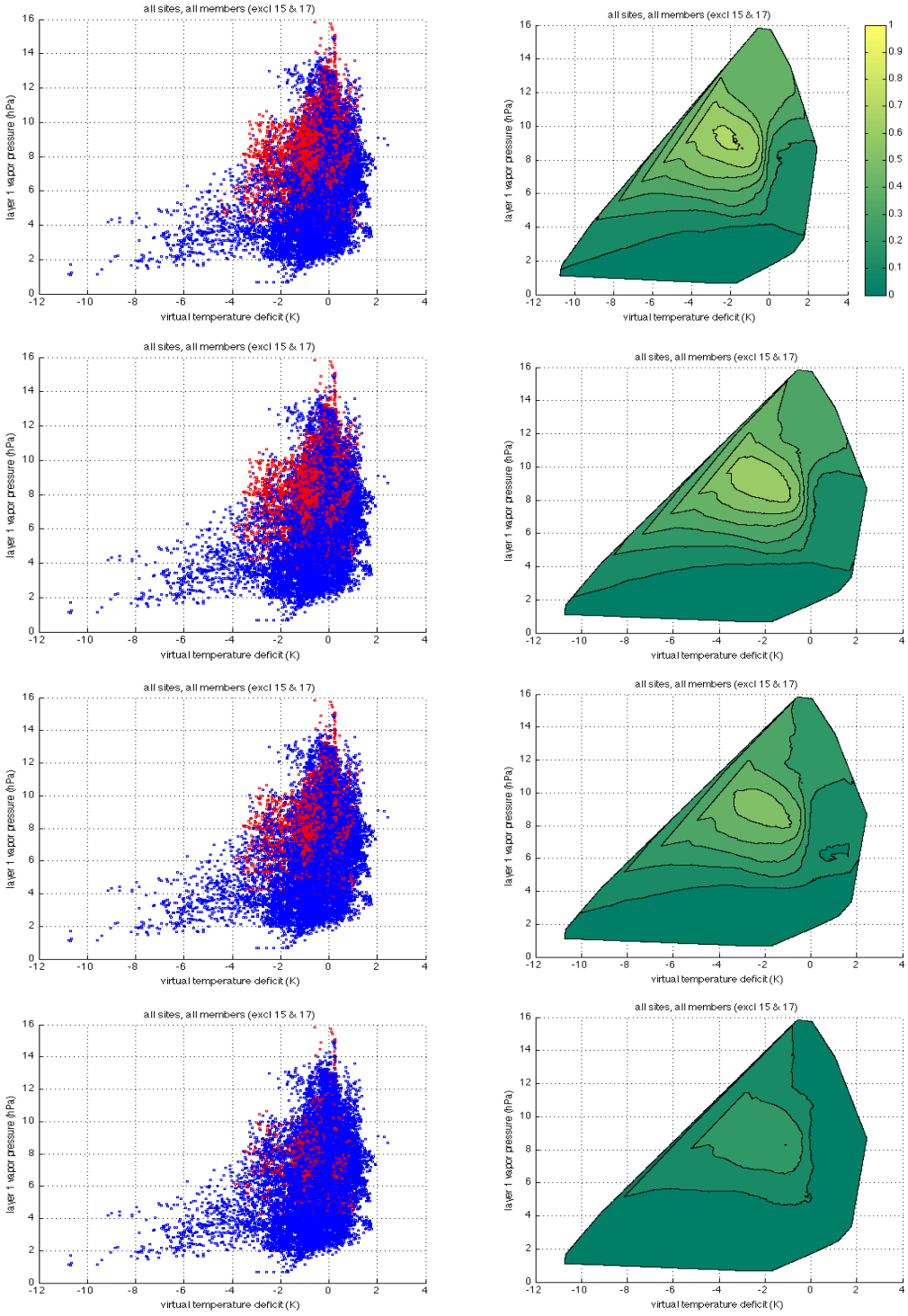

Figure 79. Same as in Figure 71, but for the all regions domain. The parameters are virtual temperature deficit and layer 1 vapor pressure. The rows correspond to each of the four β_e thresholds, increasing from top to bottom.

Figure 78 and Figure 79 are largely unchanged below 5 hPa, potentially signaling their post-processing impact on VIF predictive skill in the mountain region will be similar.

7. Joint Parameter Space, Sensitivity to Bin Size

In this framework, the degree to which the training data is overfitted is a function of the bin size. Larger bins reduce the risk of overfitting and increase the likelihood of reliability improvement, but potentially reduce resolution as the probability forecasts approach the climatological incidence. Bins that are too small and have overfitted the training data have captured unresolved high-frequency variations in the predictions rather than a systematic NWP model behavior, potentially resulting in reliability and resolution decreases.

In order to examine these impacts of bin size changes in the joint parameter space post-processing framework, predictions are tested using modified versions of the all regions joint parameter space map developed in *JP_B*. For the large bin experiment, *JP_LB*, the bin size was increased by 50%, such that each bin includes one-eighth of the total data rather then the one-twelfth figure used elsewhere. For the all regions domain used in the experiment, this results in 2490 predictions in each bin. *JP_SB* uses bins that are 33% smaller than *JP_B*, or one-eighteenth of the total data for a bin size of 1107 predictions.

Variation of post-processing maps with bin size is shown in Figure 80, with the standard bin size used in *JP_B* also included for comparison (center column). As bin size decreases, the bins reveal more fine scale structure of the space, with a wider probability range and higher overall plot variance. Cross-validation is performed using these maps to gauge the extent to which these structures represent systematic NWP behavior as opposed to overfitted training data. It will also serve to present the basic considerations regarding predictive reliability and resolution when selecting bin size or other contouring strategies in the parameter space.

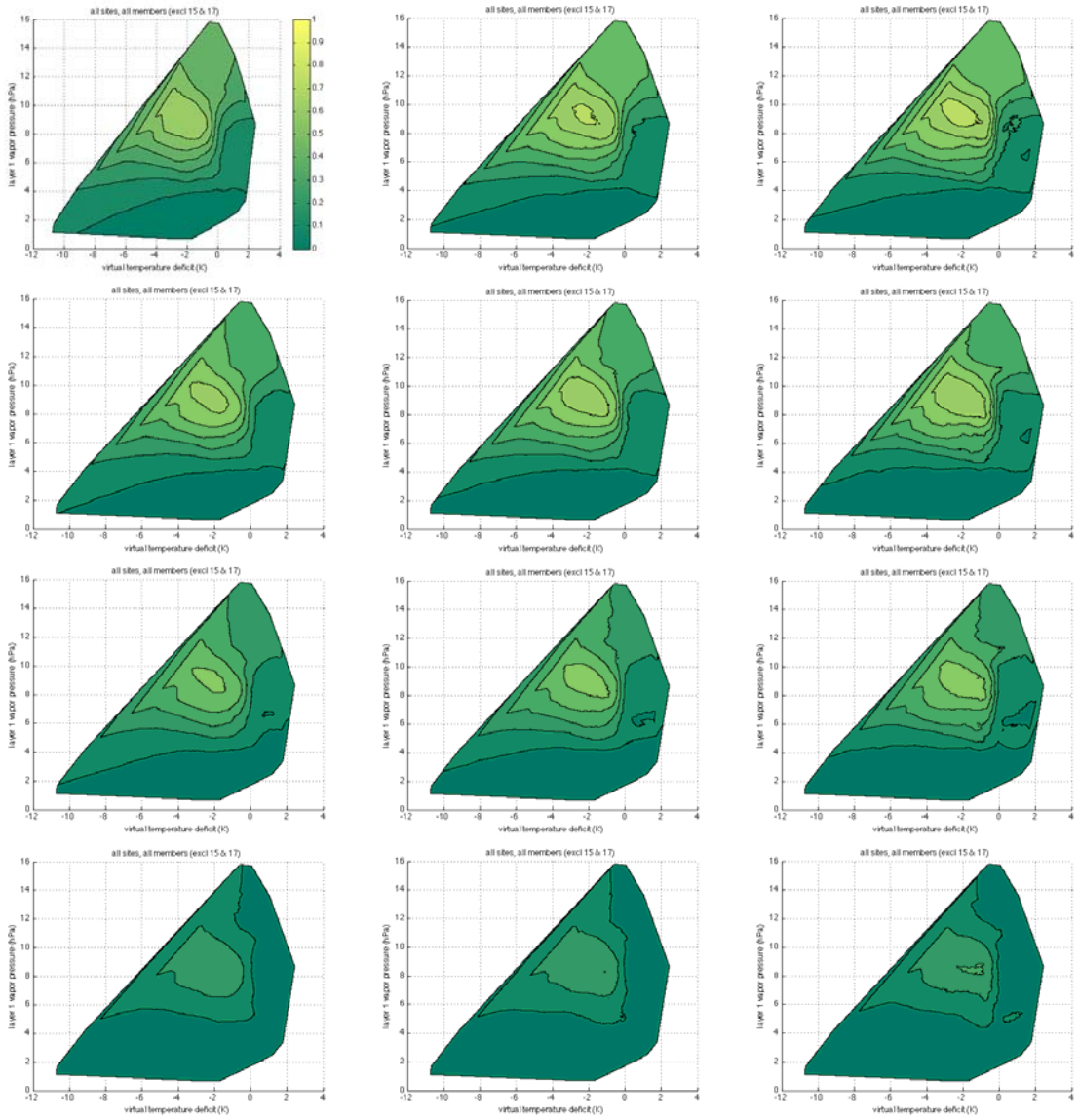

Figure 80. The joint parameter map from *JP_B* for the all regions domain, with bin sizes increased 50% (left column) and decreased 33% (right column). The center column shows unchanged bin sizes (i.e., identical to *JP_B*) for comparison. The rows correspond to each of the four β_e thresholds, increasing from top to bottom.

8. Joint Parameter Space, Best Universal

JP_U represents a best effort to maximize the transferability of the post-processing framework developed in this work. It cross-validates a parameter pair for each domain might have more transferability within it domain category because its predictive usefulness is believed to be less reliant on a particular aspect of the local

climatology than the best overall parameter pairs tested in *JP_B*. These parameter pairs are termed *universal* for this reason. Selection of these pairs also is highly subjective compared to simply identifying the largest variances as was done for *JP_B*, and is further complicated by the fact that the physical mechanisms behind the success of certain parameter pairs are not readily apparent. In addition to examining the plot variances of the parameter pairs, particular deference was given to parameter pairs using derived variables that entail a ratio (e.g., RH), difference (e.g., vapor pressure deficit), or time rate of change, as these were often more easily ascribed to reasonable physical mechanisms not heavily dependent on local climatology. In contrast, absolute variables such as vapor pressure usually appeared more likely to be associated with a localized phenomenon and were generally avoided.

The increased transferability sought in *JP_U* was not performed with inter-domain transferability in mind, but instead refers to transferability to a different locale with the same geographic region makeup, and perhaps during a different season. Therefore, the four-domain structure (coastal, valley, valley/mountain, all regions) is maintained in the development and testing of *JP_U*. As an example, *JP_U* for the valley/mountain domain is developed such that it might remain valid for a valley/mountain setting such as the Panjshar Valley/Hindu Kush Mountains of Afghanistan, but not for a coastal setting. It will be shown that the main differences among domains in *JP_U* are the probability maps themselves rather than the parameter pairs used.

It cannot be known for certain how truly universal these joint parameter maps are without a validation process involving other climatologies, which is not performed in this work. Obviously, the variances of the universal joint parameter maps are lower than those in *JP_B* (sometimes by more than 50%). However, they are presented as a practical alternative to *JP_B* for use in other climates or seasons much different from the training data.

The *JP_B* post-processing map for the valley region, which uses as it parameter pair saturation vapor pressure deficit and layer 1 vapor pressure depression, is not believed to be particularly specific to the local climatology of the test sites. Therefore, no

JP_U experiment is performed in this region. The *JP_U* experiments for the coastal, valley/mountain, and all regions domains are presented below.

a. Coastal Optimization

Figure 81 shows post-processing maps for the coastal region believed to provide more universal function than the 2-m vapor pressure predictions used in *JP_B*. *JP_U* once again leverages the more accurate 2-m predictions in this region, using 2-m RH paired with virtual temperature deficit for the joint space.

We saw in *RH_P* that 2-m RH is a reasonable predictor of fog, especially as it pertains to ruling out fog when predicted RH values are low. Output probabilities generally increased at higher predicted RH values, but topped out at only 0.252 at the highest RH predictions (for the lowest β_e) in that experiement, barely higher then climatological incidence of 0.200 for the entire plot. Figure 81 shows we might improve resolution at these high RH predictions by utilizing the predictions of virtual temperature deficit. This variable was used in the valley/mountain domain and the all regions domain of *JP_B* partly for its value in predicting radiation inversions crucial for fog in the valley region. In the coastal region, it is also believed to signaling marine boundary layer stability as determined by the air-sea temperature difference. This function was performed by the 2-m vapor pressure predictions in *JP_B*, but virtual temperature deficit appears to be an adequate substitute for this purpose that is likely less location-specific.

The mechanism by which this variable indicates stability conditions near the coast is fundamentally the same as with a radiation inversion in a valley: the 2-m temperature predictions will have values in between the layer 1 predictions and the surface (soil or sea) temperature in the member, and so negative deficits are an indication that the surface temperature is likely colder than the layer 1 temperature in the member, and a stable lower boundary layer exists. A stable boundary layer alone is not sufficient for fog in the coastal region, but Figure 81 indicates an incidence >0.4 at the lowest β_e threshold if the predicted RH is also >0.8.

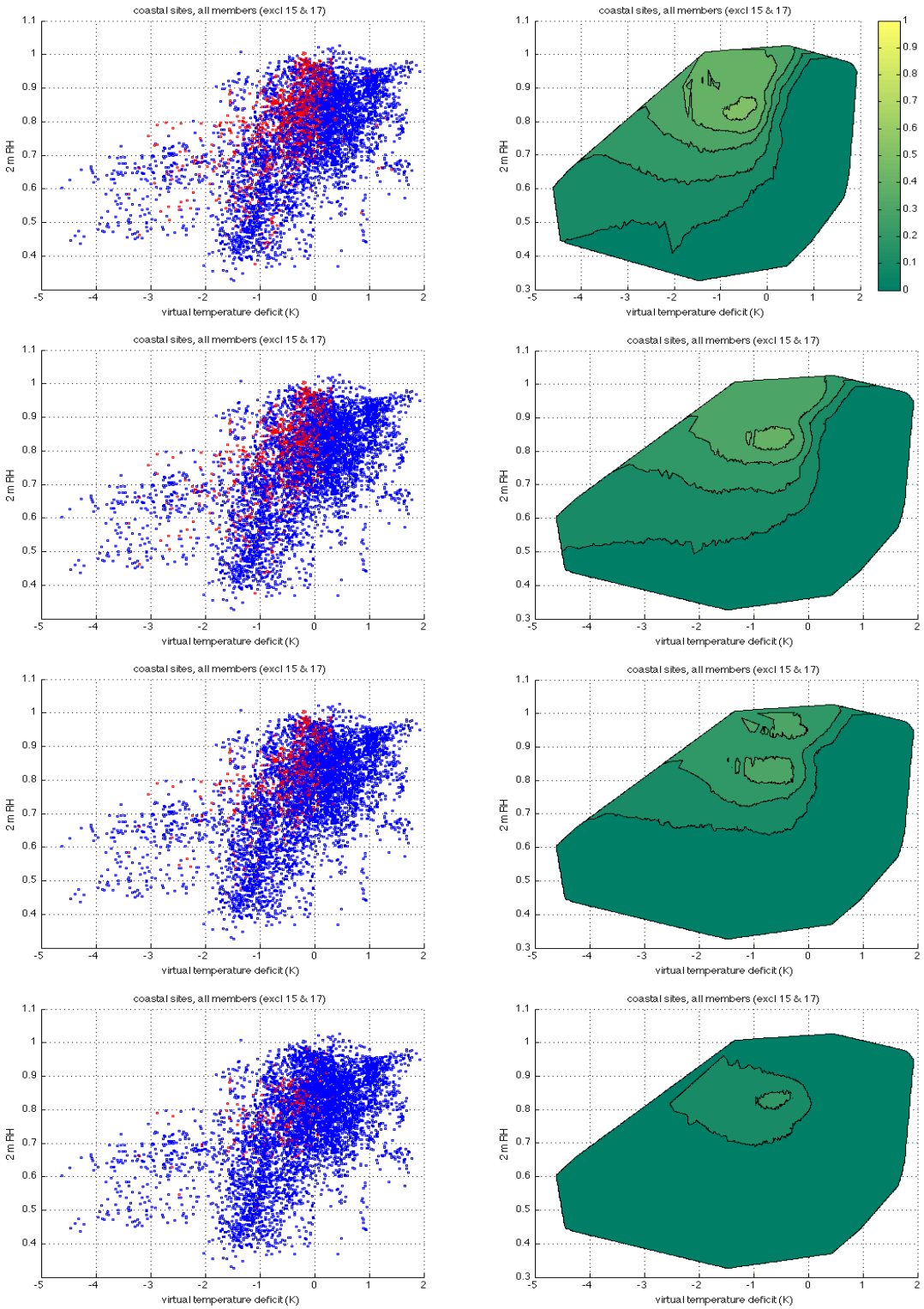

Figure 81. Same as in Figure 71, but for the coastal region. The parameters are virtual temperature deficit and 2-m RH. The rows correspond to each of the four β_e thresholds, increasing from top to bottom.

Of course, this region is heavily influenced by the stability over water, but the sites themselves are still on land, and are accordingly affected by diurnal radiative forcing. Nighttime radiation inversions certainly do exist and play a part in the predictive usefulness of virtual temperature deficit predictions. Figure 81 indicates that the incidence of fog is very low when predicted virtual temperature deficits are >0.5 K, which tend to occur with either cold outbreaks (during which the marine boundary layer is unstable) or post-sunrise radiative heating of the land.

An important consideration to the predictions at the coastal sites is that they are bi-linearly interpolated from two NWP model grid points over land and two over water. We will not explore all the implications this might have, but in regard to lower boundary layer stability predictions, they represent some mixture of the offshore marine layer structure and the terrestrial structure within a few kilometers of the coast. Whether this is a beneficial or detrimental configuration is not known, but the virtual temperature deficit predictions seems to offer some measure of the stability that, when paired with the 2-m RH predictions, provide a useful joint parameter space for post-processing. Note that the behavior of the virtual temperature deficit predictions could change if, instead of a bi-linear interpolation, the nearest grid point to the site were used thereby rendering the influence of radiative forcing stronger (if the nearest point were over land) or weaker (if it were over water).

b. Valley/Mountain Optimization

A more universal joint parameter space was sought for the valley/mountain domain that might have more certainty in its transferability. Joint parameter space that can effectively separate the mountain predictions from the valley predictions are generally found to provide the highest variance in output probabilities since the predictions from each region otherwise tend to dilute each other. Layer 1 RH was found to be the most promising among universal options, and is paired with virtual temperature deficit to comprise the *JP_U* test for this domain (Figure 82).

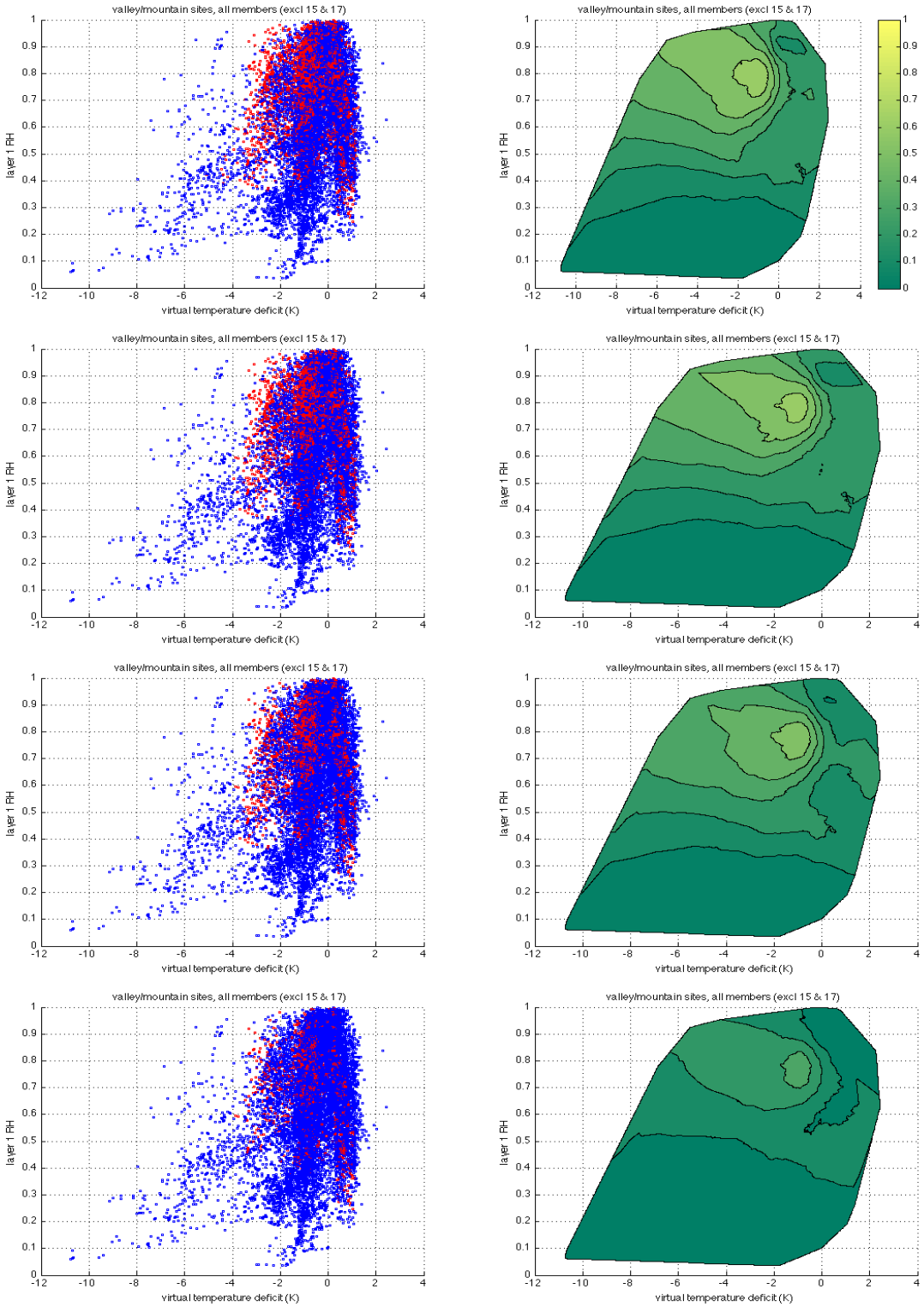

Figure 82. Same as in Figure 71, but for the valley/mountain domain. The parameters are virtual temperature deficit and layer 1 RH. The rows correspond to each of the four β_e thresholds, increasing from top to bottom.

In *JP_B*, layer 1 vapor pressure predictions in the joint parameter space served the role of parsing the mountain data from the valley data, providing a relatively undiluted portion of the space in which virtual temperature deficit predictions could be used to detect likely radiation inversions in the valley region. The questionable transferability of this space stems from the fact that the range of vapor pressures for which the inversions appear important (>6 hPa) and for which fog can be virtually ruled out (<4 hPa) would seem to be dependent on the general temperature and moisture climatology of the domain. For instance, the applicability of the *JP_B* map is not entirely clear if the background climatology were increased 5–10 K with proportional increases in moisture (as might be expected in a different locale or season). In this hypothetical scenario, perhaps the range of critical vapor pressures indicated by the map would need adjustment to account for the changes. Alternatively, it could be that mountain fog would in fact be more likely in this scenario and the *JP_B* map is reasonably applicable in assigning high probabilities prescribed by the higher vapor pressure predictions. Unlike the coastal domain, where the *JP_B* map is believed to closely dependent on local water temperature, the location-specificity of the *JP_B* map in this domain is less clear and warrants further examination.

Compared to *JP_B*, there is significant unavoidable overlap of predictions from each region in the joint parameter space of *JP_U*, resulting in its variance being 54% lower than that of *JP_B*. The degradation is most evident at upper portions of the space, where the mountain data contains a substantial amount of high RH predictions that have reduced fog probabilities by approximately 0.2–0.3 at the lowest β_e threshold (0.29 km^{-1}) compared to *JP_B*.

Still, the majority of the mountain predictions have layer 1 RH values <0.6, possibly providing adequate separation of the two regions' predictions and giving this map some merit in the combined domain for the promise of better transferability. The cross-validation results will show that moderate dilution of the post-processing map caused by overlapping of the two regions' predictions is more forgiving in the valley region than it is in the mountain region.

c. All Regions Optimization

Detecting inversions using predictions of virtual temperature deficit from the NWP model has been shown to be effective in the individual coastal and valley regions, as well as the combined domains. In *JP_B*, this parameter was paired with layer 1 vapor pressure to effectively parse the mountain predictions from the rest of the data. For *JP_U*, we use virtual temperature deficit paired with layer 1 RH (Figure 83), just as we did in the valley/mountain domain. Adding the coastal predictions to this map does not produce drastic changes to the output probabilities compared to Figure 82, with the most significant change being the lowering of probabilities when the predicted virtual temperature deficit is >0 (i.e., inversions are not predicted, which rarely result in observed fog in the coastal region). The use of layer 1 RH instead of 2-m RH (which is more accurate and generally has more predictive usefulness than layer 1 in the coastal region) is due to its better compatibility with the valley fog data. Since fog in the valley region is most likely with layer 1 RH predictions of 0.7–0.8, the negative biases of the coastal region layer 1 RH predictions cause many of its observed fog data to be located in the same portion of the space.

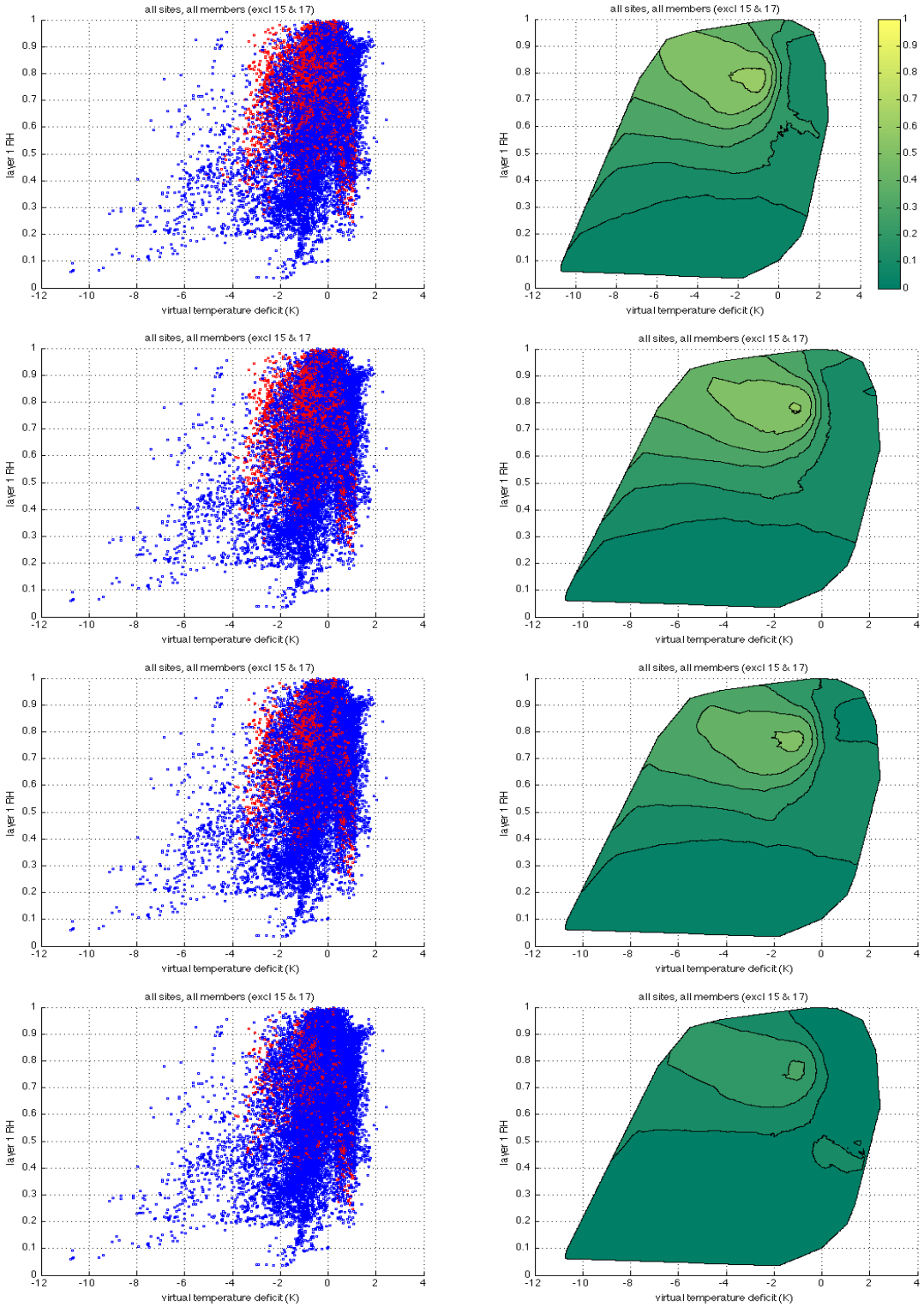

Figure 83. Same as in Figure 71, but for the all regions domain. The parameters are virtual temperature deficit and layer 1 RH. The rows correspond to each of the four β_e thresholds, increasing from top to bottom.

B. VERIFICATION METHODOLOGY

To verify the experiments, the most difficult test is sought without an unreasonable computational demand. A modified version of "leave one out" cross-validation is used, with the predictions grouped along the mode that produces the most variation in output among the groups (and therefore likely the lowest verification skill in cross-validation).

With the exception of *SCW*, each of the experiments listed in Table 6 involve a development process during which optimal thresholds or joint parameter space maps were designated based on the entire set of predictions subject to the post-processing (i.e., every prediction with $q_c \leq 8.5 \times 10^{-4}$ g m^{-3}). Cross-validation is the process of dividing the data into a developmental portion, for which the thresholds or maps are re-optimized, and a testing portion, for which the re-optimized technique can be verified on data independent of its development (Stull, 1988). This provides some indication as to how much overfitting has occurred during development, and therefore how well the technique might predict outcomes when employed with new input data.

To improve the fidelity of the verification, cross-validation can be performed multiple times, where the developmental and testing portions of the data are changed each time, and the verification results of each of these repetitions are averaged. "Leave one out" is a special case of this type of verification where the number of repetitions is equal to the number of predictions, and the testing portion of the dataset is a single prediction that changes with each repetition. The result is that each prediction is tested exactly once using developmental data from all the other predictions.

A proper leave one out cross-validation requires a tremendous computational demand for large datasets and is not feasible here. Therefore, the number of repetitions is reduced by verifying each of several *groups* of predictions exactly once using developmental data from the other groups, and averaging the results. To group the predictions, three modes were considered: groupings by member, by site, and by case day. The variance of output probabilities from the all regions domain post-processing map in *JP_B* was computed among the groups for each of the three grouping modes. This variance, as well as the probability output map for each group, are shown in Figure

84 (grouping by member), Figure 85 (grouping by site), and Figure 86 (grouping by case day – for brevity, only maps from selected case days are shown).

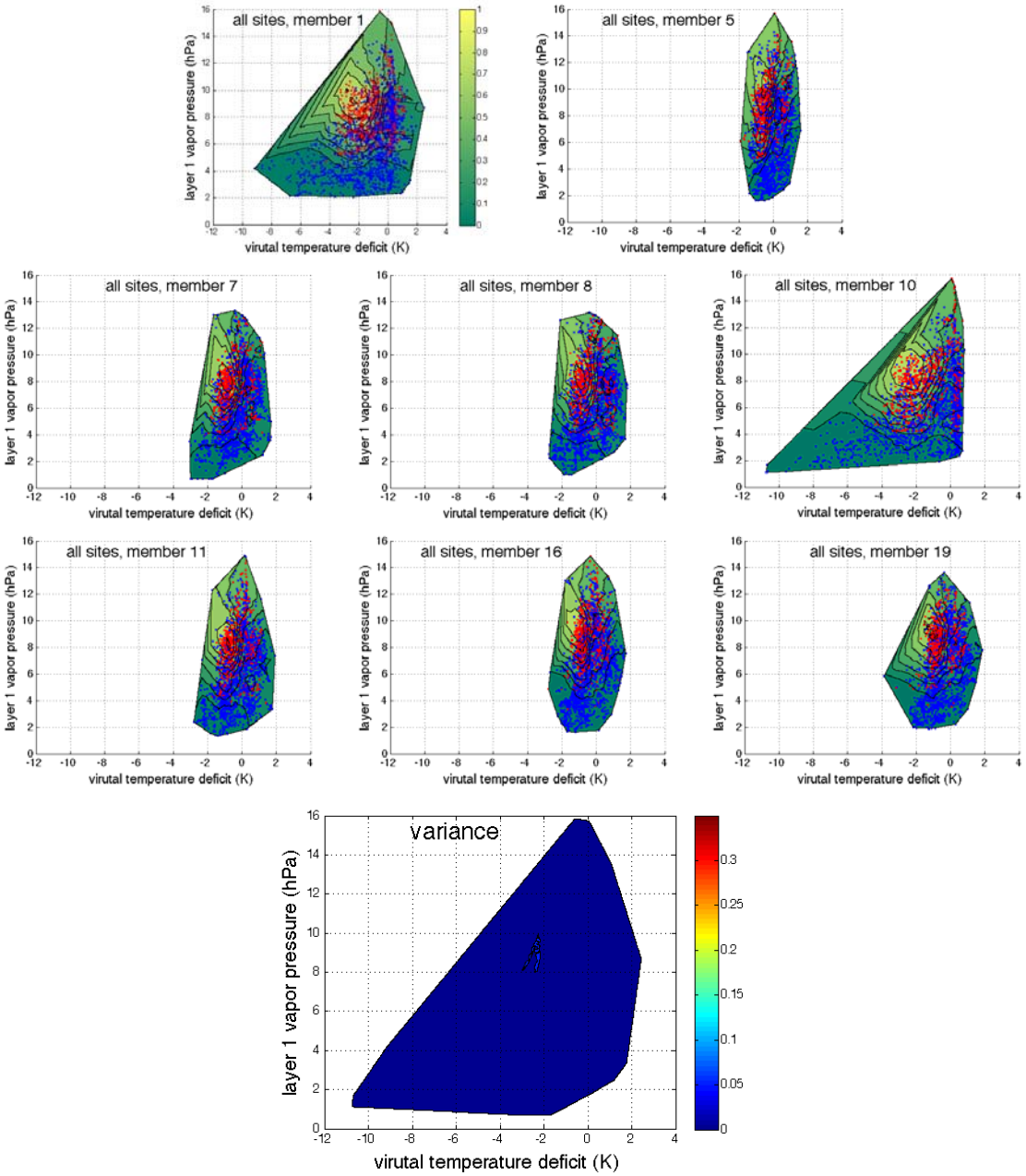

Figure 84. Observed fog (red) and no fog (blue) plotted in the all regions domain joint parameter space of *JP_B* for each member. Contouring is based on bin sizes equaling one-twelfth of the total data in each plot. The variance of the probability output among all the plots is shown in the bottom panel. The first six hours of each case are excluded.

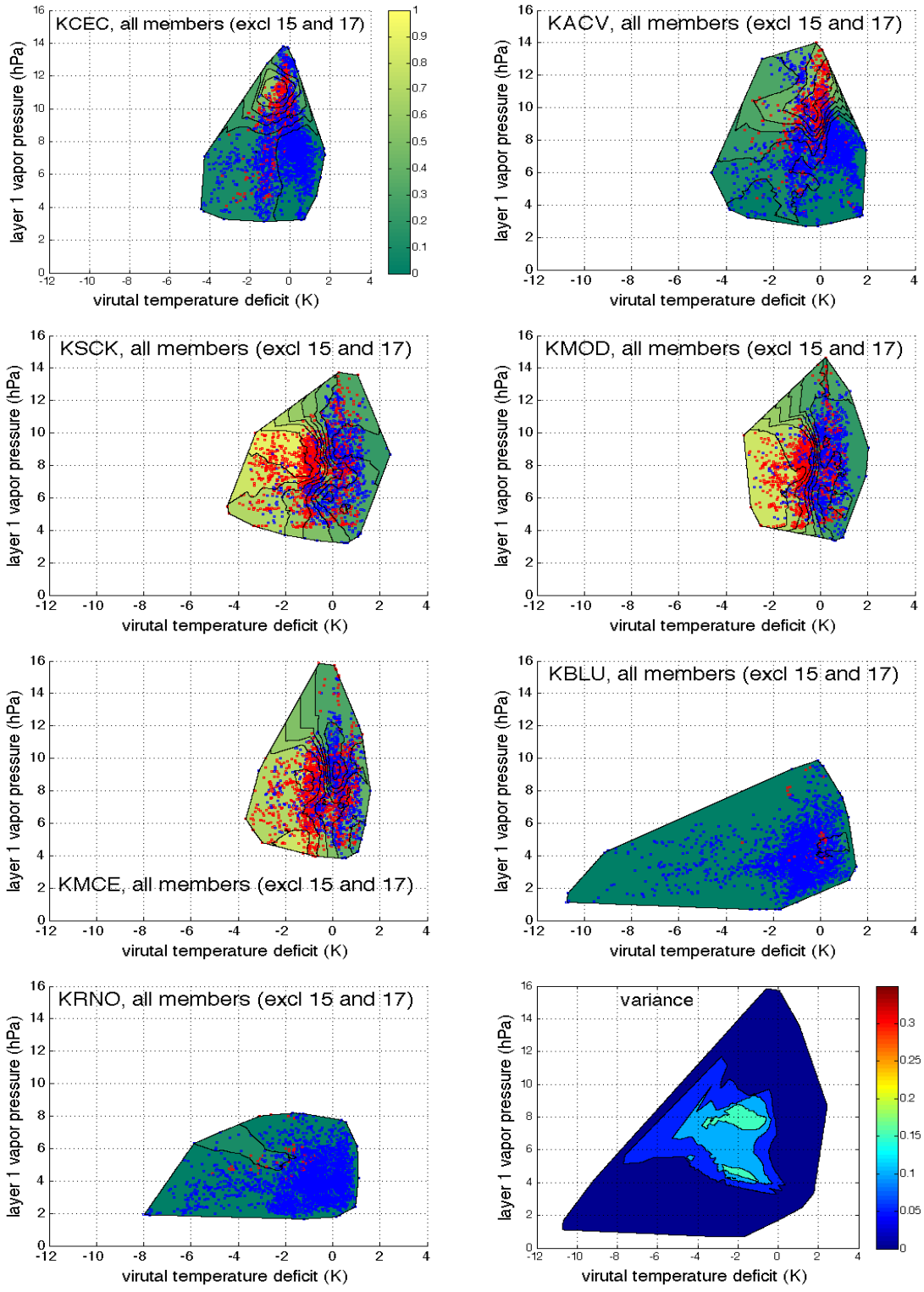

Figure 85. Same as in Figure 84, but for each site.

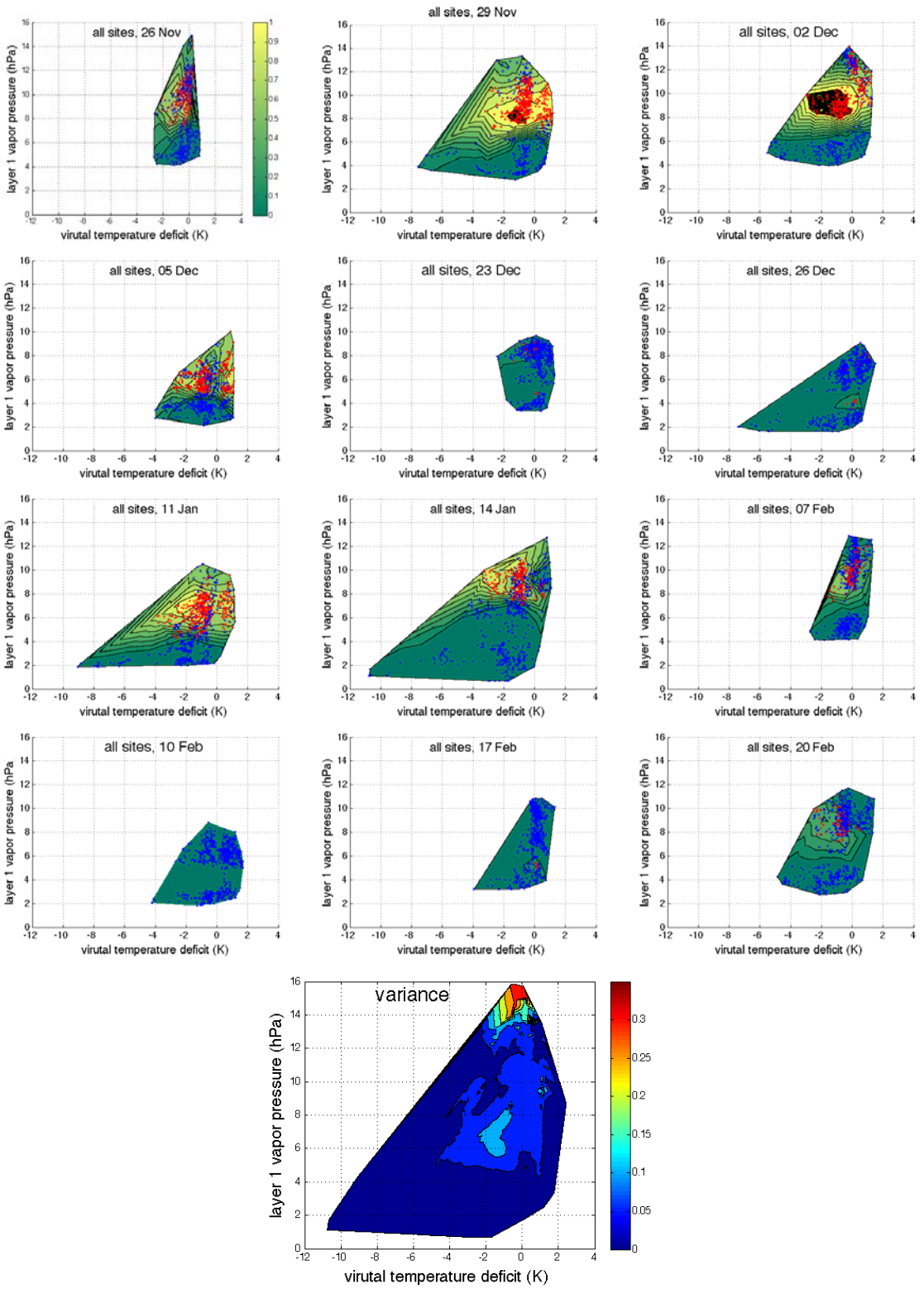

Figure 86. Same as in Figure 84, but for selected case days. The variance plot shows the variance among all 29 case days.

As is shown in the figures, rarely does the domain of predictions for any individual map cover the entire joint parameter space represented by all the data. For example, the *JP_B* map for KRNO (Figure 85) does not include any predictions, and therefore has no probability output, for layer 1 vapor pressure predictions >9 hPa. To account for this, the variance at any point in the space is calculated using only the maps producing probability output at that point; maps without data at the point were left out of the computation.

Among the three modes tested, there is comparatively low variance in probability output among the members (Figure 84), and so this mode is ruled out for grouping the predictions for cross-validation. The variance among the sites (Figure 85) has two local maxima in the space, both corresponding to predicted temperature inversions (where the virtual temperature deficit is <0). This first of these is at vapor pressures of 4–5 hPa, where the higher variance is caused by overlapping valley data (with high fog probabilities) and mountain data (with low fog probabilities). The second local maximum occurs near 8 hPa, which is dominated by predictions from the coastal and valley sites. This portion of the space produces low fog probabilities in the coastal region (where the incidence of fog does not significantly increase until predicted vapor pressure is >9 hPa), and high fog probabilities in the valley region, together accounting for the larger variance.

Variance among the case days (Figure 86) shows very high variance at predicted vapor pressures >13 hPa. However, this portion of the space represents relatively few predictions and is therefore of less importance than portions of the space with higher data density. For this reason, the increased variances near the center of the plot are of more significance, and unlike with the variances among the sites, the region of higher variances among the case days extends to the positive side of the x-axis; that is, when predictions of vapor pressure depression are >0. These typically correspond to low fog probabilities associated with post-sunrise heating or (in the coastal region) cold air outbreaks. However, there are several cases (e.g., 29 Nov, 2 Dec, 11 Jan) when valley fog persisted past sunrise, well after the predicted inversion was destroyed, creating high fog probabilities in those cases and increasing the variance in the output probabilities in that

portion of the joint parameter space. Since this portion of the plot also has high data density, the increased variance there is significant.

In order to measure the total variance of the entire joint parameter space for each mode, weighted by data density, the variance at the location of every prediction in the joint space was summed and averaged. The results are shown in Table 9. Although the groupings by site produced a larger area of high variance near the center of the joint parameter space, the variance among the case days is higher in the portions of high data density, resulting in the highest overall variance among the three modes. The same calculation was performed on each mode using the coastal domain map and the valley/mountain domain map from *JP_B*[5], with the case day mode producing the largest variance in each domain.

Table 9. Total variance of probability output for individual *JP_B* joint parameter space maps when grouped along each of the three modes. The variance of each map is computed by averaging the variances at the location of each prediction in the joint space. The data for the coastal domain and valley/mountain domain includes predictions from members 15 and 17.

Mode	All Regions Domain	Coastal Domain	Valley/Mountain Domain
Grouping by Member	0.0064	0.0038	0.0051
Grouping by Site	0.0575	0.0013	0.0740
Grouping by Case Day	0.0623	0.0552	0.0862

Based on these results, "leave one out" cross-validation is performed along the case day mode, such that each case day is verified using the post-processing technique that was optimized with data from the other 28 case days. This is true for all aspects of the optimization for each experiment in Table 6; for example, in *BiasRH_D*, the bias *and* the optimal threshold are computed for each repetition from the 28 case days of developmental data prior to verifying the one case day of testing data. Using this same approach for all the experiments permits valid comparison among techniques. The single

[5] The variance calculation in these two domains includes predictions from members 15 and 17. It is believed their removal would not convincingly change the conclusion that the largest variance is achieved when the predictions are grouped by case day.

exception is *SCW*, for which no cross-validation is required because there is no optimization or training of the technique. For this experiment, the technique is verified by simply applying it to all the predictions.

VI. RESULTS

The results of the cross-validation, which are presented separately for each region, are given in Figures 87–100. Table 10 summarizes the organization of the results among these figures.

To facilitate comparison among the experiments, each figure contains the plotted results from all the experiments, using the symbols and line types given in Table 6 (for convenience, Table 6 is reprinted here as Table 11). Discussion will mainly focus on the RPSS (Figures 87 and 88) and the verification results at the lowest β_e threshold corresponding to a daytime visibility of 6.5 mi (Figures 89–91), but the results at the other three β_e thresholds are also included in the suite of figures and are referenced when notable.

In a few instances, BSSs for certain experiments are significantly lower (values <-3) than the majority of the results shown, and these are often not plotted or only partially plotted. This is especially common at the higher β_e thresholds (2.75 and 0.875 mi daytime visibility) in the mountain region, where several of the techniques performed poorly. Instead, results of these poorest-performing experiments are adequately captured by their verification at other β_e thresholds, as well as the RPSSs shown in Figure 87, which includes all the experiments for each region.

Table 10. Summary of results figures.

Figure Number	Region	Description
87	All	RPSS across all four β_e thresholds in each region, zoomed out to show all data
88	All	Same as above, but zoomed in to show detail for highest-performing experiments
89	Coastal	Reliability, resolution, uncertainty, and BSS at lowest β_e threshold (0.29 km^{-1})
90	Valley	
91	Mountain	
92	Coastal	Reliability, resolution, uncertainty, and BSS at second β_e threshold (0.41 km^{-1})
93	Valley	
94	Mountain	
95	Coastal	Reliability, resolution, uncertainty, and BSS at third β_e threshold (0.68 km^{-1})
96	Valley	
97	Mountain	
98	Coastal	Reliability, resolution, uncertainty, and BSS at fourth β_e threshold (2.10 km^{-1})
99	Valley	
100	Mountain	

Table 11. (Reprint of Table 6) Summary of post-processing techniques tested, with symbols used in figures 87-100. All the techniques are first developed and tested without regional specificity, and some are then refined for specific regions or region combinations, which are listed.

Symbol	Name	Description	Optimization Domains
	Cntrl	Unaltered NWP predictions	N/A
	SCW	Small, non-zero cloud water values	All regions
	RH_D	RH threshold, deterministic	All regions, coast, valley, valley/mountain
	BiasRH_D	RH threshold with 2-m temperature bias correction, deterministic	All regions, coast, valley, valley/mountain
	RH_P	RH, probabilistic	All regions, coast, valley, valley/mountain
	BiasRH_P	RH with 2-m temperature bias correction, probabilistic	All regions, coast, valley, valley/mountain
	JP_B	Joint parameter space, best overall	All regions, coast, valley, valley/mountain
	JP_LB	Joint parameter space, large bins	All regions
	JP_SB	Joint parameter space, small bins	All regions
	JP_U	Joint parameter space, best universal	All regions, coast, valley/mountain
Line Type Used in Results to Denote Domain Optimization			
	All regions domain		
	Individual coast or valley domain		
	Combined valley/mountain domain		

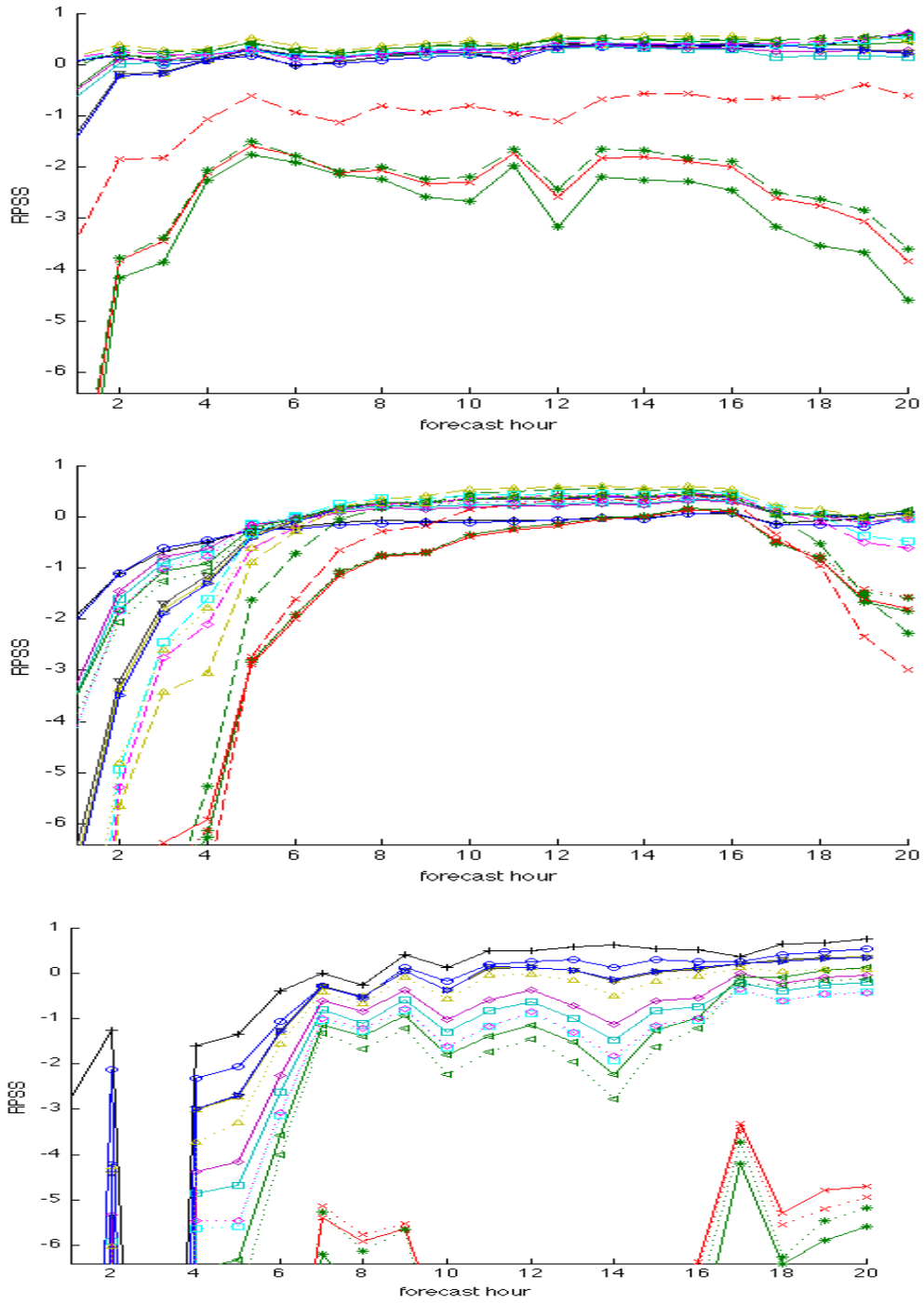

Figure 87. Cross-validation Ranked Probability Skill Scores in the coastal (top), valley (center), and mountain (bottom) regions for each experiment. Plotted symbols are used according to Table 6. Solid lines indicated experiments optimized for all regions, dashed lines (in the coastal and valley regions) are optimized for that specific region, and dotted lines (in the valley and mountain regions) are optimized for the valley/mountain domain.

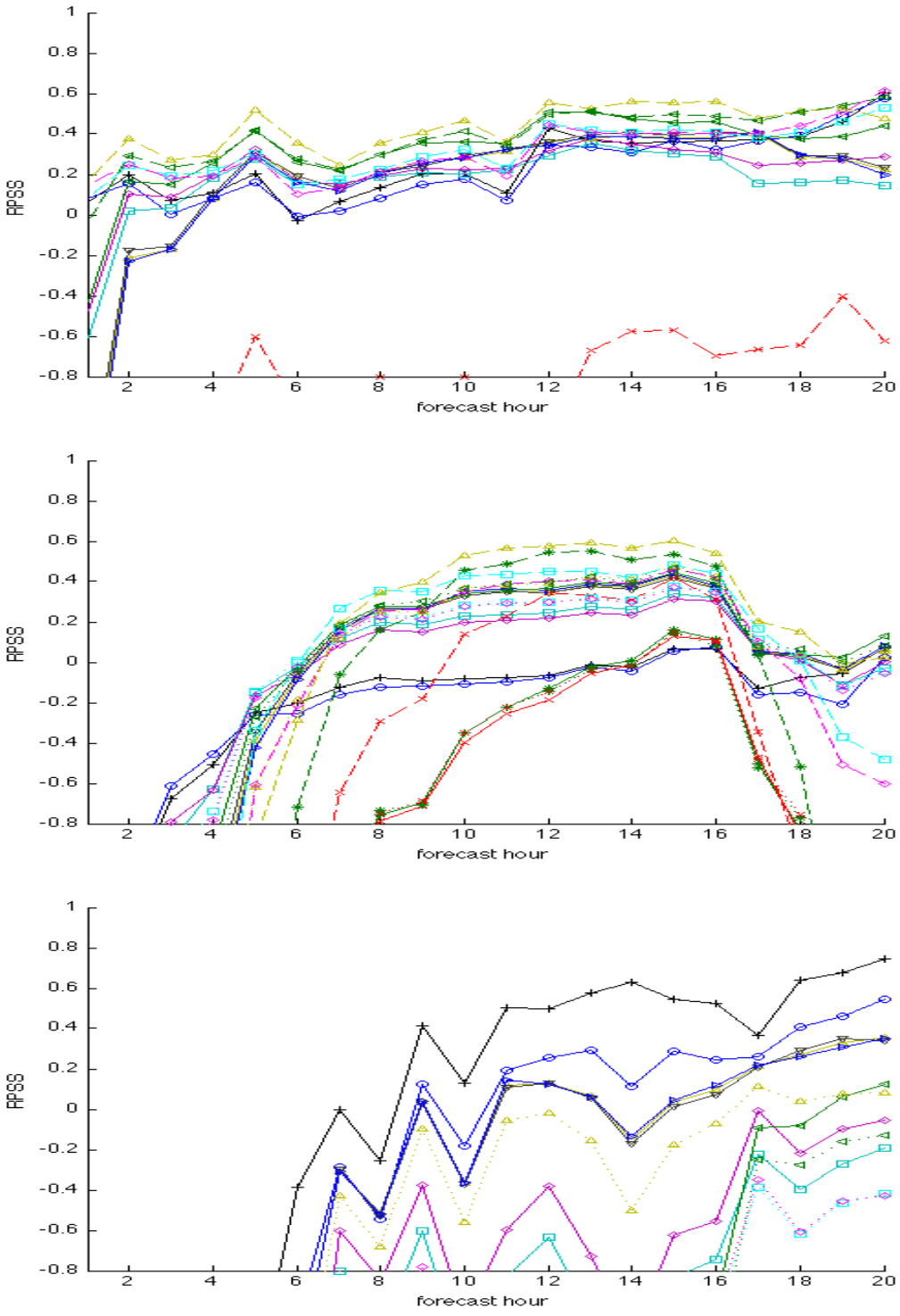

Figure 88. Same as Figure 87, but zoomed in to show more detail for the best-performing experiments.

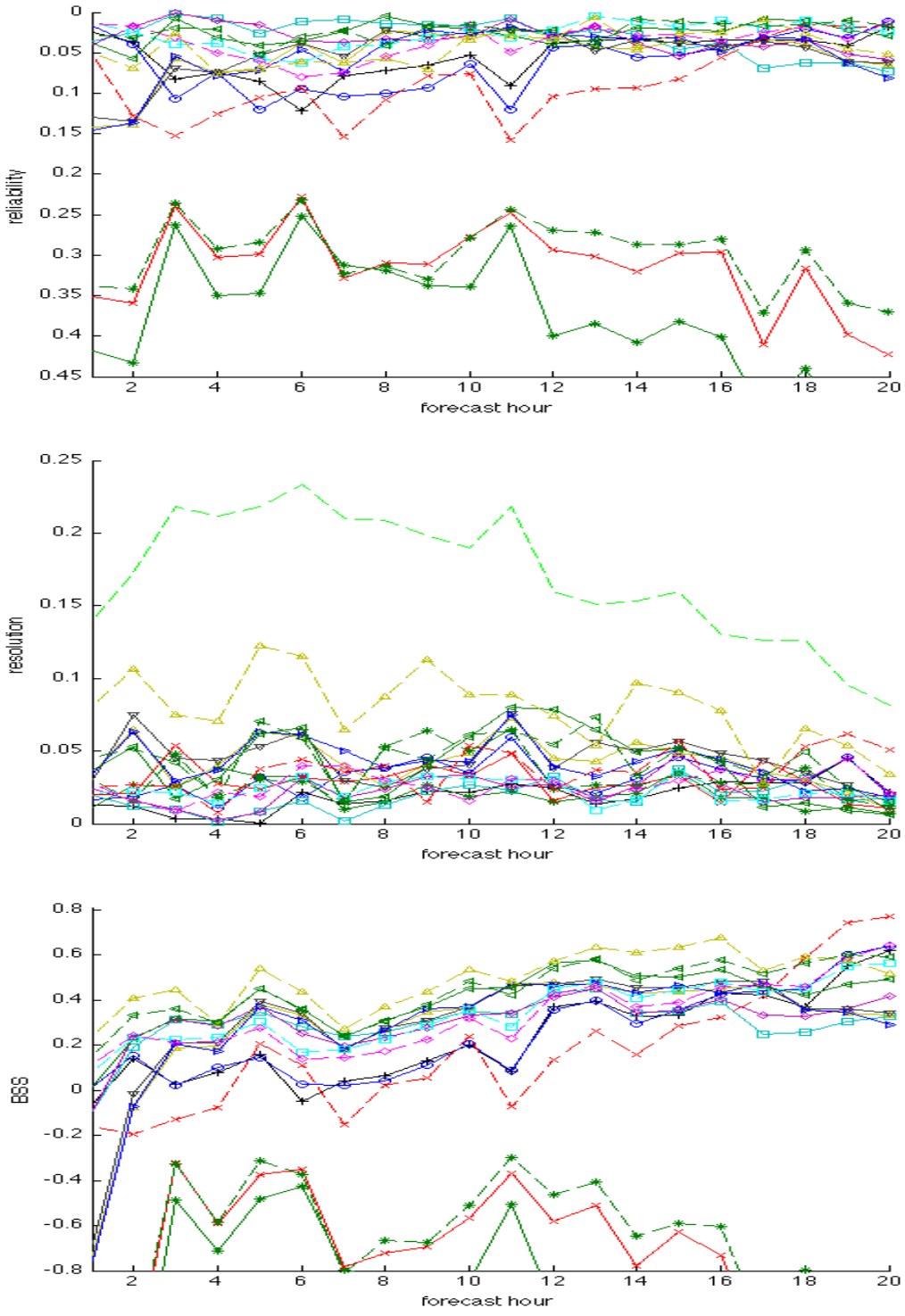

Figure 89. Cross-validation reliability (top), resolution (center), and Brier Skill Score (bottom) at the lowest β_e threshold (0.29 km^{-1}) in the coastal region for each experiment. In the center panel, the uncertainty is indicated with the dashed light green line.

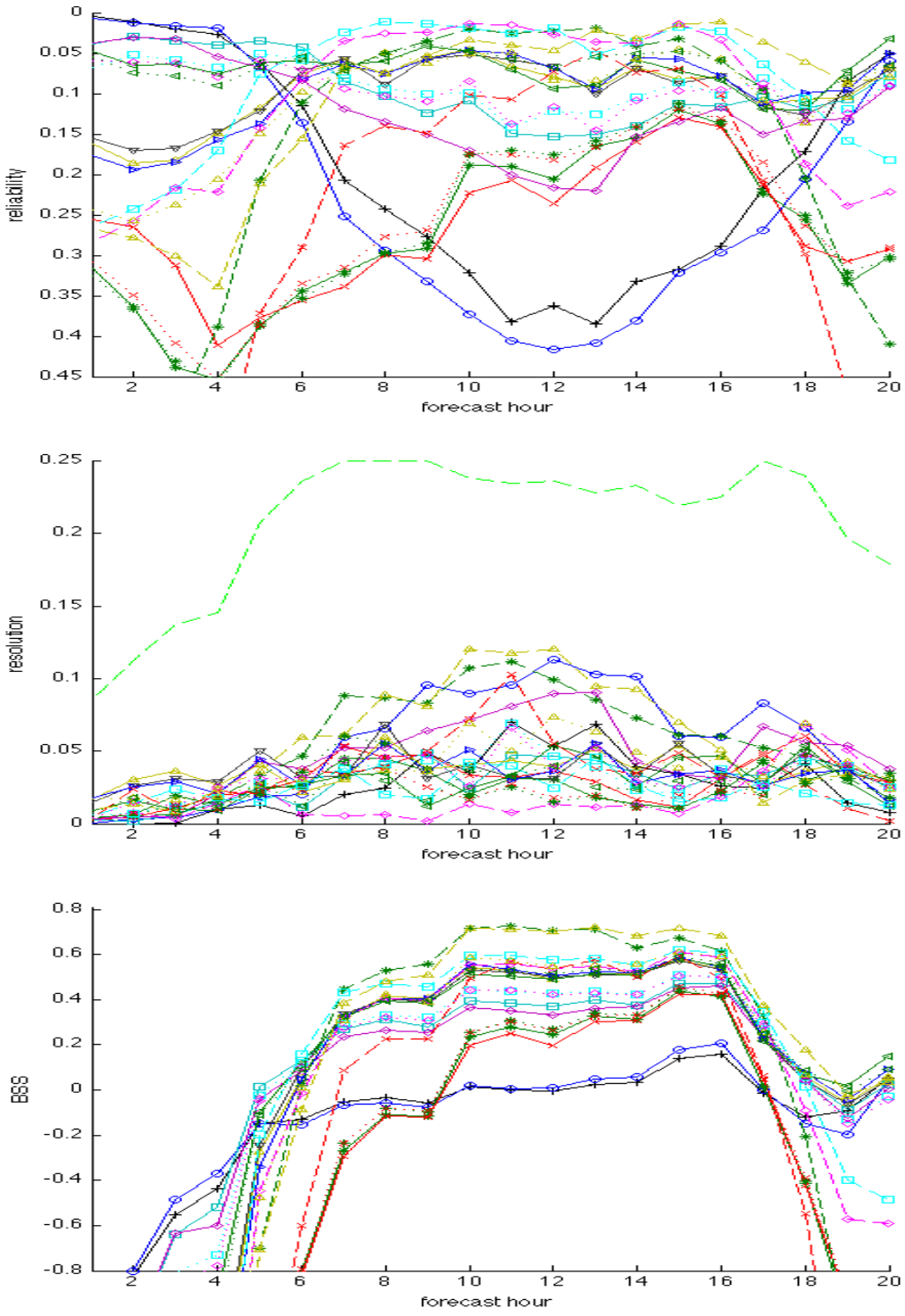

Figure 90. Same as in Figure 89, but for the valley region.

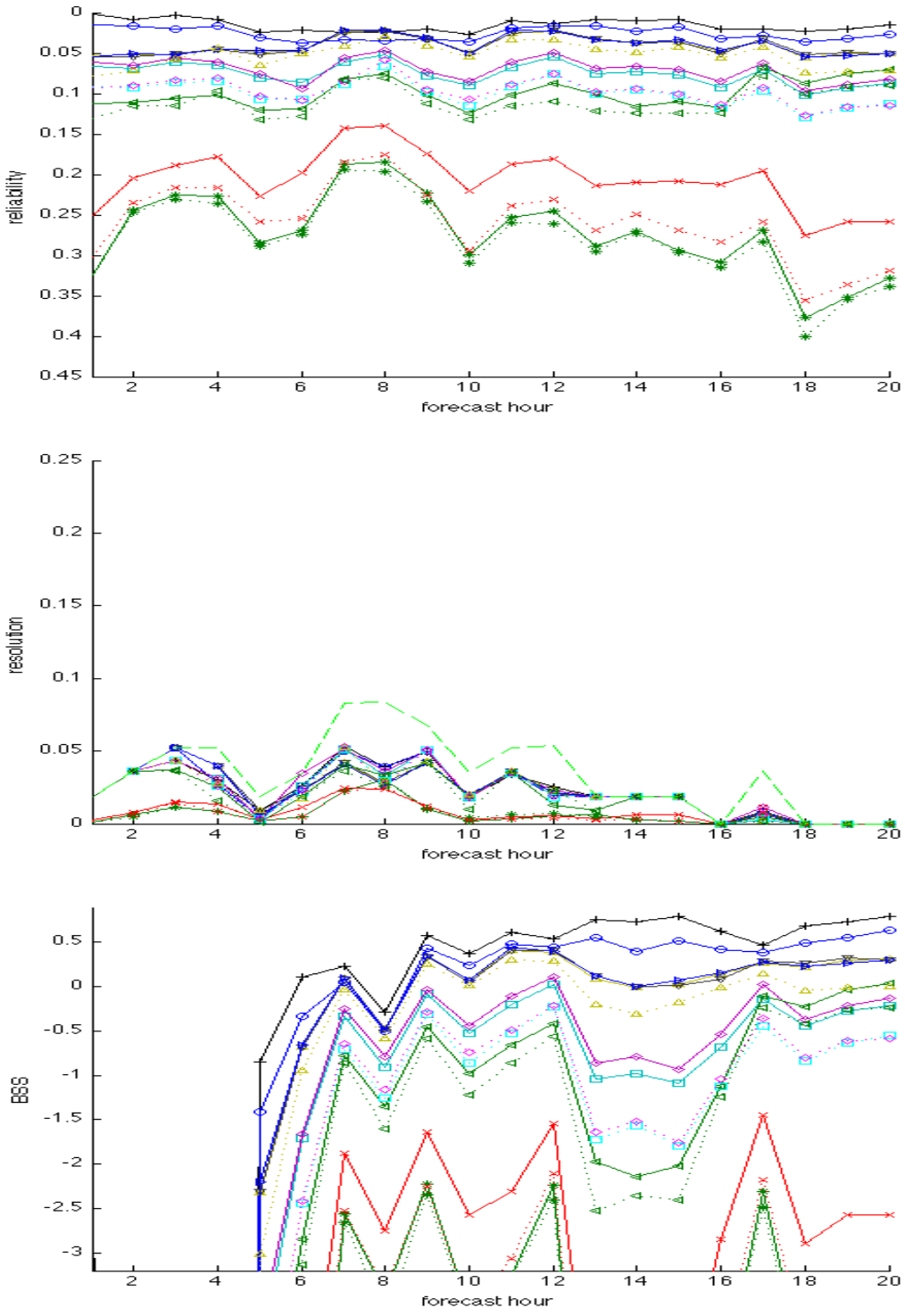

Figure 91. Same as in Figure 89, but for the mountain region. Note that in the bottom panel, the y-axis extends to lower values than in Figure 89 and Figure 90 to accommodate especially poorly-performing experiments in this region.

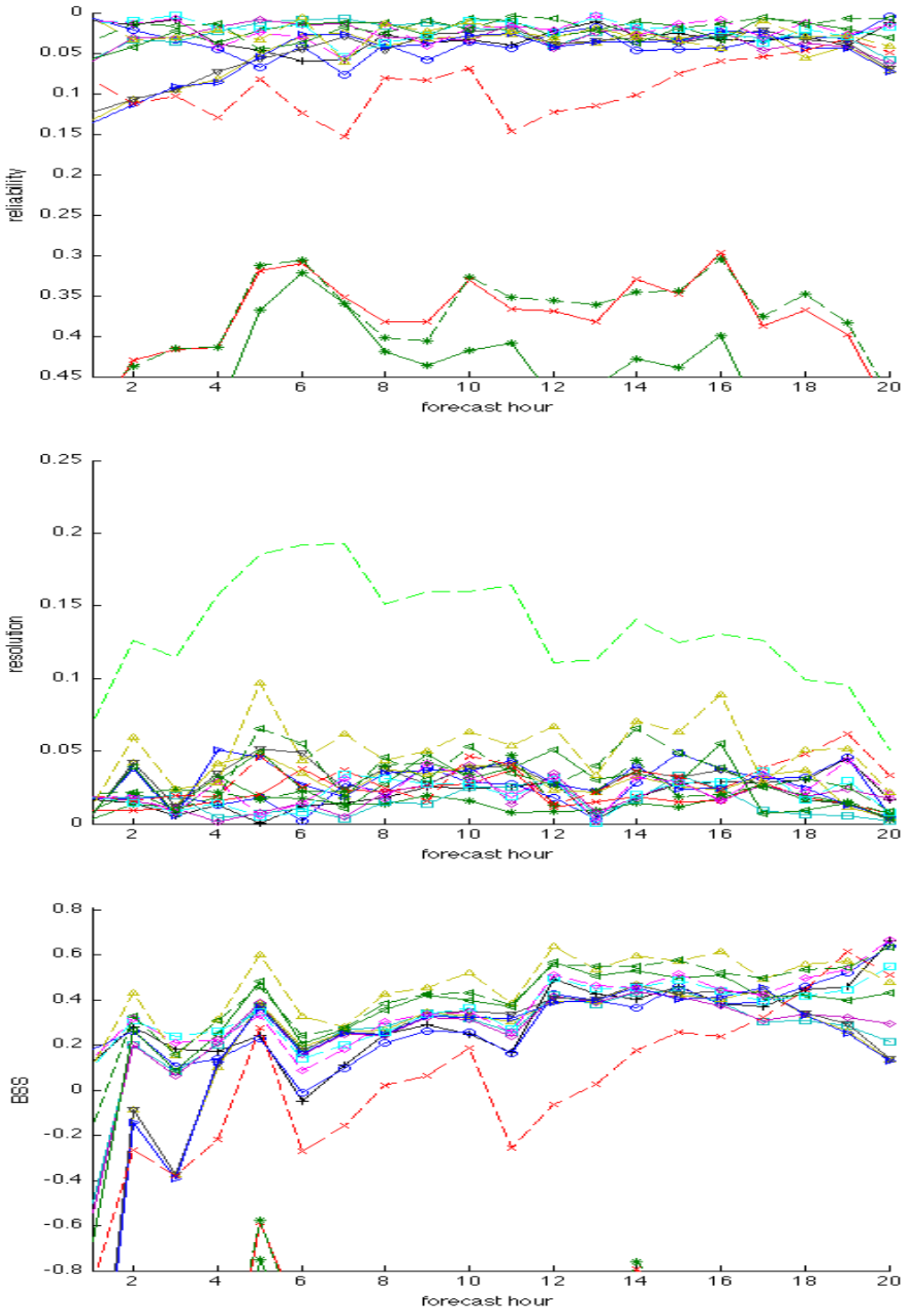

Figure 92. Same as in Figure 89 (coastal region results), but at the second β_e threshold (0.41 km^{-1}).

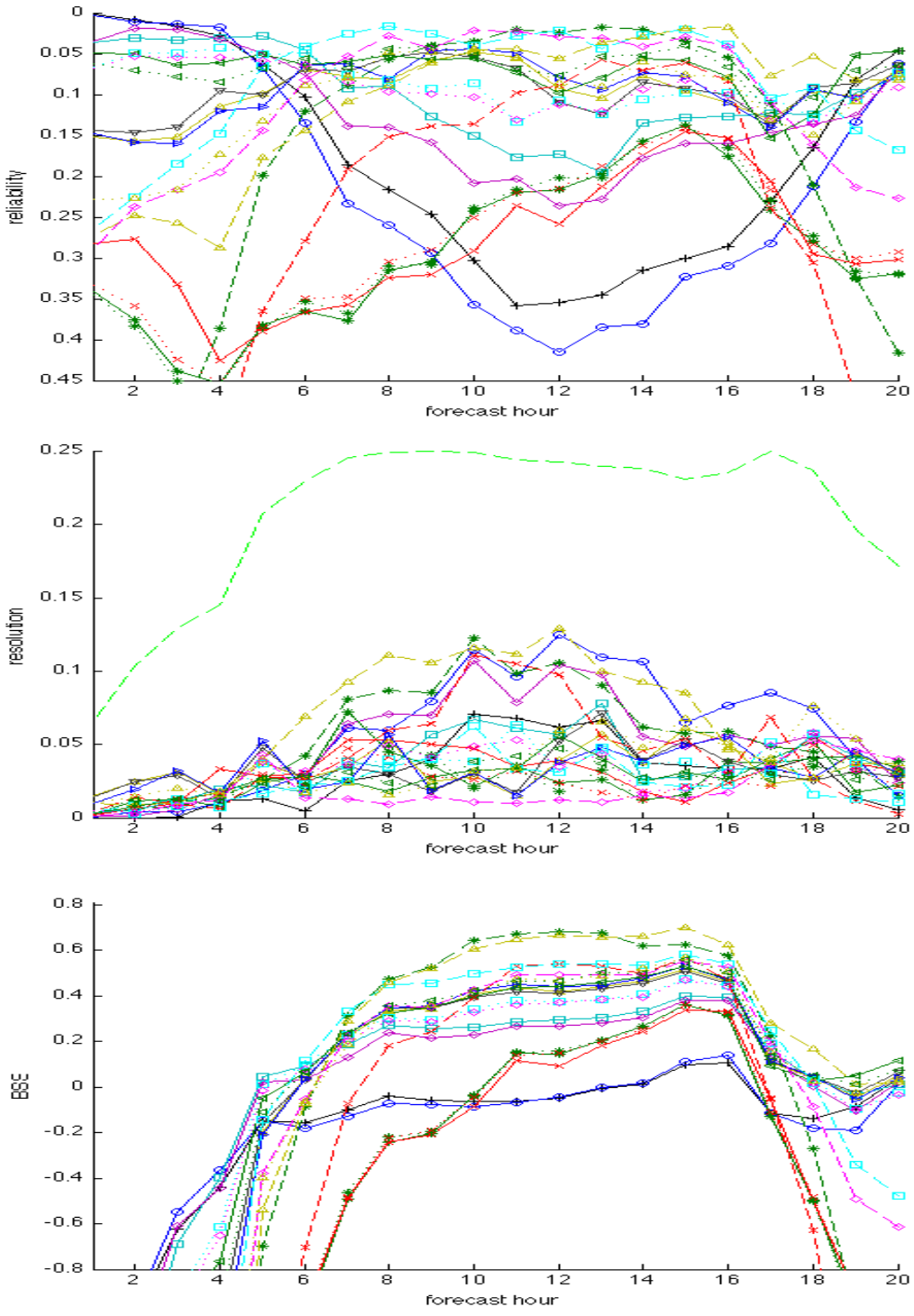

Figure 93. Same as in Figure 90 (valley region results), but at the second β_e threshold (0.41 km^{-1}).

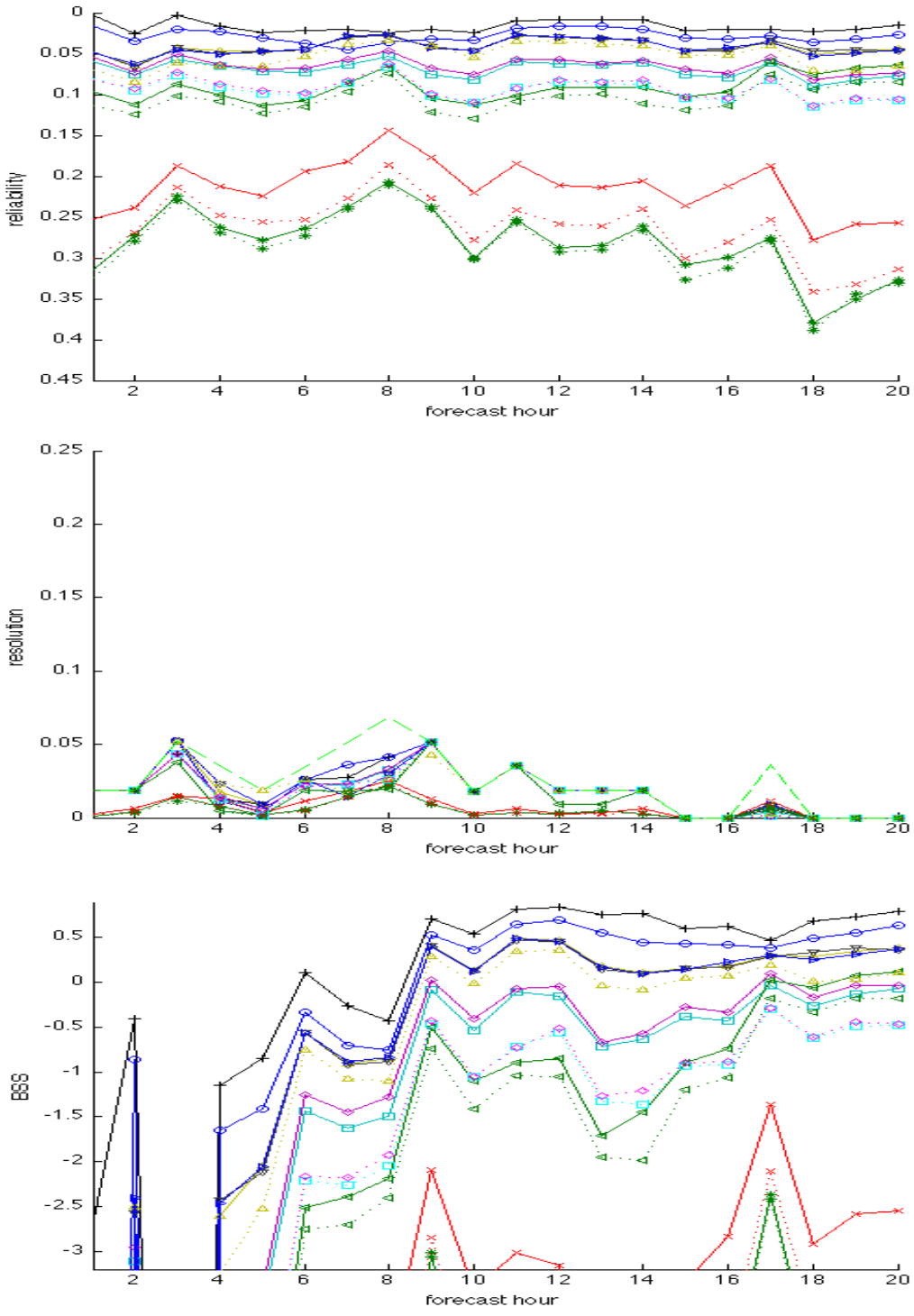

Figure 94. Same as in Figure 91 (mountain region results), but at the second β_e threshold (0.41 km^{-1}).

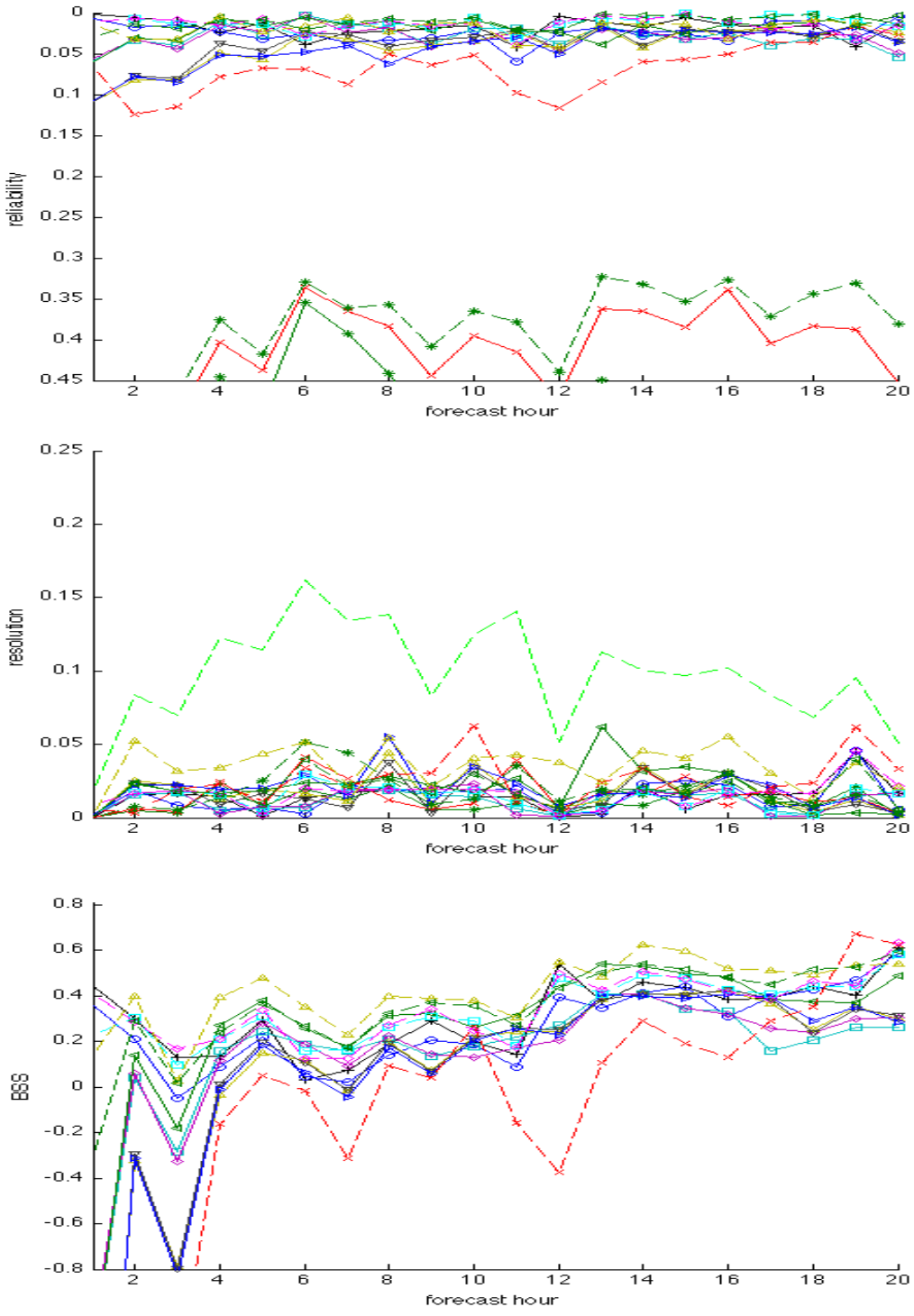

Figure 95. Same as in Figure 89 (coastal region results), but at the third β_e threshold (0.68 km^{-1}).

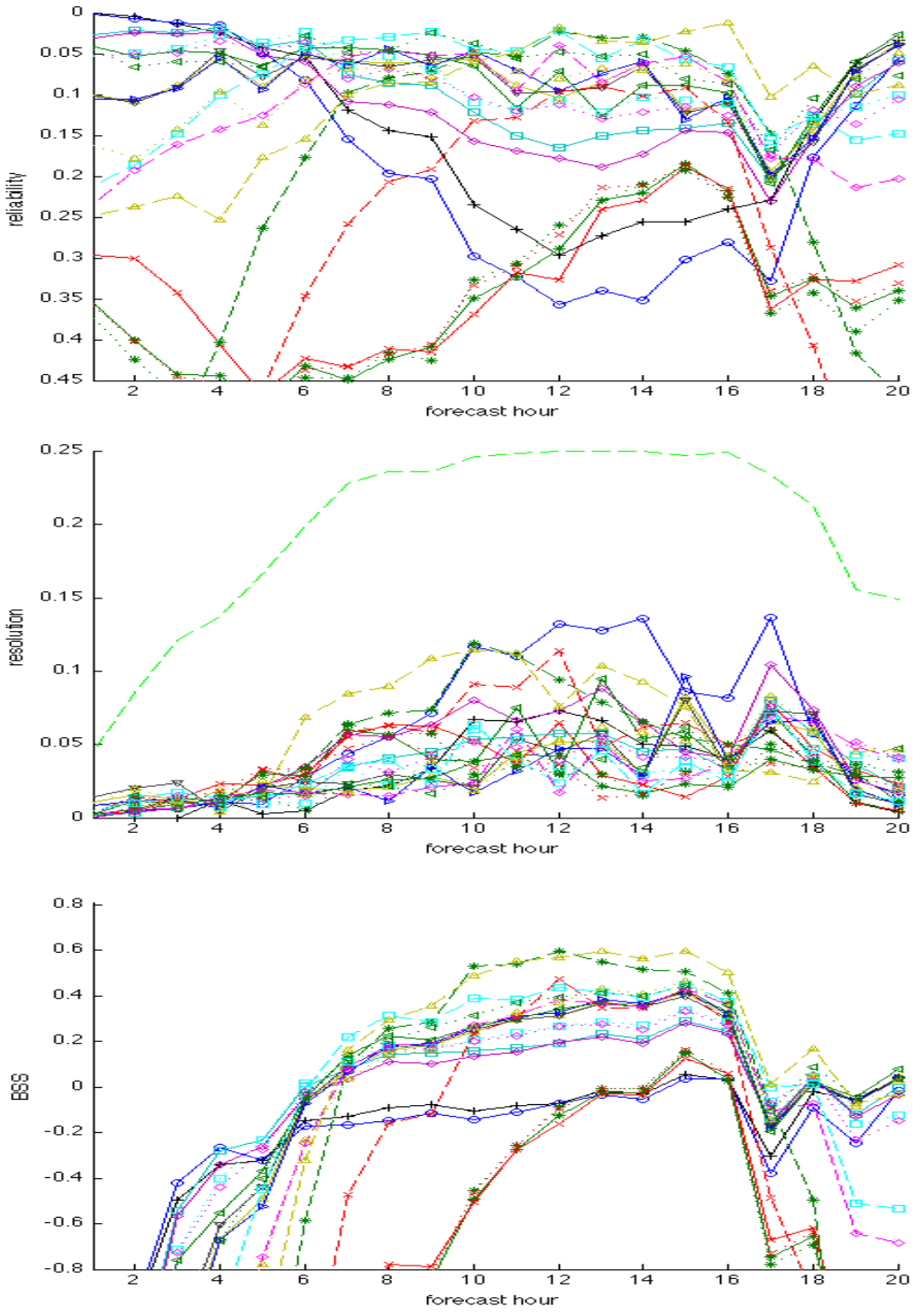

Figure 96. Same as in Figure 90 (valley region results), but at the third β_e threshold (0.68 km^{-1}).

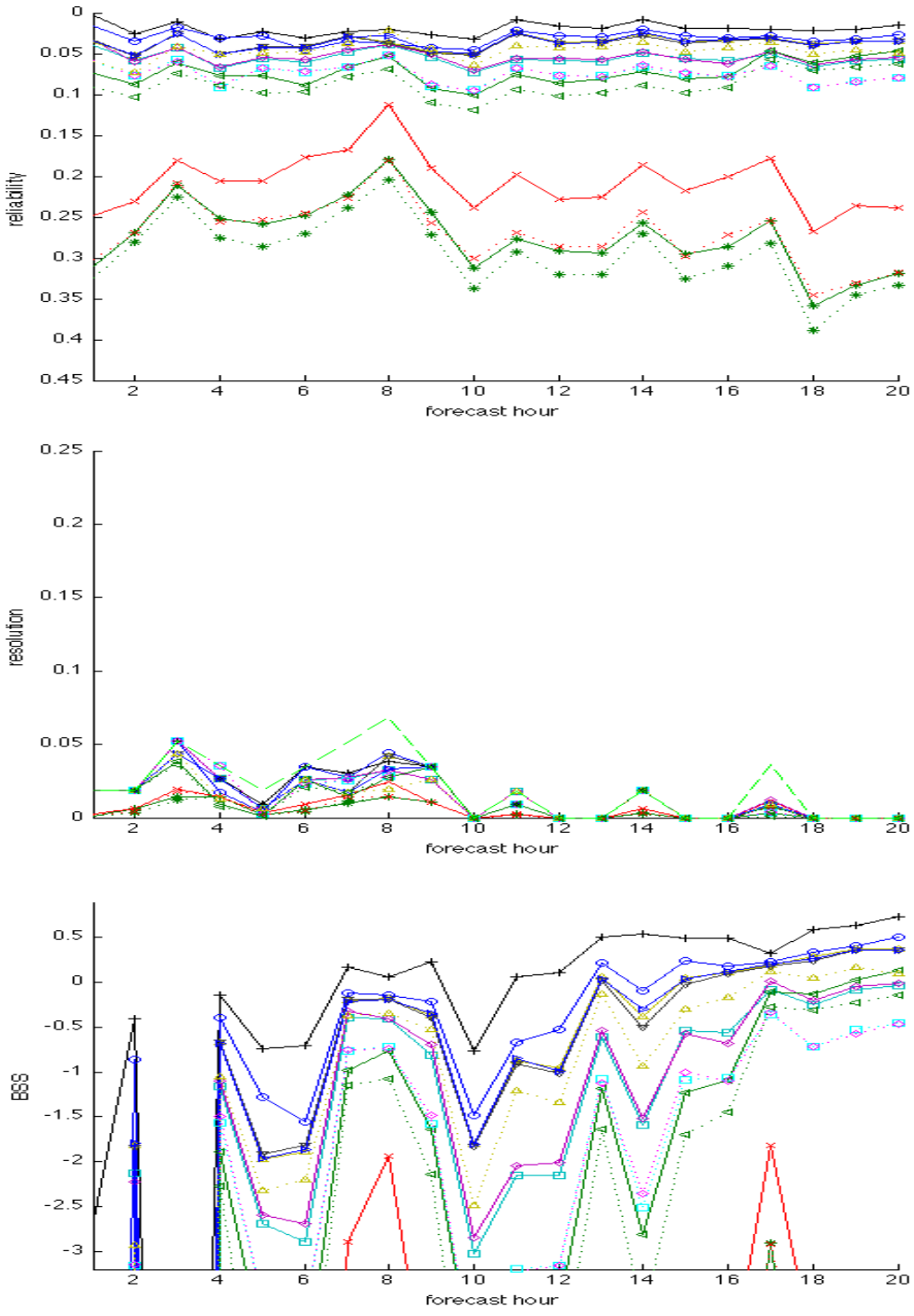

Figure 97. Same as in Figure 91 (mountain region results), but at the third β_e threshold (0.68 km^{-1}).

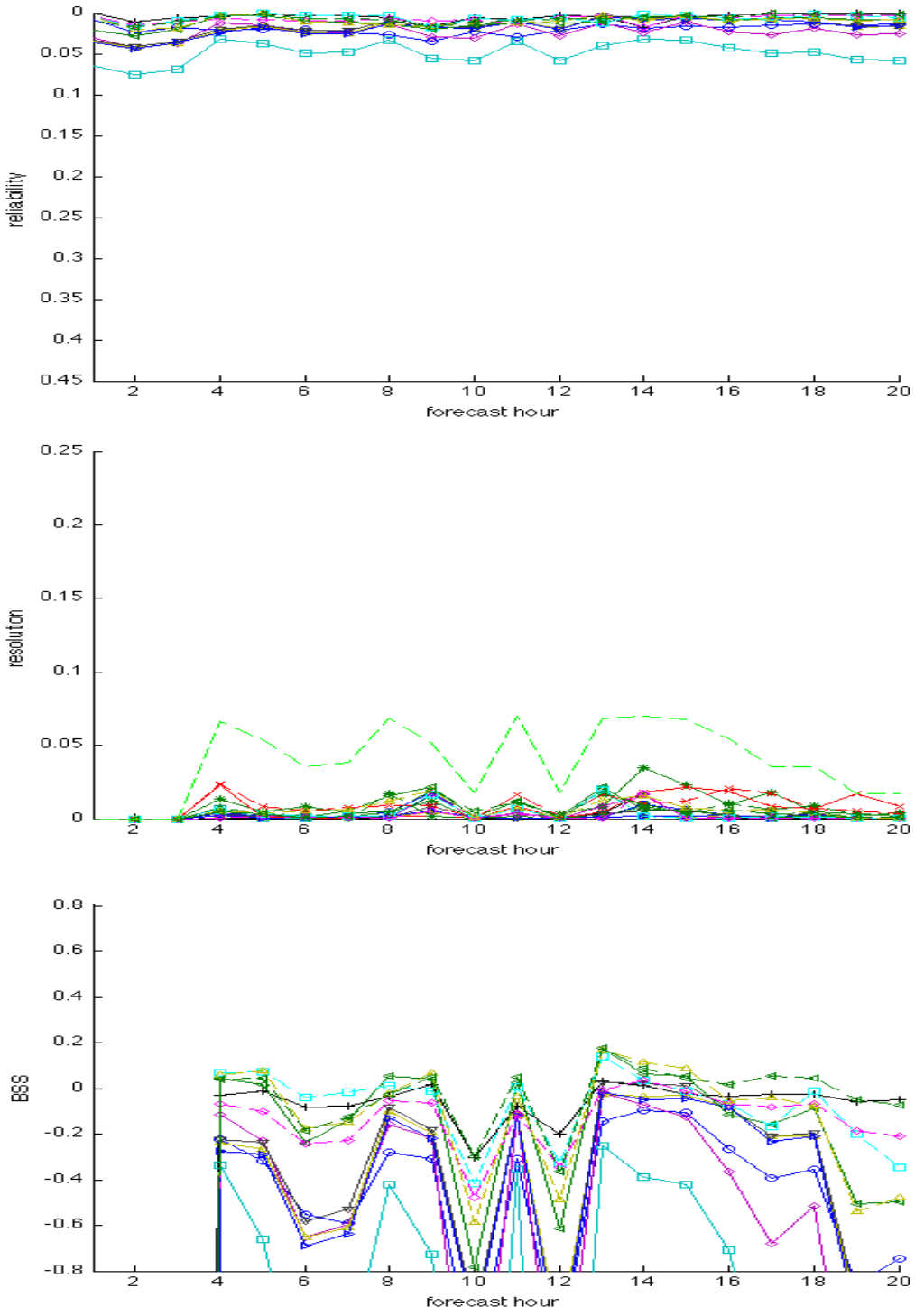

Figure 98. Same as in Figure 89 (coastal region results), but at the fourth β_e threshold (2.10 km^{-1}).

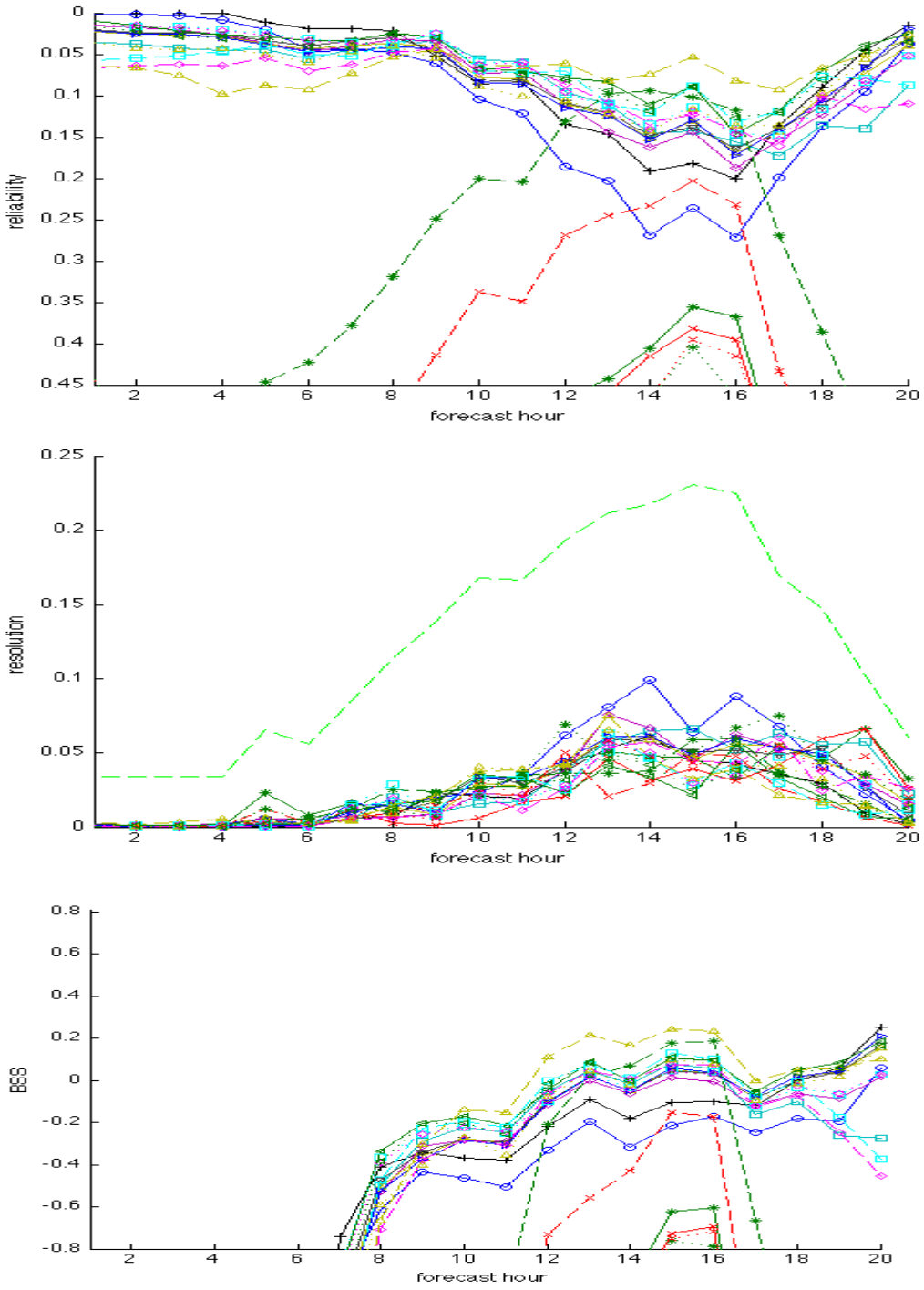

Figure 99. Same as in Figure 90 (valley region results), but at the fourth β_e threshold (2.10 km^{-1}).

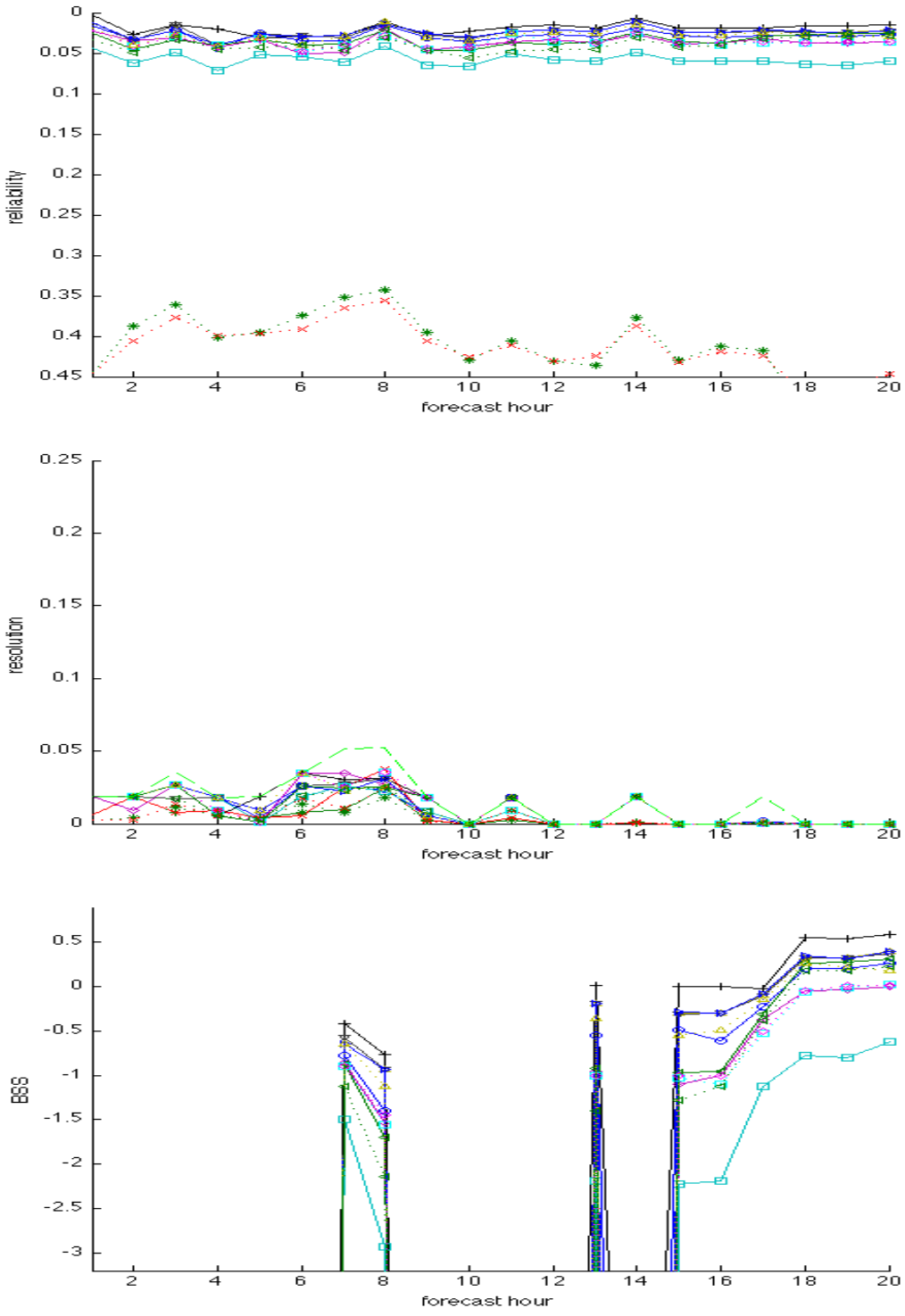

Figure 100. Same as in Figure 91 (mountain region results), but at the fourth β_e threshold (2.10 km^{-1}).

1. Overview and Comparison to *Cntrl*

We will first make some general observations about the results, and then examine each experiment in more detail in the next sections. Figure 88 indicates that most of the techniques tested in this work add some degree of skill to the stochastic ensemble predictions in the coastal and valley regions. In the coastal region, the improvement is evident at most forecast hours and is achieved via a combination of reliability and resolution increases at each β_e threshold (Figures 89, 92, 95, and 98). Reliability increases are not surprising since the NWP predictions have a negative q_c bias and each post-processing technique can only maintain or increase (but never decrease) the probability of β_e exceedance for any given forecast hour. Resolution improvement is more encouraging because it suggests the post-processing technique is effective at making larger upward probability adjustments to prediction corresponding to observed fog cases than those corresponding to observed no-fog cases. In contrast, if a technique indiscriminately increases probabilities, it might improve reliability but will not improve resolution, similar to what would be produced by a purely statistical bias correction to the final predicted probabilities from the ensemble

All of the techniques except *SCW* improved prediction skill in the valley region from 9–17 h (Figures 90, 93, 96, and 99), which corresponds to the period of highest observed fog incidence and least reliability of the unaltered NWP predictions. Reliability improvements are readily obtained by simply increasing the probabilities during this period, leading to a large portion of the skill increase for many of the experiments. Some, but not all, of the techniques also produced resolution improvements. The post-sunrise hours are characterized by a split in the results, with some of the joint-parameter techniques able to maintain a reliability (and skill) advantage over *Cntrl*, while most of the single-parameter RH techniques have lesser skill due to reliability decreases as the observed fog incidence decreases.

None of the techniques examined produce appreciable skill increases in the mountain region at any β_e threshold (Figures 91, 94, 97, and 100). Although modest resolution improvements are evident in some experiments, reliability decreases for each

experiment at every hour for every β_e threshold. This confirms the supposition that the framework in which these techniques exists—namely, making only upward adjustments to β_e probabilities when fog is not predicted by the member—are ill-suited for use in the mountain region because fog is relatively rare and the NWP predictions lack the negative q_c bias present in the other regions. Consequently, the overall viability of each technique can only be examined in the context of reconciling skill improvements in the coastal and/or valley regions with skill reductions in the mountain region.

2. *SCW*

In general, the resulting skill of *SCW* deviates only slightly from *Cntrl* in the coastal and valley regions, with slightly larger skill reductions at increasing β_e thresholds. However, closer examination reveals that the technique produces resolution improvement that was counteracted by reliability decreases, particularly at lower β_e thresholds. This is especially evident in the valley region during the overnight hours (Figure 90), where *SCW* occasionally has the highest resolution of any experiment.

These results indicate the small, non-zero q_c predictions are more likely to exist during observed *no-fog* cases. The upward adjustments of output probabilities in *SCW* are disproportionately applied to observed no-fog cases, which causes reliability reductions but resolution improvements. Since the probability adjustment prescribed by *SCW* rarely exceeds 0.2, the trend of all the metrics throughout the forecast period closely mimics *Cntrl* (e.g., low reliability overnight followed by post-sunrise increases in the valley region), unlike many of the other experiments. But the mechanism behind these small, non-zero q_c predictions, which do appear to have predictive usefulness for fog, deserves further examination. It does not appear to represent a systematic behavior of WRF but rather the behavior of two specific members using the Ferrier microphysics scheme.

The resolution improvements produced by *SCW* must be attained via a presently unclear linkage to observed fog incidence. Recall that over 99% of the small, non-zero q_c predictions from these two members have q_c values $\leq 1.68 \times 10^{-9}$ g m^{-3} (about six orders of magnitude less than the lowest verification threshold of 8.5×10^{-4} g m^{-3}). With such

small values, in addition to the fact these predictions are negatively correlated to observed fog compared to q_c predictions exactly equal to zero, it is unlikely they are a purposeful fog prediction from the NWP model. This represents a promising research path, which might explore the physical linkage between small, non-zero q_c predictions in the Ferrier scheme and observed low fog incidence, and more broadly examine which microphysics schemes are best suited for fog prediction.

In the mountain region, *SCW* resulted in the smallest skill decreases of any experiment, caused by small but consistent decreases in reliability coupled with mostly unchanged (or in some cases, slightly higher) resolution (Figures 91, 94, 97, and 100). The comparatively strong performance of *SCW* is attributed to the relatively modest probability adjustments associated with this technique, as well as the fact that small, non-zero q_c predictions occur with less frequency (compared to zero q_c forecasts) in this region compared to the coastal and valley regions (Figure 22–Figure 24).

3. *RH_D* and *BiasRH_D*

Using a single 2-m RH threshold as a deterministic fog predictor for each member, as was done in *RH_D* and *BiasRH_D*, generally performed poorly compared to other experiments (Figure 87). As implied by *JP_U* (Figure 81), RH predictions alone can be a useful predictor of fog in the coastal region, but are significantly more skillful when paired with a second parameter such a virtual temperature deficit. Without such a pairing and in a deterministic framework, *RH_D* and *BiasRH_D* still produced modest resolution improvements over *Cntrl*, with higher resolution achieved when the critical RH threshold is optimized for the region (dashed lines) as opposed to all the regions (solid lines), as shown in Figure 89 for example. Any resolution gain is more than offset by reliability decreases, which is attributed to extremely aggressive probability adjustments that assign an exceedance probability of 1 (at the lowest β_e threshold of 0.29 km^{-1}) to over half of the member predictions subject to post-processing. Although the unaltered NWP predictions have a strong negative q_c bias in the coastal region, *RH_D* and *BiasRH_D* are insufficiently discerning, affecting the majority of observed fog cases as well as too many no-fog cases.

The results show that the 2-m temperature bias correction employed in *BiasRH_D* had a positive effect in the coastal region in both reliability (Figure 89) and overall skill (Figure 87), with the largest effect over *RH_D* when tuned specifically for the region. Recall that the regional tuning in *BiasRH_D* includes not only the critical threshold, but also the bias correction itself, which is more than 2.5 times larger in the coastal region than in any other domain tested. The improvements produced by the bias correction can only be due to RH predictions from the WRF slightly below the critical threshold prior to the bias correction that were adjusted above the threshold after the correction. This disproportionately affects predictions at lower temperatures, since their RH values will increase more given a fixed downward temperature correction. Therefore, the improvement of *BiasRH_D* over *RH_D* indicates that RH predictions just below the critical RH threshold are disproportionally likely to be associated with fog at colder predicted temperatures. The opposite is true in the valley and mountain regions, which have more modest 2-m temperature bias corrections but where *RH_D* generally outperforms *BiasRH_D* (Figure 87). Regardless, even with the reliability improvements achieved by the temperature bias correction in the coastal region (Figure 89), its RPSS is still well below zero at all hours and lower than other experiments (Figure 87).

In the valley region, both experiments lead to RPSS >0 during some overnight hours, regardless of the domain used for optimization (Figure 88). This is remarkable considering the optimal RH threshold is a reverse classifier when optimized for the valley region (dashed lines), but not when optimized for the all regions domain (solid lines) or valley/mountain domain (dotted lines). The result illustrates the severity of the negative q_c bias in the valley region during the overnight hours when the observed fog incidence is highest; simply increasing the probabilities by even a crude technique yields skill improvements via reliability increases (Figure 90). Beyond 8 h appreciable resolution improvements (Figure 90) are only achieved by *RH_D* and *BiasRH_D* when optimized for the valley region (which employs the reverse classifier).

The performance of *RH_D* with valley region optimization is particularly noteworthy, with an RPSS that exceeds *Cntrl* from 8–17 h (Figure 88), and is among the top performing experiments during this period in terms of both RPSS and resolution at

the lowest β_e threshold (0.29 km^{-1}, Figure 90). However, none of the deterministic RH techniques perform well after sunrise in the valley region, with BSSs dropping well below zero after 17 h regardless of the optimization domain (Figure 90).

In the mountain region, RPSSs for both *RH_D* and *BiasRH_D* are well below zero, making these techniques unviable for indiscriminate use across a variety of geography within a model domain (Figure 87). In a clearly defined valley region or for a point forecast where overnight radiation fog is a concern, *RH_D* could be justified as a very simple fog classifier for overnight predictions to apply to members not already predicting fog.

4. *RH_P* and *BiasRH_P*

Conceptually, the use of probabilistic post-processing should outperform a corresponding deterministic framework since it should more thoroughly sample the prediction error compared to the sampling achieved by the 10 individual member predictions. This is supported by the results of *RH_P* and *BiasRH_P* in the coastal region, where the RPSSs of these two experiments are significantly higher than their deterministic counterparts, *RH_D* and *BiasRH_D* (Figure 87). The probability adjustments prescribed by *RH_P* and *BiasRH_P* in this region are generally within +/- 0.15 of the climatological incidence of fog for the entire subset of data subject to post-processing, and they produce only small resolution improvements compared to *Cntrl* (Figure 89). No clear resolution advantage over *RH_D* and *BiasRH_D* is evident. However, their reliability is superior to *Cntrl* at most β_e thresholds, and significantly higher than the reliability of their deterministic counterparts at all β_e thresholds (Figure 89). Since the forecast probability map for all regions optimization is similar to that for coastal optimization (Figures 68-70), the coastal region results show small sensitivity to the domain optimization for *RH_P* and *BiasRH_P* compared to domain sensitivity in *RH_D* and *BiasRH_D* (Figure 87).

In contrast, large differences between the forecast probability map of all regions optimization and valley optimization were evident in Figures 68-70, and are reflected in the reliabilities of *RH_P* and *BiasRH_P* in the valley region when optimized for the

various domains (Figure 90). When optimized for the all regions domain or the valley/mountain domain, overnight predictions of low RH have smaller upward probability adjustments than predictions with high RH, which is exactly the opposite of what is observed in the valley region and of what is prescribed when valley optimization is used. This has the effect of producing comparatively lower reliabilities, but the impact on resolution is small or even slightly positive compared to when optimized for the valley domain. Overall, *RH_P* and *BiasRH_P* for all optimizations have higher reliabilities (Figure 90) and RPSS (Figure 88) than *Cntrl* from 5-18 h, and mostly higher RPSS than their deterministic counterparts. The exception is *RH_D* with valley optimization, which outperforms the probabilistic RH techniques during the overnight hours. While the overnight differences are small between the deterministic and probabilistic RH techniques in the valley region, the probabilistic techniques offer a clear advantage after sunrise. *Cntrl* slightly outperforms the probabilistic techniques after sunrise, but significantly outperforms the deterministic techniques, whose skill decreases drastically during this period.

The impact of the 2-m temperature bias correction in the *BiasRH_P* experiments compared to the *RH_P* experiments is less than in *BiasRH_D* compared to *RH_D* (Figure 88). This is because in a probabilistic framework bias correction typically alters output probabilities by only a few percent instead of deterministically changing a prediction to a fog forecast if the RH threshold is exceeded (i.e., changing the probability from 0 to 1). As in the deterministic framework, the probabilistic RH experiments show no clear pattern as to whether the bias correction aids in the final predictive skill, exhibiting mixed results depending on the region, forecast hour, and optimization domain. In general, bias correction in this work can be quite important, particularly in a deterministic framework, but without further examination the precise impact on any given forecast is inconclusive.

As with *RH_D* and *BiasRH_D*, *RH_P* and *BiasRH_P* do not achieve positive RPSS at any hour in the mountain region (Figure 87), and so a universal application is not viable without first pre-defining region categories and excluding mountainous regions from the post-processing.

5. *JP_B* and *JP_U*

The primary advantage of *JP_B* and *JP_U* over the single-parameter techniques of earlier experiments is their post-sunrise performance in the coastal and valley regions, which maintains equal or better skill compared to *Cntrl* in contrast to the skill decreases seen in most previous experiments (Figures 87 and 88). This results from the virtual temperature deficit predictions, which offers an additional degree of freedom such that the output probability adjustments can be appropriately scaled back during post-sunrise heating. For the optimization domains in these two experiments that do not use virtual temperature deficit as a parameter, a similar parameter (saturation vapor pressure deficit in *JP_B* with valley domain optimization, and time rate of change of 2-m virtual temperature in *JP_B* with coastal domain optimization) is used that serves a similar function.

JP_B and *JP_U* produce higher skill than *Cntrl* for the entire period between 7-17 h in both the coastal and valley regions (Figure 88). However, during the overnight hours they have only marginally higher skill than some of the single-parameter techniques in these regions. The exception is *JP_B* with region-specific (i.e., coast or valley) optimization, which achieves the highest skill of any experiment in each respective region at nearly all hours. In the coastal region, *JP_U* with all regions optimization performs just as well as *JP_B* optimized for the same domain, indicating there is no clear advantage to using layer 1 vapor pressure predictions instead of layer 1 RH as a predictive parameter. Since RH is considered a more universal (i.e., transferable) parameter than vapor pressure, this is a promising finding. For coastal-only applications, the use of 2-m RH (used in *JP_U* with coastal optimization) instead of layer 1 RH (used in *JP_U* with all regions optimization) in the joint space produces a slight skill advantage after sunrise, but otherwise the affect is minimal. The skill improvements achieved by *JP_B* and *JP_U* are produced by both reliability and resolution gains in the valley region at most β_e thresholds, with these gains diminishing after sunrise but remaining competitive with *Cntrl*.

Similar to the results in the coastal region, *JP_B* seems to offer no appreciable advantage over *JP_U* when using all regions optimization in the valley region, ever after sunrise. Even when using valley/mountain optimization in these experiments, there is little difference in the results from when using all regions optimization (which uses the same joint parameter pairs as valley-mountain optimization), suggesting that the addition of coastal region predictions to this joint parameter space has little effect on the valley region output probabilities. Significant skill improvements over *Cntrl* in the valley region are obtained mostly via reliability improvements with the exception of *JP_B* with valley optimization, which also produces significant resolution gains during the overnight hours. Since the parameter pair used in *JP_B* with valley optimization (consisting of saturation vapor pressure deficit and layer 1 vapor pressure depression) is also believed to be universal (i.e., transferable to other valley regions outside the testing locale), this is clearly the most viable post-processing technique among those tested for valley-only applications.

Thus far, none of the experiments tested have achieved skill gains or even positive RPSSs in the mountain region. *JP_U* produces positive skill only during the last few hours of the runs, yet is still significantly less skillful than *Cntrl*. *JP_B*, when optimized for the all regions domain, is also less skillful than *Cntrl* but does manage positive skill beyond 10 h. We can only conclude that *JP_B* is the only acceptable framework for use in the mountain region in the sense that it does the least harm to the existing NWP model skill while still outperforming persistence. It may also carry substantial risk of being location-specific. Because there is no acceptable universal parameter pair that produces positive skill in the mountains, the best alternative is to not employ any of the post-processing techniques developed in this work in the mountain region. It should be noted that generally the joint parameter techniques did not destroy resolution in the region, but all of the experiments (with their upward adjustments to fog probability) resulted in reliability decreases due to the very low incidence of fog in the subset of data subject to post-processing, as well as the already high reliability of *Cntrl*.

6. *JP_LB* and *JP_SB*

The bin sizes in *JP_LB* are more than double the size of those in *JP_SB* (2490 versus 1107 predictions, respectively), yet the two experiments produce BSSs that vary only slightly from each other or from *JP_B*. Conceptually, to the extent that they do not overfit the training data, smaller bins are preferable because they leverage finer details of the joint parameter space to provide more predictive resolution at the expense of a some reliability. As bin size is decreased to the point that resolution gains no longer offset reliability losses in cross-validation, the bins have overfitted the training data and there is no benefit to reducing the bin size further.

Results from these experiments show that there is no consistent reliability or resolution advantage for *JP_LB* or *JP_SB* at any β_e threshold compared to *JP_B*, with only subtle signals in certain regions and hours. For example, *JP_SB* has slightly lower reliability, resolution, and BSS than *JP_LB* and *JP_B* at β_e = 0.29 km^{-1} after 15 h in the valley region (Figure 90), perhaps indicating minor overfitting. But any differences are small or negligible, allowing us to conclude that this particular joint parameter space has little sensitivity to bin size within the range of bin sizes tested. We suspect any sensitivity to bin size is more important when smaller bins are used, but as this work aims to develop a post-processing framework that is transferable the use of conservatively large bins is appropriate until further testing or a proper optimization can be performed. These results suggest there is no single optimal bin size for all scenarios, as overfitting appears to emerge sooner in certain regions and forecast hours as bin size is decreased.

In addition to altering the bin size, other binning strategies exist that might better capture signals in the joint parameter space. The strategy used in this work of having a fixed number of predictions for the bins was selected for its relative simplicity and apparent effectiveness after some preliminary testing. However, a more sophisticated strategy was also considered that assigned a weighted influence of each prediction based on its distance, r, from the prediction of interest in the parameter space. The weighting itself, defined as $1/r^x$, was found to be extremely sensitive to the choice of x; a conservative value of x = 1 produced results with virtually no resolution, while x = 2

clearly resulted in overfitting. With further refinement, this or other binning strategies might improve the results achieved here.

7. Summary and Additional Discussion

The joint parameter techniques outperform all other techniques during the post-sunrise hours in the coastal and valley regions. The expansion from single-parameter RH techniques to the joint parameter space permits one of the parameters in the joint parameter techniques to be used to identify the switch from a nighttime to a daytime regime. This is crucial for preventing rapid skill decreases post-sunrise, because the nature of the NWP model error is different before and after sunrise. The single-parameter techniques are not able to discern the switch from night to day, but produce overnight skill that is competitive or slightly better than the joint parameter techniques.

Virtual temperature deficit is used in most of the joint parameter techniques to serve as a delineator between night and day. This parameter is favored over more obvious choices such as temperature or temperature change because it appears to have predictive usefulness for forecasting the presence of low-level inversions. In addition, it is proposed to have the added benefit of indicating the stability condition of the marine boundary layer, which is also crucial for fog prediction near the coast.

The results show that distinguishing coastal regions from valley regions for the purposes of post-processing is not necessary to achieve skill improvements in both regions. This is because *JP_U* produces similar results in the coastal region whether it has coastal optimization or all regions optimization, and produces significant skill improvement in the valley region. To achieve even greater skill in valley-only applications, *JP_B* with valley optimization is prescribed, whose parameters are universal and which offers the largest skill improvement of any technique.

When using all regions optimization, *JP_B* did not achieve appreciably higher skill than *JP_U* in the coastal or valley regions. Both of these experiments use virtual temperature deficit as one of the joint parameters, but it is paired with layer 1 vapor

pressure in *JP_B* and layer 1 RH in *JP_U*. The substitution appears to only affect skill in the mountain region, which is lower in *JP_U*.

The success of *JP_U* supports the finding of Hippi et al. (2010). Using temperature, moisture, and wind measurements near the surface and at 500 m elevation at two stations in Finland, they showed that the two best fog predictors were the temperature difference between the surface and 500 m, and surface RH. We extended this finding to the NWP model predictions space, showing that using predictions of virtual temperature deficit and layer 1 RH as fog predictors also accounts for the error characteristics of the NWP model.

None of the techniques tested in this work improve the already skillful unaltered NWP model predictions in the mountain region, and except for the non-universal joint parameter space of *JP_B*, none of the techniques even produce sustained positive skill in this region. Therefore, if applying one of these post-processing techniques to a large geographical domain that includes mountainous topography, it is appropriate to pre-define the mountainous region and exclude it from the post-processing. The boundaries of such a region would seem to be defined arbitrarily, and perhaps a better approach is to gradually decrease the influence of post-processing as the topography transitions to mountainous from some other region category. Either way, further research is needed to develop more objective criteria that can discern a mountain region and its characteristic NWP model behavior from other regions.

For all the probabilistic experiments (*RH_P*, *BiasRH_P*, *JP_B*, *JP_LB*, *JP_SB*, and *JP_U*) conservatively large bins were used to minimize the risk of overfitting the training data. This is likely to have sacrificed some resolution in the results, which can be obtained using smaller bins. The results of *JP_LB* and *JP_SB* indicate overall reliability, resolution, and skill have low sensitivity to bin size in the range of bin sizes used, so larger variations are indicated to draw more definitive conclusions regarding the optimal bin size. Absent a formal bin size optimization, the use of a larger bin size is preferable in the sense it appears to make the exact choice of bin size rather irrelevant to the results.

In addition to using large bins to minimize overfitting of the training data, several other measures were taken in this work to attempt to maintain as much transferability as possible in the prediction framework. These include 1) restricting the use of predictors to those with a clear thermodynamic linkage to fog, and excluding those whose linkage might be speculative or vary by location, 2) seeking joint parameter pairs that are believed to possess a high universal quality, and 3) performing cross-validation along the mode with highest variance in post-processing output. Despite these measures, this study encompasses only a single winter season at seven sites, and merely lays the groundwork for a larger validation of its findings before its true transferability can be known.

THIS PAGE INTENTIONALLY LEFT BLANK

VII. SUMMARY, RECOMMENDATIONS, AND FUTURE WORK

A. SUMMARY AND ADDITIONAL DISCUSSION

The goal of this research was to investigate the viability of a new framework for producing short-term (≤20 h) probabilistic VIF predictions using existing mesoscale ensemble output suitable for use in data-denied areas away from existing airfields. The 4-km grid spacing, 10-member WRF ensemble used was constructed to closely match the specifications of the AFWA MEPS.

Two distinct sources of error were investigated in fog prediction using the ensemble. The first was error in the q_c predictions, which existed as a large negative bias in the coastal and valley regions due to excessive zero or near-zero q_c forecasts from each WRF member at the expense of predictions of light fog with visibilities 1–7 mi. The predictions in all regions also had highly bimodal distributions such that most of the fog predictions were for heavy fog with visibility <0.875 mi. The bimodality of the predictions was higher than the bimodality of the observations in the coastal and valley regions, but reasonably matched the bimodality of observations in the mountain region.

The second source of error stemmed from the conversion of q_c to β_e, which was sensitive to several unmodeled quantities including droplet size distribution. To sample the uncertainty in the conversion of q_c to β_e, we built a parametric visibility parameterization based on the estimated uncertainty in field measurements from Kunkel (1984) and Gultepe et al. (2006). Predictions in the range of visibilities of interest (approximately 1–7 mi) were found to have negligible sensitivity to visibility parameterization error due to the highly bimodal distribution of the q_c predictions from WRF. In the visibility range of interest, error in the q_c predictions from WRF was therefore the primary source of error.

Despite the highly bimodal q_c predictions and strong negative q_c bias, the stochastic q_c predictions from the ensemble were generally skillful compared to persistence in the coastal region but unskillful in the valley region. The mountain region q_c predictions did not exhibit large bias and were the most skillful of any region beyond 7

h. After forecast initialization at 1600 LT, skill generally increased overnight in each region, then increased more slowly after sunrise through the end of the runs at 20 h (1200 LT).

In the coastal and valley regions, the negative q_c bias was traced to a negative RH bias, which was primarily caused by a warm bias that was worse during the overnight hours. There was very little q_v bias in either region except after sunrise, when a negative bias was present.

In the coastal region, the 2-m temperature and q_v biases were equal to or greater than the biases at layer 1, but the predictions at 2 m had lower error variances. In the valley region, 2-m temperature predictions had less warm bias and a moist q_v bias compared to layer 1, with slightly lower error variances. The 2-m predictions in the mountain region were significantly worse than the layer 1 predictions, with larger biases and error variances.

Post-processing of the WRF predictions focused on identifying and leveraging alternative aspects of the NWP model output with predictive usefulness for fog. The strategy did not pursue site-specific calibration, but maintained a measure of transferability by targeting only systematic error characteristics of the WRF predictions, and using only aspects of the predictions with a close and recognizable physical link to fog.

Given the nature of the q_c prediction error from WRF (large negative bias, highly bimodal distribution), the post-processing strategy made upward adjustments to the probability of β_e exceedance (at four measured thresholds) for individual members predicting zero or negligible q_c. This simplified the strategy since adjustments were only made in one direction (upward), and potentially preserved the skill already achieved by the raw WRF predictions. This strategy was not well-suited to the mountain region since it had different error characteristics (small moist bias, highest overall skill) than the coastal and valley regions. All tested methods lack skill improvement in this region, where the predictions are already highly skillful. A strategy with the capability to adjust β_e exceedance probabilities up or down is better suited for potential skill improvement in the mountain region.

A single-parameter method using 2-m RH predictions to predict β_e threshold exceedance generally decreased skill in the coastal region, and increased overnight skill in the valley region. In the valley region, overnight fog was less likely with high RH, and more likely with RH well below saturation. This was because the warm bias and negative RH bias was worse on nights when fog is likely to form. These biases also tended to be present at initialization prior to overnight fog forming, and not present prior to a night without fog.

In the coastal region, the single-parameter method was significantly more skillful when applied probabilistically to each member rather than deterministically, producing comparable skill to the raw WRF predictions. In the valley region, the deterministic single-parameter method was just as skillful as the probabilistic method overnight, with the probabilistic framework being significantly more skillful after sunrise. Applying a 2-m temperature bias correction to the predictions prior to using the single-parameter RH methods had a positive small impact for the deterministic method in the coastal region, but had little impact otherwise.

The expansion of the single-parameter methods to a framework utilizing joint parameters from the member predictions was performed by first testing hundreds of joint parameter pairs for viability. In each of four domains (coastal, valley, valley/mountain, and all regions), two parameter pairs were selected for full evaluation. The best overall parameter pair was the one that produced the highest predictive resolution in the training data, but often (except in the valley domain) possessed predictive usefulness specific to the local climatology of the test sites. The best universal parameter pair was the one with the highest predictive resolution among those possessing transferability to other locations with the same domain category.

The universal parameter pairs invariably included a moisture parameter such as RH or vapor pressure depression, and a low-level stability parameter. Compared to the single-parameter methods, this joint parameter framework produced similar or slightly worse results during night, but much better results after sunrise, when predictive skill was difficult to achieve due to the higher skill of the persistence reference forecast. The physical mechanism behind the improvement was the use of the low-level stability

parameter as an axis in the joint space, which indicated the likelihood of a low-level inversion, which when present generally indicated higher β_e exceedance probabilities (depending on the value of the second parameter in the space). When an inversion was not predicted by the member, fog was rare. The inversions themselves were often due to radiative cooling of the ground, which normally ended shortly after sunrise and, if predicted by the MEPS member, moved the prediction to a different portion of the joint space with appropriately modified β_e exceedance probabilities for a post-sunrise (or otherwise unstable) regime. Low-level inversions were also be due to downward heat flux at the sea surface, which was indicative of a stable marine boundary layer and favorable fog condition for coastal sites.

For coastal region post-processing using the best universal parameter pair, there was very little advantage to using a coastal optimization (which used parameters of virtual temperature and 2-m RH) instead of all regions optimization (which simply replaced the 2-m RH with layer 1 RH). 2-m RH provided slightly higher skill after sunrise due to the lower error variances at 2-m compared to layer 1 in this region. Both parameter pairs increased skill over the raw WRF predictions.

Skill in the valley region was improved over the raw predictions by also using the best universal parameter pair with all regions optimization. In the valley region, the layer 1 RH predictions were favored over the 2-m RH predictions for predictive resolution, despite increased dispersion present in the 2-m predictions. The dispersion was due to a wide spread of q_v biases among the individual members, which was less desirable than dispersion generated from increased error variance among consistently-biased members, and actually blurred the predictive signal in the 2-m predictions.

For valley-only applications such as a small model domain or a point forecast, even greater skill was produced using the best overall parameter pair with valley optimization. This parameter pair, which includes saturation vapor pressure deficit and layer 1 vapor pressure depression, was also universal in the sense it is reasonably transferable to other valley-like domains.

Making a bias correction prior to applying the joint parameter framework showed a minimal impact and was generally unnecessary. This is particularly true if the bias

correction produced a mostly linear response in the parameter pair, which would only cause the probability forecast map to shift along an axis but not change the post-processing outcome.

When a joint parameter map is developed, the degree of overfitting of the training data is related the size of the bin used to compute the observed fog incidence at each point in the joint space. We selected a conservatively large bin equal to one-twelfth of the total dataset to minimize the risk of overfitting. When the bin size was increased 50% and decreased 33%, there was little change in the results, indicating low sensitivity to bin size when the bins are large. Greater predictive resolution may be possible with significantly smaller bins, but reliability and resolution will suffer if bins are decreased too aggressively and overfitting occurs.

The implications of optimizing a post-processing routine on a subset of the data (i.e., predictions without fog), does not mean it was necessarily optimized to produce maximum skill when verified using the entire data set (i.e., when the post-processed data was combined with the member predictions that produced fog on their own). However, the magnitude of the negative q_c bias was large enough that 93.7% of the raw WRF predictions did not predict fog and were subject to post-processing. Any degradation that might occur from the minor difference between datasets was likely small. The largest impact might be in the valley region, which had the largest proportion of its total predictions not subject to post-processing (because the member predicted fog on its own).

1. *Broader Implications*

This research has laid a path for a simple post-processing routine that can be easily applied to deterministic or ensemble output to improve visibility forecasting in fog in a coastal or valley geographical region without the need for any observational record. It has revealed several systematic deficiencies of WRF predictions relevant for fog forecasting, and demonstrated that applying a conservative statistical element that is not heavily site-specific can improve the skill of the predictions. Furthermore, it has identified a physically-based mechanism for the predictive usefulness of the post-processing, which considers both the error characteristics of the predictions and the fog

dynamics, such that it can be properly interrogated and refined as needed for other locales or as NWP model improvements are made.

Since the WRF ensemble used for this research closely resembles the AFWA MEPS, the post-processing framework developed here could be used to add skill to predictions of surface visibility restrictions due to fog.

Several systematic deficiencies of the WRF predictions were identified that might help inform future model development. In addition, member-specific behavior revealed in this work could assist in evaluating physics suites unique to each member.

A broader verification using different test sites and seasons is needed to better gauge the transferability of the techniques developed in this work.

B. RECOMMENDATIONS

To further verify the AFWA MEPS fog prediction improvement produced by this framework, experimental testing in a new model domain and/or different season is recommended for the best universal joint parameter (*JP_U*) framework with all regions optimization. *JP_U* offers the best balance of skill improvement with the potential for transferability to other like regions (i.e., other model domains with coastal and/or valley geography). It can be applied indiscriminately to both coastal and valley geography without the need to pre-define these regions and apply separate post-processing schemes. The forecast probability map (Figure 83) utilizes WRF predictions of virtual temperature deficit and layer 1 RH as its parameter pair.

For valley-only applications, greater skill was achieved by using *JP_B* with valley optimization, which is also considered highly transferable. It uses a forecast probability map with a parameter pair of saturation vapor pressure deficit and layer 1 vapor pressure depression (Figure 76). Further experimentation is warranted to verify the results in a different locale and/or season.

For a model-generated point forecast at the coast using this framework, there appears to be benefit to bilinearly interpolating the model data from the four surrounding gridpoints, which incorporates low-level stability predictions (via predictions of virtual temperature) from two points over land and two points water that are important to the success of the framework. In contrast, using model data from only the nearest grid point

might degrade the performance of the post-processing since the stability condition will be dominated by either terrestrial or marine conditions depending on where the grid point is located.

For coastal-only applications, slightly greater skill was achieved and should be tested using virtual temperature deficit and layer 1 RH as a parameter pair according to the forecast probability map of *JP_U* with coastal optimization (Figure 81).

No post-processing presented in this work is recommended in a mountain geographic region, as it did not improve skill.

Using nonlinear regression to fit an expression to the joint parameter forecast probability maps was not part of this work, and is not needed for implementation as values can be interpolated from the map itself. However, it might be recommended if the interpolation is computationally expensive in an operational setting.

For any operational fog forecasting or fog verification study, it is crucial to be aware of the two algorithms used by ASOS to produce visibility observations (see Table 2 and accompanying discussion). When the algorithm is switched near sunrise or sunset, reported visibility can quickly be reduced by half (if switching from night to day) or doubled (if switching from day to night). Since the abrupt adjustment is not associated with any change in meteorological conditions other than ambient light, it is easily overlooked in forecasting and research.

C. FUTURE WORK

Future enhancements to a fog post-processing framework might produce a forecast PDF of β_e rather than probabilities of exceedance at four fixed β_e thresholds as is done here. A PDF is preferable because it provides the entire uncertainty profile of the prediction, including the probability of exceedance at any given threshold rather than at predetermined thresholds. Significant challenges exist to produce a β_e PDF, including whether a reasonable curve of β_e distribution can be drawn from the members that predict fog on their own. This research suggests it cannot, which allowed us to ignore the PDF shape and use democratic voting to verify predictions at each β_e threshold since most predictions are either well above or below all thresholds. An alternative approach is to fit the q_c predictions from the members to a fixed, predetermined distribution shape

informed by the climatological distribution, which is not Gaussian (Figure 18. Also Chmielecki and Raftery 2011). The post-processing framework would also have to be refined to provide a PDF (or at least PDF parameters) rather than an exceedance probability as was done in this work. Additionally, uncertainty in the visibility parameterization used to convert q_c to β_e could also be considered since it might make an important contribution to the PDF shape that was ignored in this work after showing it did not affect verification at four thresholds.

The unaltered MEPS members produced sufficiently high skill in the mountain region q_c predictions that it is questionable as to how much more skill can be added, even with a more sophisticated post-processing strategy. Instead, an examination of WRF error and q_c skill in the transition zones between region categories might lead to some objective criteria as to where these boundaries begin and end, or perhaps how they transition from one to the others. This permits the post-processing to be easily excluded from these areas. Without this information, mountain regions must simply be arbitrarily identified and avoided, with little understanding as to what constitutes a mountain region.

During MEPS development, H11 experimented with adding a form of stochastic backscatter to the model integrations, which is a way to represent model uncertainty from interactions with unresolved scales (Berner et al. 2009). At that time, it added beneficial dispersion to the ensemble wind and temperature predictions. It was ultimately not used in MEPS or in this work, but could improve the performance of this post-processing frameowork since it would produce larger dispersion in the post-processed forecast probabilities. Most of the layer 1 and 2-m thermodynamic variables examined in this work are underdispersive, which ultimately decreases the skill of the predictions and this framework. Hacker and Snyder (2012, personal communication) are preparing to test the impact of this capability on fog predictions.

Additional skill might be produced with the framework presented in this work simply by using smaller bins. The results of this work show low sensitivity to rather aggressive bin size changes, perhaps suggesting the bins could be significantly reduced to improve resolution and perhaps reliability before overfitting occurs (manifest as declining reliability and resolution). A larger testing dataset and robust cross-validation is

suggested for this purpose as it will inform how small the bins can be made before overfitting occurs. The results in this work suggest overfitting does not occur across all predictions at once, but affects certain regions at certain forecast hours before others. It is possible that with smaller bin sizes, post-processed skill decreases in the mountain region could be reduced or eliminated, negating the need to pre-define and exclude these regions from the framework and easing the framework's operational employment.

Closer examination of the small non-zero q_c predictions is warranted since results indicate that, compared to the prediction of exactly zero q_c, they are disproportionally more likely during observed no fog. The mechanism behind this predictive usefulness is not understood. Nearly all of these small non-zero q_c predictions are produced by the only two ensemble members using the Ferrier microphysics scheme, where they occur in >10% of the total predictions.

Examining WRF predictions above layer 1 might provide additional predictive usefulness to be leveraged. This is particularly true given the inherent numerical challenges at layer 1, which is heavily influenced by information passed vertically from the land surface and surface layer below that is not necessarily seamlessly integrated into the model grid (Thompson, 2012 personal communication). This phenomenon is analogous to the horizontal edge of a local area model, where boundary condition information being passed horizontally into the domain might negatively affect predictions at the edge as the modeled atmosphere conforms to the new resolution, physics suites, etc. There are obvious disadvantages to using higher model layers, one being that fog is heavily influenced by surface conditions and higher model layers are further disconnected from the surface information. But given the WRF systematic q_c errors at layer 1, the potential benefits of using predictions at a higher layer may outweight the drawbacks, especially if utilized in the joint parameter space where they could be paired with predictions from a lower layer to leverage any useful predictive signal for fog that may exist.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A. POTENTIAL FOR 850-hPa WIND DIRECTION AS A HEAVY FOG PREDICTOR IN COASTAL REGION

As one of the parameters evaluated for use in a parameter pair for the joint parameter space technique, 850-hPa wind direction predictions generally did not exhibit high predictive usefulness. One prominent exception is with heavy fog prediction at the highest β_e threshold of 2.1 km^{-1} (0.875 mi daytime visibility), for which 850-hPa wind predictions paired with 2-m vapor pressure predictions (Figure 101) produced the highest plot variance of any parameter pair, indicating that it may provide resolution specifically for predicting heavy fog.

As was discussed in the *JP_B* experiment, the predictive usefulness of 2-m vapor pressure predictions is tied to the stability condition. This parameter is paired with predicted 850-hPa wind direction to form the joint parameter space shown in Figure 101. The top row of the figure displays the data as in previous joint parameter plots, with heavy fog missed opportunities plotted in red and heavy fog correct rejections plotted in blue. The data indicate heavy fog is significantly more likely when the 850-hPa wind direction is predicted to be northerly or northeasterly. The forecast probabilities indicated by the contouring of this data are relatively modest, with a maximum value of 0.2–0.3. However, considering that the variance of the joint parameter plots for any parameter pair are generally much lower for heavy fog prediction than for prediction at the lower β_e thresholds, the forecast probabilities indicated in Figure 101 provide a better separation between occurrences and non-occurrences than any other parameter pair examined. For comparison at this β_e threshold, *JP_B* with coastal optimization (Figure 74 bottom row) and *JP_U* with coastal optimization (Figure 81 bottom row) both produced forecast probability maps with smaller areas of forecast probabilities >0.2, which implies less resolution in the predictions.

A physical explanation for the potential skill gained from 850-hPa wind direction predictions for predicting heavy fog is not fully explored in this work, but two possible

links are put forth. Subjectively, the 850-hPa wind direction predictions have small error during these events, so heavy fog appears to be more likely with observed (not just predicted) northerly or northeasterly 850-hPa winds.

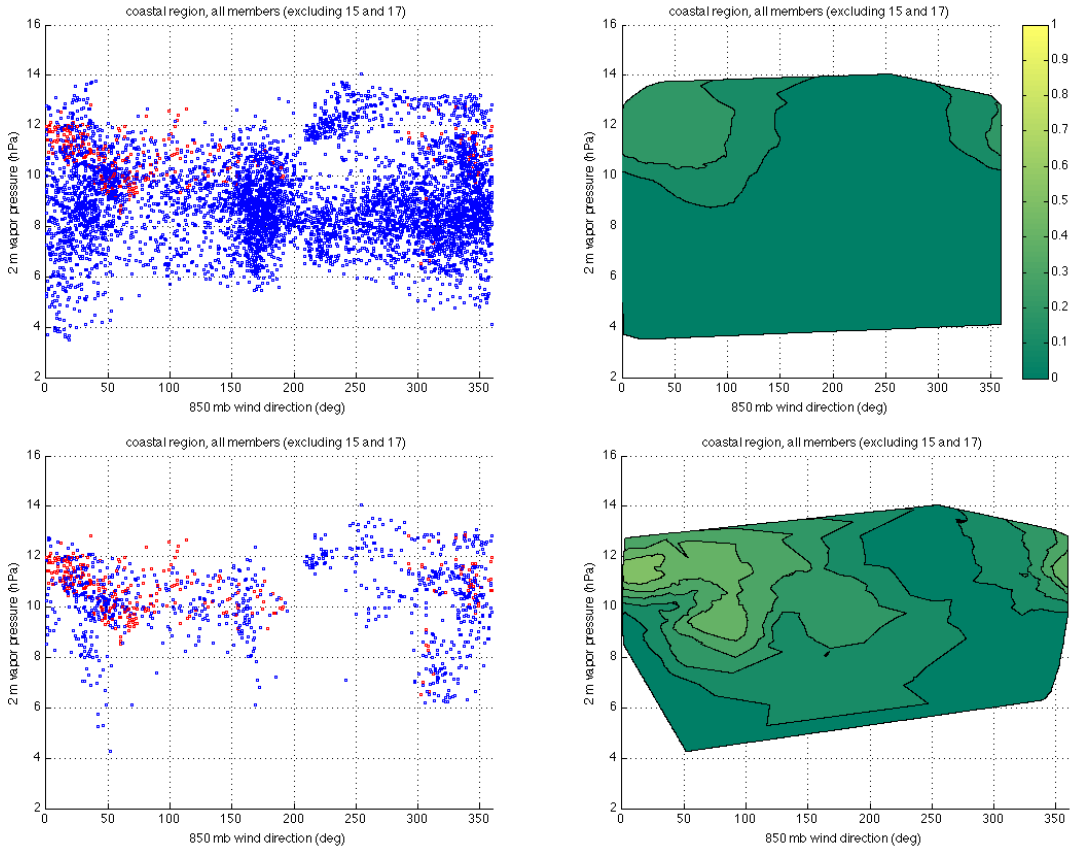

Figure 101. Same as Figure 71, but using 850-hPa wind direction predictions and 2-m vapor pressure predictions as the parameter pair. The top row distinguishes heavy fog missed opportunities (red) from heavy fog correct rejections (blue). The bottom row distinguishes heavy fog missed opportunities (red) from light fog missed opportunities (blue), and therefore displays the conditional probability of a heavy fog event given the occurrence of an unforecast (light or heavy) fog event. Heavy fog is defined as exceeding the highest β_e threshold of 2.1 km^{-1}, corresponding to daytime visibility of 0.875 mi. Light fog is defined as exceeding the lowest β_e threshold of 0.29 km^{-1}, corresponding to daytime visibility of 6.5 mi.

As a first possible link, northerly or northeasterly 850-hPa winds seem to provide the ideal conditions for radiation fog at these sites. These heavy fog events are characterized by calm or very weak northeasterly low-level winds, which are created with a surface high pressure center overhead or just offshore of the coastal sites. With weak vertical tilting, a high pressure center at 850-hPa would be expected just offshore of these

sites, creating northeasterly flow at this level. Upper air analysis during times of the heavy fog events indeed show an 850-hPa high pressure center is often present just offshore of these sites.

A second explanation is that the mainly offshore nature of the 850-hPa winds during these events results in greater cloud condensation nuclei (CCN) at the sites, which increases N during fog events. Various work (Thompson et al. 2008, Gultepe et al. 2009b) has suggested fog existing in a maritime or generally unpolluted airmass tends to have lower values of N than fog in a continental airmass, or an airmass near an urban area. This could reasonably be extended to include wind direction near the coast, where onshore winds are expected to advect lower N values from the maritime environment than offshore winds with a continental origin. The importance of N is that for a volume with a given q_c, many smaller droplets have a larger total cross-sectional area, and therefore larger β_e, than fewer larger droplets (Koenig 1971, Brenguier et al. 2000, Gultepe et al. 2006). Gultepe measured the relationship during RACE and found it was more precise than when N is ignored:

$$\beta_e = 3.904(q_c \cdot N)^{0.6473} \tag{13}$$

Depending on the airmass, recommended values of N vary in the literature between extremes of 40 cm^{-3} to over 300 cm^{-3}, a range that produces β_e changes that span several thresholds used in this work for a given q_c. In order to effectively use equation (13) for VIF prediction, more precise q_c predictions are needed from WRF without excessive zero q_c predictions. However, even without the benefit of more accurate q_c predictions, predictive information about N may have a role in a post-processing strategy. Predictions of N could be explicit from the WRF itself or deduced from other model variables.

Perhaps a more appropriate use of information regarding N is to predicts conditional fog severity. This concept is demonstrated in the bottom row of Figure 101, which used the same parameter pair as the top row applied to different datasets. As in the top row, the red points represent heavy fog missed opportunities. However, instead of plotting these cases with all other predictions that (correctly) do not include heavy fog, they are plotted against missed opportunities for all other (non-heavy) fog events. The

probabilities in the plot therefore provide the conditional probability of heavy fog, given the occurrence of any unforecast fog event. Forecast probabilities are as high as 0.5, an indication this parameter pair might have significant predictive value for identifying high probability of conditional heavy fog. When properly used in a post-processing strategy, this conditional probability would be multiplied by the probability of all fog (i.e., at the lowest β_e threshold), as determined by another parameter pair better suited for that task (for example, the parameter pair used in *JP_U)*.

The major advantage of using conditional probabilities in post-processing is that it can leverage certain parameters that have predictive usefulness for fog severity, but not necessarily for the presence of fog. *N* may be one of these, but several other parameters could also have this trait.

The specific example used in this appendix is intended to illustrate the potential uses of wind direction, *N*, and conditional probabilities in post-processing, but Figure 101 should not be considered a fully evaluated post-processing map since it has not been cross-validated. Additionally, this post-processing map is not universal (i.e., has little transferability) since the predictive usefulness of wind direction is likely to be highly dependent on several site-specific characteristics, including orientation of the coastline.

LIST OF REFERENCES

AFWA Model Analysis Team, cited April 2011: Forecast Products. [Available online at https://weather.afwa.af.mil:443/host_home/DNXM/ABOUTMM5/toc/products/index.html (CAC reqd)].

American Meteorological Society, cited 2012: Glossary of Meteorology. [Available online at http://amsglossary.allenpress.com/glossary].

Bang, C.-H., 2006: An evaluation of the visibility and fog forecasts using the WRF. *The KMS Spring conference*, Gwangju, Korea, 27-28 April.

Bankert, R. L, and M. Hadjimichael, 2007: Data mining numerical model output for single-station cloud-ceiling forecast algorithms. *Wea. Forecasting*, **22**, 1123–1131.

Barber, D., and H. Huhdanpaa, 1996: The quickhull algorithm for convex hulls. *ACM Trans. Math. Software*, **22**, 469–483.

Belfort Instrument, 2005: Automated Weather Observing System (AWOS) System Description. [Available online at http://www.digiwx.com/_media/client/pdf/digiwx/AWOS_SystemDescription.pdf].

Berner, J., G. J. Shutts, M. Leutbecher, T. N. Palmer, 2009: A Spectral Stochastic Kinetic Energy Backscatter Scheme and Its Impact on Flow-Dependent Predictability in the ECMWF Ensemble Prediction System. *J. Atmos. Sci.*, **66**, 603–626.

Bott, A. and T. Trautmann, 2002: PAFOG—A new efficient forecast model of radiation fog and low-level stratiform clouds. *Atmos. Res.*, **64**, 191–203.

Bremnes, J. B., and S. C. Michaelides, 2007: Probabilistic visibility forecasting using neural networks. *Pure Appl. Geophys*., **164**, 1365–1382.

Brenguier, J., H. Pawlowska, L. Schuller, R. Preusker, J. Fischer, and Y. Fouquart (2000), Radiative properties of boundary layer clouds: Droplet effective radius versus number concentration, *J. Atmos. Sci*., **57**, 803– 821.

British Atmospheric Data Center, 2006: Belfort Model 6230A Visibility Sensor. [Available online at http://badc.nerc.ac.uk/data/cardington/instr_v7/Belfort.html].

Chmielecki, R.M. and A.E. Raftery (2011). Probabilistic Visibility Forecasting Using Bayesian Model Averaging. *Monthly Weather Review* **139**, 1626--1636.

Clark, P. A., Harcourt, S. A., Macpherson, B., Mathison, C. T., Cusack, S., and Naylor, M., 2008: Prediction of visibility and aerosol within the operational Met Office Unified Model. I: Model formulation and variational assimilation, *Q. J. Roy. Meteor. Soc*., **134**, 1801–1816,

Dai, Aiguo, Kevin E. Trenberth, Thomas R. Karl, 1999: Effects of Clouds, Soil Moisture, Precipitation, and Water Vapor on Diurnal Temperature Range. *J. Climate*, **12**, 2451–2473.

Dai, A., J. Wang, R. H. Ware, and T. V. Hove, 2002: Diurnal variation in water vapor over North America and its implications for sampling errors in radiosonde humidity. *J. Geophys. Res.*, **107**, 4090.

Dejmal, Karel and Josef Novotny, 2011: Application of Fog Stability Index for significantly reduced visibility forecasting in the Czech Republic. Preprints, *Ninth International Conference on Fluid Mechanics and Aerodyniamics*, Florence, Italy, Association Mechanical Engineering and The World Scientific and Engineering Academy and Society, 317–320.

Delaunay, D., 1934: Sur la sphère vide. *Izv. Akad. Nauk SSSR*, **7**, 793–800.

Douglas, C. A., and R. L. Booker, 1977: *Visual Range: Concepts, Instrumental Determination, and Aviation* Applications. NBS Monogr., No. 159, Department of Commerce, 362 pp.

Duntley, S.Q., The reduction of apparent contrast by the atmosphere. *Journal of the Optical Society of America*. **38 i2**, 179–191.

Eckel, F. A., and C. F. Mass, 2005: Aspects of effective mesoscale, short-range ensemble forecasting. Wea. Forecasting, **20**, 328–350.

Eldridge, Ralph G., 1971: The Relationship Between Visibility and Liquid Water Content in Fog. *J. Atmos. Sci.*, **28**, 1183–1186.

Europa Technologies, Google, INEGI. "California." Map. *Google Maps*. Web. 23 Apr 2011.

Fast, Jerome D., William I. Gustafson, Elaine G. Chapman, Richard C. Easter, Jeremy P. Rishel, Rahul A. Zaveri, Georg A. Grell, Mary C. Barth, 2011: The Aerosol Modeling Testbed: A Community Tool to Objectively Evaluate Aerosol Process Modules. *Bull. Amer. Meteor. Soc.*, **92**, 343–360.

Federal Aviation Administration, 2012: Pilot/Controller Glossary (P/CG). [Available online at http://www.faa.gov/air_traffic/publications/atpubs/pcg/].

Freeman, L.E., Perkins, J.S. Meteorological Techniques, AFWA/TN-98/002, 242 pp. [Available online at http://www.comptus.com/PDF/AFWA_TN_98-002.pdf].

Garland, J .A., 1971: Some fog droplet size distributions obtained by an impaction method. *Quart. J. Roy. Meteor. Soc*., **97**, 483–494.

----------, J. R.. Branson and I. C. Cox, 1973: A Study of the contribution of pollution to visibility in a radiation fog. *Atmos. Environ*., **7**, 1079–1092.

Geiszler, D. A., J. Cook, P. Tag, W. Thompson, R. Bankert, and J. Schmidt, 2000: Evaluation of Ceiling and Visibility Prediction: Prelimary Results Over California Using the Navy's Couple Ocean/Atmosphere Mesoscale Prediction System (COAMPS). Preprints, *Ninth Conference on Aviation,Range, and Aerospace Meteorology*, Orlando, FL, American Meteorological Society, 334–338.

Ghirardelli, Judy E., and Bob Glahn. (2010) The Meteorological Development Laboratory's Aviation Weather Prediction System. *Weather and Forecasting* **25**, 1027–1051

Glahn, Harry R., Dale A. Lowry, 1972: The Use of Model Output Statistics (MOS) in Objective Weather Forecasting. *J. Appl. Meteor.*, **11**, 1203–1211.

Guidard V. and D. Tzanos, 2007: Analysis of fog probability from a combination of satellite and ground observation data. *Pure Appl. Geophys*., **164**, 1207–1220.

Gultepe, I and G. A. Isaac, 2004: Aircraft observations of cloud droplet number concentration: Implications for climate studies. *Quart. J. Roy. Meteor. Soc*., **130A**, 2377–2390.

----------, M. D. Müller, Z. Boybeyi, 2006: A New Visibility Parameterization for Warm-Fog Applications in Numerical Weather Prediction Models. *J. Appl. Meteor. Climatol.*, **45**, 1469–1480.

----------, M. D. Muller, and Z. Boybeyi, 2007: Fog research: A review of past achievements and future perspectives. *Pure Appl. Geophys*., **164**, 1121–1159.

---------- and J.A. Milbrandt, 2007. Microphysical Observations and Mesoscale Model Simulation of a Warm Fog Case during FRAM Project. *Pure Appl. Geophys*., **164**, 1161–1178.

----------, B. Hansen, S. G. Cober, G. Pearson, J. A. Milbrandt , S. Platnick, P. Taylor, M. Gordon, and J. P. Oakley, 2009a: The Fog Remote Sensing and Modeling Field Project. *Bull. Amer. Meteor. Soc.*, **90**, 341–359.

----------, P. Minnis, J. Milbrandt, S. G. Cober, G.A. Isaac, C. Flynn, L. Nguyen, and B. Hansen, 2009b: Visibility forecasing for warm and cold fog conditions observed during FRAM field projects. Preprints, *2009.Aviation, Range and Aerospace Meterol. Spec. Symp. on Weather-Air Traffic Management integration. AMS Annual Mtg.*, Phoenix, AZ, Amer. Meteor. Soc., P1.22

Hacker, J., S.-Y. Ha, C. Snyder, J. Berner, F. Eckel, E. Kuchera, M. Pocernich, S. Rugg, J. Schramm, and X. Wang, 2011a: The U.S. Air Force Weather Agency's mesoscale ensemble: Scientific description and performance results. *Tellus A*, **63**, 625–641.

----------, W. Ryerson, K. Nielsen, M. Jordan, 2011b: Probabilistic ceiling and visibility predictions for the Air Force Weather Agency. Final Report, *2011 Air Force Weather Agency Annual Review*. Unpublished.

Hall, Timothy J., Rachel N. Thessin, Greg J. Bloy, Carl N. Mutchler, 2010: Analog Sky Condition Forecasting Based on a k-nn Algorithm. *Wea. Forecasting*, **25**, 1463–1478.

Hansen, B., 2007: A fuzzy logic-based analog forecasting system for ceiling and visibility. *Wea. Forecasting*, **22**, 1319–1330.

Herzegh, P., G. Wiener, R. Bankert, R. Bateman, S. Benjamin, J. Brown, J. Cowie, M. Hadjimichael, M. Tryhane, and B. Weekley, 2006: Development of FAA national ceiling and visibility products: Challenges, strategies and progress. *Proceedings of the 12th Conference on Aviation Range and Aerospace Meteorology*. American Meteorological Society.

Hilliker, Joby L., Girish Akasapu, George S. Young, 2010: Assessing the Short-Term Forecast Capability of Nonstandardized Surface Observations Using the National Digital Forecast Database (NDFD). *J. Appl. Meteor. Climatol.*, **49**, 1397–1411.

Hippi, M., I. Juga, A. Rossa, F. Zardini, F. Domenichini, 2010: ROADIDEA – Road weather and model development – friction model and fog model. Report, Finnish Meteorological Institute and the Regional Agency for Environmental Prevention and Protectino of Veneto, 41 pp.

Jacobs, A. J. M., and N. Maat, 2005: Numerical guidance methods for decision support in aviation meteorological forecasting. *Wea. Forecasting*, **20**, 82–100.

Kalnay, Eugenia, 2003: *Atmospheric Modeling, Data Assimilation and Predictability.* Cambridge University Press, Cambridge, United Kingdom, 341 pp.

Keith, R., S. M. Leyton, 2007: An experiment to measure the value of statistical probability forecasts for airports. *Wea. Forecasting*, **22**, 928–935.

Koenig, L. R., 1971: Numerical experiments pertaining to warm-fog clearing. *Mon. Wea. Rev.*, **9**, 227–241.

Koschmieder, H., 1924: Theorie der horizontalen Sechtweite (Theory of horizontal visibility). *Beitr. Phys. freien Atmos.*, **12**, 33–53, 171–181.

Kuchera, cited April 2011: Weather Ensemble Technical Details. [Available online at https://www.intelink.gov/wiki/Weather_Ensembles_Technical_Details (pswrd reqd)].

Kumai, Motoi, 1973: Arctic Fog Droplet Size Distribution and Its Effect on Light Attenuation. *J. Atmos. Sci.*, **30**, 635–643.

Kunkel, Bruce A., 1984: Parameterization of Droplet Terminal Velocity and Extinction Coefficient in Fog Models. *J. Climate Appl. Meteor.*, **23**, 34–41.

Lim, Kyo-Sun Sunny, Song-You Hong, 2010: Development of an Effective Double-Moment Cloud Microphysics Scheme with Prognostic Cloud Condensation Nuclei (CCN) for Weather and Climate Models. *Mon. Wea. Rev.*, **138**, 1587–1612.

Liou, K.N., 2002: *An Introduction to Atmospheric Radiation.* 2nd edition, Academic Press, San Diego, 583 pp.

Marzban, Caren, Stephen Leyton, Brad Colman, 2007: Ceiling and Visibility Forecasts via Neural Networks. *Wea. Forecasting*, **22**, 466–479.

Mathworks, Inc., The, 2009: Delauney. [Available online at http://www.mathworks.com/help/techdoc/ref/delaunay.html].

Middleton, W. E. Knowles, 1954: *Vision Through the Atmopshere.* University of Toronto Press, 250 pp.

Morrison, H., J. A. Curry, V. I. Khvorostyanov, 2005: A New Double-Moment Microphysics Parameterization for Application in Cloud and Climate Models. Part I: Description. *J. Atmos. Sci.*, **62**, 1665–1677.

National Climatic Data Center, cited 2011: Online Climate Data Directory. [Available online at http://lwf.ncdc.noaa.gov/oa/climate/climatedata.html].

National Data Buoy Center, cited 2012: Website (Available online at http://www.ndbc.noaa.gov/].

National Oceanic and Atmospheric Administration, Department of Defense, Federal Aviation Administration, United States Navy, 1998: Automated Surface Observing System, (ASOS) User's Guide. [Available online at http://www.nws.noaa.gov/asos/pdfs/aum-toc.pdf].

National Weather Service Office of Meteorology, 1999: Automated Surface Observing System. [Available online at http://www.nws.noaa.gov/ost/asostech.html].

Office of the Federal Coordinator for Meteorological Services and Supporting Research, 2005: Federal Meteorological Handbook No. 1 – Surface Weather Observations and Reports. [Available online at http://www.ofcm.gov/fmh-1/fmh1.htm].

Pinnick, R. G., D. L. Hoihjelle, G. Fernandez, E. B. Stenmark, J. D. Lindberg, G. B. Hoidale, and S. G. Jennings, 1978: Vertical structure in atmospheric fog and haze and its effect on visibility and infrared extinction. *J. Atmos. Sci.*, **35**, 2020–2032.

Pruppacher, H. R. and J. D. Klett, 1996: *Microphysics of Clouds and Precipitation*. Springer, 976 pp.

Rasmussen, R. M., J. Vivekanandan, J. Cole, B. Myers, and C. Masters, 1999: The estimation of snowfall rate using visibility. *J. Appl. Meteor.*, **38**, 1542–1563.

Rogers, Eric, Thomas Black, Brad Ferrier, Ying Lin, David Parish, and Geoffrey DiMego, 2005: Changes to the NCEP Meso Eta Analysis and Forecast System: Increase in resolution, new cloud microphysics, modified precipitation assimilation, modified 3DVAR analysis. [Available online at http://www.emc.ncep.noaa.gov/mmb/mmbpll/eta12tpb/].

Rogers, R. R., and M. K. Yau, 1989: *A Short Course in Cloud Physics*. Butterworth–Heinemann, Bulington, MA, 304 pp.

Seifert, A, 2009: Physical Parameterizations II: Cloud Microphysics and Subgrid-Scale Cloudiness. *COSMO/CLM Training Course*, Langen, Germany. German Meteorological Service. [Available online at http://www.clm-community.eu/dokumente/upload/e8a46_Training09_Cloud_Physics_Seifert.pdf].

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X. Y. Huang, W. Wang, and J. G. Powers, 2008: A description of the advanced research WRF Version 3, Technical Report TN-475, National Center for Atmospheric Research.

Stoelinga, Mark T., Thomas T. Warner, 1999: Nonhydrostatic, Mesobeta-Scale Model Simulations of Cloud Ceiling and Visibility for an East Coast Winter Precipitation Event. *J. Appl. Meteor.*, **38**, 385–404.

Stull, Roland B., 1988: *An Introduction to Boundary Layer Meteorology*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 666 pp.

Tardif, R., 2007: The impact of vertical resolution in the explicit numerical forecasting of radiation fog: A case study. *Pure Appl. Geophys*., **164**, 1221–1240.

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev*., **136**, 5095–5115.

Tomasi, C., and F. Tampieri, 1976: Features of the proportionality coefficient in the relationship between visibility and liquid water content in haze and fog. *Atmosphere*, **14**, 61–76.

Vislocky, R. L., and J. M. Fritsch, 1995: Improved Model Output Statistics forecasts through model consensus. *Bull. Amer. Meteor. Soc.,* **76,** 1157–1163.

Waynant RW, Ediger MN, 2000: *Electro-Optics Handbook, Second Edition.* McGraw-Hill, Inc., New York, NY, 992 pp

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sci- ences—An Introduction*. International Geophysics Series, Vol. 59, Academic Press, 467 pp.

F. Yang, 2010: Neural network metamodeling for cycle time-throughput profiles in manufacturing. *Proceedings of European Journal of Operational Research*, 172–185.

Zhou, B., 2006: Asymptotic solution of liquid water content equation at equilibrium state of radiation fog. NCEP Office Note 450, National Centers for Environmental Prediction, 41 pp.

----------, J. Du, B. S. Ferrier, J. McQueen, and G. DiMego, 2007: Numerical forecast of fog—Central solutions. Preprints, *22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction*, Park City, UT, Amer. Meteor. Soc., 8A.6.

----------, and B. S. Ferrier, 2008: Asymptotic analysis of equilibrium in radiation fog. *J. Appl. Meteor. and Climatol*., **47**, 1704–1722.

----------, J. Du, J. McQueen and G. DiMego, 2009: Ensemble forecast of ceiling, visibility and fog with NCEP Short-Range Ensemble Forecast System (SREF) *Aviation, Range and Aerospace Meteorology Special Symposium on Weather-Air Traffic Management Integration, 89th AMS Annual Meeting*, Phoenix, AZ, 11–15, January 2009.

----------, J. Du, 2010: Fog Prediction from a Multimodel Mesoscale Ensemble Prediction System. *Wea. Forecasting*, **25**, 303–322.

----------, J. Du and G. DiMego, 2010: Introduction to the NCEP Very Short Range Ensemble Forecast System (VSREF), *14th Conf. on Aviation, Range, and Aerospace, 90th AMS Annual Meeting*, Atlanta, GA, 17–21, January.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Joshua Hacker
   Naval Postgraduate School
   Monterey, CA

4. Steven Rugg
   Air Force Weather Agency
   Offutt AFB, NE

5. Evan Kuchera
   Air Force Weather Agency
   Offutt AFB, NE

6. Ismail Gultepe
   Environment Canada
   Downsview, Ontario

7. William Ryerson
   Naval Postgraduate School
   Monterey, CA