Stony Brook University

Climatology and Predictability of Cool-Season High Wind Events in the New York City Metropolitan and Surrounding Area

A Thesis Presented

by

Michael Layer

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Master of Science

in

Marine and Atmospheric Science

(Concentration - Meteorology)

Stony Brook University

May 2014

Stony Brook University

The Graduate School

Michael Layer

We, the thesis committee for the above candidate for the

Master of Science degree, hereby recommend

acceptance of this thesis.

Brian A. Colle – Thesis Advisor
Professor, Marine and Atmospheric Science

Sultan Hameed – Second Reader
Professor, ITPA Director, Marine and Atmospheric Science

Jeffrey S. Tongue – Third Reader
Science and Operations Officer, National Weather Service, Upton, NY

This thesis is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Thesis

Climatology and Predictability of Cool-Season High Wind Events in the New York City Metropolitan and Surrounding Area

by

Michael Layer

Master of Science

in

Marine and Atmospheric Science

(Concentration – Meteorology)

Stony Brook University

2014

Damaging wind events not associated with severe convective storms or tropical cyclones can occur over the Northeast U.S. during the cool season and can cause significant problems with transportation, infrastructure, and public safety. These non-convective wind events (NCWEs) events are difficult for operational forecasters to predict in the NYC region as revealed by relatively poor verification statistics in recent years. This study investigates the climatology of NCWEs occurring between 15 September and 15 May over 13 seasons from 2000-2001 through 2012-2013. The events are broken down into three distinct types commonly observed in the region: pre-cold frontal (PRF), post-cold frontal (POF), and nor'easter/coastal storm (NEC) cases. Relationships between observed winds and some atmospheric parameters such as 900 hPa height gradient, 3-hour MSLP tendency, low-level wind profile, and stability are also studied. Overall, PRF and NEC events exhibit stronger height gradients, stronger low-level winds, and

stronger low-level stability than POF events. Model verification is also conducted over the 2009-2014 time period using the Short Range Ensemble Forecast system (SREF) from the National Centers for Environmental Prediction (NCEP). Both deterministic and probabilistic verification metrics are used to evaluate the performance of the ensemble during NCWEs. Although the SREF has better forecast skill than most of the deterministic SREF control members, it is rather poorly calibrated, and exhibits a significant overforecasting, or positive wind speed bias in the lower atmosphere.

Table of Contents

List of Tables

List of Figures

ensemble mean, calculated over the first forecast period (1-24 h). Biases were calculated for 39 land and water sites and interpolated to an evenly-spaced grid.

Fig. 26. Vertical profiles of temperature bias (in K) for the (a) SREF vs. RAP and (b) SREF vs. ACARS for the first forecast period (1-24 h) and (c) SREF vs. RAP and (d) SREF vs. ACARS for the third forecast period (51-72 h). Error bars are plotted at the 95% confidence interval.

Fig. 27. Brier Skill Scores of the old SREF control members (5) referenced to the full ensemble, calculated over increasing 10 meter wind speed thresholds (in m s^{-1}), for the (a) first forecast period (1-24 h) and (b) third forecast period (51-72 h). Error bars are plotted at the 95% confidence interval.

Fig. 28. Brier Skill Scores of the new full SREF ensemble referenced to each of the three SREF control members, calculated over increasing 10 meter wind speed thresholds (in m s^{-1}), for the (a) first forecast period (1-24 h) and (b) third forecast period (51-72 h). Error bars are plotted at the 95% confidence interval.

Fig. 29. Reliability diagrams for the old SREF 10 meter wind at three wind speed thresholds (10, 14, and 18 m s^{-1}) for the (a) first forecast period (1 to 24 h) and (b) third forecast period. The blue diagonal line represents a perfect forecast.

Fig. 30. Reliability diagrams for the new SREF 10 meter wind at three wind speed thresholds (10, 14, and 18 m s^{-1}) for the (a) first forecast period (1 to 24 h) and (b) third forecast period. The blue diagonal line represents a perfect forecast.

Fig. 31. Rank histogram for the (a) old SREF and (b) new SREF 10 meter wind, calculated over the first forecast period (1 to 24 h).

Chapter 1: Introduction

a. Background

Damaging wind events that occur in the absence of severe convection or tropical cyclones are referred to as non-convective wind events (NCWEs). For the New York City (NYC) National Weather Service (NWS), a High Wind Warning (HWW) is issued when the sustained winds reach or exceed 18.0 m s^{-1} (35 kts or 40 mph) and/or the wind gusts reach or exceed 25.7 m s^{-1} (50 kts or 58 mph). Since these (NCWEs) can be associated with a variety of weather phenomena ranging from heavy rain or snow to clear skies, there is less public awareness for these events as compared with severe convection or hurricanes. NCWEs occur primarily during the cool season, and often cause problems with transportation, infrastructure, and public safety. Research focusing on NCWEs will increase public awareness, improve public safety, and encourage authorities to plan more accordingly in advance of such events.

Ashley and Black (2008) analyzed the fatalities associated with NCWEs over the U.S. from 1980 to 2005. They found that the number of NCWE related fatalities (616) exceed those associated with tropical cyclones (181) and are nearly equivalent with those associated with straight-line thunderstorm winds (696). Most of the NCWE fatalities are associated with vehicles (43%) and boats (25%). When standardizing the NCWE fatalities by state area, Connecticut, New Jersey, and New York have the greatest number of normalized fatalities over the Continental U.S. Ashley and Black (2008) presented several physical and social explanations for this maximum of fatalities, noting that the Northeast U.S. is a favored region for strong mid-latitude cyclones, possesses a high population density, is rather heavily forested, and is in close proximity to large bodies of water. They also noted that fatalities may be increased due to the relatively low awareness of NCWEs, which are normally associated with extratropical cyclones

as opposed to hurricanes and severe thunderstorms.

NCWEs can be difficult to forecast due to the myriad of factors influencing the transport of high winds to the surface. Over a six year period from 2007-2013 there were a total of 422 county zone-based HWWs issued across the NYC NWS region (which includes all stations within the red outline in Fig. 1), 299 of which were not verified, yielding a false alarm ratio (FAR = false alarms / (false alarms + hits)) of 0.71 (NWS Performance Management). During the same period, there were a total of 191 observed events (individual reports of winds reaching warning criteria and/or wind damage), 150 of which occurred while a HWW was in effect at the time and location of the report, yielding a probability of detection (POD = hits / (hits + misses)) of 0.79. Given the dangers of NCWEs as referenced to in Ashley and Black (2008), it is in the interest of the research community, forecasting community, and the general public to explore ways to improve these poor verification statistics.

b. Past studies of high wind climatology and mechanisms

There have been many studies focusing on the climatology of NCWEs and their physical mechanisms over different regions, including the Great Lakes and Midwest U.S. (Lacke et al. 2007; Knox 2004; Crupi 2004; Durkee et al. 2012), western New York state (Niziol and Paone 2000), Texas (Schultz and Meisner 2009), the Pacific Northwest U.S. (Mass and Dotson 2010), the northern Plains U.S. (Kapela et al 1995; Kurtz 2010), the northeast U.S. (Lee and Girodo 1997; Asuma 2010), Alaska (Lynch et al. 2003), and western Europe (Browning 2004). Knox et al. (2011) conducted a comprehensive review of previous literature on NCWEs, looking at the historical and cultural significance of such events, as well as their climatological characteristics across many areas of the globe. The authors highlight four possible mechanisms for the development of strong low-level winds, including topography (Crupi 2004; Mass and Dotson

2010), isallobaric wind (Crupi 2004; Niziol and Paone 2000), tropopause folds (Browning and Reynolds 1994; Schultz and Meisner 2009), and sting jets (Browning 2004), but noted that the dominance of these mechanisms is case dependent. In addition, some of these mechanisms are not particularly relevant to the Northeast U.S., such as sting jets, which typically occur in association with the bent-back front of landfalling oceanic cyclones along the western coast of Europe (Browning 2004) and the Pacific Northwest U.S. (Steenburgh and Mass 1996).

Kapela et al. (1995) conducted a synoptic analysis of a post-cold frontal NCWE on 10 January 1990 over the Northern Plains and produced a checklist for the issuance of wind hazard products for the NWS, which was then applied to a set of twenty cases. A number of parameters were evaluated, including surface pressure gradient, 500 hPa vorticity and its advection, surface isallobars, subsidence between 500 hPa and 1000 hPa, cold air advection through the depth of the troposphere, specifically centered around 850 hPa, lapse rates in the lower troposphere, dry slot signatures on satellite imagery, 300 hPa jet position and strength, directional wind shear from the surface up to levels above the boundary layer, and surface geostrophic wind. A numerical scale was generated for the events based on the magnitudes of the aforementioned parameters, and was then used to separate the events into four categories: weak, moderate, strong, and watch/warn. The checklist of operational clues assisted forecasters in judging the potential severity and impact of impending high wind events. While the checklist might be useful for post-cold frontal NCWEs, it would need to be adjusted for pre-cold frontal or nor'easter NCWEs, which have significantly different synoptic conditions.

Although some mechanisms are dependent on the type of weather system producing high winds, many of the aforementioned studies argue that low-level stability and mean sea-level pressure (MSLP) evolution in conjunction with sufficiently strong low-level winds are key

factors in contributing to the momentum transfer of high winds to the surface (Kapela et al. 1995; Lee and Girodo 1997; Niziol and Paone 2000; Crupi 2004; Schultz and Meisner 2009; Mass and Dotson 2010; Asuma 2010). Therefore, the analysis of wind and thermal profiles is essential in forecasting NCWEs. However, additional smaller scale meteorological phenomena can arise and generate high surface winds with little or no warning. Gravity wave formation occurring on the cold (north and west) side of coastal storms (Bosart and Sanders 1986) and low-topped convection associated with narrow cold-frontal rain bands (NCFRs; Van Den Broeke et al. 2005) can produce locally higher surface winds, further complicating forecasts.

c. Wind forecasting, verification, and modeling studies

Hart and Forbes (1998) used hourly model-generated soundings from the Eta and Mesoscale Eta models at 48 km and 20 km grid spacing, respectively, to forecast non-convective wind gust probabilities (at no greater than 24 h lead times) for three thresholds (>30 mph (13 m s^{-1}), >40 mph (18 m s^{-1}), >50 mph (22 m s^{-1}) for 12 stations across the mid-Atlantic and Northeast U.S. They found that the wind speed at the second model level above the surface (height not specified) was the most reliable in predicting observed surface wind gusts, and by comparing several thousand model forecast and observation pairs, developed regression equations to forecast gust probabilities. Hart and Forbes also emphasized the influence of boundary layer stability on the vertical momentum transport, showing that the false alarm rate increases as static stability increases.

Whereas Hart and Forbes took a statistical approach to wind gust estimation, Brasseur (2000) developed a physical approach, considering the energy balance between turbulent kinetic energy and buoyancy, which was applied to a period of high cyclone activity over Western Europe between January and March 1990. In order to address the uncertainty of gust estimation,

Brasseur's method predicted an estimated value surrounded by a lower bound determined by varying the turbulent kinetic energy contribution and an upper bound determined by the maximum boundary layer wind speed. The width of the bound was influenced by the boundary layer height and static stability. Although the method accurately predicted daily maximum wind gusts within an error margin of 5 m s^{-1}, the hourly temporal evolution of wind gusts was highly dependent on the model fields. Lippman (2012) evaluated the performance of three different wind gust estimation methods, one empirical and two physical, including the method described in Brasseur (2000), and discovered that the empirical method, which simply takes into account the 2-s gust and 10-min mean wind, outperforms these two physical methods (Lippman 2012).

Many model simulation studies have been conducted for areas of complex terrain and sharp temporal variations in surface winds, including western Nevada (Cairns and Corey 2002), New Mexico (Rife and Davis 2005; Rife et al. 2009), and the Gulf of California (Cohen and Cangialosi 2009). Rife and Davis (2005) hypothesized that typical deterministic verification metrics penalize small spatial and temporal errors in mesoscale forecasts too harshly, since they do not discriminate among models with different abilities to represent small-scale features. They suggest feature or object-based verification as better alternatives. Casati and Ebert (2009) conducted an intercomparison study of several different spatial forecast verification methods, including neighborhood (Ebert 2009), scale-separation, feature-based, and field deformation methods. Although these metrics are typically used to evaluate precipitation forecasts, they could likely be applied to surface winds, which also exhibit distinct spatial patterns.

Errors in initial conditions (ICs) and other physical parameterizations can also impact the low-level wind forecasts. For example, Lackmann (2001) analyzed the evolution of potential vorticity (PV) in extratropical cyclones and its relation to the low level jet. He showed that the

cold-frontal maximum PV contributes to 15-40% of the strength of the low level jet. Consequently, errors in the representation of diabatic processes through model physics parameterizations, especially in cyclones containing heavy precipitation, can produce errors in the modeled low-level jet strength and potentially affect surface wind forecasts, particularly if conditions favor vertical mixing in the boundary layer (Brennan et al. 2008). IC uncertainty can also lead to large errors in coastal wind forecasts for extratropical cyclones. For example, Langland et al. (2002) showed that upstream errors over the Pacific Ocean resulted in poor forecasts a few days later for the January 2000 cyclone along the U.S. East coast.

Although the uncertainties originating from the ICs and model physics suggest the need for ensemble model forecasts to better predict NCWEs, only a few of these studies have taken an ensemble approach to probabilistically verify low-level or near-surface winds. Most ensemble studies for winds have been based in Europe (Pinson and Hagedorn 2012), in association with the expansion of the wind power industry in that region. Thorarinsdottir and Johnson (2012) used a non-homogeneous Gaussian regression applied to 48-hr ahead forecasts of wind speed obtained from the University of Washington mesoscale ensemble to probabilistically forecast daily maximum sustained wind, probability of wind gust, and daily maximum wind gust over the Pacific Northwest U.S. Applying this framework resulted in significant improvements in predictive performance over the raw ensemble or climatological reference forecasts. Zhang and Zheng (2003) evaluated the diurnal cycles of surface winds in the central U.S. as simulated by five different planetary boundary layer (PBL) parameterizations in the fifth-generation Pennsylvania State University (PSU)–National Center for Atmospheric Research (NCAR) Mesoscale Model (MM5). Although the study was conducted under fair conditions with light

winds, it still proves useful in determining relationships between surface winds and thermal profiles, as well as systematic biases in particular PBL schemes.

d. Motivation

There has been relatively limited research on the predictability of NCWEs over the Northeast U.S. These events have proven difficult for operational forecasters to predict for the NYC region as revealed by the relatively poor POD, FAR, and CSI. Finally, the atmospheric models and ensembles have not been verified extensively for these types of wind events over the Northeast U.S. It is imperative to improve the predictability of NCWEs through the use of both conceptual models and ensemble numerical weather prediction (NWP) models.

There are several outstanding questions that need to be addressed:

1. Are there common synoptic setups/regimes in which cool-season NCWEs occur in the NYC Metropolitan and surrounding area?
2. What atmospheric variables and mechanisms can be used to explain the production and transport of high winds to the surface in past events, and how can they be used to aid in the prediction of future events?
3. Can an ensemble forecast system improve the prediction of NCWEs as compared with deterministic forecasts?
4. How does the predictability of NCWEs change based on synoptic regime or event strength? What factors might be responsible for these predictability differences?

Chapter 2: Data and Methods

a. Datasets

The North American Regional Reanalysis (NARR) (Mesinger et al. 2006) and Meteorological Terminal Air Report (METAR) observations from 25 Automated Surface Observation System (ASOS) sites are used as the observational datasets for the climatology part of the study. Five-minute METARs, along with archived HWWs available through the National Climatic Data Center (NCDC) back to 2001 April are used to define and classify forecasted and observed NCWEs. The NARR data are then used to analyze the meteorological conditions during the events. Observational datasets used for the model verification part of the study include the Rapid Update Cycle (RUC; Benjamin et al 2004) 13-km analyses, Rapid Refresh (RAP; Brown et al 2011; Weygandt et al 2011) 13-km analyses, which replaced the RUC on 1 May 2012, METARs from 25 ASOS sites, Aircraft Communications Addressing and Reporting System (ACARS) profiles at three of 25 ASOS sites, five Automated Weather Observing System (AWOS) sites, one unclassified land site, and observations from eight buoy sites (Fig. 1). ACARS profiles of temperature and winds are spatially limited to the three major NYC airports: Newark (EWR), John F. Kennedy (JFK), and LaGuardia (LGA), labeled on Fig. 1. Vertical profiles centered at the airports are compiled by taking observations from landing and departing airplanes within a half hour window centered on the hour, or fifteen minutes either side of the hour. Unfortunately, ACARS data is not available for some NCWEs given inclement weather.

The Short-Range Ensemble Forecast (SREF) system from the National Centers for Environmental Prediction (NCEP) is evaluated for cool-season NCWEs in the New York City metropolitan area. SREF was chosen for this study because it incorporates a blend of different model cores, model physics, and initial condition perturbations to increase the range of possible

forecast outcomes (Du et al. 2009). This study uses the previous version of SREF that was operational from late October 2009 to August 2012 (Du et al. 2009), and the new version operational from August 2012 to present (January 2014). The old SREF consisted of 21 total ensemble members, with four different model cores: the Weather Research and Forecasting (WRF)-Advanced Research WRF (ARW), WRF-Nonhydrostatic Mesoscale Model (NMM), Eta, and Regional Spectral Model (RSM). The WRF-ARW, WRF-NMM, and RSM cores each contain a total of five members, including one control member, two negative perturbation members, and two positive perturbation members. The Eta core contains a total of 6 members, with two control members, each with different model physics, two positive perturbation members, and two negative perturbation members.

The current version of SREF also contains 21 ensemble members, but it now contains the following three model cores: the WRF-ARW, WRF-NMM, and the newly added NOAA Environmental Modeling System (NEMS) based Nonhydrostatic Multiscale Model on the B-grid (NMMB) model (Janjic et al., 2005; Janjic and Black, 2007). Each core contains seven members, including one control, and three positive and negative perturbation members. Horizontal grid spacing for the old SREF was 32 km for all model cores except the WRF-ARW, which was 35 km. Horizontal grid spacing for the new SREF was reduced to 16 km for all model cores. The old and new SREF data was utilized at 25 hPa spacing from 1000 hPa to 650 hPa, and 100 hPa increments from 600 hPa up to 300 hPa, along with near-surface variables such as 10-m wind and 2-m temperature. The PBL parameterization schemes in the old SREF are unique to each model core, with the WRF-ARW employing the Yonsei University (YSU) PBL scheme (Hong et al. 2006), the WRF-NMM employing the Mellor-Yamada-Janjic (MYJ) PBL scheme (Janjic 2001), the ETA employing the Mellor-Yamada (MY) PBL scheme (Mellor and Yamada 1982),

and the RSM employing the Medium-Range Forecast (MRF) PBL scheme (Hong and Pan 1996). The new SREF uses almost exclusively the MYJ PBL scheme, except for the NMMB n2 and p2 members, which use the Global Forecast System (GFS) PBL scheme (Han & Pan 2011). Two of four daily SREF runs are used, one early morning run initialized at 0900 UTC and one late afternoon run initialized at 2100 UTC, are used in this study. From these two run cycles, we use the model forecast hours in one-hour increments from initialization to six hours after initialization and then in three-hour increments from six hours to 72 hours after initialization.

b. High wind event definitions and categorizations

The cool season in this study is defined from 15 September to 15 May in order to account for anomalous early and late-season extratropical cyclone events. The period of record for the climatology section includes 13 cool seasons from 2000-2001 through 2012-2013, while the period of record for the model verification section includes five (two incomplete) cool seasons from 2009-2010 through 2013-2014. Since the old SREF system was upgraded on 28 October 2009, verification for the 2009-2010 season begins on 0900 UTC 31 October 2009. The old SREF verification runs from the 2009-2010 through 2011-2012 seasons. The SREF was upgraded during summer 2012, so verification for the new SREF runs from the 2012-2013 to the 2013-2014 (only up to 15 January 2014). Observed NCWEs are defined as at least one of the 25 ASOS sites in Fig. 1 reporting winds reaching warning criteria (sustained two-minute average wind speed $\geq$ 35 kts and/or wind gust $\geq$ 50 kts) for any period of time. Since NCWEs are not tied to smaller-scale convective storms, there should be a good chance that warning criteria can be satisfied at one of the 25 ASOS sites. NWS warning criteria follows the same sustained wind and gust thresholds, except that the sustained winds must persist for at least one hour. This definition was not followed in this study because setting the criteria so that every five-minute observation

within an hour must reach the 35 kt sustained wind threshold would result in a very small sample size, likely including only a few of the most significant events. Warning criteria observations separated by ≥ 9 h are counted as separate events, except in cases of regime change, such as the passage of a cold front. Observed events are classified as either hits or misses based on whether or not a NWS HWW was in effect at the time. False alarm NCWEs are defined as no observations reaching warning criteria during which a warning is in effect.

All NCWEs are also broken down into three different groups based on their regime type: pre-cold frontal (PRF), post-cold frontal (POF), and nor'easter/coastal storm (NEC), which is determined based on the track and evolution of the cyclone. A PRF NCWE occurs when the region experiences high winds ahead of a cold front, typically within the warm sector, of a cyclone passing to the north and west of the area. A POF NCWE occurs when the region of interest experiences high winds behind the cold front of a cyclone passing north of the area. A NEC NCWE occurs when the region of interest experiences high winds in association with a cyclone passing to the south and/or east of the area. Snapshots from 48 h before the start of the event, every 12 h until 24 h after the start of the event, are shown to give a clearer representation of the three different event types. Motivation for binning NCWEs into different groups stems from the desire to identify common synoptic conditions and possible mechanisms contributing to the development of high surface winds, determining if high winds occur under unique synoptic environments, and if their predictability changes by regime.

c. Structural evolution of wind events

Structural evolution for the three types of NCWEs is illustrated by generating composites of a number of important atmospheric variables using the NARR reanalysis data. 500 hPa geopotential height and standardized anomalies and mean sea-level pressure (MSLP) evolutions

are plotted over the continental scale, and 900 hPa height gradient and three-hour MSLP tendencies are plotted over the regional scale. Standardized anomalies of 500 hPa height are calculated to communicate the magnitude of their departure from climatological mean values over the cool season, defined over the 15 September through 15 May time period. Wind and thermal profiles at the local scale are vertically interpolated in order to gain insights into the low-level jet height and strength, PBL height, and low-level stability. Composites are intended to serve as pattern-recognition or analog devices to provide weather forecasters with advanced notice of impending NCWEs.

d. Model verification methodology

Surface observations at 39 land and water sites are used to verify 10-m wind, while the RUC/RAP and ACARS data are used to verify vertical profiles of wind and temperature in the SREF. Scalar verification metrics such as bias (mean error) and mean absolute error (MAE) are calculated to quantify the deterministic accuracy of the SREF, and the Brier Skill Score (BSS) and its components (reliability and uncertainty) help quantify the probabilistic skill. Calculating the BSS for the SREF ensemble referenced to each of the deterministic forecasts (five SREF control members for the old version, three SREF control members for the new version) communicate the probabilistic improvement in forecast skill of the SREF ensemble referenced to the deterministic forecasts. Point verification is conducted by implementing a bilinear interpolation to the observation sites. Verification statistics are averaged over all 39 land and water sites for 10 m winds, and over three ACARS sites (EWR, JFK, and LGA) for vertical profile verification. ACARS data does not exist at specific model levels and is interpolated vertically to match the vertical resolution of the SREF and RUC/RAP grids. Unfortunately, in some instances where ACARS data may be sparse, interpolation results in the creation of

unrepresentative values. In order to prevent the introduction of erroneous data into the ACARS profiles, any data displaced ≥ 25 hPa from the model levels is not interpolated. Verification is separated into three forecast periods to convey trends and differences in predictability as a function of lead time: first period (1-24 hour lead time), second period (27-48 hour lead time), and third period (51-72 hour lead time). Verification is also separated into three different wind direction groupings: west to north (270 to 360 degrees), north to southeast (0 to 135 degrees) and southeast to west (135 to 270 degrees). This is done to determine if model performance differs by synoptic regime. Statistical significance is calculated at the 95% confidence interval for all scalar verification and BSS plots. Error bars are determined by using bootstrapping to randomly resample the data points 1000 times.

Chapter 3: Composite Results

a. Distribution of events

Using the 25 ASOS stations in Fig. 1, there were 61 observed NCWEs from the 2000-2001 through 2012-2013 seasons, including 14 PRF, 33 POF, and 14 NEC events (Table 1). There was only one observed event in both September and May, three events in both November and April, seven events in both October and March, 10 events in February, 13 events in January, and 16 events in December. Nearly 64% of the events occurred during the three-month climatological winter period of December through February. PRF and NEC NCWEs are most commonly observed in the month of December, while POF NCWEs are most commonly observed in January (Fig. 2). Hits, missed events, and false alarms were determined by comparing the observations to issued NWS HWWs. Unfortunately, since warning data is only available back to April 2001, five observed events occurring before then could not be verified as either hits or misses. Out of 56 observed events from the 2001-2002 through 2012-2013 seasons, there were 41 misses and only 15 hits. There were also 12 false alarm cases during this period, including 1 PRF, 4 POF, and 7 NEC events (Table 2). The statistics calculated includes: false alarm ratio (FAR = false alarms / (false alarms + hits), Probability of Detection (POD = hits / (hits + misses), and critical success index (CSI = hits / (hits + misses + false alarms). These statistics yield a FAR of 0.44, a POD of 0.27, and a CSI of 0.22. Part of the reason for this poor verification might be that HWWs are issued for sustained winds of $\geq$ 35 kts persisting for at least one hour, whereas in the study we classified a NCWE as occurring even if only one five-minute observation at any ASOS site in the area of interest reached the sustained wind threshold. As mentioned in the previous section, setting the criteria to have every five-minute observation reach the 35 kt sustained wind threshold would likely result in a very small sample size. Five-

minute observations available from NCDC were also chosen over hourly ones because higher sustained winds and/or gusts occurring in between hourly observations would be missed. Using hourly METARs (at least one hourly METAR meeting the 35 kt sustained wind and/or 50 kt gust thresholds) to classify events yielded only 26 observed events (13 hits and 13 misses) during this same time period, with a FAR of 0.48, POD of 0.5, and CSI of 0.34 (not shown). Table 3 gives some insight into how the observed events were reached in this study. When looking at all events, the majority (62%) were reached by sustained wind only, whereas only two events were reached by gust only. The majority of PRF NCWEs were reached by both sustained wind and gust meeting warning criteria, while NEC events were evenly balanced between being reached by sustained wind only (50%) and both sustained wind and gust (43%). However, nearly 80% of all POF events were reached by sustained wind only (Table 3), a result contributed to by using a more lenient criteria (five minutes vs. one hour) and the fact that some POF events are reached by low-end HWW criteria winds (maximum wind speed of 35 kts and maximum gust of only 40-45 kts).

Overall, NEC NCWEs exhibit the highest FAR (0.58) and POF NCWEs exhibit the lowest POD (0.13) and CSI (0.12) (Table 4). Verification statistics were also computed over three four-season time periods to decipher any trends. Table 5 shows that the worst verification was over the 2001-2005 period (2001-2002 through 2004-2005 cool seasons). Statistics were also computed for each individual season, and the best verification occurred during the 2012-2013 season, which included Hurricane Sandy in late October 2012. There were four hits, one missed event, and two false alarms, yielding a FAR of 0.33, POD of 0.8 and CSI of 0.57 (not shown) during the 2012-2013 cool season. Table 6 also gives some interesting insights into the spatial variability of high winds in NCWEs. When looking at the ASOS sites of maximum observed

sustained wind and gust, the highest winds are clearly observed in urban and coastal locations (JFK, LGA, EWR), likely due to a combination of reduced frictional drag and greater near-surface heating. For POF NCWEs (33 events), maximum winds are almost exclusively observed at the three major NYC terminals of JFK, LGA, and EWR (Table 6). For observed PRF (14 events) and NEC NCWEs (14 events), the site of maximum wind is more evenly distributed, but is generally confined JFK and LGA for PRF and LGA for NEC events (Table 6). However, the inland site HPN occasionally observes the highest wind in events with northerly or southerly winds, suggesting that winds may be enhanced when blowing parallel to the Hudson Valley. For false alarm NEC cases, maximum winds tend to be focused further east, especially at sites GON and FOK (Table 2).

b. Spatial climatology

Composite 500 hPa heights and MSLP, along with their standardized anomalies referenced to the climatological mean from 15 September through 15 May for 1979 through 2012, illustrate the synoptic scale evolution of the three different types of NCWEs. The period extends from 48 hours before the start of the event (t-48) to 24 hours after the start of the event (t+24) every 12 hours. PRF NCWEs exhibit a 500 hPa trough initially over the interior western U.S. and Rocky Mountains at t-48 (Fig. 3a), which deepens to around -3 standard deviations greater than the climatological mean and pivots to a negative tilt over the Great Lakes by t-0 (Fig. 3b). Meanwhile, there is an anomalous upper-level ridge over +1 standard deviation from the climatological mean occurring over the western Atlantic and eastern Canada. Surface low pressure develops over the lower Mississippi Valley at t-36 to t-24 (Fig. 4a) and moves northeastward through the eastern Great Lakes region at t-0 (Fig 4b). The cyclone deepens as it passes northwest of the New York City metropolitan area during that time period, thus placing it

within the warm sector. Surface high pressure is centered to the east of the region, strengthening slightly as it retreats into the western Atlantic (Figs. 4a, b).

POF NCWEs are preceded by a generally zonal flow over the eastern half of the U.S. at t-48 (Fig. 5a), with a weak positively tilted 500 hPa trough extending from the Northern Plains into the Rocky Mountains. The trough translates eastward, becoming negatively tilted over southeastern Canada and the Northeast U.S. and deepens to below -2.5 standard deviations from the climatological mean by t-0 (Fig. 5b). The MSLP evolution has a northern stream dominant system moving west to east from the Great Lakes at t-24 (Fig. 6a) to New England at t-0 (Fig. 6b), deepening rapidly as it passes north of the local area by t-0 (Fig. 6b), and then turning northeastward into the Canadian Maritimes as occlusion begins by t+12 (not shown). A large high pressure system dives southeastward from the northern Plains into the Tennessee Valley behind the exiting cyclone from t-12 to t+12.

NEC NCWEs exhibit a relatively similar 500-hPa height evolution as the PRF events, except that the trough digs further south and east and its base is less sharp. At t-48, a trough is present over the Central U.S. (Fig. 7a), which translates eastward, deepening and becoming negatively tilted by t-12 to t-0 (Fig. 7b), when the trough extends from the Great Lakes into the Southeast U.S. The bagginess at the base of the trough suggests the existence of a closed low in some cases. Looking at individual cases, a closed 500 hPa low was present in seven of 14, or 50% of the cases, between t-24 and t-0. Negative height anomalies exceed -2.5 standard deviations from the climatological mean over the Southeast U.S. and Mid-Atlantic, and positive anomalies exceed +1 standard deviation over far eastern Canada by t-0 (Fig. 7b). MSLP evolution reveals a large blocking high pressure system that is present ahead of the storm, extending from interior southeast Canada into northern New England from t-48 to t-0. At t-24

(Fig. 8a), low pressure begins to develop off the Atlantic side of the Southeast U.S. coast. The cyclone tracks north-northeastward, deepening as it travels from coastal North Carolina at t-12 to southeast of Long Island at t-0 (Fig. 8b) to just east of Cape Cod at t+12 (not shown).

Northeast U.S. regional composites of mean 900 hPa geopotential height and gradient (Figs. 9a-c) and 3-hour MSLP tendency (Figs. 10a-c) are calculated over all hours that HWW criteria winds are observed. The 900-hPa height gradient is plotted instead of the MSLP gradient in order to reduce the terrain influence on the values, although some terrain influence is still evident in the vicinity of the highest mountain ranges. MSLP tendency is plotted to illustrate potential contributions from the isallobaric wind, which blows from maximum pressure rises to maximum pressure falls. A tighter MSLP tendency gradient correlates to a stronger isallobaric wind. Comparing the 900 hPa height gradient plots between the three event types shows that the gradient is stronger for PRF (15-25 m/100 km) (Fig. 9a) and NEC NCWEs (15-20 m/100 km) (Fig. 9c) than for POF NCWEs (12-17 m/100 km) (Fig. 9b). These plots suggest that stronger 900-hPa winds are observed in PRF and NEC NCWEs than POF NCWEs. Plotting composite 900 hPa winds indeed verifies this hypothesis; the maximum 900 hPa winds over the region of interest is ~65 kts for PRF events, ~55 kts for NEC events, and ~40-45 kts for POF events (not shown).

The MSLP tendency plots show the strongest pressure changes are observed in PRF NCWEs, with maximum pressure falls approaching -6 hPa/3 h over central New England (Fig. 10a). The orientation of the tendency contours suggests an isallobaric enhancement of southerly winds across Long Island and southern New England. POF NCWEs exhibit maximum pressure rises up to +3.5 hPa/3 h over Pennsylvania into western and central New York State (Fig. 10b). The meridional orientation of the tendency contours suggests an isallobaric enhancement of

westerly winds across a large portion of the Northeast U.S., particularly from NYC and Long Island through New England. NEC NCWEs exhibit maximum pressure falls to -3.5 hPa/3 h east-northeast of the metropolitan area, centered just east of Cape Cod (Fig. 10c). However, the orientation of the contours suggests an isallobaric wind from the southwest, which may act to reduce the total wind because it opposes the low-level northeasterly flow observed during these events. The isallobaric contribution may be very sensitive to the track of NEC events, since one can imagine that a cyclone tracking further west would likely result in a northeast to southwest isallobaric wind, which would act to enhance the surface winds over the area.

c. Vertical composites

Mean vertical profiles of temperature, potential temperature, and wind speed are generated for all hours that warning criteria winds are observed using the NARR (Figs. 11a-c), and interpolated to the site(s) of maximum wind and gust in each event to illustrate the characteristic local meteorological factors in observed NCWEs. Thin lines plotted on either side of the mean represent the highest and lowest value observed for each level, not necessarily observed during the same event. Comparing the profiles between the three types of NCWEs illustrates similar profiles between PRF and NEC events, and distinct differences between these two types and POF events. PRF and NEC NCWEs have a distinct low-level jet (LLJ) structure, with a wind maximum between 800 and 900 hPa (Figs. 11a, c), which is consistent with the stronger height gradients shown in the regional composite plots. The mean LLJ is slightly stronger for PRF (65 kts) than NEC events (60 kts), although the maximum wind is greater for NEC NCWEs (~100 kts) than for PRF events (~82 kts). Low level stability tends to be highest for PRF events, as evident by the isothermal to weakly inverted temperature profile in the lowest 50 to 75 hPa and increasing potential temperature with height (Fig. 11a). The mean stability is

weaker for NEC events than PRF events, but the thermal profiles tend to be more variable for coastal storms events, as illustrated by the upper and lower bounds of temperature and potential temperature (Fig. 11c). Some events occur during isothermal to weakly inverted conditions while others occur with fairly well-mixed low levels, where temperature decreases with height and potential temperature is nearly constant with height (Fig. 11c). Wind profiles in POF events generally do not have a LLJ signature, but rather a monotonic increase with height (Fig. 11c), and low level winds on average are significantly weaker than the other two types of NCWEs. Some events do exhibit LLJs, as evident by the bulge in the upper bound wind maximum (Fig. 11b), but they tend to be more elevated (750 to 850 hPa). However, the thermal profiles illustrate much weaker low-level stability and deeper mixed layers in the POF events, with temperature decreasing rapidly all the way up to 825 hPa in some cases (Fig. 11b). Therefore, the weaker stability compensates for the weaker LLJs observed in POF versus PRF or NEC NCWEs. The wind profiles also serve as verification of the regional height gradient composite plots, which showed tighter gradients in PRF and NEC NCWEs.

d. Relationship between wind parameters

Scatterplots were generated of the maximum 900 hPa height gradient versus the maximum 10-meter two-minute average wind speed (Fig. 12a) and gust (Fig. 12b) observed in the region for each event, including observed (60 blue circles) and false alarm (12 red plus symbols) NCWEs. Trend lines were also fit to the observed data points to determine the correlation between the two variables. In order to compute the correlations, Hurricane Sandy had to be taken out of the dataset because the much higher value would skew the correlation much higher. Additionally, since the data points are not normally distributed, the correlation was calculated by computing the median of 10 groups of six points, and then fitting a trend line to

those median values. Interestingly, the correlation between the maximum height gradient and sustained wind is small (0.33) (Fig. 12a), but for wind gust is much more robust (0.84) (Fig. 12b). These results suggest that the height gradient and LLJ strength cannot solely be used to discern between an event and a non-event. The scatterplots also raise other concerns, including the fact that observed warning criteria winds can be observed with a height gradient as low as ~14 m/100 km, and yet false alarms can occur with height gradients over 25 m/100 km (Figs. 12a, b). It was hypothesized that the three-hourly temporal resolution of the NARR may have had an influence on the maximum observed height gradient. Values on the low end were though to be too low in cases of short-duration events, where maximum values might have occurred between the three-hour observations. However, after re-calculating the maximum height gradient for 18 events between 2009 and 2012 with the hourly RUC analyses, values at the lower end of the range (~14 m/100 km) did not increase. Other physical factors, such as low-level stability, isallobaric wind contribution, heavy precipitation, and other mesoscale effects, are likely influencing the downward momentum transfer of high winds to the surface. Individual correlations between maximum 3-h pressure tendency, and minimum 1000 to 900 hPa stability with maximum wind speed and gust were also calculated (not shown), but there were no significant linear correlations between these variables, suggesting that the relationship is quite complicated and variable from event to event.

Histograms showing the number of hourly observations of warning criteria winds versus time of day were created to see if diurnal factors play a significant role (Figs. 13a-c). In comparing the three histograms for the three types of NCWEs, POF events exhibit a clear diurnal signal, with a maximum of high wind observations during the daytime hours (1500-2000 UTC) and a minimum during the early morning hours (0600-1100 UTC) (Fig. 13b). PRF events also

show a maximum in observations during the daytime hours (Fig. 13a), while NEC NCWEs show a maximum during the nighttime hours (0000-0900 UTC) (Fig. 13c). The diurnal signal for the PRF and NEC NCWEs is not statistically significant given the relatively small sample size, and these events are typically occurring in conjunction with extensive cloud cover and precipitation, such that less diurnal effects would be expected. However, the diurnal signal for POF events is statistically significant, since these events are characterized by cold and dry advection in the wake of a departing low pressure and/or frontal system. Events occurring during the daylight hours may be enhanced by reduced cloud cover, which results in greater near-surface heating, low-level lapse rate steepening, and thus more efficient mixing. In contrast, reduced cloud cover during the early morning hours leads to greater near-surface cooling, resulting in the development of a low-level inversion, and thus inhibiting mixing. Another plot showing the times of valid HWWs for false alarm events versus time of day was created (not shown), because it was hypothesized that the time of day may impact the verification of warnings. But it was found that the times between observed and false alarm events were fairly similar, and so the diurnal argument cannot be used to explain the occurrence of false alarms.

The influence of heavy rainfall on downward momentum transport was also explored since it was hypothesized that heavy rain may have contributed to some missed events for PRF and NEC cases. After analyzing the five-minute METAR data by hand, only four missed events (two PRF and two nor'easters) were classified as having occurred specifically due to heavy rainfall: 11 December 2003 at ISP, 2 February 2008 at ISP, 5 March 2008 at JFK, and 8 December 2011 at FOK. All four events observed a brief but significant increase in rainfall intensity and wind, as seen in Table 7 a-d. However, it should be noted that the event occurring on 8 December 2011 was not necessarily due to heavy precipitation alone, but due to an intense

gravity wave that pushed across Long Island, evident in the rapid surface pressure drop and subsequent rise (as high as +20 hPa/h at FOK). All false alarm PRF (1) and nor'easter (7) NCWEs were found to have only light or moderate precipitation at the time of maximum wind, but since the sample size is small, it cannot be said with confidence that lack of heavy rainfall was the main culprit for those unverified warnings.

All 12 false alarm cases were also explored in detail (Table 8), including regional plots of 900 hPa height gradient and three-hour MSLP tendency, and local wind and thermal vertical profiles. Each case was classified into three categories: where the data suggested high confidence of warning level winds (green cases in Table 8), where the data suggested a possibility of warning level winds (yellow cases), and where the data suggested little to no confidence of warning level winds (red cases). These categories were created based on low-level winds and stability and the pressure tendency to get an idea of the magnitude and direction of the isallobaric wind. This revealed that the warning decision process can be quite tedious and also rather difficult for these events. The numbers shown in Table 8 are based off the NARR reanalyses, so it cannot be assumed that these are the exact numbers that forecasters were looking at during the warning decision process with some lead time before the events. Table 8 shows that five of the 12 false alarms were "surprise" busts, in which a warning seemed likely to verify based on the analysis data, while three of 12 were "on the fence", and the remaining four of 12 were more obvious busts, in which a warning seemed unlikely to verify based on the data. SREF data was available for six of 12 false alarm cases, which were manually verified by comparing the mean vertical wind profile and 10 m wind of the ensemble mean and the observations. Perhaps the biggest surprise bust event was 25 February 2011, in which winds were 50 kts at the top of the mixed layer (~900 hPa), 55 kts just above (875 hPa), and a strong isallobaric contribution in the

same direction as the low-level flow. This bust event had the largest wind speed bias in the LLJ region out of the six events verified with the SREF, with the ensemble mean forecasting winds >5 kts higher than the observations between 825 and 925 hPa (Fig. 14). Several nor'easter false alarm cases also seemed likely to verify based on the analysis data, but it should be noted that in almost all the cases, the isallobaric wind opposes the low-level flow. As for the events that seemed unlikely to verify, 2 March 2009 was the most obvious bust, since the maximum LLJ was only 42 kts, well below warning criteria, and the thermal profile was also rather stable (Figs. 15a-c). For the three events "on the fence," it is clear that a slight shift in mixing height, stability, and low-level winds would have yielded a different outcome. For example, in the 10 February 2001 false alarm case, winds of 53 kts exist just above the subsidence inversion (Figs. 16a-c), so a slight increase in mixing height would likely have brought higher winds down to the surface. For some of these bust and marginal cases, it is likely that the numerical guidance may have been too strong with the LLJ, which motivates the next section of the thesis on model validation.

Chapter 4: Model Verification

a. Deterministic verification

Verification of vertical wind profiles in the old and new SREF was completed for three land sites with abundant ACARS profile data (EWR, JFK, and LGA). The 235 (123) events chosen from October 2009 to May 2012 (September 2012 to January 2014) were those in which the 10 m sustained wind observed and/or forecasted by any SREF member at any of the 25 ASOS sites was $\geq$ 10.29 m s^{-1} (20 kts). Events persisting for less than three consecutive hours were excluded because the three-hour temporal resolution of the SREF would possibly miss shorter duration events and thus artificially skew the verification results. The ensemble was also validated using the RUC analyses, which had some notable differences as compared with using the ACARS data. Verification was averaged over three different forecast lead time periods: first forecast period (1-24 h), second forecast period (27-48 h), and third forecast period (51-72 h).

The mean observed wind and temperature profiles for the RUC and ACARS is shown to put the bias plots in perspective (Figs. 17a, b). Some differences are evident, with the RUC observing stronger winds than ACARS in the LLJ region (850 to 975 hPa), and weaker winds than ACARS throughout much of the middle atmosphere (400 to 775 hPa). Low-level wind biases are ~0.5 m s^{-1} smaller in magnitude when calculated against the RUC around 950 hPa (Fig. 18a), with the highest biases (1.5-2.5 m s^{-1}) using RUC at the 1000 hPa level. Using ACARS, the bias (model minus observations) between the SREF subgroups (ARW, Eta, NMM, and RSM) and ensemble mean reveals a 1-2.5 m s^{-1} high wind speed bias in the low levels (850 to 1000 hPa) for the 1-24 h period (Fig. 18b), with the largest bias centered between 900 and 950 hPa. The bias is largest for the WRF-ARW model core, while the WRF-NMM and Eta cores exhibit the smallest error (Figs. 18a, b). Using RUC, the bias during the 51-72 h period is similar

at the surface as earlier forecast periods (Fig. 18c), but it is < 1 m s^{-1} for all members except ARW between 850 and 975 hPa. The SREF wind speed bias in the ARW members using ACARS is similar around 950 hPa (2-3 m s^{-1}) for the 51-72 h period (Fig. 18d), while the bias in the other members and the ensemble mean decreases to 0.5-1.5 m s^{-1}. Overall, differences between ACARS and RUC in the low-levels are likely reflective of a systematic bias in the RUC, which underestimates 10 m winds when compared with raw METAR observations. RUC 10 m winds were verified against hourly METAR observations at the three ACARS sites for 18 observed NCWEs between 2009 and 2012, and were found to be as much as around 10 m s^{-1} lower than raw observations.

Mean absolute error (MAE) plots show similar patterns as the bias, with the MAE between the SREF and RUC more uniform through the lower atmosphere, with a slight maximum at 1000 hPa not seen using ACARS (Figs. 19a, b). MAE between the SREF and ACARS is larger in the LLJ region (2.5-3.5 m s^{-1} at 900-950 hPa), and the lowest MAE observed at the 1000 hPa level (~2 m s^{-1}) during the first 1-24 h forecast period (Fig. 19b). Of the four model cores in the old SREF, the WRF-ARW, which had the highest bias, also has the highest MAE, especially by the 51-72 h period (Figs. 19c, d). Due to the increasing magnitude of winds with height, the highest MAE is observed in the upper levels (above 600 hPa), which increases above 3 m s^{-1} in the first forecast period (Figs. 19a, b) and above 6 m s^{-1} by the third forecast period (Figs. 19c, d). SREF MAE with respect to both RUC and ACARS increases in magnitude with increasing lead time at all model levels, most significantly for the upper levels.

Bias of 10 m wind is also calculated for all 39 land and water observation sites and contoured on a map to convey any possible spatial variability in the SREF verification. SREF 10 m winds were adjusted for the 8 buoy sites and MTP because the anemometer heights at these

stations are lower (3.5 to 5 m), except for one buoy off the coast of southeast Connecticut (20 m). A simple exponential relationship assuming neutral stability was used to adjust the winds. The spatial bias of the ensemble mean for the first forecast period (Fig. 20a) illustrates a low bias (-1 to 1 m s^{-1}) in the immediate NYC metropolitan area, and extending northward into the Lower Hudson Valley. The highest biases are located over Suffolk County, Long Island (3-3.5 m s^{-1}), and also the suburbs north and west of NYC (2.5-3 m s^{-1}), while the lowest biases are observed over water, especially western Long Island Sound and off the southern coast of Queens and Nassau counties. The three sites with the highest biases are FOK, HWV, and WST. High model biases at these stations are likely due to a combination of the SREF treating these locations as water points and less than ideal anemometer siting. On the other hand, low biases in NYC may be due to a combination of enhanced mixing as a result of the urban heat island effect, good anemometer siting, and more reliable measurements as a result of newer anemometers and better quality control. Analysis of the biases of the four SREF subgroups shows relatively similar spatial patterns between the ARW, NMM, and RSM cores, with significant positive biases over eastern Long Island, and slightly negative biases extending from the NYC area northward into the Lower Hudson Valley. The Eta core exhibits much higher bias overall compared to the other subgroups, with positive biases >1 m s^{-1} observed throughout nearly the entire region of interest except for the buoy sites, and approximately half of the area observing biases >3 m s^{-1} (Fig. 20b). The 10 m wind speed biases did not change significantly with increasing lead time (not shown).

The SREF temperature bias is also calculated using RUC and ACARS. The mean observed temperature profiles are fairly similar between the two observation sources (Fig. 17b), and as such, the biases are also similar. A negative temperature bias exists in the low levels (850 to 1000 hPa), suggesting the model is too stable. However, this bias is small in magnitude,

generally between 0 and -1 K for all model cores in the first forecast period (Figs. 21a, b). With increasing lead time, the biases change little for all model cores except the RSM, which observes an increase in negative bias to around -1.5 K in the LLJ region by the third forecast period when verified against both RUC and ACARS (Figs. 21c, d). Since the temperature biases are still rather small, it likely does not have a significant effect on mixing efficiency. It should also be noted that the ACARS temperature data appears somewhat questionable based on these plots; not enough data was available to calculate model error at the 1000 hPa level, and a positive bias of +1 K that isn't observed when calculating error with respect to the RUC is present over a large portion of the middle atmosphere (400 to 700 hPa) when calculated with respect to ACARS (Figs. 21b, d). This discrepancy may be due to the increasing distance of airplane measurements from the airports at higher altitudes.

Comparing verification between the three regimes shows that the WRF-ARW has a significantly higher wind bias between 925 and 975 hPa than the other subgroups in W-SE and N-SE flow, suggesting that this model core overestimates the strength of LLJs in onshore flow regimes. The Eta core has a significantly lower bias than the other SREF subrgroups between 925 and 975 hPa in W-SE flow, with the bias compared to the RUC approaching zero m s^{-1} during the first forecast period. Overall, SREF MAE in the low levels is lower for the W-N flow regime (1-3 m s^{-1}) than the N-SE and W-SE regimes (2-4 m s^{-1}). The difference in temperature verification between the three regimes is more noticeable than compared to winds. Temperature biases at 925 to 1000 hPa are -1 to -2 K for W-N flow, while they are slightly positive (0 to 1 K) for N-SE flow, and in between (0 to -1 K) for W-SE flow. The RSM subgroup has the most negative near-surface temperature bias for W-N flow, and the positive slope of the line suggests it is too stable in the low levels, while for N-SE flow, the RSM has the most positive near-surface

bias, and the negative slope of the line suggests it is too unstable in the low levels. Overall, these patterns do not change significantly with increasing lead time, although the spread between model subgroups increases.

Similar wind and temperature biases were found in conducting the verification of the new SREF for 123 events with winds $\geq$ 10.29 m s^{-1} (20 kts). The same methodology is followed as in the old SREF verification, except that the RAP model analysis is used as the observational dataset because it replaced the RUC. Mean observed wind (Fig. 22a) and temperature (Fig. 22b) profiles show similar patterns as in the prior set of events, with the RAP exhibiting slightly higher winds in the low levels compared to ACARS, and then lower winds from 825 hPa up to 300 hPa. Verification plots show more similar results between the RAP and ACARS than the RUC and ACARS, lending some evidence to the hypothesis that the RUC had a systematic bias in the lower model levels, particularly at 1000 hPa. SREF biases between the RAP and ACARS during the first forecast period (Figs. 23a, b) reveal greater clustering of the different model cores, although as seen in the old SREF, the WRF-ARW core still retains the highest bias in the lower atmosphere. At around 950 hPa, it has a 2 m s^{-1} positive bias compared to the RUC (Fig. 23a), and at 975 hPa, it has a 2.5-3 m s^{-1} positive bias compared to ACARS (Fig. 23b). The WRF-ARW bias holds nearly steady or increases slightly with increasing lead time, but for the other two model cores, becomes slightly less positive by the third forecast period (Figs. 23c, d). Although the bias profiles between the RAP and ACARS have more similar shapes, the SREF bias compared to the RAP is still about 0.5-1 m s^{-1} lower in the LLJ region than compared to ACARS. The MAE plots once again show the WRF-ARW core with the highest error in the LLJ region, and the highest overall MAE in the upper levels (above 600 hPa), which increase with increasing lead time (Figs. 24a-d).

New SREF 10 m wind bias for the ensemble mean (Fig. 25) during the first forecast period (1-24 h) is fairly similar to that of the old SREF (Fig. 20a), with biases of -1 to 1 m s^{-1} in the immediate NYC metropolitan area and extending northward into the Lower Hudson Valley. The highest biases are again observed over eastern Long Island (up to ~3.5 m s-1) and the suburbs north and west of NYC (2.5-3 m s-1) (Fig. 25). The spatial pattern of small biases in and around NYC and higher biases to the east and west are consistent across all of the subgroups (ARW, NMB, NMM), with the NMB exhibiting the smallest bias area-wide (not shown)..

Mean temperature profiles between the RAP and ACARS are slightly more different than between the RUC and ACARS, with the RAP temperature warmer than ACARS in the mid-levels (500 to 800 hPa) (Fig. 22b). Temperature biases are slightly negative (0 to -1 K) in the low levels during the first forecast period when compared to both RAP (Fig. 26a) and ACARS (Fig. 26b), although over a more shallow layer (900 to 1000 hPa) than in the old SREF (850 to 1000 hPa). Biases remain fairly small with increasing lead time, becoming slightly more negative over a slightly deeper layer for the ensemble mean, WRF-ARW and WRF-NMM cores, but changing little for the WRF-NMB core (Figs. 26c, d). Overall, temperature biases in the new SREF are slightly smaller in magnitude, remaining less than -1 K against both RAP and ACARS.

Differences in wind biases between the three wind direction groupings reveals similar biases for W-N and N-SE flow, with a positive wind speed bias of 1-2 m s^{-1} in the lowest 100 hPa for the mean and three model subgroups in the first forecast period. Overall, the subgroups are more clustered together in W-N flow. A significantly more positive wind speed bias of 2-4 m s^{-1} between 925 and 975 hPa is observed in the W-SE regime, with the WRF-ARW observing the highest bias. Wind speed MAE in the low levels increases from 2-3 m s^{-1} in W-N flow to 2-3.5 m s^{-1} for N-SE flow to 2.5-4 m s^{-1} for W-SE during the first forecast period. Temperature biases

among the three regimes are similar to what was observed in the old SREF, except that the model subgroups are much more clustered. Biases of -0.5 to -1.5 K from 925 to 1000 hPa in W-N flow suggests the SREF is too stable in this regime. Biases are rather small but slightly positive for N-SE flow (~0.5 K) and slightly negative for W-SE flow (~-0.5 K). While the spread between model subgroups increases, the overall patterns change little with increasing lead time.

b. Probabilistic verification

Brier skill scores for 10-m wind at all 39 observation sites in the region of interest (Fig. 1) were calculated for the full 21-member SREF ensemble referenced to each of the five control members (ARW, Eta1, Eta2, NMM, RSM) for the old SREF and three control members (ARW, NMB, NMM) for the new SREF. The scores were calculated over 5 different wind speed thresholds, from 10 to 18 m s^{-1} in 2 m s^{-1} increments to see if the probabilistic skill changes with event strength. The Brier Skill Score (BSS) can range from -1 to 1, with values increasingly greater than zero representing increasingly greater skill that the ensemble has over the individual deterministic forecast, and values increasingly less than zero representing increasingly less skill that the ensemble has over the deterministic forecast. BSS plots for the old SREF (Figs. 27a,b) show that the ensemble has better skill over all of the control members for nearly all thresholds and at all forecast lead time periods. Overall, the BSS appears to change little from the first forecast period (Fig. 27a) to the third forecast period (Fig. 27b), with ensemble skill over the ARW control showing the most significant increase. Overall, the BSS referenced to the ARW and RSM controls increases slightly with increasing threshold, while it decreases for the two Eta controls. The WRF-NMM appears to be the most skillful control member within the SREF, with the BSS ranging generally from -0.1 to 0.1 (Figs. 27a, b), suggesting the WRF-NMM has nearly the same forecast skill as the full SREF ensemble. Larger error bars with increasing threshold

represent diminishing confidence due to a smaller sample size. BSS plots for the new SREF illustrate that the probabilistic skill of the new SREF is better than the old SREF (Figs. 28a, b), especially by the second to third forecast periods. The BSS of the full ensemble referenced to each of the three control members hold steady or increases with increasing lead time, and with increasing wind speed threshold. Once again, the ARW control member appears to be the least skilled, with the BSS of the ensemble increasing to around 0.8 by the third forecast period at the highest (18 m s^{-1}) threshold (Fig. 28b). Whereas the WRF-NMM was the most skillful control member in the old SREF, the new WRF-NMB control now has the greatest skill in the new SREF, but overall its skill is less than the NMM in the old SREF.

Although the both the old and new SREF ensemble has greater forecast skill over basically all of the deterministic control member forecasts, this was to be expected. In addition to calculating the BSS, reliability diagrams and rank histograms were constructed for 10 meter winds in order to get a better understanding of the probabilistic performance. Reliability is calculated for three 10-m wind speed thresholds (10, 14, and 18 m s^{-1}), which are roughly associated with Small Craft Advisory, Wind Advisory, and HWW criteria, respectively. Overall, the SREF exhibits poor calibration at all three thresholds and lead time periods. Reliability of the old SREF is fairly similar for the 10 and 14 m s^{-1} thresholds (Fig. 29a), with a forecast probability to observed relative frequency of about two to one, illustrating a significant overforecasting, or positive wind speed bias. Reliability at the 18 m s^{-1} threshold is more variable due to smaller samples size; although it is nearly perfect at 80% forecast probability, at 90% and 100% forecast probability (only 3 cases) for 10 m wind exceeding warning level threshold, there were no observed events. Little to no trend is observed with increasing lead time, although beyond 60% forecast probabilities, there is no sample for winds reaching or exceeding the 18 m

s^{-1} threshold (Fig. 29b).

Unfortunately, the reliability of the new SREF doesn't appear to be any better than that of the old version. At lower forecast probabilities, the reliability is actually slightly worse than the old SREF, with a forecast probability to observed relative frequency ratio of 2.5 to one (Fig. 30a). Once again, little change is observed with increasing lead time (Fig. 30b). Reliabilities at the three wind speed thresholds are fairly similar, with 18 m s^{-1} slightly worse than 10 and 14 m s^{-1}. The reliability for $\geq$ 18 m s^{-1} seems to improve at higher forecast probabilities, with an observed relative frequency of around 70% at 100% forecast probability during the first forecast period (Fig. 30a) and around 60% at 90% forecast probability during the third forecast period (Fig. 30b). Overall, the reliability diagrams indicate a strong overforecasting (positive) bias in 10 m winds for both the old and new SREF, which is supported by the positive bias in low level winds observed in the vertical profile verification plots. Rank histograms for both the old SREF (Fig. 31a) and new SREF (Fig. 31b) calculated over the first forecast period (1-24 h) confirm the poor SREF ensemble calibration illustrated by the reliability diagrams. The SREF appears to suffer from significant overforecasting, with many 10 m wind observations falling below the range of wind speeds forecasted by the ensemble members. Overall, the magnitude of the calibration and rank errors for 10 m winds seems unusually large given the wind speed biases in the low levels, which are generally in the 1 to 3 m s^{-1} range. Perhaps choosing such high wind speed thresholds allows for more observations to fall below the given thresholds, thus skewing the reliability diagrams and rank histograms. During the first forecast period, the FAR is quite high (0.63 for the old SREF and 0.61 for the new SREF), thus confirming the pattern illustrated in the rank histograms.

Chapter 5: Summary

a. Synthesis and Conclusions

Analysis of the distribution of observed NCWEs revealed temporal and spatial variations in high winds in the NYC metropolitan and surrounding area. Overall, NCWEs are most frequent during the three-month winter period from December through February, and maximum winds are typically observed in urban and coastal locations, especially at the three major airports in the metropolitan area (EWR, JFK, and LGA). The slightly more lenient event criteria used in this study compared to the NWS resulted in a higher sample of events, but also skewed the verification statistics. POF events were the most common of the three types (more than 50% of the total cases), with the majority reached by only sustained wind meeting the 35 kt warning criteria. POF events also had the lowest POD and CSI, while PRF events had the highest POD and CSI, while NEC events had the highest FAR.

Large scale synoptic evolutions of composite 500 hPa heights and anomalies and MSLP from t-48 to t+24 are intended to be used as pattern recognition or analog tools in the later short-term to medium range. They can be compared with real-time operational forecast models and provide advanced warning of possible NCWEs. By breaking the observed events into three types (PRF, POF, and NEC), the evolutions can better reveal features specific to the synoptic regimes that high winds occur within. Once within the 24 to 36 hour lead time frame when decisions need to be made regarding High Wind Watches and Warnings, the focus should shift to the regional and local composites. By this lead time, cyclone strength and position is usually well-handled, so comparing large-scale synoptic patterns will not provide much value. Regional height and/or pressure gradient and three hour MSLP tendency can be compared with the composites, depending on the track of the cyclone and the type of regime in which high winds are expected to

occur. If the magnitude of the height/pressure gradient and tendency are within the range of past observed NCWEs, the next step would be to analyze the local thermal and wind vertical profiles to determine the mixing potential. Although the LLJ strength by itself should not be used to issue a warning, since stability must also be taken into account, benchmark LLJ values can provide greater confidence of realizing a NCWE.

Additional factors should also be looked at if applicable, including the diurnal influence in the case of POF NCWEs, and heavy rainfall and/or gravity wave potential in the case of PRF and nor'easter events. Analysis of individual false alarm cases and some of the missed events reveals the full complexity and subjectivity in the forecast process for evaluating high wind potential. NWP model preference and forecast methodology changes based on the forecaster, so the verification statistics shown in this study are at least somewhat tainted by "human error." In addition, the classification of HWW level events in this study differs from that of the NWS in that a slightly looser time constraint is used, resulting in more observed warning level events and thus leading to lower verification scores. However, the climatology results still illustrate some important points. One is that the extremely low POD for POF events might be at least partially attributable to the diurnal influence. Another is that the high FAR for nor'easter cases might be attributable to both higher low-level stability and an isallobaric wind that more often than not opposes the mean low-level flow, thus hindering the realization of warning level winds at the surface, even when winds might exceed the warning criteria threshold by a significant margin just a few model levels above MSL.

Both the old and new SREF ensembles exhibit fairly a significant positive wind speed bias in the low levels (850 to 1000 hPa), even at short lead times, shedding some light on the difficulty in predicting NCWEs. 10m wind speed bias varies spatially across the area, with the

smallest biases observed around the NYC area and the highest biases observed over eastern Long Island. The SREF also exhibits a negative temperature bias in the low levels, although the magnitude of the errors is small enough that it likely does not have an impact on the downward momentum transport of high winds to the surface. Overall, the clustering of the model cores was greater in the new SREF, which was expected given the elimination of the Eta and RSM cores in the old SREF and the addition of another WRF-based core (WRF-NMB) in the new SREF. Another interesting result was the difference in model error when using the RUC/RAP versus the ACARS as the observational dataset. The differences are likely a result of a significant low wind speed bias observed at 10 m in the RUC. Verifying against the RAP seemed to increase the similarity to the ACARS error profiles, but not enough information was gathered to conclude which data source is more accurate, or if both sources have systematic biases.

As expected, BSS plots illustrate that both the old and new SREF ensemble provide a significant improvement in forecasting skill over nearly all of the deterministic control members. Overall, the WRF-ARW control has the worst skill, while the WRF-NMM control has the best skill in the old SREF, and the WRF-NMB has the best skill in the new SREF. This is backed up by the bias and MAE plots, which show that the WRF-ARW core has the highest wind speed error and the WRF-NMM and NMB have the smallest error in the low levels. Although no clear trend is observed in the old ensemble BSS with increasing lead time, the new ensemble BSS increases rather significantly with increasing lead time. However, this doesn't necessarily imply that the SREF should be weighted more heavily than other models when making a decision on whether or not to issue a HWW. Due to data availability and model lead time, the SREF was the best option for this study, but other models may be more useful in high wind event forecasting. Even though the SREF has greater forecast skill than any of the deterministic control members,

reliability diagrams and rank histograms illustrate that the ensemble itself is poorly calibrated, and has a significant overforecasting (positive) bias for 10 m winds.

b. Future Work

Several different options exist for potential future work on this topic. NCWE climatology could be expanded spatially to include a larger area of the Mid-Atlantic and Northeast U.S., including all of the major metropolitan areas from Baltimore and Washington, D.C. up to Boston. The climatology could also be expanded temporally back in time, perhaps several decades, to gather a larger sample size. Although the NARR data is available back to 1979, raw observational data would be more of a challenge to acquire. METAR data in downloadable format is only available back to 2000, but some websites of private weather companies (such as wunderground.com) do have at least hourly and special observation METARs available for some ASOS sites as far back as the late 1940s. The predictability of NCWEs can be further explored by applying post-processing to the SREF for the cases used in this study to see how well it improves the probabilistic forecast skill. Another option would be to evaluate other NWP models, particularly shorter range and higher resolution models, to see if they can better resolve the local spatial variation in high winds.

References

Ashley, W. S., and A. W. Black, 2008: Fatalities associated with nonconvective high wind events in the United States. *J. Appl. Meteor. Climatol.*, **47**, 717–725.

Asuma, J. V., 2010: Cool-season high wind events in the Northeast U.S. Thesis, Department of Atmospheric and Environmental Sciences, University at Albany, State University of New York, 132 pp.

Benjamin, S.G., and Coauthors, 2004: An hourly assimilation/forecast cycle: the RUC. *Mon. Weather. Rev.*, **132**: 495-518.

Bond, N. A., and Bernard A. Walter, 2002: Research aircraft observations of the mean and turbulent structure of a low-level jet accompanying a strong storm. *J. Appl. Meteor.*, **41**, 1210–1224.

Bosart, L. F., and F. Sanders, 1986: Mesoscale structure in the megalopolitan snowstorm of 11–12 February 1983. Part III: A large-amplitude gravity wave. *J. Atmos. Sci.*, **43**, 924–939.

Brasseur, O., 2001: Development and application of a physical approach to estimating wind gusts. *Mon. Wea. Rev.*, **129**, 5–25.

Brennan, M. J., G. M. Lackmann, and K. M. Mahoney, 2008: Potential vorticity (PV) thinking in operations: The utility of nonconservation. *Wea. Forecasting*, **23**, 168–182.

Brown, J.M., and Coauthors, 2011: Improvement and testing of WRF physics options for application to Rapid Refresh and High Resolution Rapid Refresh. Preprints, 14th Conf. on Mesoscale Processes/15th Conf. on Aviation, Range, and Aerospace Meteorology, Los Angeles, CA, Amer. Meteor. Soc., 5.5. [Available online at https://ams.confex.com/ams/14Meso15ARAM/webprogram/Paper191234.html.]

Browning, K. A., and R. Reynolds, 1994: Diagnostic study of a narrow cold-frontal rainband and severe winds associated with a stratospheric intrusion. *Q.J.R. Meteorol. Soc.*, **120**, 235–257.

Browning, K. A., 2004: The sting at the end of the tail: Damaging winds associated with extratropical cyclones. *Q.J.R. Meteorol. Soc.*, **130**, 375–399.

Cairns, M. M., and J. Corey, 2003: Mesoscale model simulations of high-wind events in the complex terrain of Western Nevada. *Wea. Forecasting*, **18**, 249–263.

Cohen, A. E., and J. P. Cangialosi, 2010: An observational and high-resolution model analysis of gale wind events in the Gulf of California. *Wea. Forecasting*, **25**, 613–626.

Crupi, K. M., 2004: An anomalous non-convective high wind episode over Upper Michigan. *National Weather Digest*, **28**, 3–12.

Du, J., G. Dimego, Z. Toth, D. Jovic, B. Zhou, J. Zhu, J. Wang, and H. Juang, 2009: Recent upgrade of NCEP short-range ensemble forecast (SREF) system. *Proc. 19th Conf. on Numerical Weather Prediction and 23rd Conf. on Weather Analysis and Forecasting*, Omaha, NE, Amer. Meteor. Soc., 4A.4. [Available online at https://ams.confex.com/ams/23WAF19NWP/techprogram/paper_153264.htm.]

Durkee, J. D., C. M. Fuhrmann, J. A. Knox, J. D. Frye, 2012: Ageostrophic contributions to a non-convective high wind event in the Great Lakes region. *National Weather Digest* **36**, pp. 26–41.

Ebert, E. E., 2009: Neighborhood verification: A strategy for rewarding close forecasts. *Wea. Forecasting*, **24**, 1498–1510.

Gilleland, E., D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert, 2009: Intercomparison of spatial forecast verification methods. *Wea. Forecasting*, **24**, 1416–1430.

Han, J., and H. Pan, 2011: Revision of Convection and Vertical Diffusion Schemes in the NCEP Global Forecast System. *Wea. Forecasting*, **26,** 520–533.

Hart, R. E., and G. S. Forbes, 1998: The use of model-generated hourly soundings to forecast mesoscale phenomena: Part II. Initial assessment in forecasting nonconvective strong wind gusts. *Wea. Forecasting*, **14**, 461 – 469.

Hong, S.-Y., and H.-L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Weather Rev.,* **124***,* 2322–2339.

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

Janjić, Z. I., 2001: Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP meso model. National Centers for Environmental Prediction Tech. Rep. 437, 61 pp.

Janjic, Z., T. Black, M. Pyle, H.-Y. Chuang, E. Rogers, and G. DiMego: High resolution applications of the WRF NMM, in: Joint Session 16, 1-21, 2005.

Janjic, Z. I. and T. Black, 2007: A unified model approach from meso to global scales, *Geophys. Res. Abstr.*, **9**, Sref–ID: 1607-7962/gra/EGU2007-A-05 025.

Kapela, A. F., P. W. Leftwich, and R. Van Ess, 1995: Forecasting the impacts of strong wintertime post-cold front winds in the northern Plains. *Wea. Forecasting*, **10**, 229-244.

Knox, J. A., 2004: Non-convective windstorms in the Midwest United States: Surface and satellite climatologies. *Extended Abstracts, 22nd Severe Local Storms Conf.,* Hyannis, MA, Amer. Meteor. Soc., P5.3.

Knox, J. A., J. D. Frye, J. D. Durkee, and C. M. Fuhrmann, 2011: Non-convective high winds associated with extratropical cyclones. *Geography Compass*, **5**, 63–89.

Kurtz, J. T., and J. Martinelli, 2010: A climatology of cold-season nonconvective wind events across the North Central Plains. Thesis, Department of Atmospheric Science, Creighton University, 108 pp.

Lacke, M. C., J. A. Knox, J. D. Frye, A. E. Stewart, J. D. Durkee, C. M. Fuhrmann, and S. M. Dillingham, 2007: A climatology of cold-season nonconvective wind events in the Great Lakes region. *J. Climate*, **20**, 6012–6022.

Lackmann, G. M., 2002: Cold-frontal potential vorticity maxima, the low-level jet, and moisture transport in extratropical cyclones. *Mon. Wea. Rev.*, **130**, 59–74.

Langland, R. H., M. A. Shapiro, and R. Gelaro, 2002: Initial condition sensitivity and error growth in forecasts of the 25 January 2000 East Coast snowstorm. *Mon. Wea. Rev.*, **130**, 957–974.

Lee, J. E., and M. Girodo, 1997: Use of the Eta model to determine the role of static stability in *forecasting surface winds. Preprints, 16th Conf. On Weather Analysis and Forecasting, Phoenix,* AZ, Amer. Meteor. Soc., 56-58.

Lippman, J., 2012: "An examination of physical and empirical approaches in forecasting nonconvective wind gusts". *Dissertations & Theses in Earth and Atmospheric Sciences.* Paper 35.

Lynch, A. H., E. N. Cassano, J. J. Cassano, and L. R. Lestak, 2003: Case studies of high wind events in Barrow, Alaska: Climatological context and development processes. *Mon. Wea. Rev.*, **131**, 719–732.

Mass, C., and B. Dotson, 2010: Major extratropical cyclones of the Northwest United States. Part I: historical review, climatology, and synoptic environment. *Mon. Wea. Rev.*, **138**, 2499–2527.

Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys.*, **20**, 851–875.

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360.

National Weather Service, 2012: NWS Performance Management. [Available online at https://verification.nws.noaa.gov/.]

National Weather Service, 2014: Technical Implementation Notice 12-30, Amended. [Available online at http://www.nws.noaa.gov/os/notification/tin12-30sref_aaa.htm]

Niziol, T. A., and T. J. Paone, 2000: A climatology of non-convective high wind events in western New York state. NOAA Tech. Memo. NWS ER-91, 36 pp.

Pinson, P., and R. Hagedorn, 2012: Verification of the ECMWF ensemble forecasts of wind speed against analyses and observations. *Met. Apps*., **19**: 484–500.

Rife, Daran L., and C. A. Davis, 2005: Verification of temporal variations in mesoscale numerical wind forecasts. *Mon. Wea. Rev.*, **133**, 3368–3381.

Rife, Daran L., C. A. Davis, and J. C. Knievel, 2009: Temporal changes in wind as objects for evaluating mesoscale numerical weather prediction. *Wea. Forecasting*, **24**, 1374–1389.

Schultz, J., and B. Meisner, 2009: The 24 February 2007 north Texas dust storm: an impact weather event. *National Weather Digest*, **33**, 165–184.

Steenburgh, W. J, and C. F. Mass, 1996: Interaction of an Intense Extratropical Cyclone with Coastal Orography. *Mon. Wea. Rev.*, **124**, 1329–1352.

Thorarinsdottir, T. L., and M. S. Johnson, 2012: Probabilistic wind gust forecasting using nonhomogeneous gaussian regression. *Mon. Wea. Rev.*, **140**, 889–897.

Van Den Broeke, M.S., D. M. Schultz, R. H. Johns, J. S. Evans and J. E. Hales, 2005: Cloud-to Ground Lightning Production in Strongly Forced, Low-Instability Convective Lines Associated with Wind Damage. *Wea. Forecasting*, **20**, 517-530.

Weygandt, S. S., and Coauthors, 2011: Evaluation of the National Center for Environmental Prediction (NCEP) implementation version of the Rapid Refresh and its skill in providing short-term guidance for aviation hazards. Preprints, 15th Conf. on Aviation, Range, and Aerospace Meteorology, Los Angeles, CA, Amer. Meteor. Soc., 5.4. [Available online at https://ams.confex.com/ams/14Meso15ARAM/webprogram/Paper191213.html.]

Wilks, D. S., 2006*: Statistical Methods in the Atmospheric Sciences, 2nd edn. Academic Press:* Burlington, MA*;* 627 pp.

Tables and Figures

Date	Event Type	Max Wind Site	Max Wind	Max Gust Site	Max Gust	Verification
12/12/00	POF	EWR	42	EWR	54	n/a
12/17/00	PRF	EWR	35	EWR	43	n/a
12/18/00	POF	JFK	40	JFK	47	n/a
03/22/01	NEC	LGA	36	LGA	42	n/a
03/22/02	POF	JFK	36	JFK	43	Miss
02/12/03	POF	JFK	35	JFK	40	Miss
09/19/03	NEC	LGA	35	LGA	44	Miss
10/15/03	POF	JFK	39	JFK	47	Miss
11/13/03	POF	JFK	43	LGA	54	Hit
11/29/03	POF	JFK	37	JFK	46	Miss
12/11/03	NEC	ISP	36	ISP	44	Miss
12/15/03	NEC	BDR	38	BDR	46	Miss
01/14/04	POF	JFK	35	JFK	40	Miss
11/05/04	POF	BDR	35	BDR	46	Miss
12/01/04	POF	JFK	42	JFK	48	Miss
01/23/05	NEC	HVN	35	GON	50	Miss
03/09/05	POF	JFK	41	JFK	50	Miss
04/03/05	PRF	LGA	38	LGA	50	Miss
10/25/05	NEC	LGA	35	LGA	42	Hit
01/15/06	NEC	HPN	40	HPN	54	Miss
01/18/06	PRF	LGA	36	HPN	50	Hit
01/18/06	POF	JFK	35	JFK	43	Miss
01/22/06	POF	LGA	35	LGA	45	Miss
02/17/06	POF	EWR	37	EWR	49	Hit
10/20/06	POF	LGA	36	LGA	44	Miss
10/28/06	PRF	GON	38	JFK	47	Hit
10/29/06	POF	EWR	38	EWR	47	Miss
12/01/06	PRF	JFK	36	JFK	44	Hit
03/06/07	POF	JFK	39	JFK	46	Miss
12/16/07	NEC	BDR	35	BDR	44	Miss
12/17/07	POF	LGA	35	LGA	47	Miss
01/30/08	POF	LGA	38	JFK	46	Miss
02/02/08	PRF	ISP	34	ISP	51	Miss
02/10/08	POF	JFK	38	JFK	50	Miss
03/05/08	PRF	JFK	43	JFK	53	Miss
03/09/08	POF	JFK	38	JFK	45	Hit
02/12/09	POF	EWR	43	EWR	52	Hit
10/07/09	POF	JFK	37	JFK	49	Miss
12/29/09	POF	JFK	36	ISP	49	Miss
01/03/10	POF	HPN	35	HPN	46	Miss
01/25/10	PRF	GON	37	HPN	54	Hit

01/29/10	POF	JFK	35	JFK	43	Miss
03/14/10	NEC	JFK	45	JFK	65	Hit
05/08/10	POF	EWR	37	EWR	48	Miss
12/01/10	PRF	LGA	36	TEB	51	Miss
12/27/10	NEC	JFK	42	HPN	58	Hit
12/27/10	POF	LGA	38	LGA	49	Miss
02/14/11	POF	JFK	36	JFK	46	Miss
02/19/11	POF	JFK	42	JFK	52	Miss
02/25/11	PRF	JFK	36	JFK	44	Miss
04/17/11	PRF	ISP	35	ISP	48	Miss
12/08/11	NEC	FOK	40	FOK	60	Miss
12/28/11	PRF	GON	40	HVN	55	Miss
01/13/12	POF	EWR	36	EWR	45	Miss
04/23/12	NEC	LGA	40	LGA	47	Miss
10/29/12	NEC	LGA	56	ISP	78	Hit
12/21/12	PRF	LGA	37	LGA	53	Hit
01/31/13	PRF	BDR	43	LGA	54	Hit
01/31/13	POF	EWR	39	JFK	51	Miss
02/09/13	NEC	GON	34	GON	52	Hit

Table 1. List of 61 observed events from the 2000-2001 through 2012-2013 cool seasons. The ASOS sites of maximum sustained wind and gust (in kts) are shown.

Date	Event Type	Max Wind Site	Max Wind	Max Gust Site	Max Gust
02/10/01	POF	JFK	32	LGA	41
10/16/02	NEC	LGA	31	FOK	40
04/16/07	NEC	GON	34	MMK	46
11/03/07	NEC	GON	28	GON	36
01/01/09	POF	JFK	34	JFK	41
03/02/09	NEC	GON	28	FOK	39
10/01/09	PRF	FRG	30	LGA	44
02/26/11	POF	JFK	31	EWR	44
10/30/11	NEC	FOK	31	FOK	43
02/25/12	POF	JFK	34	GON	44
11/08/12	NEC	GON	31	FOK	42
12/27/12	NEC	LGA	34	GON	44

Table 2. List of 12 false alarm cases from 2001 through the 2012-2013 cool seasons. The ASOS sites of maximum sustained wind and gust (in kts) are shown.

Event Type	Sustained Only	Gust Only	Both
All (61)	38 (62%)	2 (3%)	21 (34%)
PRF (14)	5 (36%)	1 (7%)	8 (57%)
POF (33)	26 (79%)	0 (0%)	7 (21%)
NEC (14)	7 (50%)	1 (7%)	6 (43%)

Table 3. Observed events reached by sustained wind only (≥35 kt maximum wind speed but <50 kt maximum wind gust), gust only (<35 kt maximum wind speed but ≥50 kt maximum wind gust), and both sustained wind and gust (≥35 kts maximum wind speed and ≥50 kt maximum wind gust).

Event Type	Hits	Misses	False Alarms	FAR	POD	CSI
PRF	6	7	1	0.14	0.46	0.43
POF	4	26	4	0.5	0.13	0.12
NEC	5	8	7	0.58	0.38	0.25

Table 4. List of events by three synoptic types (PRF, POF, NEC), and their verification type for the 2001-2002 through 2012-2013 cool seasons. Also shown is the FAR, POD, and CSI for each event type.

Time period	Hits	Misses	False Alarms	FAR	POD	CSI
2001-2005	1	13	1	0.5	0.07	0.07
2005-2009	7	12	4	0.36	0.37	0.30
2009-2013	7	16	6	0.46	0.30	0.24

Table 5. NCWE verification type and FAR, POD, and CSI over three four-season time periods.

Site	All Events		PRF		POF		NEC	
	Wind	Gust	Wind	Gust	Wind	Gust	Wind	Gust
BDR	4	3	1	0	1	1	2	2
EWR	8	7	1	1	7	6	0	0
FOK	1	1	0	0	0	0	1	1
GON	4	2	3	0	0	0	1	2
HPN	2	5	0	2	1	1	1	2
HVN	1	1	0	1	0	0	1	0
ISP	3	5	2	2	0	1	1	2
JFK	23	23	3	4	18	18	2	1
LGA	15	13	4	3	6	6	5	4
TEB	0	1	0	1	0	0	0	0

Table 6. Number of observed events for each ASOS site in which the maximum wind and/or gust was observed, broken down by all events and the three types (PRF, POF, NEC). Highlighted rows indicate the site(s) experiencing the highest frequency of maximum wind/gust.

		Time	Wind Direction	Wind Speed	Wind Gust	Rainfall Intensity	MSLP
(a)	12/11/03 ISP	1610	110	22	31	RA	29.21
		1615	110	20	31	+RA	29.20
		1620	120	23	35	+RA	29.20
		1625	110	36	44	+RA	29.19
		1630	130	20	44	+RA	29.19
		1635	120	22	36	+RA	29.18
		1640	130	23	30	+RA	29.18
		1645	130	24	30	+RA	29.18
		1650	130	21	29	-RA	29.17
(b)	2/02/08 ISP	0100	140	23	30	RA	29.67
		0105	150	24	35	RA	29.66
		0110	150	20	35	+RA	29.68
		0115	170	31	51	+RA	29.69
		0120	190	28	51	-RA	29.70
		0125	190	22	37	-RA	29.71
		0130	190	20	30	-RA	29.71
(c)	3/5/08 JFK	0910	140	29	35	-RA	29.46
		0915	150	34	44	-RA	29.47
		0920	160	35	46	-RA	29.49
		0925	170	38	53	+RA	29.50
		0930	170	32	53	-RA	29.50
		0935	170	36	46	-RA	29.51
		0940	180	32	46	-RA	29.51
		0945	170	34	43	-RA	29.52
		0950	170	31	39	-RA	29.52
(d)	12/8/11 FOK	0725	310	30	42		29.38
		0730	300	40	56	-RA	29.40
		0735	310	33	56	+RA	29.42
		0740	300	35	46	+RA	29.44
		0745	300	31	48	RA	29.46
		0750	300	38	60	RA	29.48
		0755	300	30	60	+RA	29.50
		0800	310	28	48	+RA	29.50
		0805	310	31	48	+RA	29.49
		0810	320	26	45	+RA	29.50
		0815	320	30	44	RA	29.49
		0820	330	27	40	RA	29.49
		0825	330	20	41	-RA	29.49

Table 7. 5-Minute METAR data, including time (UTC), wind direction (degrees), wind speed and gust (kts), rainfall intensity, and MSLP (in. Hg), for four missed events induced by heavy rain: (a) 12/11/03, (b) 2/2/08, (c) 3/5/08, and (d) 12/08/11. Rows highlighted in yellow represent observations meeting warning criteria.

Event Date	Event Type	Mixing height (hPa)	Highest wind at or below mixing height (kts)	Level of maximum LLJ strength (hPa)	Maximum LLJ strength (kts)	Maximum pressure tendency (hPa/ 3 hr) and direction of isallobaric wind	Maximum 10 m wind and gust (kts)
10 Feb 2001	POF	875	44	850	53	+6 NW-SE	32 G 41
16 Oct 2002	NEC	950	52	900	57	-8 NW-SE to SW-NE	31 G 40
16 Apr 2007	NEC	Inversion	n/a	900	66	-7 n/a	34 G 46
3 Nov 2007	NEC	950	47	925	54	-6 SW-NE	28 G 36
31 Dec 2008	POF	875	43	850	47	+8 W-E	34 G 41
2 Mar 2009	NEC	950	36	925	42	-3 N-S to SW-NE	28 G 39
1 Oct 2009	PRF	Isothermal	n/a	900	56	-3 n/a	30 G 44
25 Feb 2011	POF	900	50	875	55	+12 W-E	31 G 44
30 Oct 2011	NEC	950	51	900	55	-6 WSW-ENE	31 G 43
25 Feb 2012	POF	850	40	775	50	+4 SSW-NNE	34 G 44
8 Nov 2012	NEC	925	52	875	66	-3 n/a	31 G 42
27 Dec 2012	NEC	950	47	900	62	-10 NNE-SSW to SW-NE	34 G 44

Table 8. Synoptic analysis of 12 false alarm cases. Mixing height refers to the level at which the temperature decreases with height up to, and/or the level at which theta is steady up to. If temperature is steady or increasing with height, n/a is used for the highest wind within the mixed layer. The direction of isallobaric wind is inferred from the MSLP tendency plots. In some cases, the direction of the isallobaric wind shifts during the event. If there is little to no MSLP tendency gradient over the area, the isallobaric wind has little to no contribution, and is represented by n/a.

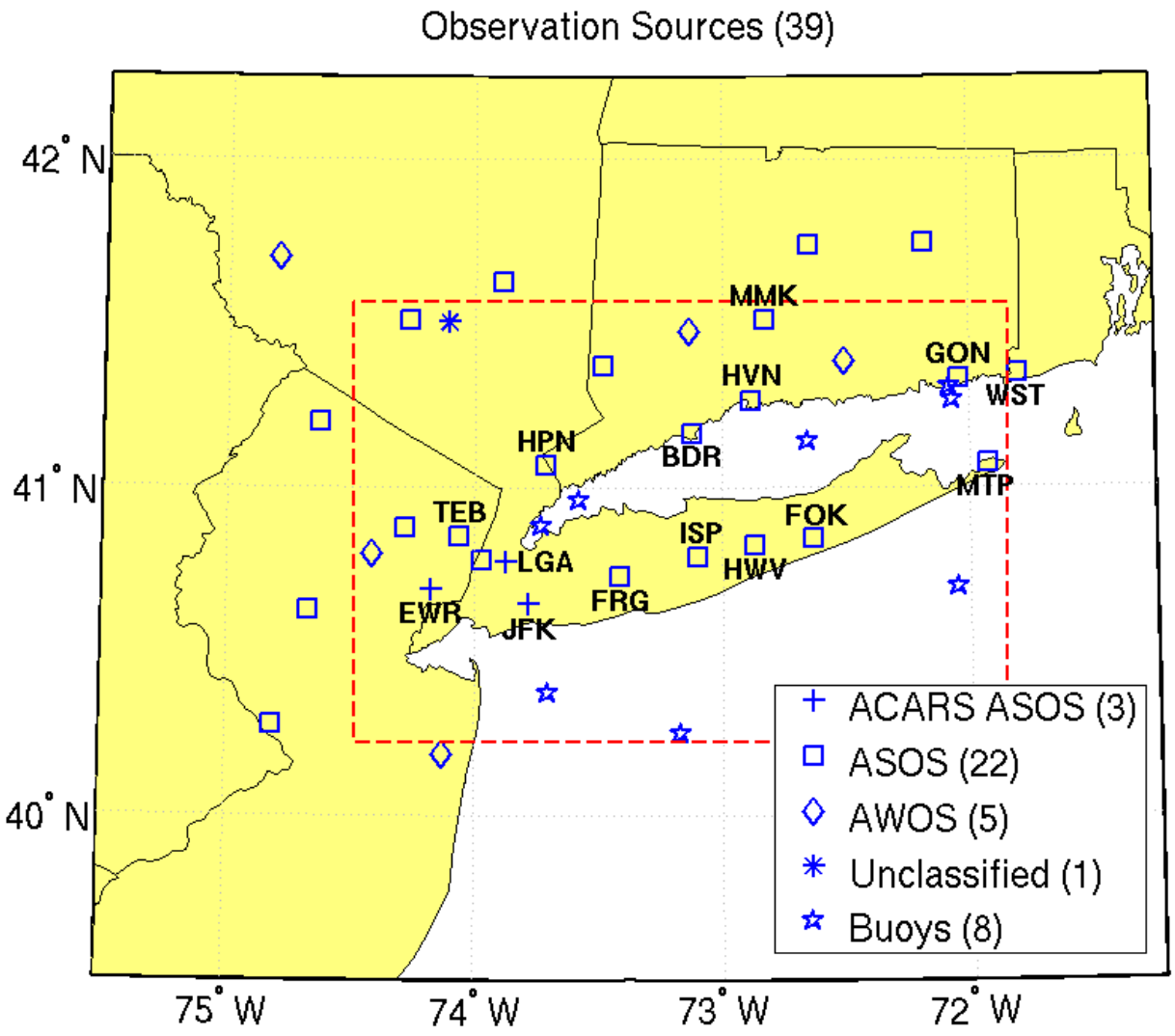

Fig. 1. Regional view showing all 39 observation sites. 25 ASOS sites were used to classify the HWW level observed events and false alarms. All 39 sites were used in probabilistic verification of 10 meter winds. The dashed red box includes all sites within the NYC NWS region.

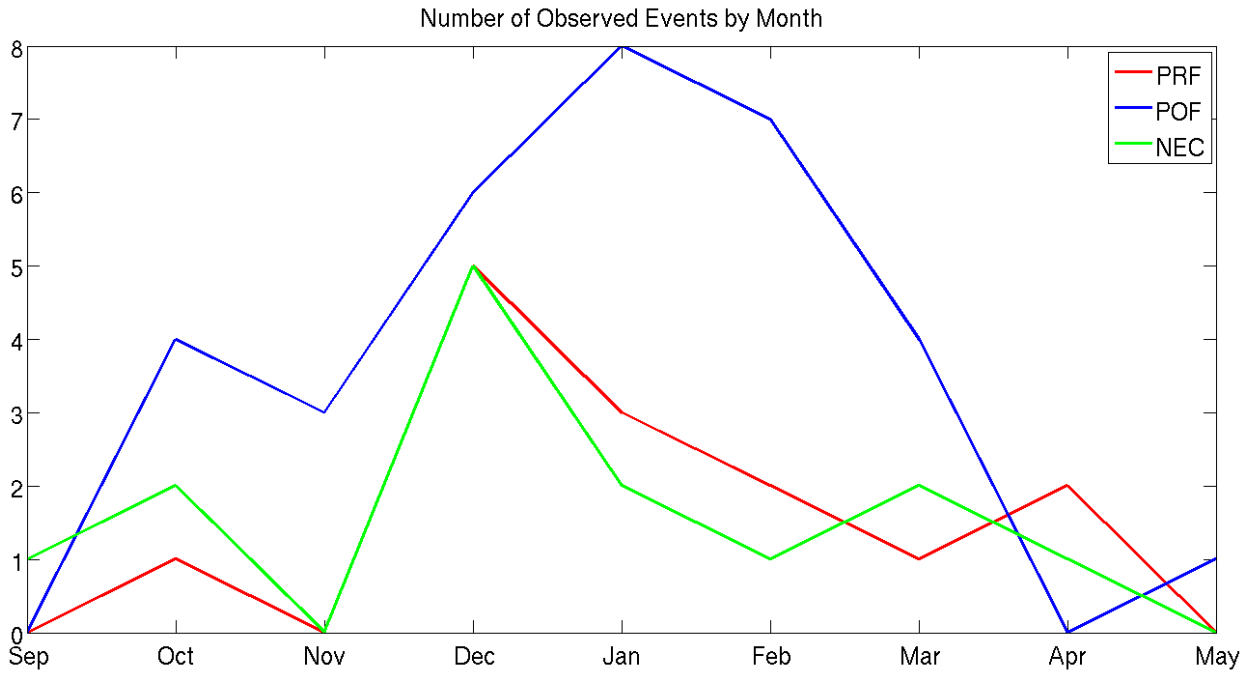

Fig. 2. Number of observed events for each type by month, summed over the 2000-2001 through 2012-2013 cool seasons.

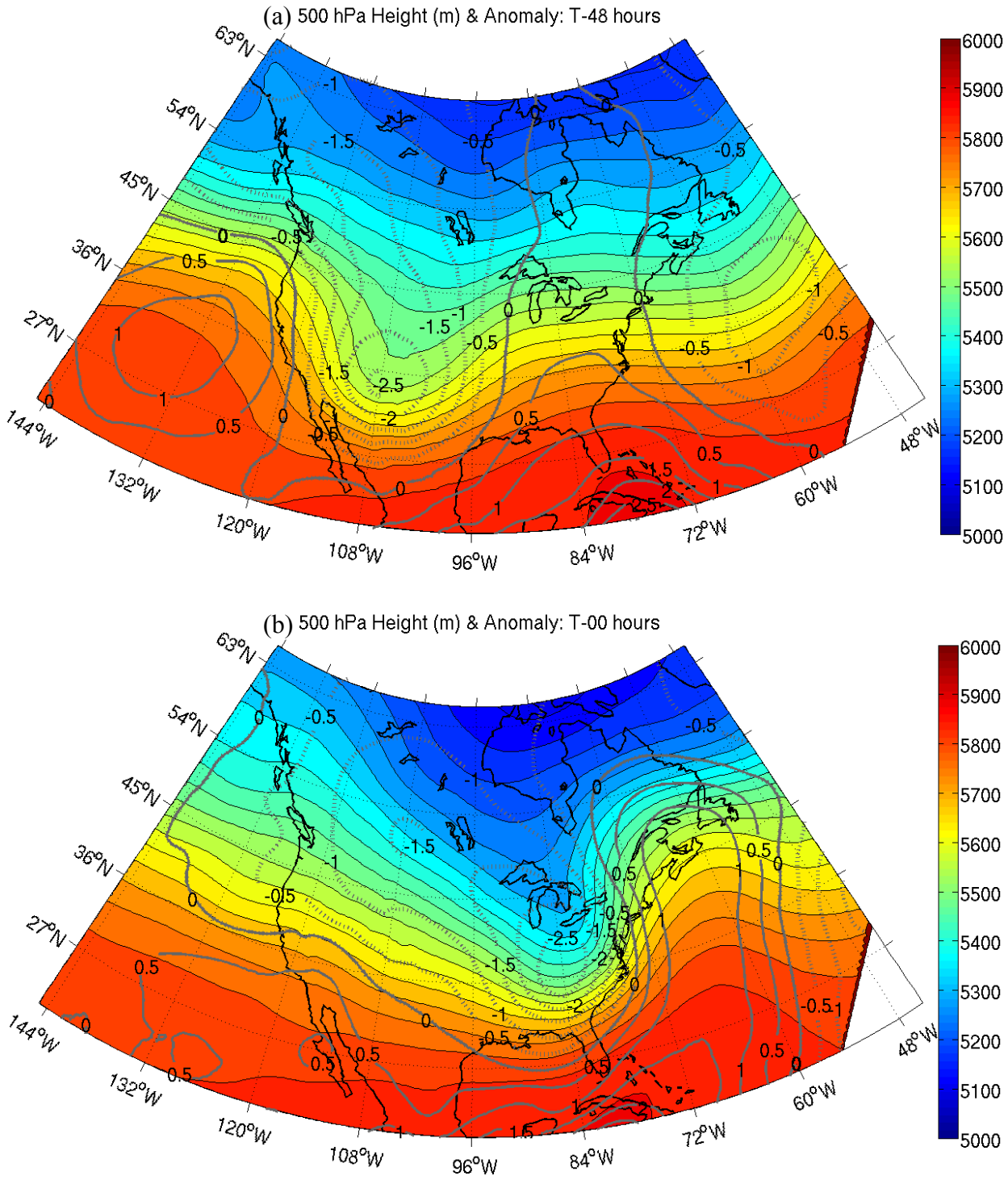

Fig. 3. 500 hPa height (in m) composite and standardized anomalies referenced to climatology for (14) observed PRF NCWEs, at (a) t-48 and (b) t-0 hours from the start of the event.

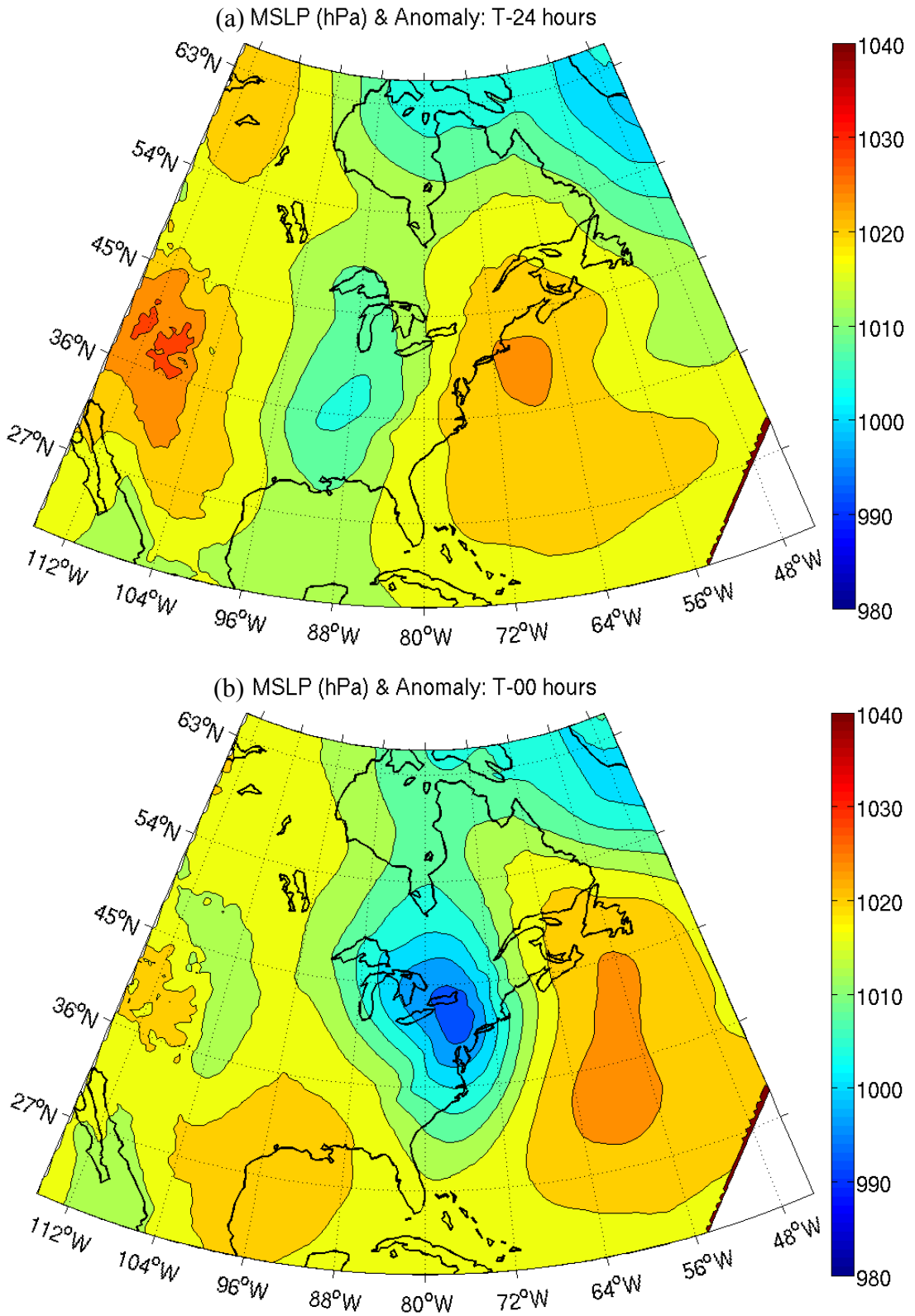

Fig. 4. MSLP (in hPa) composite for (14) observed PRF NCWEs, at (a) t-24 and (b) t-0 hours from the start of the event.

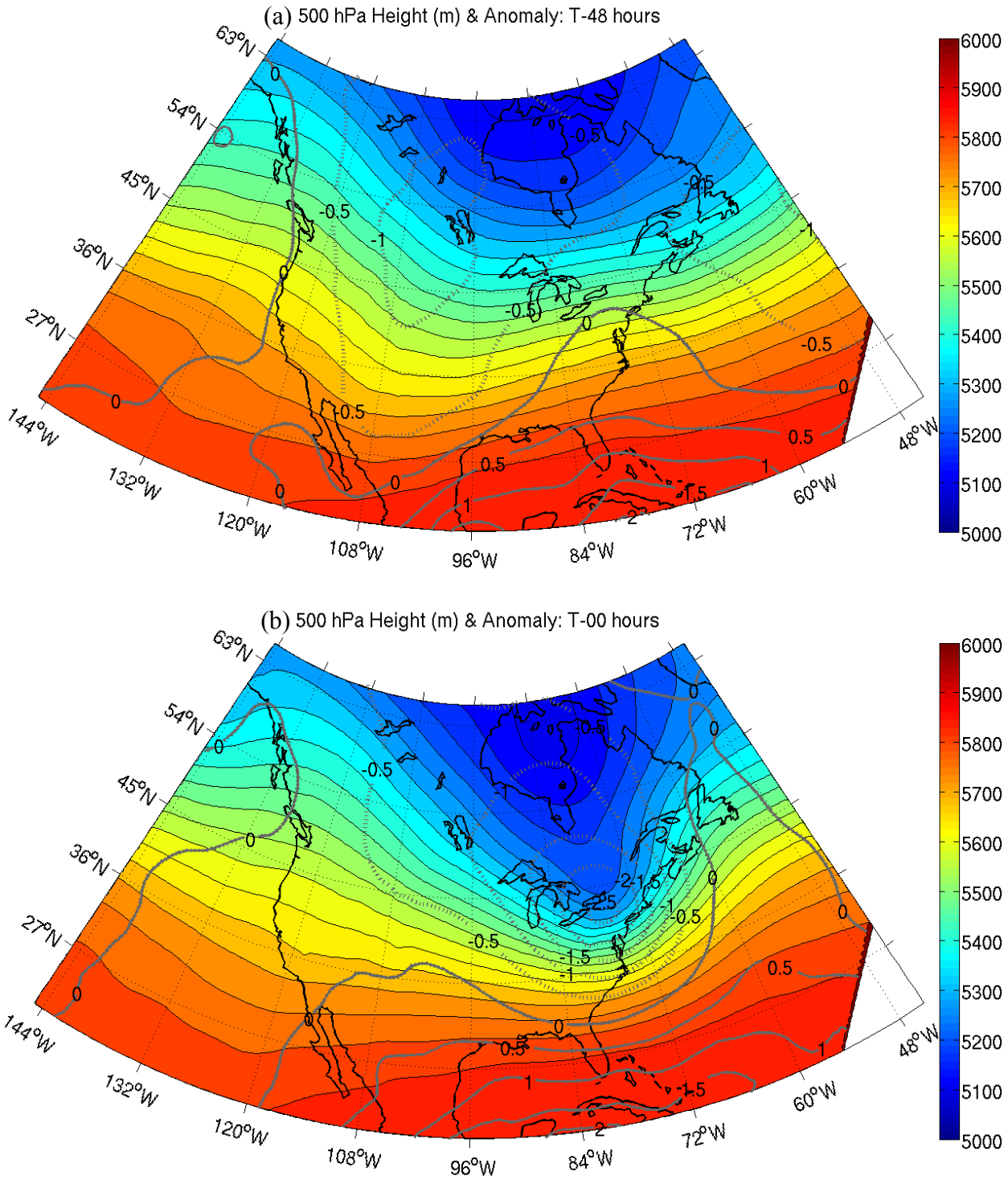

Fig. 5. 500 hPa height (in m) composite and standardized anomalies referenced to climatology for (33) observed POF NCWEs, at (a) t-48 and (b) t-0 hours from the start of the event.

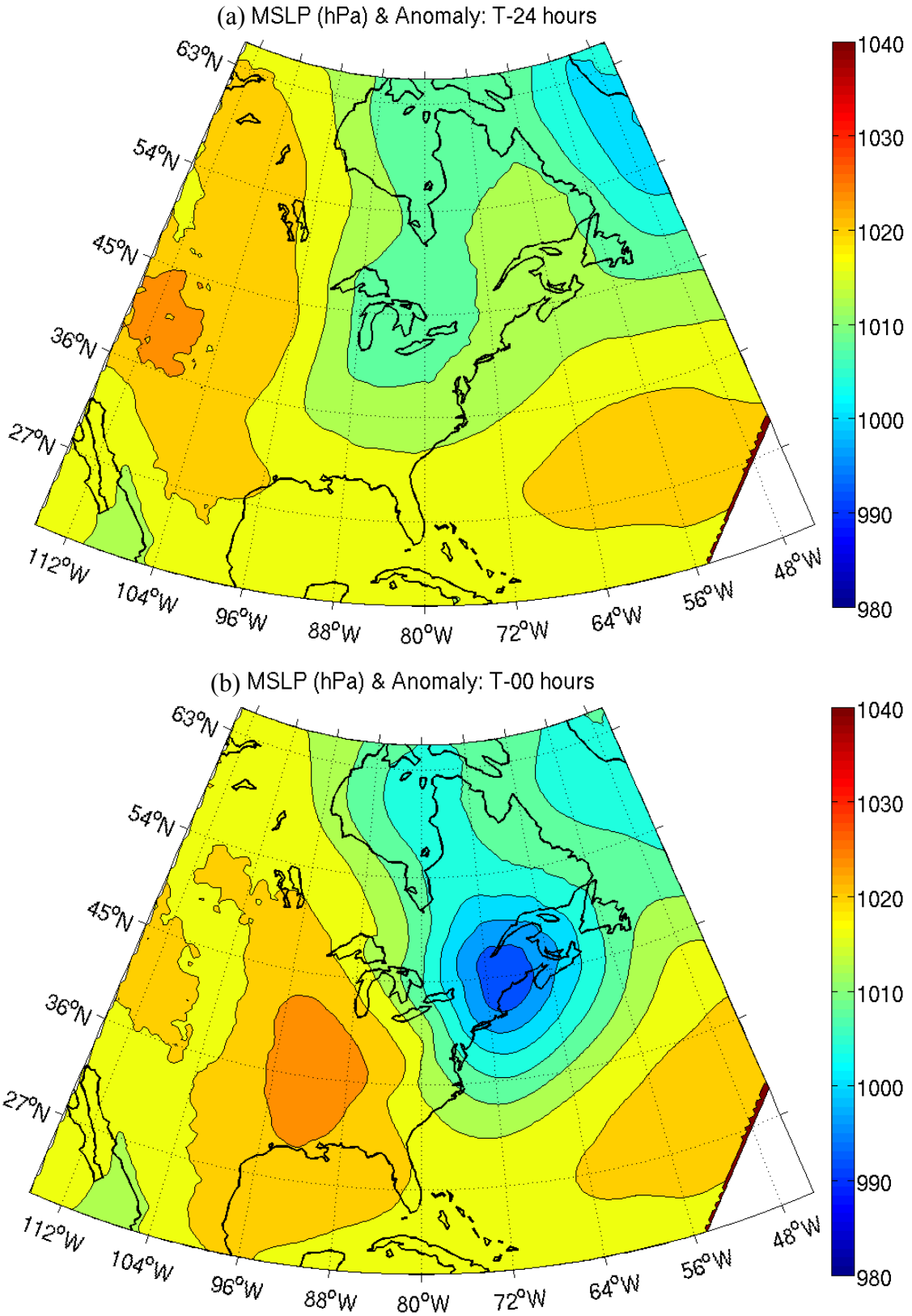

Fig. 6. MSLP (in hPa) composite for (33) observed POF NCWEs, at (a) t-24 and (b) t-0 hours from the start of the event.

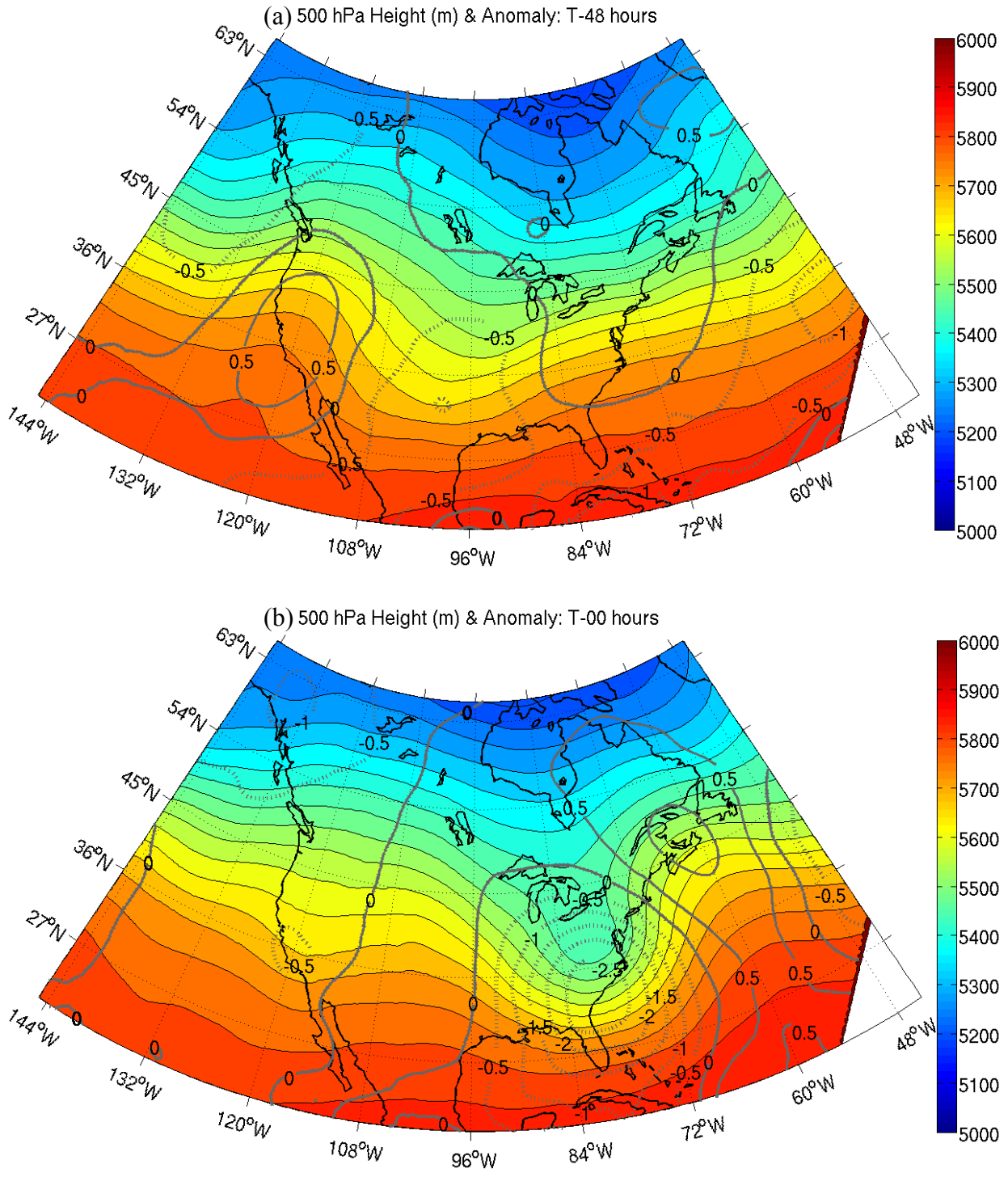

Fig. 7. 500 hPa height (in m) composite and standardized anomalies referenced to climatology for (14) observed NEC NCWEs, at (a) t-48 and (b) t-0 hours from the start of the event.

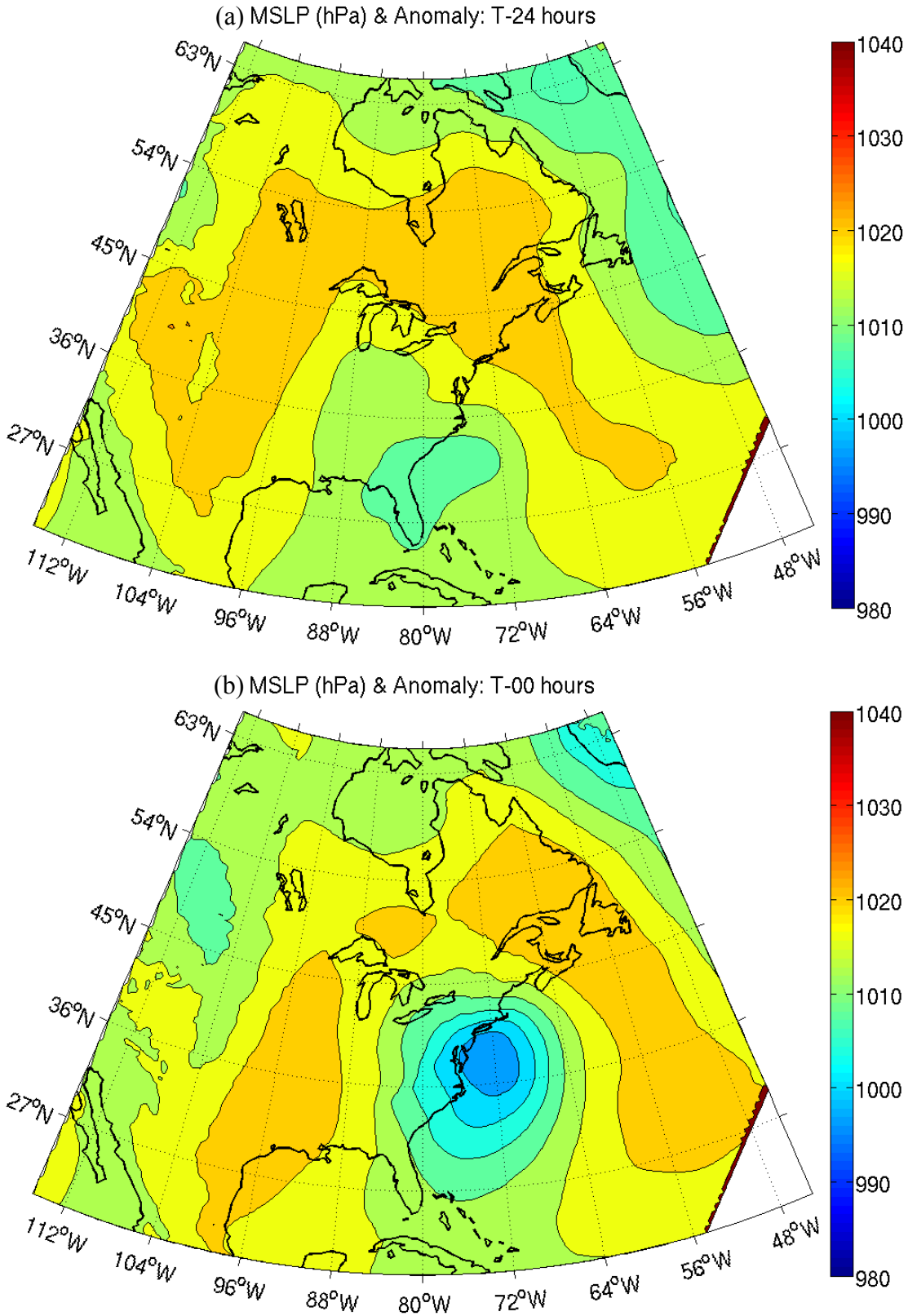

Fig. 8. MSLP (in hPa) composite for (14) observed NEC NCWEs, at (a) t-24 and (b) t-0 hours from the start of the event.

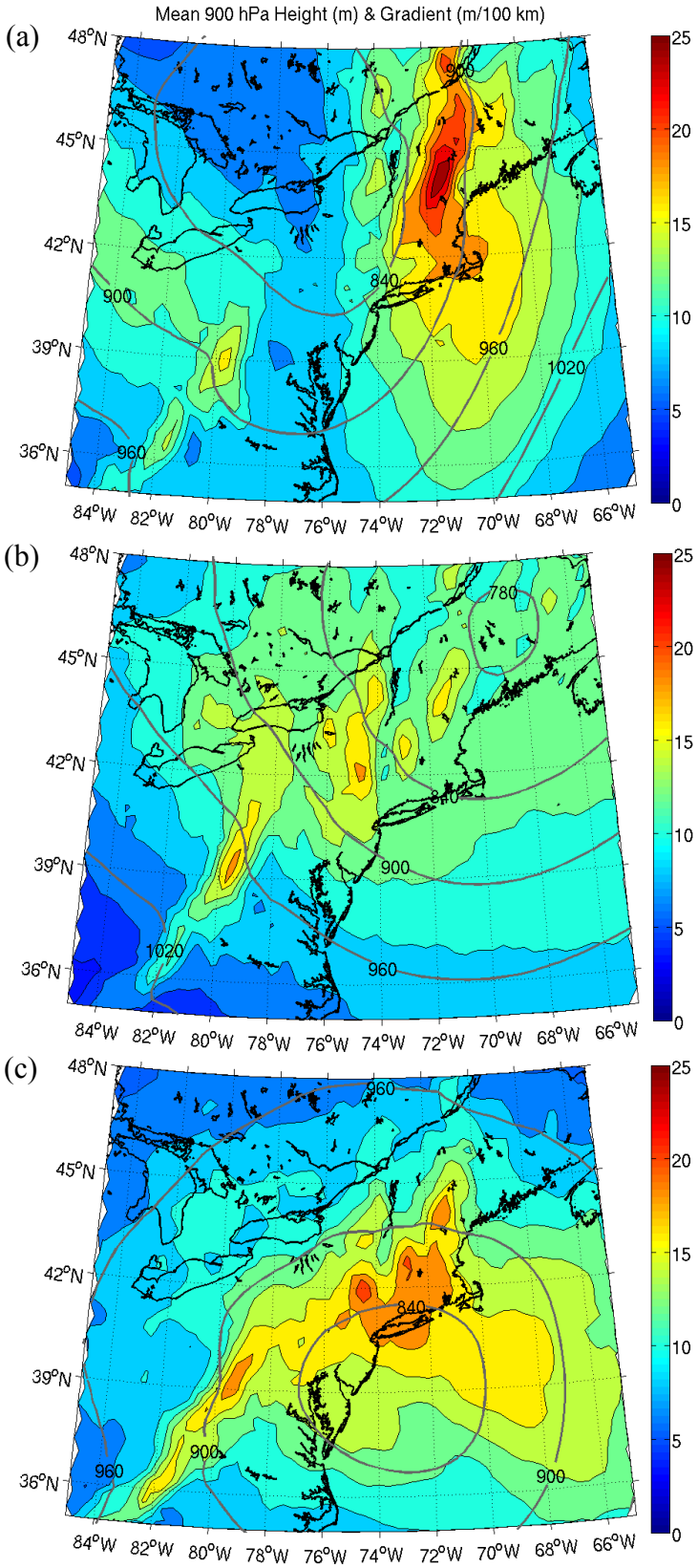

Fig. 9. Mean 900 hPa height (in m) and gradient (in m/100 km) composite of all observed (a) PRF, (b) POF, and (c) NEC NCWEs.

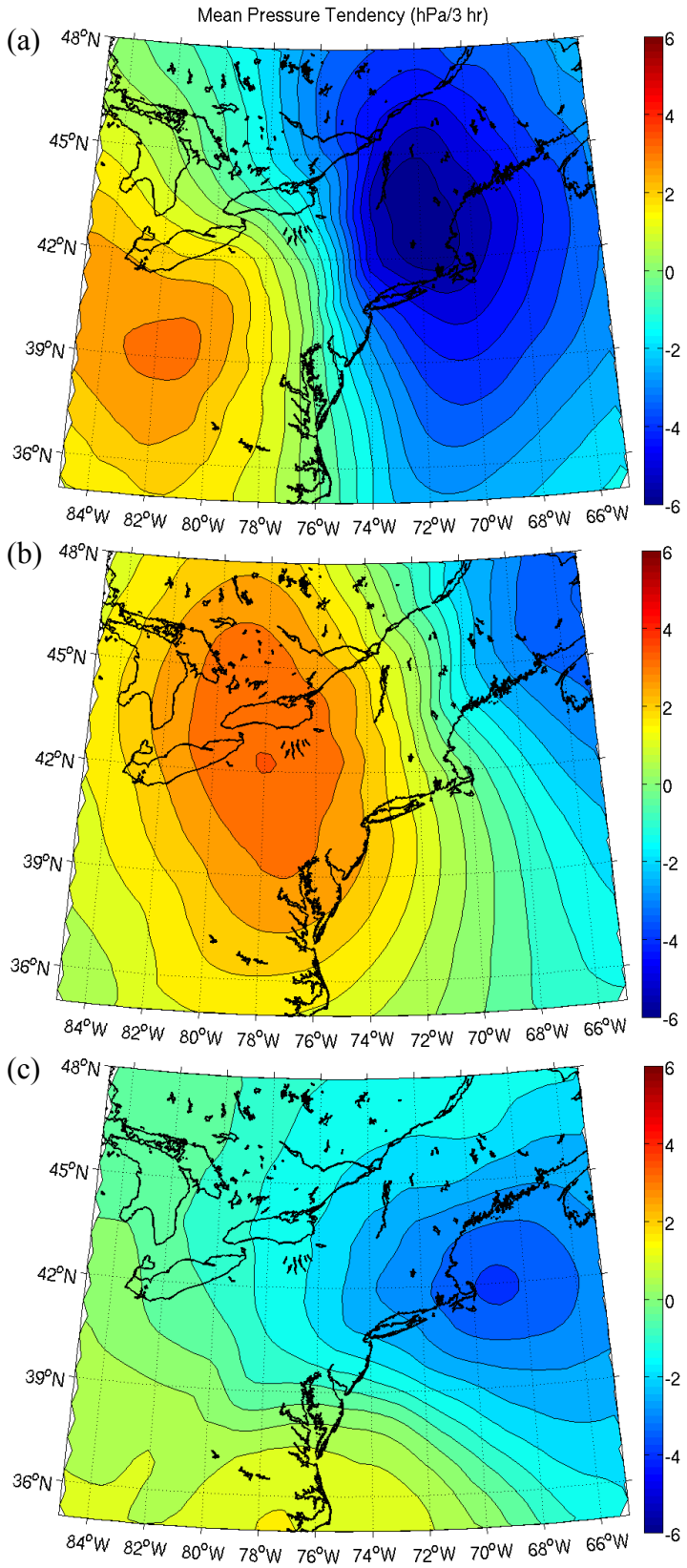

Fig. 10. MSLP tendency (in hPa/3 h) composite of all observed (a) PRF, (b) POF, and (c) NEC NCWEs.

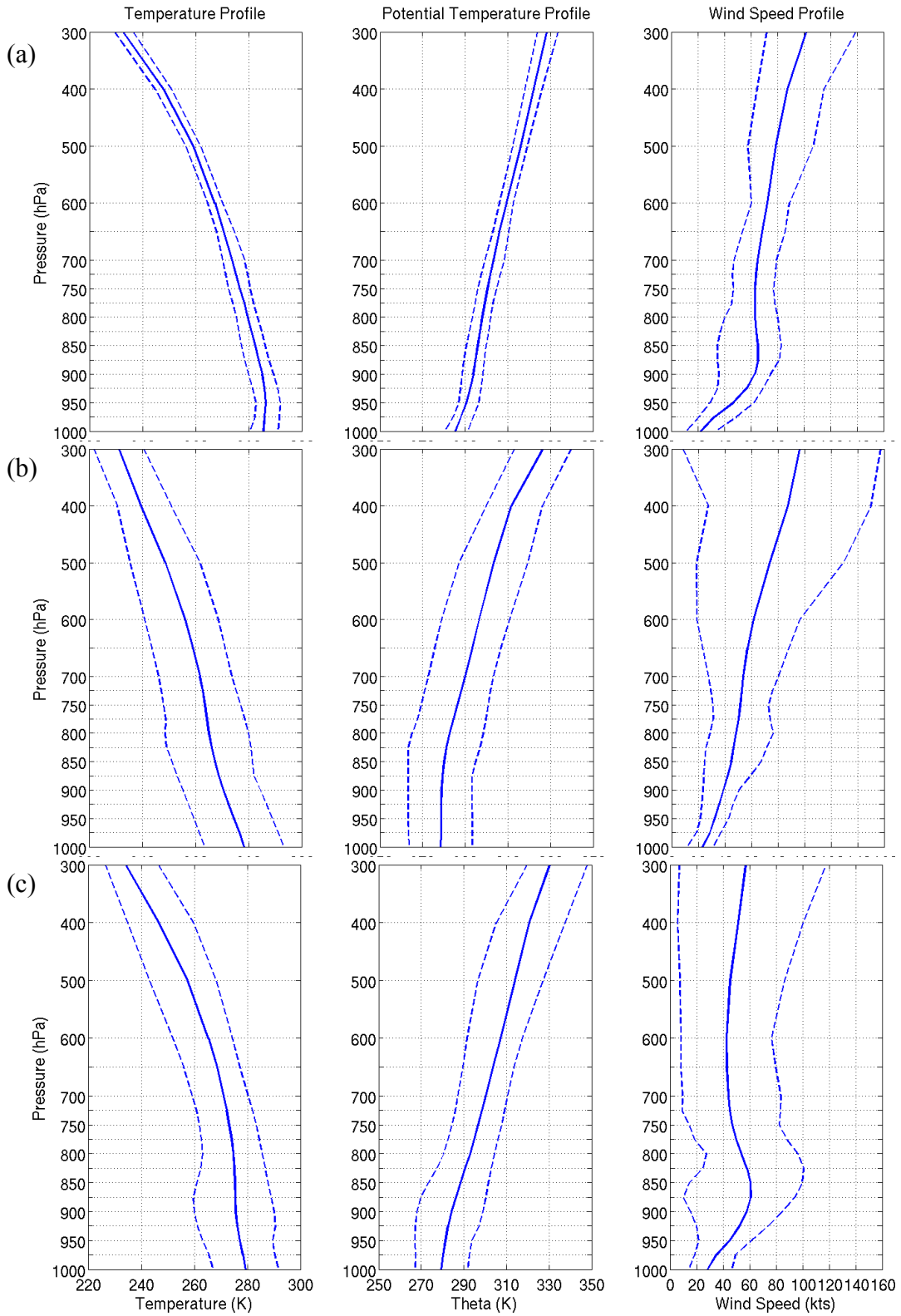

Fig. 11. Vertical profiles of temperature and potential temperature (in K), and wind speed (in kts), compositing all hours of observed HWW criteria for all observed (a) PRF, (b) POF, and (c) NEC NCWEs. Profiles are interpolated to the site of maximum wind and gust for each event.

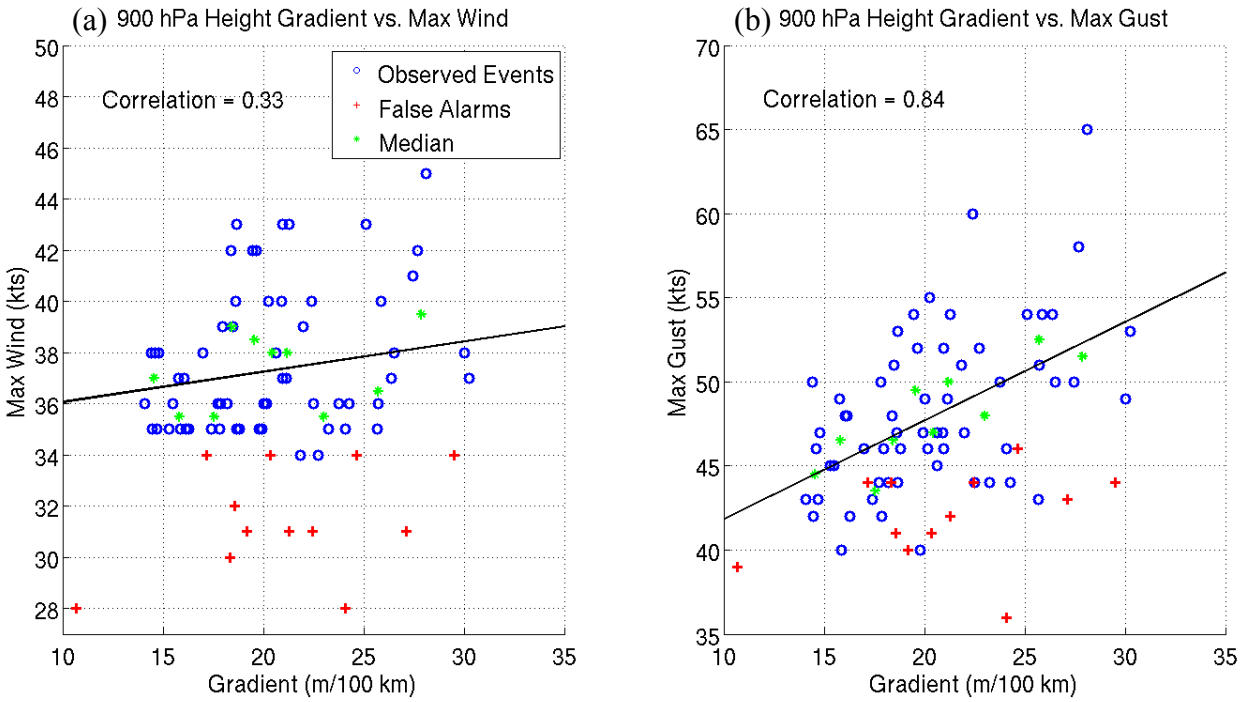

Fig. 12. Maximum 900 hPa height gradient (in m/100 km) observed over the region versus maximum (a) sustained wind speed and (b) wind gust (in kts) observed at any of the 25 ASOS stations during each event. Blue circles represent observed events, red '+' symbols represent false alarms, and green asterisks represent the median value of 6 observed data points. The black line is a linear correlation between the two variables, calculated from the median values.

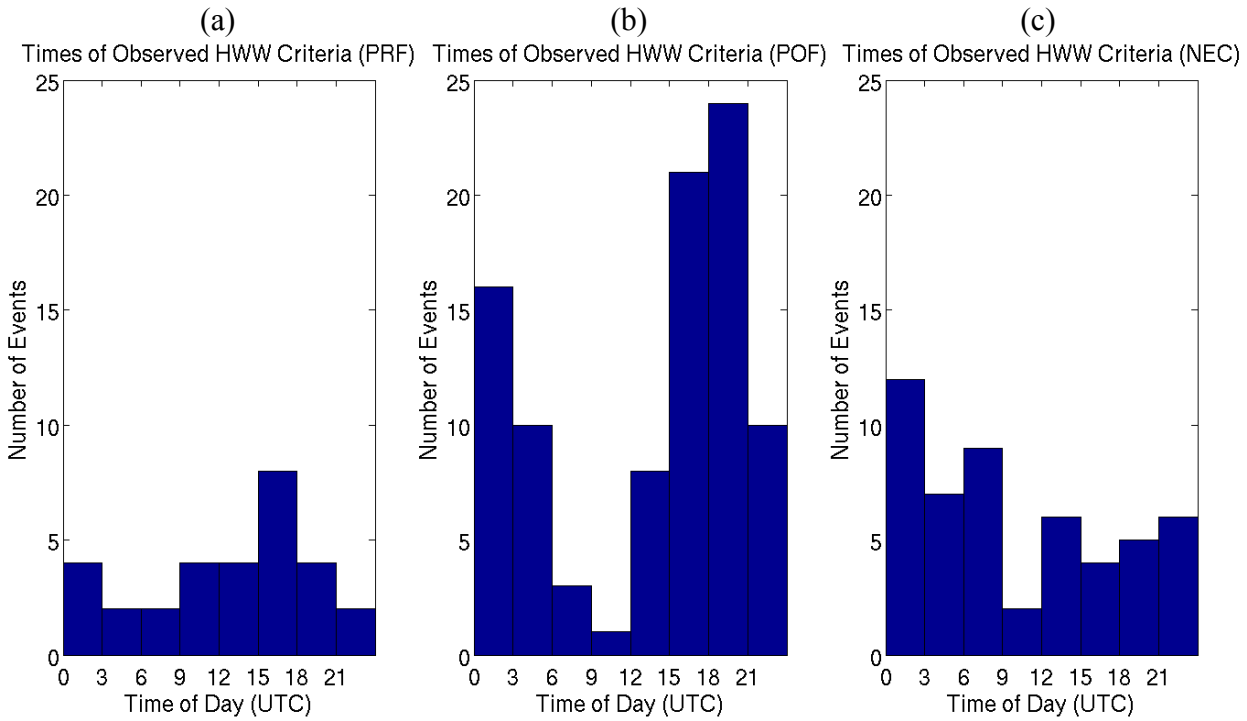

Fig. 13. Histograms of times of observed HWW criteria based on METAR observations for (a) PRF, (b) POF, and (c) NEC NCWEs. Bins are calculated in three-hour increments.

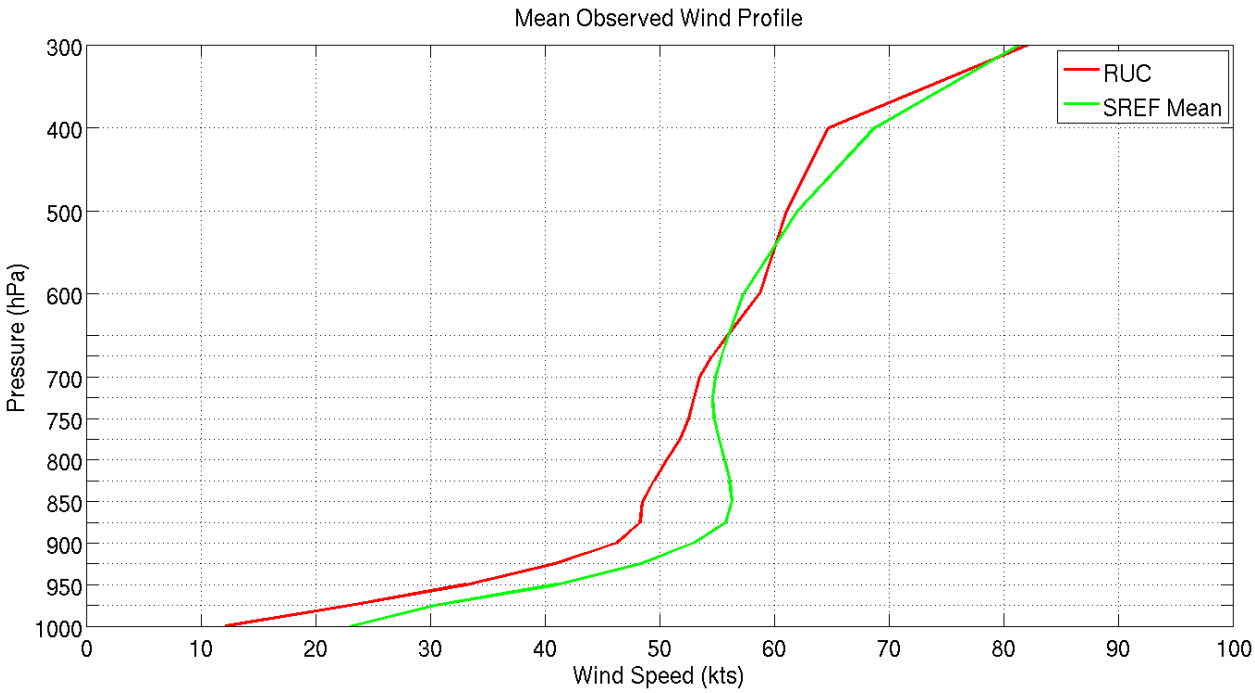

Fig. 14. Mean wind profiles (in kts) of the RUC and SREF ensemble mean for the 25 February 2011 POF false alarm case.

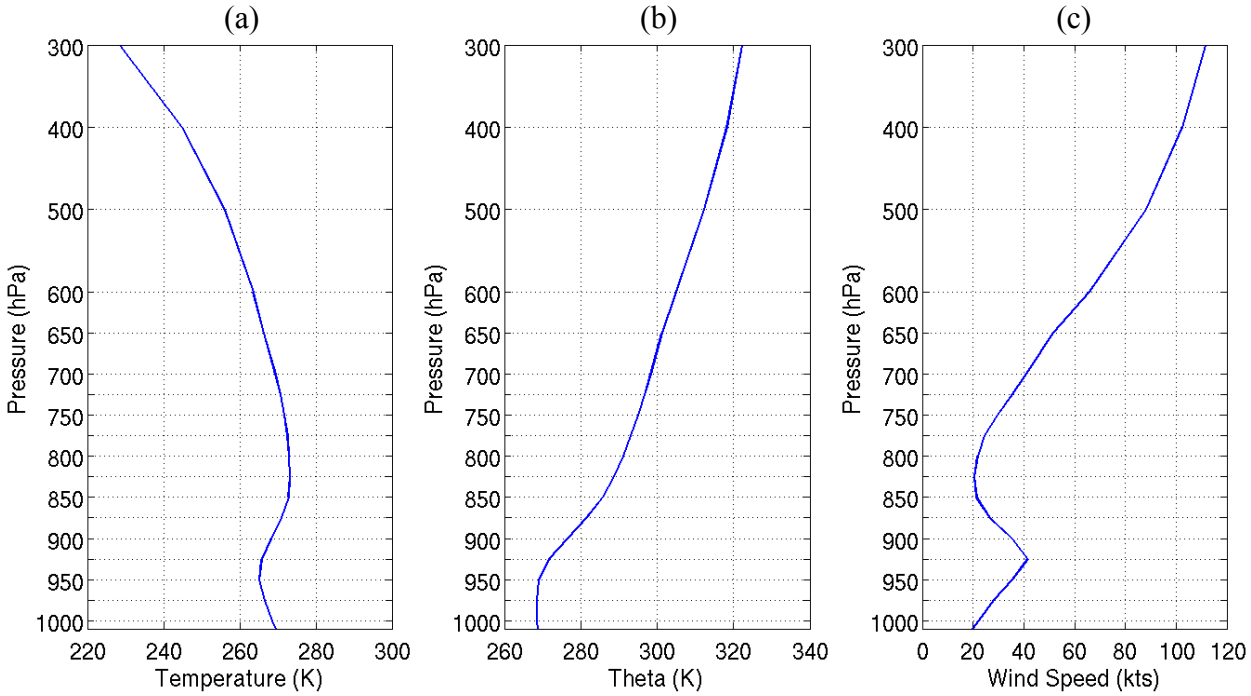

Fig. 15. Vertical profiles of (a) temperature (in K), (b) theta (in K), and (c) wind speed (in kts), interpolated at the site (GON) and time (1200 UTC) of maximum wind during the 2 March 2009 NEC false alarm case.

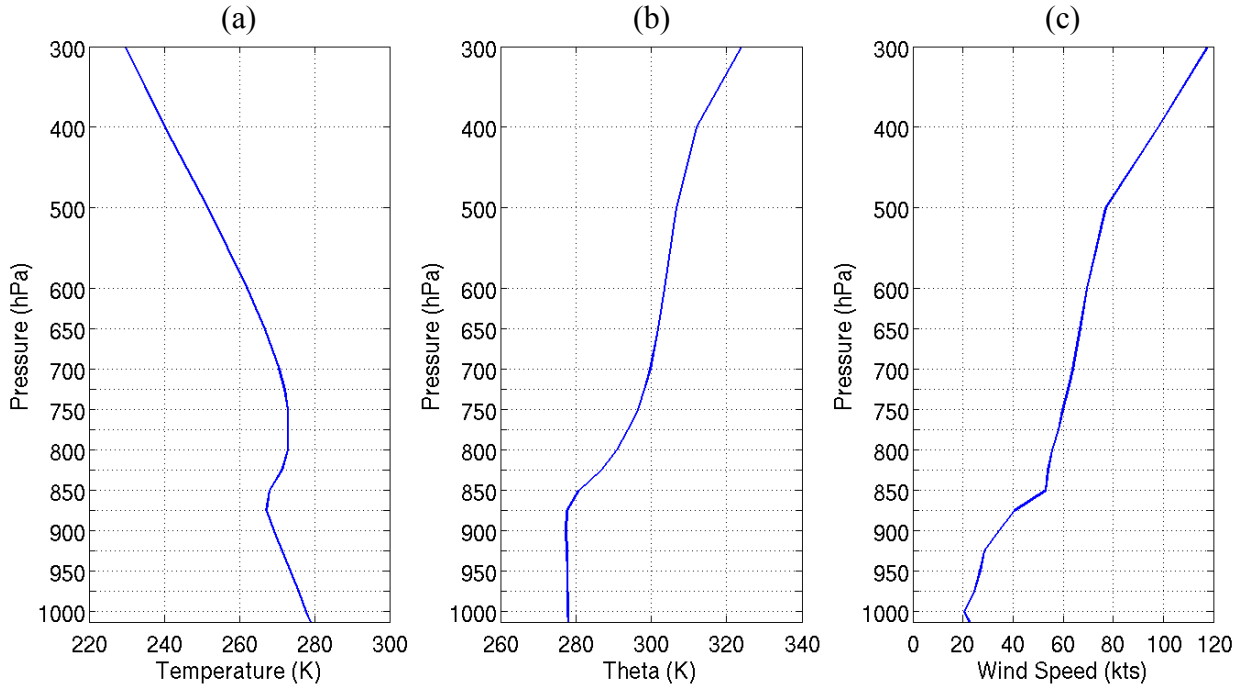

Fig. 16. Vertical profiles of (a) temperature (in K), (b) theta (in K), and (c) wind speed (in kts), interpolated at the site (LGA) and time (1800 UTC) of maximum wind during the 10 February 2001 POF false alarm case.

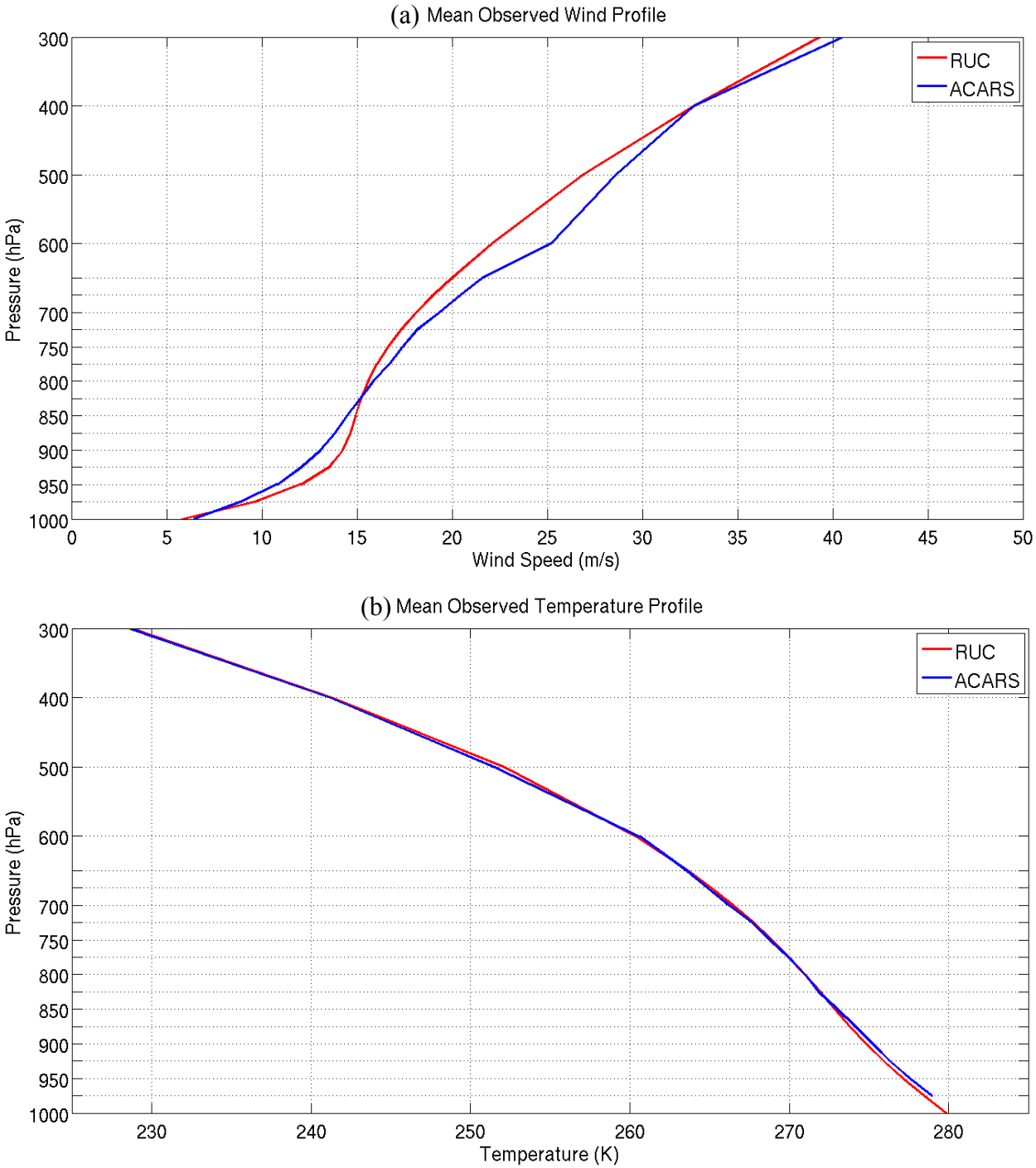

Fig. 17. Mean observed RUC and ACARS (a) wind (in m s^{-1}) and (b) temperature profiles for 235 events with wind speed ≥ 10.29 m s^{-1}.

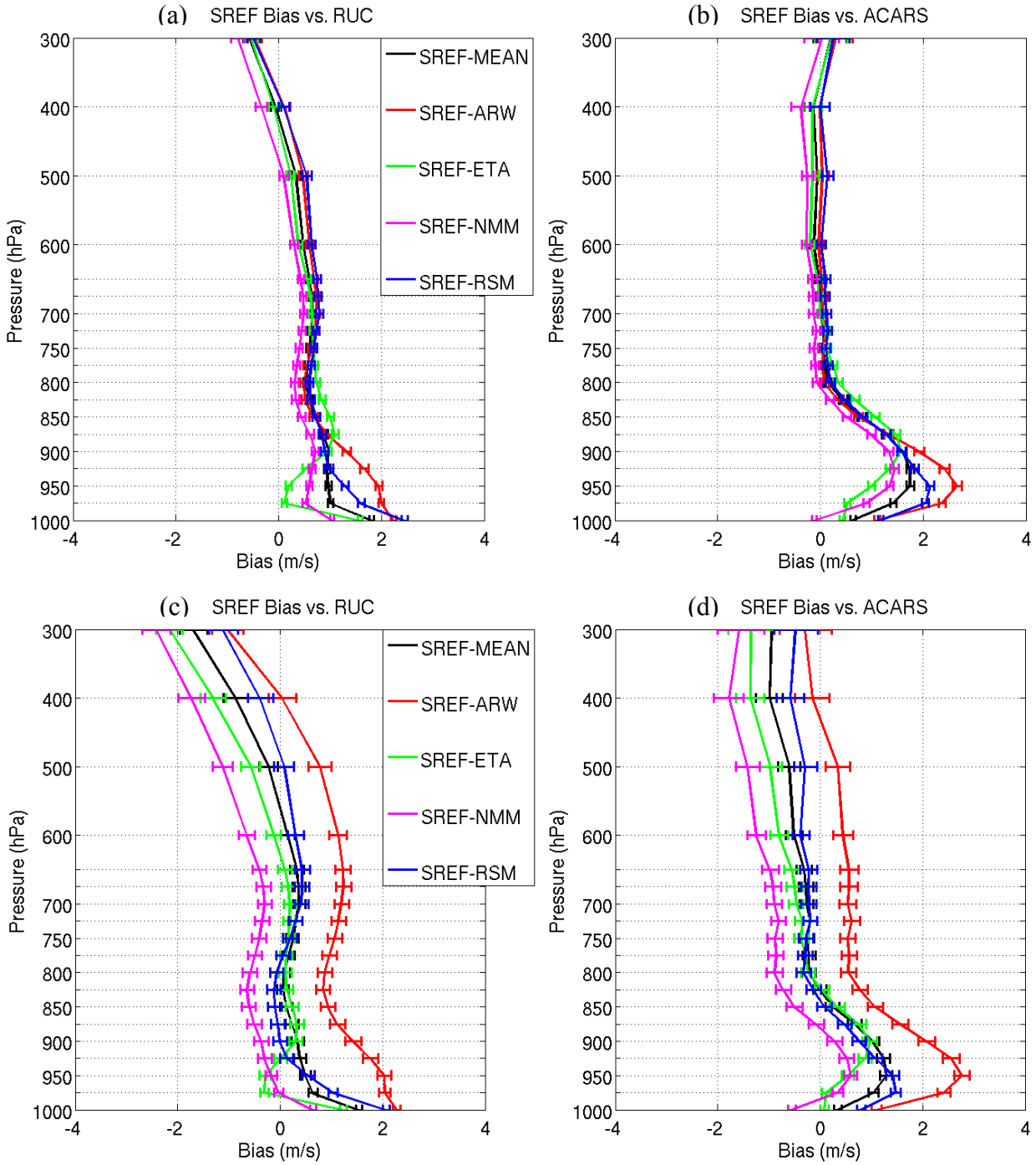

Fig. 18. Vertical profiles of wind speed bias (in m s^{-1}) for the (a) SREF vs. RUC and (b) SREF vs. ACARS for the first forecast period (1-24 h) and (c) SREF vs. RUC and (d) SREF vs. ACARS for the third forecast period (51-72 hours). Error bars are plotted at the 95% confidence interval.

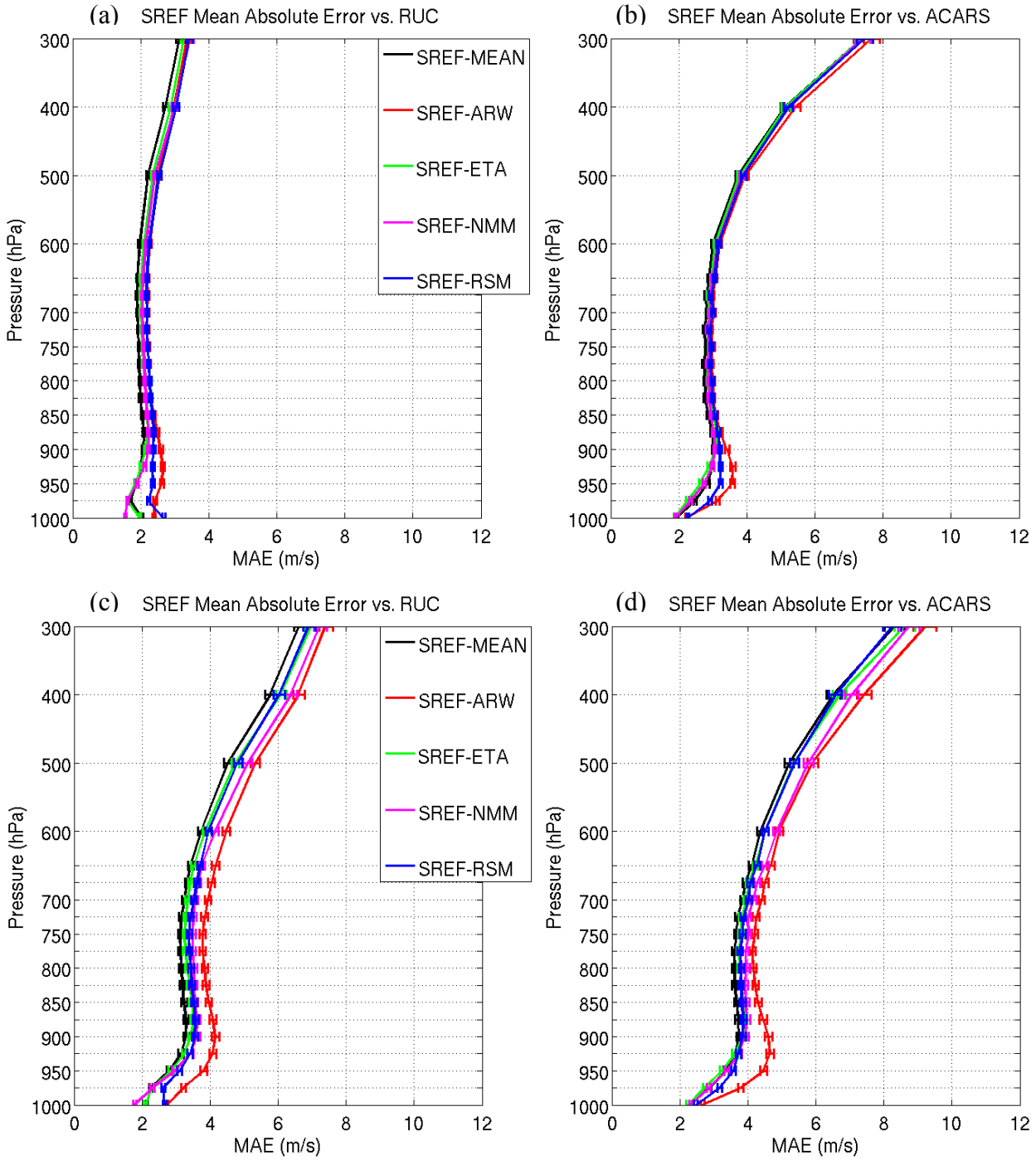

Fig. 19. Vertical profiles of wind speed mean absolute error (in m s^{-1}) for the (a) SREF vs. RUC and (b) SREF vs. ACARS for the first forecast period (1-24 h) and (c) SREF vs. RUC and (d) SREF vs. ACARS for the third forecast period (51-72 h). Error bars are plotted at the 95% confidence interval.

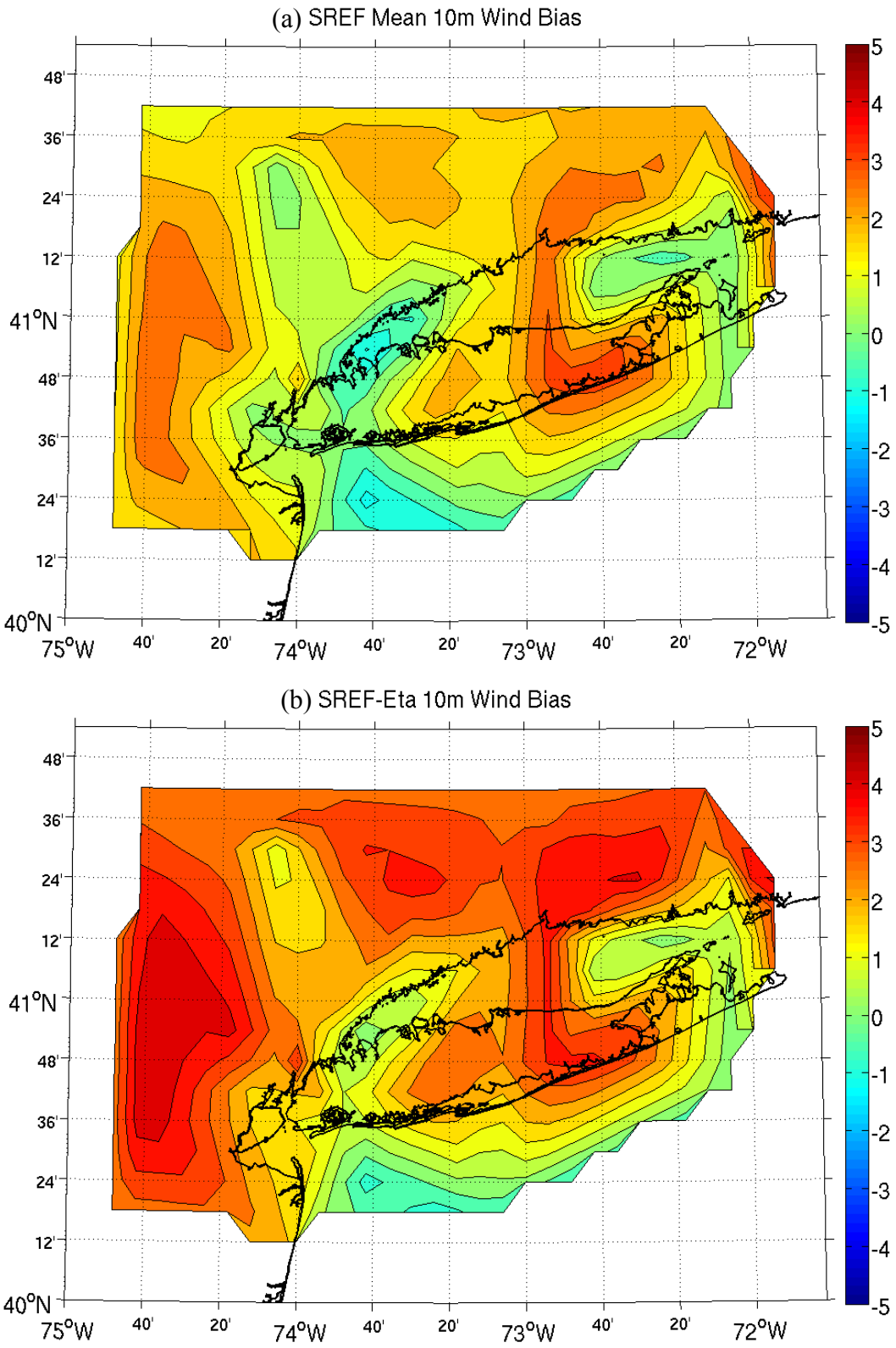

Fig. 20. SREF 10 m wind speed bias (in m s^{-1}) compared to METAR observations for the (a) ensemble mean and (b) Eta model core, calculated over the first forecast period (1-24 h). Biases were calculated for 39 land and water sites and interpolated to an evenly-spaced grid.

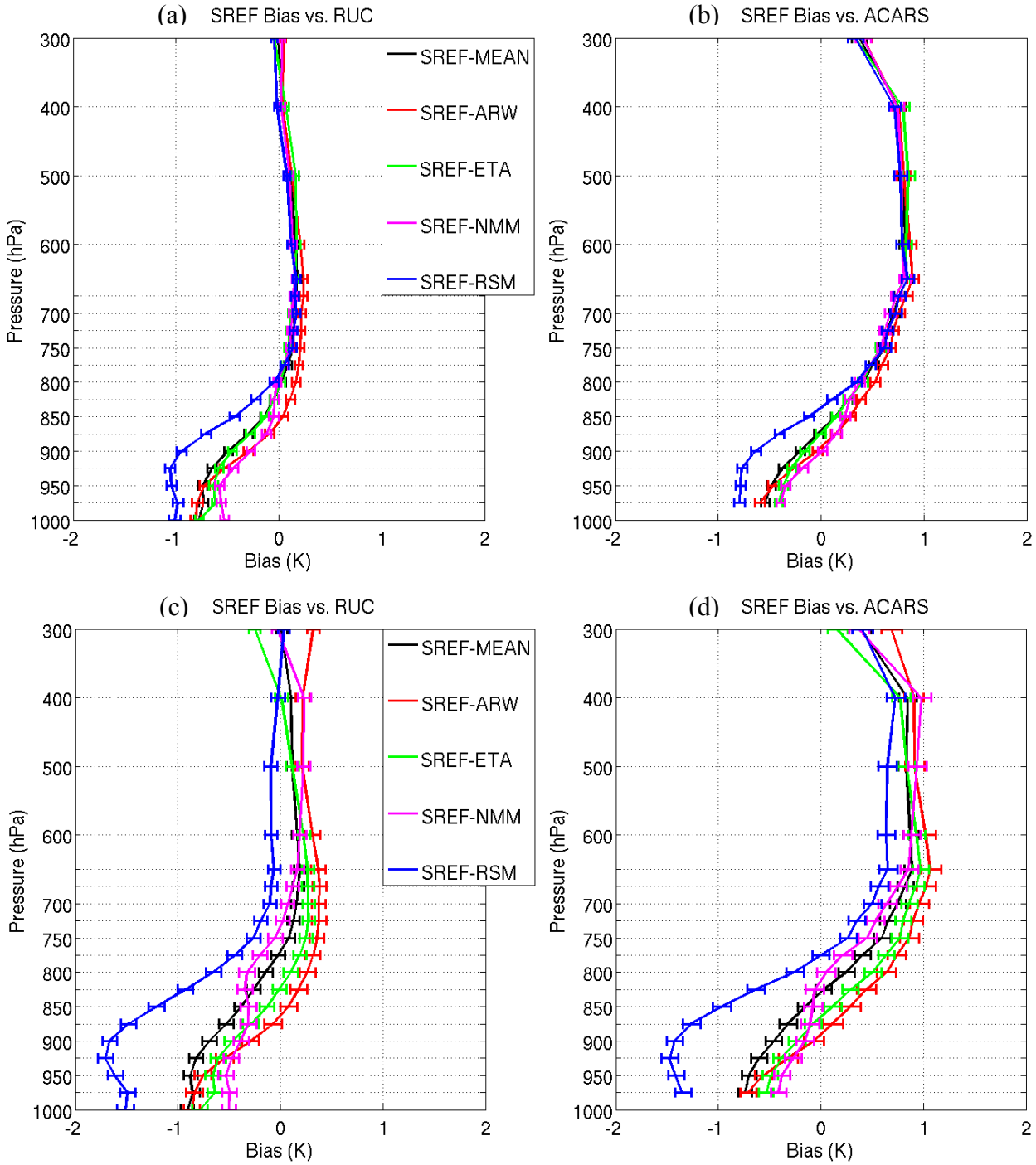

Fig. 21. Vertical profiles of temperature bias (in K) for the (a) SREF vs. RUC and (b) SREF vs. ACARS for the first forecast period (1-24 h) and (c) SREF vs. RUC and (d) SREF vs. ACARS for the third forecast period (51-72 h). Error bars are plotted at the 95% confidence interval.

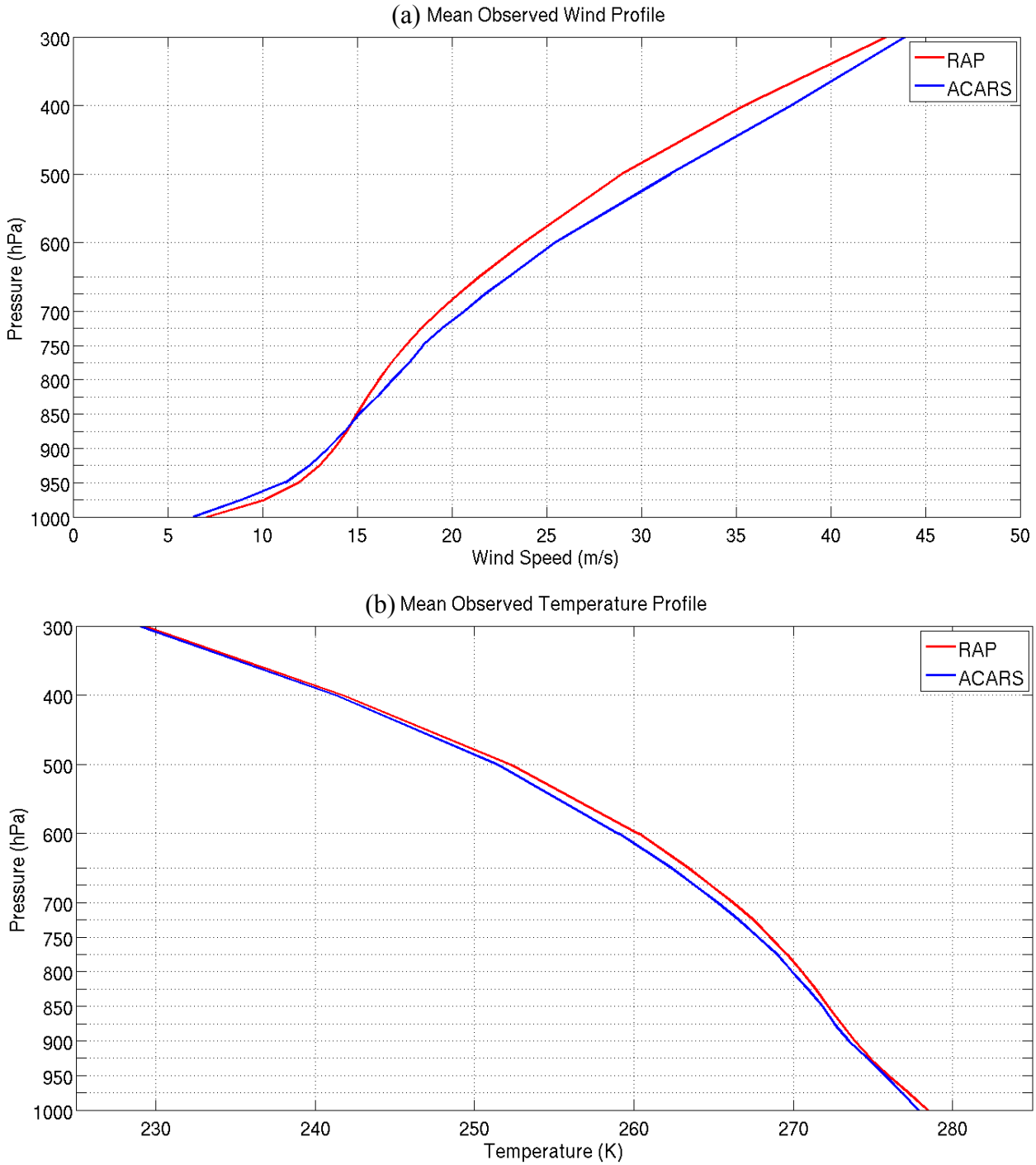

Fig. 22. Mean observed RAP and ACARS (a) wind (in m s^{-1}) and (b) temperature (in K) profiles for 123 events with wind speed ≥ 10.29 m s^{-1}.

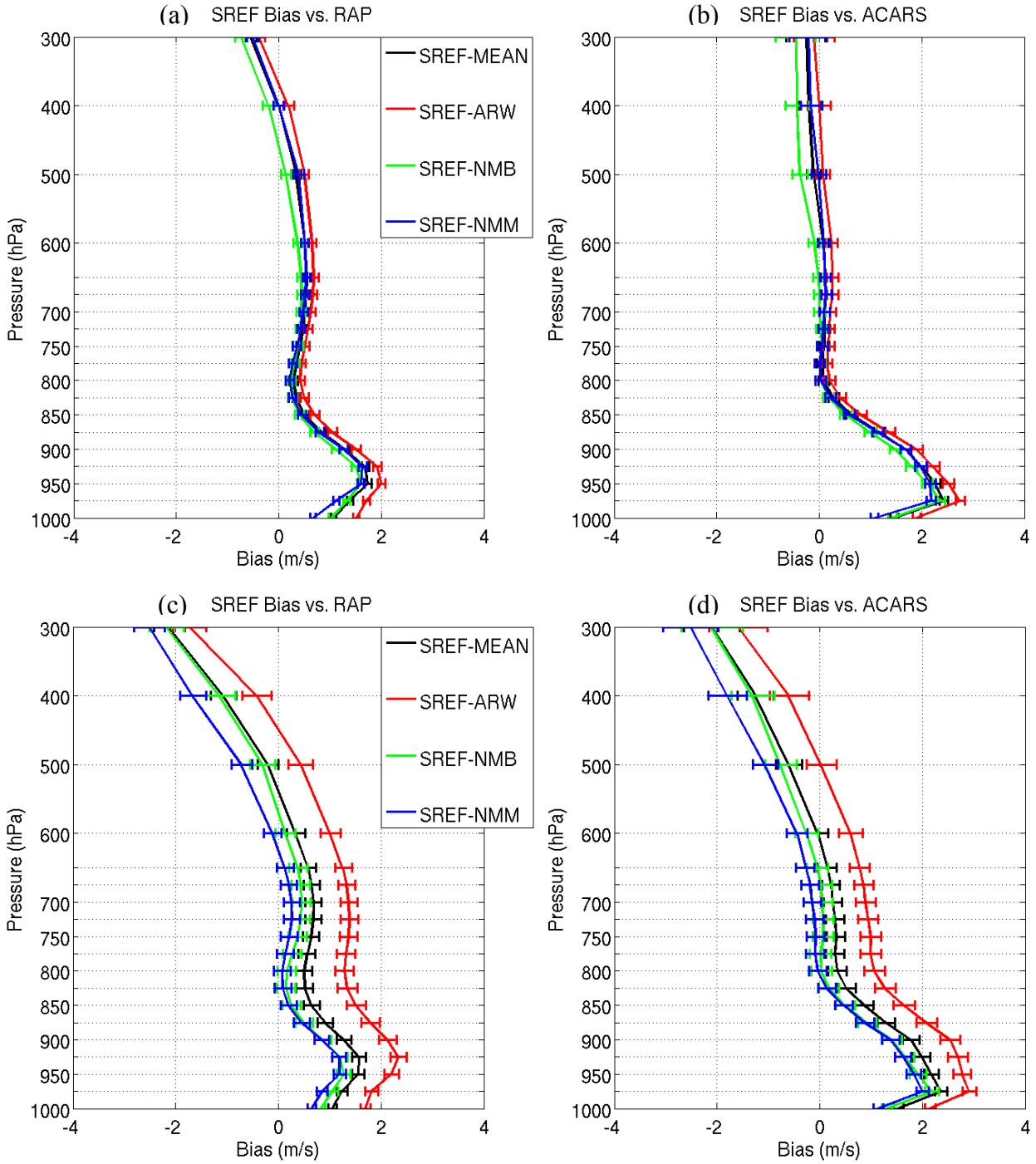

Fig. 23. Vertical profiles of wind speed bias (in m s^{-1}) for the (a) SREF vs. RAP and (b) SREF vs. ACARS for the first forecast period (1-24 h) and (c) SREF vs. RAP and (d) SREF vs. ACARS for the third forecast period (51-72 h). Error bars are plotted at the 95% confidence interval.

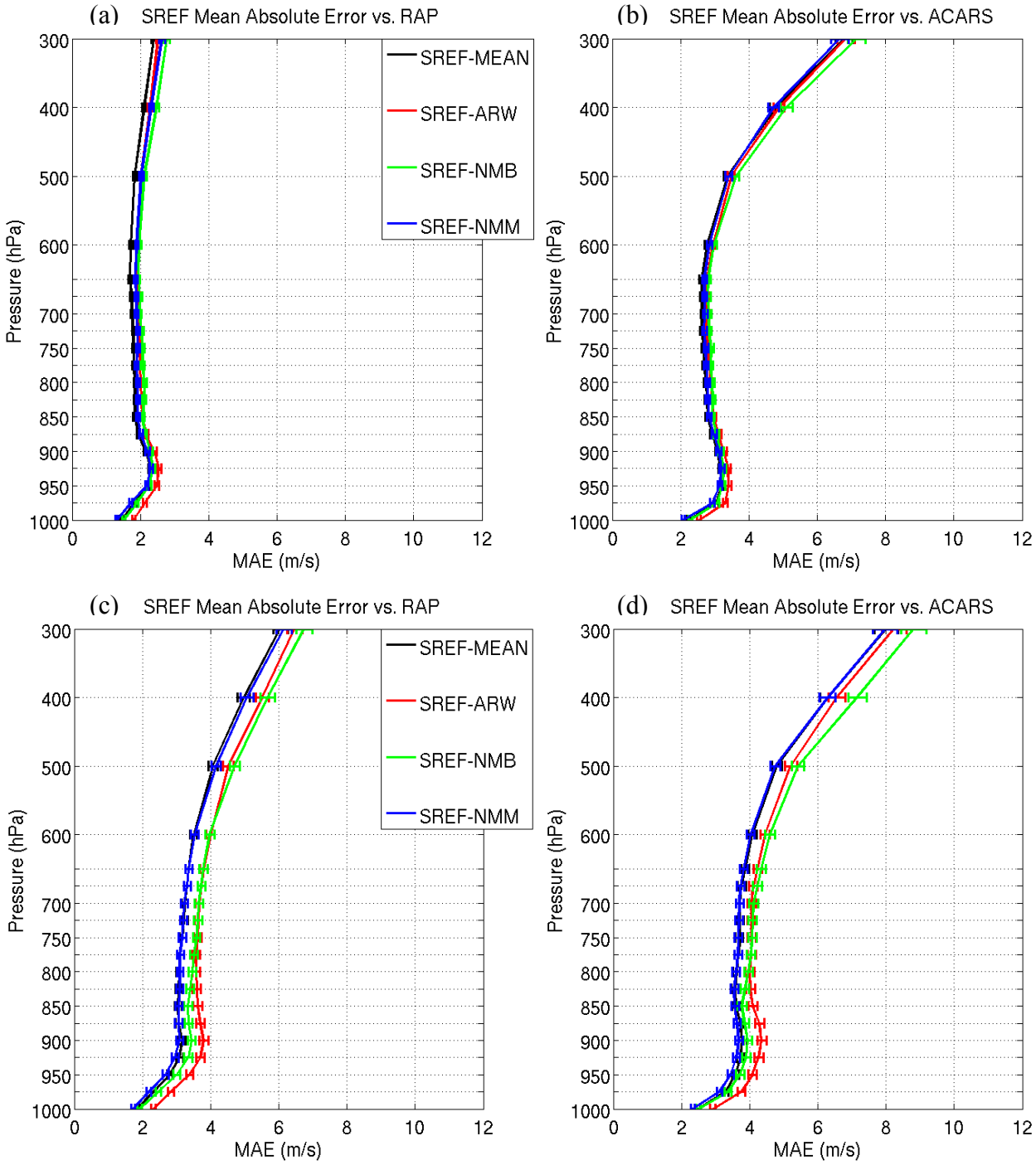

Fig. 24. Vertical profiles of wind speed mean absolute error (in m s^{-1}) for the (a) SREF vs. RAP and (b) SREF vs. ACARS for the first forecast period (1-24 h) and (c) SREF vs. RAP and (d) SREF vs. ACARS for the third forecast period (51-72 h). Error bars are plotted at the 95% confidence interval.

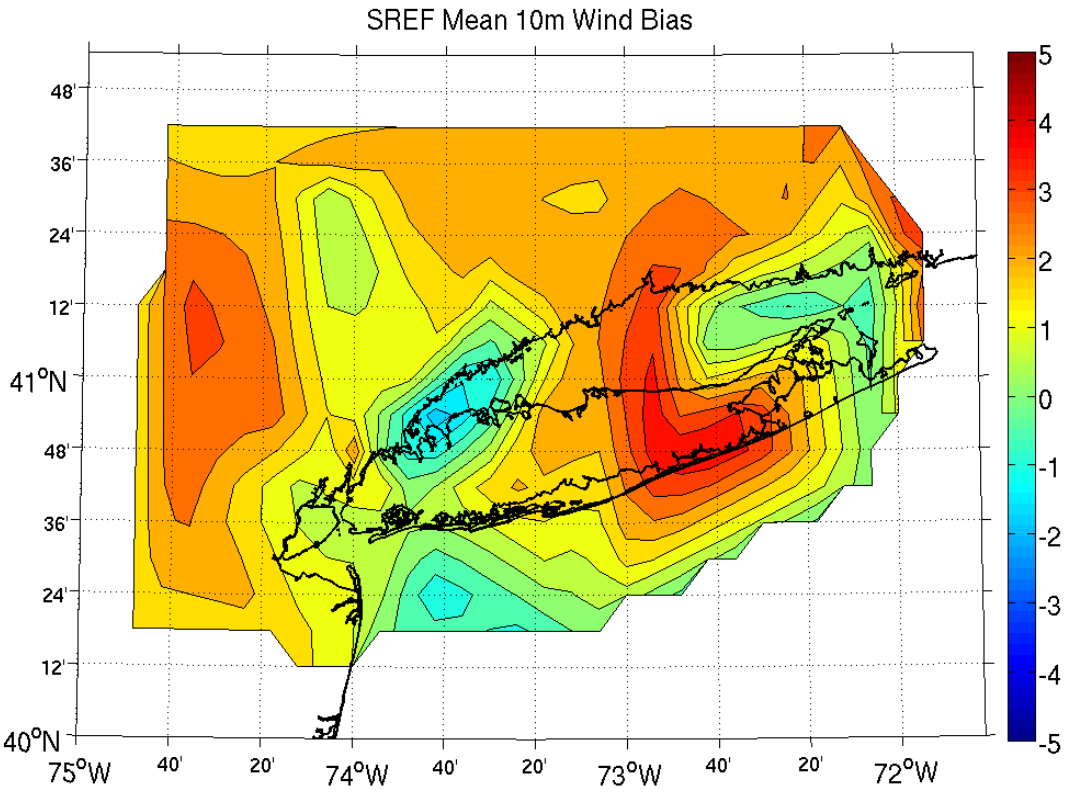

Fig. 25. SREF 10 m wind speed bias (in m s^{-1}) compared to METAR observations for the ensemble mean, calculated over the first forecast period (1-24 h). Biases were calculated for 39 land and water sites and interpolated to an evenly-spaced grid.

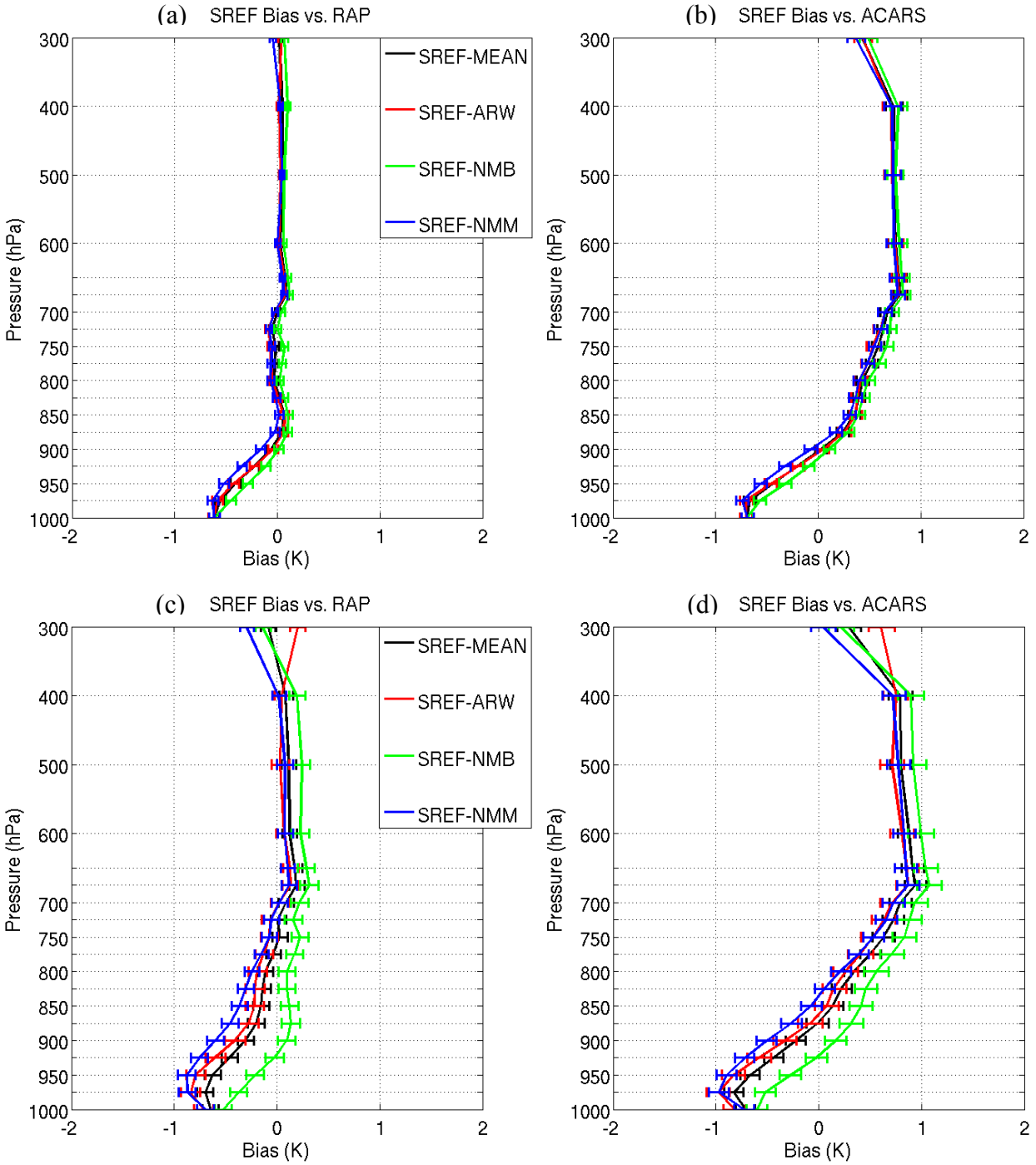

Fig. 26. Vertical profiles of temperature bias (in K) for the (a) SREF vs. RAP and (b) SREF vs. ACARS for the first forecast period (1-24 h) and (c) SREF vs. RAP and (d) SREF vs. ACARS for the third forecast period (51-72 h). Error bars are plotted at the 95% confidence interval.

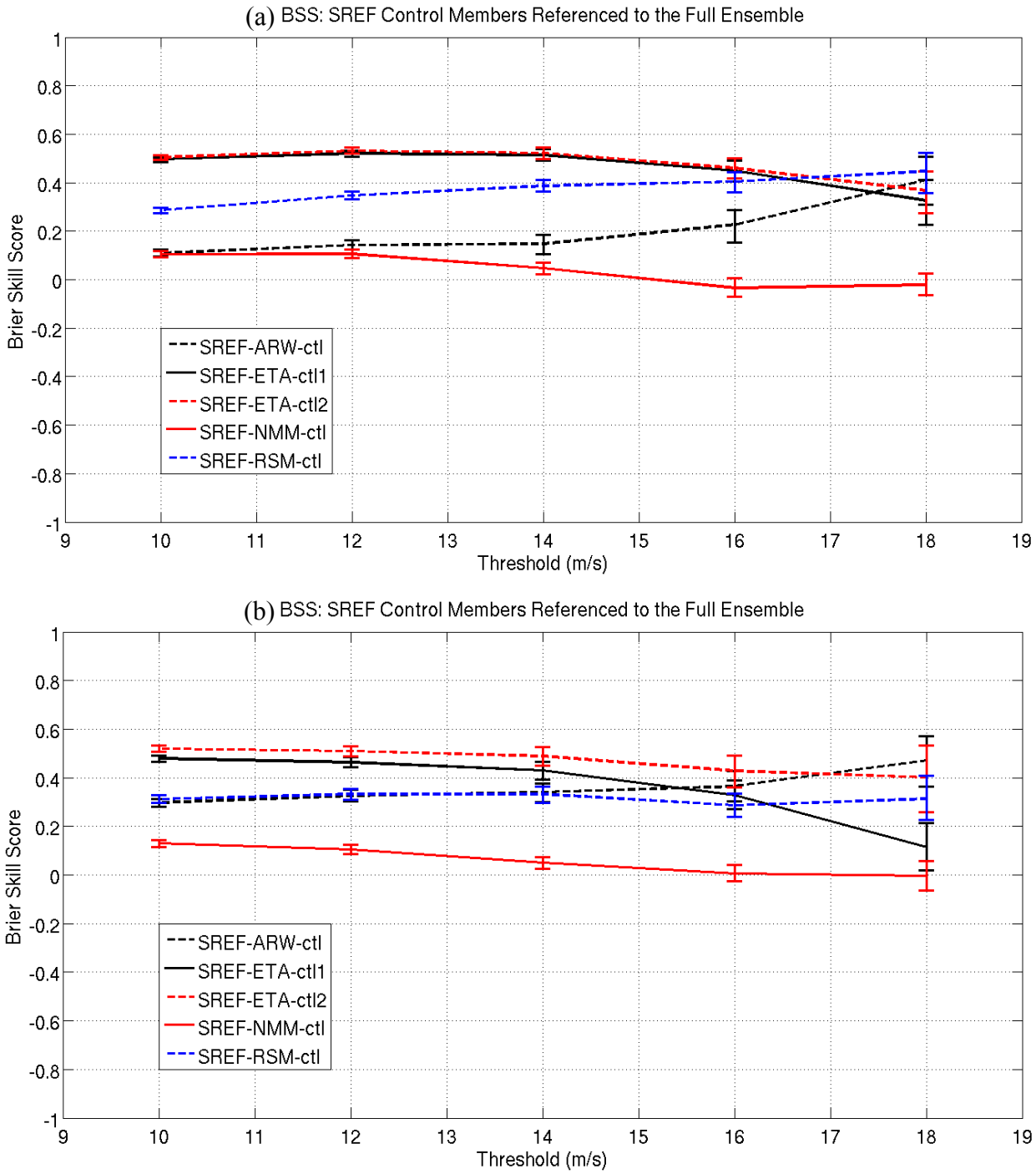

Fig. 27. Brier Skill Scores of the old SREF control members (5) referenced to the full ensemble, calculated over increasing 10 meter wind speed thresholds (in m s^{-1}), for the (a) first forecast period (1-24 h) and (b) third forecast period (51-72 h). Error bars are plotted at the 95% confidence interval.

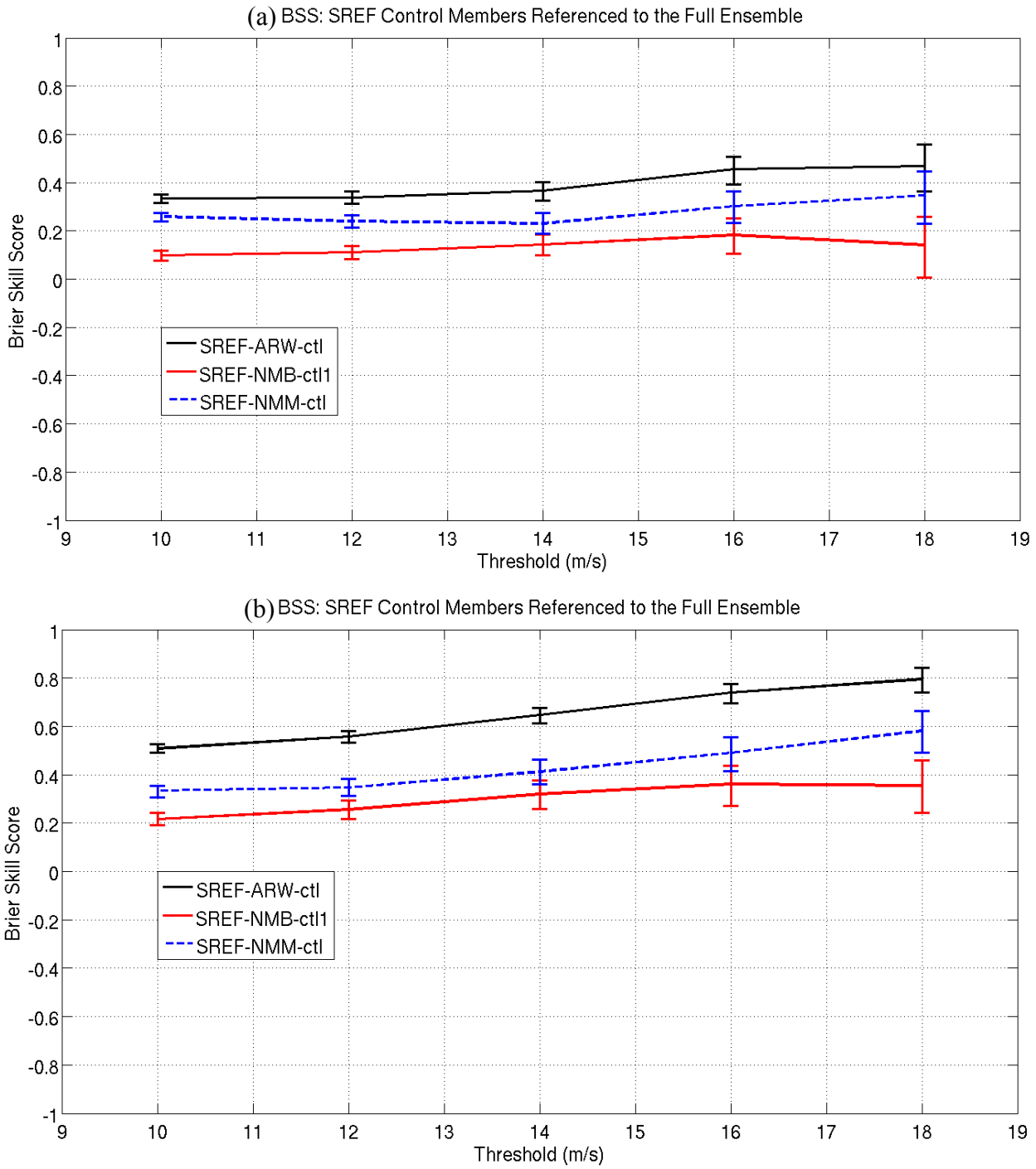

Fig. 28. Brier Skill Scores of the new full SREF ensemble referenced to each of the three SREF control members, calculated over increasing 10 meter wind speed thresholds (in m s^{-1}), for the (a) first forecast period (1-24 h) and (b) third forecast period (51-72 h). Error bars are plotted at the 95% confidence interval.

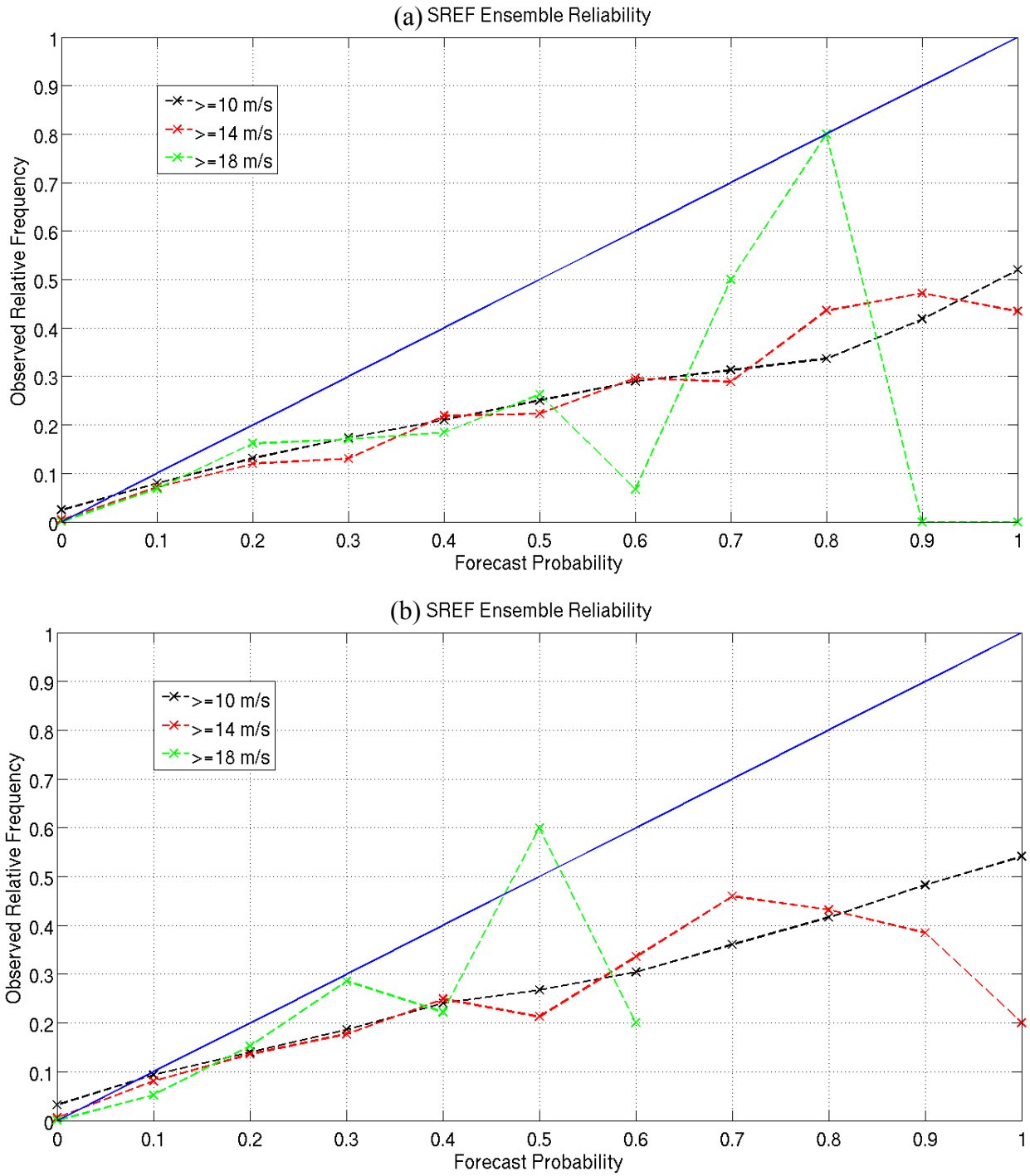

Fig. 29. Reliability diagrams for the old SREF 10 meter wind at three wind speed thresholds (10, 14, and 18 m s^{-1}) for the (a) first forecast period (1 to 24 h) and (b) third forecast period. The blue diagonal line represents a perfect forecast.

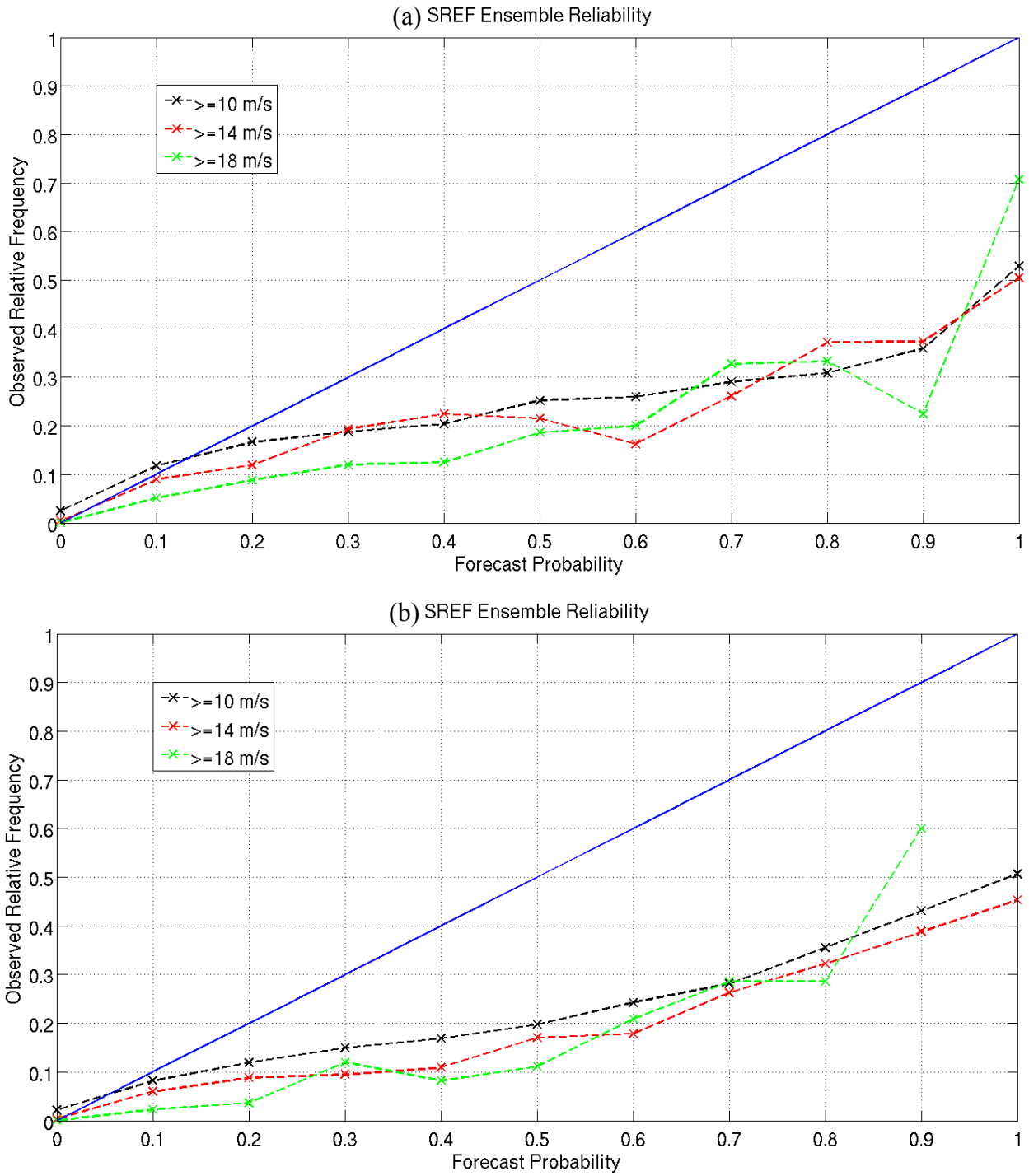

Fig. 30. Reliability diagrams for the new SREF 10 meter wind at three wind speed thresholds (10, 14, and 18 m s^{-1}) for the (a) first forecast period (1 to 24 h) and (b) third forecast period. The blue diagonal line represents a perfect forecast.

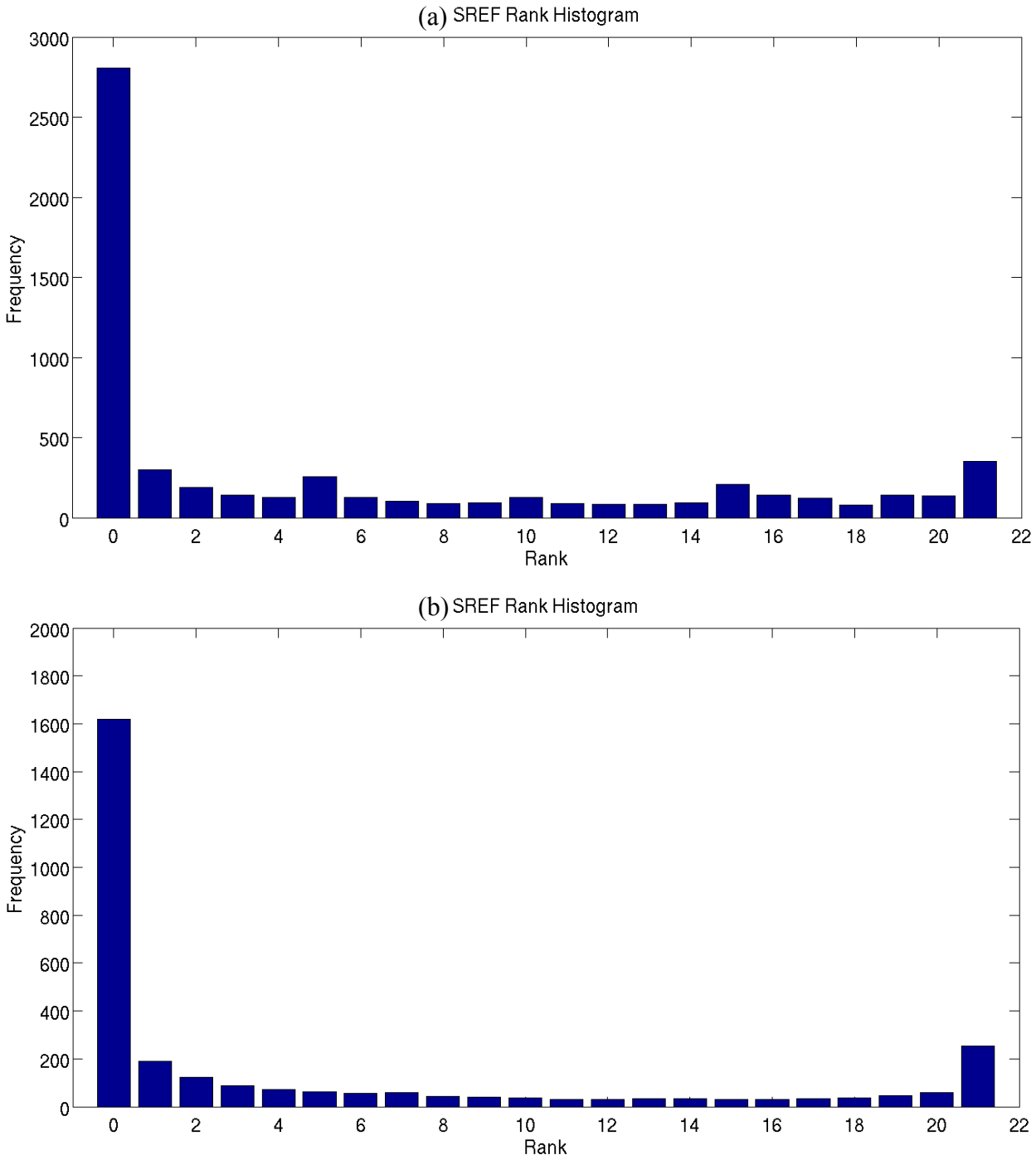

Fig. 31. Rank histogram for the (a) old SREF and (b) new SREF 10 meter wind, calculated over the first forecast period (1 to 24 h).