

**Journal of Applied Meteorology and Climatology**  
**Improved Nowcasts By Blending Extrapolation and Model Forecasts**  
 --Manuscript Draft--

<b>Manuscript Number:</b>	
<b>Full Title:</b>	Improved Nowcasts By Blending Extrapolation and Model Forecasts
<b>Article Type:</b>	Article
<b>Corresponding Author:</b>	Yunsung Hwang, M.S. University of Oklahoma Norman, OK UNITED STATES
<b>Corresponding Author's Institution:</b>	University of Oklahoma
<b>First Author:</b>	Yunsung Hwang, M.S.
<b>Order of Authors:</b>	Yunsung Hwang, M.S.  Adam J. Clark  Valliappa Lakshmanan  Steven E. Koch
<b>Manuscript Classifications:</b>	5.108: Radars/Radar observations; 7.016: Forecast verification/skill; 7.024: Forecasting techniques; 7.036: Nowcasting; 7.040: Numerical weather prediction/forecasting
<b>Abstract:</b>	<p>Planning and management of commercial airplane routes to avoid thunderstorms requires very skillful and frequently updated 0 - 8 h forecasts of convection. National Oceanic and Atmospheric Administration's High Resolution Rapid Refresh (HRRR) model is well suited for this purpose, initialized hourly and providing explicit forecasts of convection out to 15 h. However, because of difficulties with depicting convection at the time of model initialization and shortly thereafter (i.e., model spin-up), relatively simple extrapolation techniques, on average, perform better than the HRRR at 0 - 2 h lead times. Thus, recently developed nowcasting techniques blend extrapolation-based forecasts with numerical weather prediction (NWP) model based forecasts, heavily weighting the extrapolation forecasts at 0 - 2 h lead times and transitioning emphasis to the NWP-based forecasts at the later lead times.</p> <p>In this study, a new approach to apply different weights to blend extrapolation and model forecasts based on intensities and forecast times is applied and tested. An image processing method of morphing between extrapolation and model forecasts to create nowcasts is described and the skill compared to extrapolation forecasts and forecasts from the HRRR model. The new approach is called "Salient cross-dissolve" (Sal CD), which is compared to a commonly used method called "Linear cross-dissolve" (Lin CD). Examinations of forecasts and observations of the maximum altitude of echo top heights <math>\geq 18 \text{ dBZ}</math> and measurement of forecast skill using neighborhood-based methods shows that Sal CD significantly improves upon Lin CD, as well as the HRRR model at 2 - 5 h lead times.</p>
<b>Suggested Reviewers:</b>	

[Click here to download Cost Estimation and Agreement Worksheet: JAMC\\_cost\\_YHWANG.pdf](#)

<sup>1</sup> Improved Nowcasts By Blending Extrapolation and  
<sup>2</sup> Model Forecasts

<sup>3</sup> Yunsung Hwang<sup>1,2,3\*</sup>,

Adam J. Clark<sup>1,2</sup>,

Valliappa Lakshmanan<sup>4</sup>,

Steven E. Koch<sup>2,3</sup>

---

\*Corresponding author: Yunsung Hwang, University of Oklahoma, Norman, OK, yunsung.hwang@ou.edu

<sup>1</sup> Cooperative Institute of Mesoscale Meteorological Studies, University of Oklahoma <sup>2</sup> National Severe  
Storms Laboratory, Norman, OK <sup>3</sup> School of Meteorology, University of Oklahoma <sup>4</sup> Climate Corporation,  
Seattle, WA

## ABSTRACT

Planning and management of commercial airplane routes to avoid thunderstorms requires very skillful and frequently updated 0 - 8 h forecasts of convection. National Oceanic and Atmospheric Administration's High Resolution Rapid Refresh (HRRR) model is well suited for this purpose, initialized hourly and providing explicit forecasts of convection out to 15 h. However, because of difficulties with depicting convection at the time of model initialization and shortly thereafter (i.e., model spin-up), relatively simple extrapolation techniques, on average, perform better than the HRRR at 0 - 2 h lead times. Thus, recently developed nowcasting techniques blend extrapolation-based forecasts with numerical weather prediction (NWP) model based forecasts, heavily weighting the extrapolation forecasts at 0 - 2 h lead times and transitioning emphasis to the NWP-based forecasts at the later lead times.

In this study, a new approach to apply different weights to blend extrapolation and model forecasts based on intensities and forecast times is applied and tested. An image processing method of morphing between extrapolation and model forecasts to create nowcasts is described and the skill compared to extrapolation forecasts and forecasts from the HRRR model. The new approach is called “Salient cross-dissolve” (Sal CD), which is compared to a commonly used method called “Linear cross-dissolve” (Lin CD). Examinations of forecasts and observations of the maximum altitude of echo top heights  $\geq 18$  dBZ and measurement of forecast skill using neighborhood-based methods shows that Sal CD significantly improves upon Lin CD, as well as the HRRR model at 2 - 5 h lead times.

## 1. Introduction

Federal Aviation Administration (FAA) Operations Network (OPSNET) data show that more than 70% of the National Airspace System (NAS) reportable delays are contributed by convective weather (Sheth et al. 2013). Air traffic is routed around anticipated locations of convective weather systems, forcing aircraft to take large deviations. Accurate nowcasts

29 are, therefore, critical to reducing the number of such delays. While the strategic time frame  
30 for flight operations is only 8 h, longer-term flight planning requires up to 12 h forecasts  
31 of variables containing information on convection such as echo top heights and vertically  
32 integrated liquid (VIL) (Robinson et al. 2008; Pinto et al. 2010).

33 The focus of recent research in providing support for flight planning has been on developing  
34 improved weather products and making better use of probabilistic data, which benefits  
35 various participants in air traffic management (Fahey et al. 2006). Other research has fo-  
36 cused on operational concepts for managing strategic traffic flow, including examination of  
37 how improved weather data can aid traffic management initiatives efficiently (Song et al.  
38 2008).

39 For short-term prediction of convection for route-planning applications, frequently up-  
40 dating high-resolution forecasts of convection are needed (i.e., nowcasts). To address this  
41 need, since about the early 1990s, various nowcasting techniques have been developed that  
42 rely on extrapolation of observed convection as depicted by radar-derived fields (Dixon and  
43 Wiener 1993; Li et al. 1995; Germann and Zawadzki 2002, 2004; Mandapaka et al. 2012). Al-  
44 though oftentimes quite skillful at 1 to 2 h lead times, the extrapolation-based methods suffer  
45 from the obvious shortcoming that they are not able to depict rapidly changing conditions  
46 associated with processes such as convection initiation, dissipation, and changing intensities  
47 and movements. For changing conditions, a rapidly updated numerical weather prediction  
48 (NWP) model with high enough resolution to provide explicit forecasts of convection is  
49 necessary (Stratman et al. 2013).

50 To test NWP model forecasts for nowcasting applications, several recent studies have  
51 compared the forecast skill of NWP models to extrapolation-based methods at very short  
52 lead times. For example, Mandapaka et al. (2012) compared precipitation forecasts from  
53 an algorithm known as McGill Algorithm for Precipitation nowcasting using Lagrangian  
54 Extrapolation (MAPLE; Germann and Zawadzki 2002) to high-resolution NWP-based fore-  
55 casts from the Consortium for Small-scale Modeling model (COSMO2) system (<http://>

56 `cosmo-model.org`). They found that on average the MAPLE forecasts had higher skill  
57 during the first 2.5 h of the forecast, after which the COSMO2 forecasts performed better.  
58 Similarly, examining precipitation forecasts, Lin et al. (2005) found that extrapolation-based  
59 predictions were more skillful than four different NWP models, on average, up to about 6  
60 h lead times. The lower skill during the first few hours of the NWP-based predictions oc-  
61 curred because of difficulties in depicting small-scale convective features in their model initial  
62 conditions, and then correctly evolving these features (i.e., the model “spin up” problem).  
63 The “crossover” time - i.e., when the NWP-based forecasts become better - is earlier in the  
64 Germann and Zawadzki (2002) study because they used a more advanced NWP system with  
65 a more sophisticated data assimilation scheme that assimilated radar-derived rainfall fields.  
66 In theory, as sophisticated high-resolution data assimilation methods continue to improve,  
67 NWP-based forecasts may eventually become more skillful than extrapolation-based meth-  
68 ods at all lead times. However, while the extrapolation methods remain more skillful at  
69 short lead times, seamless 0 - 8 h predictions may be obtained by blending extrapolation  
70 with NWP-based forecasts, with the extrapolation forecasts heavily weighted during the first  
71 few hours and the heavier weights transitioning to the NWP-based forecasts at later lead  
72 times.

73 To address the nowcasting problem using this blending approach, the FAA collaborated  
74 with the Massachusetts Institute of Technology Lincoln Laboratory (MIT LL), the National  
75 Center for Atmospheric Research (NCAR) Research Applications Laboratory (RAL), and  
76 National Oceanic and Atmospheric Administration (NOAA) Earth Systems Research Labo-  
77 ratory (ESRL) Global Systems Division (GSD) to develop a system known as Consolidated  
78 Storm Prediction for Aviation (CoSPA; Wolfson et al. 2008; Dupree et al. 2009). CoSPA  
79 was aimed to provide information by blending extrapolation-based forecasts and NWP-  
80 based forecasts for lead times up to 8 h. The High-Resolution Rapid Refresh (HRRR;  
81 <http://ruc.noaa.gov/HRRR/>) model, an hourly updated 3-km grid-spacing convection-  
82 permitting modeling system, developed by NOAA/ESRL/GSD that became operational 30

83 September 2014, was utilized as the model forecast. The blending method of CoSPA consists  
84 of three steps: 1) calibration of the HRRR data to remove intensity biases, 2) application of  
85 a spatial correction to align the HRRR fields with observations and 3) weighted averaging  
86 of the extrapolation and HRRR fields (Pinto et al. 2010).

87 The method of obtaining weighted averages for blending in CoSPA is based on applying  
88 time-varying weights to the extrapolation and HRRR fields. The extrapolation has more  
89 weight (close to 1.0) at the shorter lead times and decreases gradually at the longer lead  
90 times (approaching 0.0). The calibration and the spatial offsets are applied based on the  
91 most up-to-date radar mosaic. Pinto et al. (2010) provides additional details on the blending  
92 procedure used in CoSPA, as well as verification results for a prototype version of CoSPA  
93 during the summers of 2008 and 2009

94 Comparing to extrapolation, as well as raw and calibrated HRRR forecasts, Pinto et al.  
95 (2010) find that the forecast skill of CoSPA, as measured by the critical success index (CSI),  
96 follows that of extrapolation during the first 2 to 3 h and then converges toward the skill  
97 of the HRRR model during the last 6 to 8 h. During a 3 h time window centered around  
98 forecast 4 h, which was the time at which model skill exceeded that of extrapolation, the  
99 margin by which the skill of CoSPA forecasts exceeded the skill of the next most skillful  
100 forecast was highest. Similar results from application of CoSPA during July 2012 can be  
101 found in Sun et al. (2014).

102 Although the blending method used in CoSPA shows promising results, biases near 0.6  
103 within the 3 to 4 h forecast period indicate a systematic under-prediction in the areal converge  
104 of convection during this time. This systematic under-prediction can likely be partially  
105 explained by the fact that weights are close to 0.5 for both extrapolation and HRRR  
106 fields during this time. Thus, in the case of slight displacements between areas of forecast  
107 convection in the HRRR and extrapolation fields after spatial correction, the fields would be  
108 reduced by half giving lower overall values in two different locations.

109 The purpose of this study is to address this underestimation problem using a new blending

approach that considers intensity in addition to forecast lead-time in the computation of weights. The new blending approach is compared to one in which weights are only dependent on forecast time. Although this time-weighted-only blending approach is less sophisticated than that used in CoSPA (e.g., no attempts are made to correct for the intensity biases of the HRRR forecasts), the comparisons with our newly developed blending approach should serve as a useful proof-of-concept for application in more advanced nowcasting systems.

The technique to apply different weights based on time and intensities is described and its results are compared to that of a time-weighted-only averaging in section 2 along with a description of data and methodology. The results are discussed in section 3. Finally, discussions and ideas for future work are presented in section 4.

## 2. Data and Methodology

### a. Dataset

In this study, forecasts and observations of 18 dBZ echo top heights are examined, which are defined as the maximum altitude at which reflectivity exceeds 18 dBZ. The observed echo top heights are estimated from the Weather Surveillance Radar 1988 Doppler (WSR-88D) data using the highest elevation angle that detects reflectivity over 18 dBZ (Lakshmanan et al. 2013). Echo top heights were chosen for examination because they are one of the parameters that determine the availability of a flight route in recently developed convective weather avoidance models (Matthews and DeLaura 2010; Sheth et al. 2013). For verification purposes, 18 dBZ echo top heights computed from the WSR-88D network covering the Contiguous United States (CONUS) are used as truth. Four different sets of 8 h forecasts are evaluated, which are all initialized at 1800 UTC. This particular initialization time was chosen because it is a few hours before the typical maximum in the diurnal cycle of convection and, thus, precedes by a few hours the largest potential impacts on flight routing. Forecasts on 24 days during the period 15 May to 13 June 2013 were examined (the dates 20, 28, 29 May

and 4 and 7 June were excluded because of missing data). The four forecasts consisted of 1) the HRRR model, 2) extrapolated observations, 3) a blending of extrapolated observations and the HRRR referred to as Lin CD, and 4) another blending of extrapolated observations and the HRRR called Sal CD. Details on the four forecasts are discussed in the following sections.

*b. HRRR*

The HRRR model is a convection-allowing model, which generates convection without convective parametrization, covering the Contiguous United States (CONUS) with 3-km grid spacing and is nested within the parent model domain of the 13-km grid-spacing Rapid Refresh (RAP; Brown et al. 2011; Weygandt et al. 2011) model. The RAP provides initial and boundary conditions and assimilates radar reflectivity observations through a diabatic digital filter initialization (Huang and Lynch 1993). The HRRR model is based on the Advanced Research core of the Weather Research and Forecasting (WRF) model (ARW; Skamarock et al. 2008) with the following WRF physics options: 1) Goddard shortwave radiation scheme (Chou and Suarez 1994), 2) Rapid radiative transfer model longwave radiation scheme (Mlawer et al. 1997), 3) RUC Smirnova land surface model (Smirnova et al. 1997), 4) Mellor-Yamada-Nakanishi-Niino (MYNN) boundary layer parameterization (Nakanishi and Niino 2004) and 5) Thompson mixed-phase microphysics scheme (Thompson et al. 2008).

The RAP and HRRR assimilate data hourly using Gridpoint Statistical Interpolation (GSI). The HRRR utilizes 3-km data assimilation to include detailed observational information using GSI. During a preforecast hour, observed radar reflectivities replace latent heating fields for the HRRR at 15 minute intervals. Moreover, at the beginning of the forecast, a 3-km non-variational cloud analysis and hydrometeor analysis using radar reflectivities are used to obtain additional information about rain and snow mixing ratios.

159    *c. Extrapolated observations*

160    The high spatial and temporal resolution of weather radar data has enabled the devel-  
161    opment of several nowcasting techniques based on extrapolation methods (Mandapaka et al.  
162    2012). In these methods the movement of storm cells is typically estimated by matching  
163    radar echoes between two successive radar images. A storm cell is typically defined as a  
164    region of reflectivity that exceeds a threshold (usually 35 or 40 dBZ). Examples of track-  
165    ing and extrapolation algorithms found in the literature include Continuity Of Tracking  
166    Radar Echoes by Correlation (COTREC; Li et al. 1995), which used variational methods  
167    (Sasaki 1958, 1970) to skillfully predict the movement of storm cells 20 minutes in advance.  
168    The Thunderstorm Identification, Tracking, Analysis and Nowcasting (TITAN; Dixon and  
169    Wiener 1993) system was created to optimally match storm cells successive radar images us-  
170    ing a linear programming optimization approach called the Hungarian Method. The McGill  
171    Algorithm for Precipitation nowcasting method using Lagrangian Extrapolation (MAPLE)  
172    (Germann and Zawadzki 2002, 2004) applies variational approaches to provide improved  
173    extrapolation.

174    For extrapolation forecasts, we used the segmotion algorithm that is implemented in the  
175    Warning Decision Support System Integrated Information (WDSS-II; Lakshmanan et al.  
176    2006). In this technique, thunderstorms are identified at different scales using the extended  
177    watershed approach (Lakshmanan et al. 2009). The image is “flooded” starting from the  
178    global maximum. The flooding level is slowly decreased so that flooding can proceed at  
179    lower and lower levels and the entire area covered by water flowing from a single maximum  
180    to a predetermined size (this size varies by scale) forms a thunderstorm. Storms identified in  
181    consecutive images are associated based on a greedy optimization algorithm (Lakshmanan  
182    and Smith 2010) that tries to optimize the match based on projected storm location and  
183    a cost function based on continuity of the maximum value. The motion vector derived  
184    from storm associations is interpolated onto the full grid using an inverse distance weighting  
185    scheme (Lakshmanan et al. 2003). These motion vectors, one for each scale, are then matched

186 to the size of the objects being extrapolated and the time period of extrapolation and used  
187 to extrapolate the current echo top grid into future time steps.

188 *d. Image Morphing*

189 In image processing literature, creating intermediate images to provide a smooth transi-  
190 tion between a pair of images is called morphing. Morphing consists of three image processing  
191 steps: warping, cross-dissolving and unwarping. Because the same entities in the two images  
192 may be slightly displaced, the process of warping attempts to align the objects in the two  
193 fields. This is typically done through a process of coordinate transformation by choosing  
194 the coordinate transformation at which a cost function is minimized. The cost function  
195 balances two concerns: that the warping is as small as possible while the difference between  
196 the warped version of the first image and the second image is also as small as possible.

197 The second image processing step (“cross-dissolving”) is to blend the warped version of  
198 the first image with the second image with different weights chosen to obtain a series of  
199 intermediate images. For example, an intermediate image that is some fraction  $w$  of the way  
200 ( $w < 1$ ) between the two images may be obtained by assigning a weight of  $w$  to each pixel  
201 in the warped version of the first image and a weight of  $(1 - w)$  to the corresponding pixel  
202 in the second image. Such a linear weighting scheme is not the only possible choice. In this  
203 paper, we will employ a saliency-based weighting scheme (discussed later).

204 The third step is to unwarped the blended image to add back the alignment difference  
205 between the pair of images being morphed. This is achieved by applying a weighted inverse  
206 of the warping function to the blended image. Thus, if the coordinate transform to warp  
207 the first image to the second was  $f(x, y)$ , the transformation applied to the blended image  
208 is  $(1 - w)f^{-1}(x, y)$  where  $w$  is the weight of the first image in the blended image.

209 With an appropriately chosen warping function, it is possible to simplify the process above  
210 into two steps: (1) warp the first image by  $wf(x, y)$  and the second image by  $(1 - w)f^{-1}(x, y)$   
211 and (2) cross-dissolve the two warped images to obtain the morphed image.

212 1) LINEAR CROSS-DISSOLVE

213 Linear cross-dissolve is a commonly-employed blending method that computes the weighted  
214 average of two aligned images pixel-by-pixel. For example, if one tries to combine two images  
215  $I_1$  (the extrapolation) and  $I_2$  (the model forecast), the cross-dissolve of the images,  $C(x, y)$   
216 can be represented as

$$C(x, y) = wI_1(x, y) + (1 - w)I_2(x, y) \quad (1)$$

217 where  $I_1(x, y)$  is the intensity of the pixel  $(x, y)$  in the first image and  $I_2(x, y)$  for the  
218 second image. Assuming that there are six time steps (0, 1, 2, 3, 4 and 5 h) between  $I_1$   
219 and  $I_2$ , at 0 h the  $C(x, y)$  is the same as  $I_1$  since  $w = 1.0$  and  $1 - w = 0.0$ . At 3 h,  
220  $C(x, y) = 0.6 \times I_1(x, y) + 0.4 \times I_2(x, y)$  gives slightly more weight to  $I_1$ . For morphing  
221 extrapolation and model fields,  $w$  for a linear cross-dissolve is shown in Fig. 1a and  $1 - w$   
222 Fig. 1b. It should be noted that blending weights ( $w$ ) are independent of the intensities  
223  $I_1(x, y)$  and  $I_2(x, y)$ .

224 For linear cross-dissolve, the same weights are applied to images at a certain fraction of  
225 time for all intensities. Essentially, features from  $I_1$  fade out as features from  $I_1$  fade in.  
226 Features present in both images fade from their presentation as in  $I_1$  to their presentation  
227 as in  $I_2$ .

228 Idealized examples of linear cross-dissolve for a line of discrete storm cells are shown in  
229 Fig. 2. There are six time steps (0, 1, 2, 3, 4 and 5 h) of three convective cells in the  
230 illustration. The cells are moving to the east at a constant speed as shown in the first row of  
231 Fig. 2. The top cell has not developed at 0 h but develops at 1 h and increases in intensity  
232 until 5 h. The center cell decreases in intensity throughout the idealized forecast period  
233 while the bottom cell maintains constant intensity. Extrapolation captures only the center  
234 and bottom cells from the observation at 0 h and extrapolates them to the east without  
235 changing the intensities. The bottom cell is well captured by extrapolation because it is  
236 unchanging in time. However, the change in intensity of the center cell is not captured by

237 the extrapolation. On the other hand, it is not possible to extrapolate the top cell since it  
238 was not present in the observations at 0 h. In the illustration, it is assumed that the model  
239 forecast simulates only the top and the bottom cells, and with lower intensities.

240 The blend of the extrapolation and model forecast using linear cross-dissolve (Lin CD)  
241 is shown in the fourth row of Fig. 2. Extrapolated images are weighted higher than the  
242 model forecasts close to the beginning of the forecast time and the opposite weighting is  
243 employed approaching the end of the forecast time. Lin CD captures all three cells even as  
244 the extrapolation and model forecast depict only two cells each. However, the top and center  
245 cells tend to have weaker intensities compared to that of the observation. For example at  
246 2 h ( $w = 0.6$  and  $1 - w = 0.4$ ), only the center cell is present in the extrapolated image  
247 and therefore, the intensity in Lin CD is decreased to 60 % of the original values. Similarly,  
248 the top cell in Lin CD at 2 h obtains 40 % of the model forecast intensity at the same time  
249 step. As exemplified above, Lin CD is simple and computationally efficient. However, Lin  
250 CD dampens the amplitude of features by applying constant weights.

251 2) SALIENT CROSS-DISSOLVE

252 A method of maintaining the features from multiple images considering the saliency (or  
253 importance) of different intensities was developed and applied to image blending (Grundland  
254 et al. 2006). The goal of that study was to preserve color and contrast while blending multiple  
255 images with different resolutions. Salience contrast and color in that study refer to the  
256 informative aspects of the image as far as human vision is concerned. In this study, we define  
257 salience as the locations of strong cells (in terms of normalized intensities). Consequently, the  
258 composite image using saliency-based cross-dissolve is defined using the following equation.

$$S(x, y) = w_s(w, r(x, y))I_1(x, y) + (1 - w_s(1 - w, r(x, y)))I_2(x, y) \quad (2)$$

259 where  $S$  is composite image of  $I_1$  and  $I_2$  and the weight  $w_s$  is a two-dimensional function of  
 260 weight ( $w$ ) and the ranked salience ( $r(x, y)$ ), where  $w_s$  is calculated using following equation.

$$w_s(w, r) = \frac{1}{2} \left( \frac{wr}{wr + (1-w)(1-r)} + \frac{\sqrt{r^2 + w^2}}{\sqrt{r^2 + w^2} + \sqrt{(1-r)^2 + (1-w)^2}} \right) \quad (3)$$

261 Compared to the linear weights of the top panel of Figs. 1a and b,  $w_s$  allows the blended  
 262 product to preserve pixel intensities with time if they are strong enough based on the  $r$  value  
 263 [See Figs. 1c and d].  $r(x, y)$  is calculated using:

$$r(x, y) = \Phi(N_1(x, y) - N_2(x, y)) \quad (4)$$

264 where  $\Phi(x)$  is a cumulative density function (i.e.,  $\Phi(\text{Min}(N_1(x, y) - N_2(x, y))) = 0$  and  
 265  $\Phi(\text{Max}(N_1(x, y) - N_2(x, y))) = 1$ ) and  $N_n(x, y)$  is the normalized intensity of the image,  
 266  $N_n(x, y) = \frac{N_n(x, y)}{\text{Max}(N_n(x, y))}$  where  $n$  is the number of images ( $n = 1$  and 2 in this study).

267 If the strongest cell is only in  $I_1(x, y)$  and not in  $I_2(x, y)$  at a location  $(x, y)$  then  
 268  $\Phi((N_1(x, y) - N_2(x, y)))$  is close to 1.0 because  $N_1(x, y)$  is close to 1 and  $N_2(x, y)$  is close to  
 269 0. In contrast, if the strongest cell is only in  $I_2(x, y)$ , then  $\Phi(N_1(x, y) - N_2(x, y))$  is close  
 270 to 0.0 because  $N_1(x, y) - N_2(x, y)$  is close to -1 at the location  $(x, y)$ . It should be noted  
 271 that  $r(x, y)$  is not the intensity itself.  $r(x, y)$  shows how close the pixel is to the maximum  
 272 intensity difference of  $N_1(x, y) - N_2(x, y)$  (i.e.,  $r(x, y)=1$ ) or the minimum intensity difference  
 273 of  $N_1(x, y) - N_2(x, y)$  (i.e.,  $r(x, y)=0$ ).

274 The composite image  $S(x, y)$  of the extrapolation and model forecasts using salient cross-  
 275 dissolve (Sal CD) is shown in the fifth row of Fig. 2. Sal CD simulates three cells better than  
 276 Lin CD especially at 2 h ( $w = 0.6$  and  $1 - w = 0.4$ ) and at 3 h ( $w = 0.4$  and  $1 - w = 0.6$ )  
 277 because the higher intensities in OBS are retained. For example at 2 h, the center cell has  
 278 high  $r$  close to 1 where  $w_s$  would be close to 0.9 (the point where  $w = 0.6$  and  $r = 1$  in Fig.  
 279 1c) and the bottom cell has low  $r$  close to 0 where  $1 - w_s$  would be close to 0.9 (the point  
 280 where  $1 - w = 0.4$  and  $r = 0$  in Fig. 1d). Thus, both the middle and bottom cells keep high  
 281 intensities in Sal CD. Additionally, the center cell is shown in Sal CD at 5 h while it is not

282 shown in Lin CD at 5 h. It is possible to obtain the center cell even when the weight for  $I_1$   
283 is zero ( $w = 0$ ) at 5 h because  $w_s$  can be 0.5 if the intensity is close to 1.

284 *e. Statistical Evaluation*

285 We employed two methods to evaluate the forecasts over the 24 days that data was  
286 available. The first evaluation method is the neighborhood (NE) method with a radius of 20  
287 km. Convection is defined as 18 dBZ echo top heights  $\geq$  9 km ( $\approx$  30000 feet). The lowest  
288 echo top height considered dangerous for an airplane is typically 25,000 feet (Matthews and  
289 DeLaura 2010). However, commercial airplanes usually fly at 30,000 feet to 40,000 feet,  
290 which was why 9 km was chosen as the threshold for convection in this study. Utilizing the  
291 neighborhood approach, a “hit” is defined when forecast convection is located within 20 km  
292 of observed convection. A “miss” is where there is no forecast convection within 20 km of  
293 observed convection. A “false alarm” is defined as a forecast for convection but no observed  
294 convection within 20 km. Finally, a “correct null” is when convection is neither forecast nor  
295 observed within 20 km. This methodology for computing neighborhood-based contingency  
296 table elements follows that of Clark et al. (2010).

297 The second evaluation method is the route-based segments (RO-seg) method. Routes  
298 are obtained from a list of 26,606 preferable routes in the database of National Airspace  
299 System Resources (NASR) (<https://nfdc.faa.gov/xwiki/bin/view/NFDC/56+Day+NASR+Subscription>). Each route is an ordered set of waypoints (37,736 points in CONUS) from  
300 the departure airport to the arrival airport. Segments consist of two waypoints of which the  
301 average length is 439.65 km with standard deviation of 489.65 km. There are 6,981 segments  
303 in preferred routes when overlapped segments are excluded and they are used as route-based  
304 segments. Based on the guidelines for horizontal spacing from the FAA, airplanes should be  
305 at least 3 to 5 nautical miles apart depending on the altitude. In this study, a 10 nautical  
306 miles wide (5 nautical miles from the airplane) jetway is considered. Taking the width of the  
307 jetway into account, if there is a convective pixel (18 dBZ echo top height over 9 km) closer

than 10 km (to the line linking two waypoints) then the segment in the route is determined  
 to be closed. With this method, a hit is defined as both routes in the model forecast and  
 observation being closed at the same segment within 20 km. A miss is the case where there is  
 a convective pixel in the closed segment in the route in observation but there is no convective  
 pixel within 20 km. A false alarm is defined as a convective pixel from the route in the model  
 forecasts is in the closed segment but no convective pixel from observation within 20 km. A  
 correct null is the case when there are open segments in routes in the model forecasts and  
 observation [See the bottom right panel of the Fig. 3b].  
 [308] [309] [310] [311] [312] [313] [314] [315]

To illustrate the RO-seg method for a specific case, open route segments are depicted  
 as blue lines in Fig. 4. Compared to the NE method, the RO-seg method is advantageous  
 for aviation applications because it only considers areas within flight routes. A histogram of  
 route segment lengths is presented in Fig. 5b. Most of the lengths are shorter than 500 km.  
 Also, the routes are not distributed uniformly across the CONUS as shown in Figs. 5c and  
 d. The routes are denser in the northeast and there are more east-west oriented routes than  
 north-south.  
 [316] [317] [318] [319] [320] [321] [322]

Contingency tables (Wilks 2011) of neighborhood and RO-seg methods are constructed  
 from the cumulative hits ( $a$ ), misses ( $b$ ), false alarms ( $c$ ), and correct rejections ( $d$ ) at each  
 forecast hour from all 24 cases. From the contingency tables, the Probability of Detection  
 (POD), False Alarm Ratio (FAR), Bias (BIAS), and Equitable Threat Score (ETS) are  
 calculated using the equations below:  
 [323] [324] [325] [326] [327]

$$\text{POD} = \frac{a}{a + b} \quad (5)$$

$$\text{FAR} = \frac{c}{a + c} \quad (6)$$

$$\text{BIAS} = \frac{a + c}{a + b} \quad (7)$$

$$\text{ETS} = \frac{a - c_h}{a + b + c - c_h} \quad (8)$$

where  $c_h$  is the number hits expected by chance and is calculated as  $c_h = \frac{(a + b) \times (a + c)}{a + b + c + d}$ .  
 BIAS indicates whether the forecast underestimates ( $<1.0$ ) or overestimates ( $>1.0$ ) areal  
 [328] [329] [330] [331] [332]

333 coverage with a perfect score of 1.0. The ETS measures the portion of observed and/or  
334 forecast events that were correctly predicted and is adjusted for hits associated with random  
335 chance. The ETS has a range of  $-\frac{1}{3}$  to 1 with a perfect score of 1 and negative values for an  
336 unskilled forecast. Previous studies (e.g., Hamill 1999) point out that comparisons of ETS  
337 from competing forecasts may be misleading if their biases are different. Thus, in some cases  
338 it is important to apply a bias adjustment to equalize the biases of the competing systems  
339 and obtain a more equitable comparison. Herein, a bias-adjustment is applied to the results  
340 presented in section 3.

341 *f. Determination of statistical significance*

342 The resampling methodology described by Hamill (1999) is applied to determine whether  
343 differences in ETS between the Sal CD forecasts and the other sets of forecasts are statisti-  
344 cally significant. For each set of comparisons at each forecast hour, resampling was repeated  
345 10000 times. For application to this study, the Hamill (1999) method involves computing  
346 a test statistic using the difference in ETS between Sal CD and the forecast to which it is  
347 compared. Then, a distribution of resampled test statistics is created by randomly choosing  
348 the Sal CD or other forecast for each case and then summing the contingency table elements  
349 over all cases. The location of the test statistic within the distribution of the resampled test  
350 statistics determines whether the differences are statistically significant.

351 **3. Results**

352 *a. Example case*

353 To illustrate qualitatively the typical performance characteristics of the various forecast-  
354 ing methods, a representative case with forecasts initialized 1800 UTC 8 June 2013 is pre-  
355 sented in Fig. 6. The synoptic weather regime associated with this case was characterized by

356 an amplifying mid-tropospheric short-wave trough that moved southeastward from northeast  
357 Wyoming to south-central Kansas during the 1200 to 0000 UTC period on 8 June. Ahead  
358 of this trough at 1900 UTC, a cold front stretched from south central Nebraska through  
359 western Kansas. As this cold front moved south and east into an increasingly unstable air  
360 mass, storms began to initiate at about 2000 UTC along the front. By 2100 UTC the storms  
361 had congealed into a line, which expanded while moving south and east. At 0200 UTC, the  
362 last forecast hour considered, a broken line of storms stretched from southwest Iowa, through  
363 eastern Kansas, into northwest Oklahoma and into the Texas panhandle. The storms at the  
364 southern end of the line in the Texas panhandle were the most intense.

365 The hourly forecasts and corresponding observations of 18 dBZ echo top heights for this  
366 case are shown in Fig. 6. According to the observations (OBS), a strong squall line with  
367 high echo top heights developed in central Kansas and moved to the east. The extrapolation  
368 (EXT) captures the movements of cells present at the starting time properly, but does  
369 not show the development of this strong storm cell. HRRR depicts the strong storm cell  
370 throughout the time clearly after 3 h. Lin CD captures the features from EXT and HRRR,  
371 however the intensities are underestimated compared to OBS. Sal CD shows the best results  
372 compared to earlier times of HRRR and later times of EXT and all of Lin CD. Sal CD  
373 captures the features from EXT and HRRR by showing the development and the movement  
374 of the center cell successfully. Compared to the results of Lin CD, Sal CD shows better  
375 results by keeping the intensities from HRRR and adding more information at 8 h.

376 *b. Statistical Evaluation*

377 POD, FAR, BIAS and ETS computed from contingency table elements defined using a  
378 20 km radius for EXT (a black line), HRRR (a blue line), Lin CD (a green line) and Sal  
379 CD (a red line) computed at each forecast hour over the 24 cases are shown in Fig. 7. The  
380 skill of EXT quickly drops with increasing forecast lead time and EXT performs better than  
381 HRRR until 3 h. HRRR skill scores drops slightly with increasing lead time, but in general

382 remain more constant than the other forecasts. Lin CD generally performs worse than EXT  
383 and HRRR from 3 to 6 h while Sal CD performs best overall with respect to POD and ETS.  
384 It should be noted that Lin CD has very low BIAS ( $< 0.3$ ) especially from 4 to 6 h (i.e.,  
385 underestimation). Sal CD performs particularly well relative to the other forecasts during  
386 the 2 to 5 h lead times. The largest differences in ETS and POD between Sal CD and the  
387 other forecasts coincides within the time that ETS and POD from EXT and HRRR “cross”,  
388 which indicates that, instead of utilizing EXT and HRRR individually, combining those data  
389 can improve the forecast.

390 Skill scores of POD, FAR, BIAS and ETS using the RO-seg method averaged over the  
391 24 days are shown in Fig. 8. The skill scores are similar to that of the NE method, however,  
392 ETS of Sal CD converges to that of HRRR at 4 h (it converged at 5 h for the NE method).  
393 ETS of HRRR using the RO-seg method shows better results than the NE method from  
394 3 to 8 h, which is likely related to the RO-seg method considering a more restricted area  
395 compared to the NE method.

396 Because bias can impact comparisons of ETS by sometimes giving the forecast with a  
397 higher bias an artificially inflated score, a bias correction procedure is applied following  
398 similar methods to Jenkner et al. (2008) and Clark et al. (2011). The corrections are only  
399 applied to EXT, HRRR, and Sal CD. Lin CD is excluded from bias correction because at  
400 some forecast hours, especially the 3 to 6 hour range, biases were as low as 0.25 (Fig. 7c)  
401 and correcting for the bias would have resulted in a drastically different appearing forecast.  
402 Biases for Sal CD, EXT, and HRRR were all clustered around 1.0, thus, the bias correction  
403 only results in a minor adjustment to the forecasts that serves to make the ETS comparisons  
404 more equitable.

405 The bias correction is applied by finding the average bias of Sal CD, EXT, and HRRR at  
406 each forecast hour. Then, using the distribution of 18 dBZ echo top heights, a new threshold  
407 that gives the average bias is computed. The new thresholds are slightly different among the  
408 three sets of forecasts, but have the same areal coverage and, thus, the ETSs computed from

409 these new thresholds are not impacted by differences in bias. The bias corrected comparisons  
410 are shown in Fig. 9.

411 *c. Statistical significance*

412 Using the methodology by Hamill 1999, distributions of differences in resampled ETS  
413 at each forecast time are calculated and the range between the 2.5 and 97.5 percentiles of  
414 these distributions is used to illustrate statistically significant differences. Those ranges are  
415 represented as error bars in Fig. 10. If the compared forecasts are outside of the range of  
416 error bars, the improvement is significant.

417 In the comparisons between Sal CD and EXT (Figs. 10a and b), the Sal CD scores are  
418 significantly better than EXT at all forecast hours. In the Sal CD and HRRR comparisons  
419 (Figs. 10c and d), Sal CD has significantly better scores up until forecast hour 4 using the  
420 NE method and until forecast hour 3 using the RO-seg method, after which the scores begin  
421 to converge. Finally, for the Sal CD and Lin CD comparisons (Figs. 10e and f), Lin CD is  
422 significantly better at forecast hour 1 for both the NE and RO-seg methods, while Sal CD  
423 is significantly better at forecast hours 3 to 7 using both methods.

424 **4. Discussion and future work**

425 *a. Discussion*

426 A new technique to blend extrapolation and model forecasts was developed and evaluated  
427 using observations and forecasts over 24 days from mid-May to mid-June 2013. In general,  
428 blending techniques using weighted averaging apply constant weights for both extrapolation  
429 ( $w$ ) and model forecasts ( $1 - w$ ) at each forecast lead time. For example,  $w=1.0$  is applied  
430 to the extrapolation and  $1 - w=0.0$  is applied to the model forecast at the beginning of  
431 the forecast and  $w$  decreases gradually to  $w=0.0$  at the end of the forecast. The weighted

averaging (“Linear cross-dissolve”) has a problem producing underestimated blended values during the middle of the forecast, where both  $w$  and  $1 - w$  are close to 0.5 if the forecasts are displaced. To mitigate this problem, the model forecast and extrapolation fields can be aligned before weights are applied, however displacements remain even after this alignment.

In order to further improve the blending results, a technique called “Salient cross-dissolve” is applied in this work. Two-dimensional weights ( $w_s$ ) based on the differences between normalized intensities from the extrapolation and model forecast are determined for each forecast hour (as a function of time fraction;  $w$ ). The novelty of salient cross-dissolve is preserving the values either in the extrapolation and model forecasts if they are high enough. For example, if there are two convective cells in both extrapolation and model forecasts in the middle of forecast lead time, salient cross-dissolve tends to shrink the cells by applying different weights (i.e., higher weights are applied to higher values and lower weights are applied to lower values which preserves most of the high-valued pixels and eliminates many of the low-valued pixels) while linear cross-dissolve cuts every value in half. Salient cross-dissolve showed better results than those of linear cross-dissolve in this study. Instead of “fading out” in linear cross-dissolve,  $w_s$  enables the pixels with strong intensities to be preserved in salient cross-dissolve resulting in more pixels with higher values.

For the forecast evaluations, a new method called the route-based segments method, which considers airplane routes, was developed and tested with comparisons made to a neighborhood-based method. Both methods gave very similar results indicating superior performance for the forecasts using Salient cross-dissolve, particularly during forecast hours 2 - 5 h.

#### 454 b. Future work

The contribution of the work is adding additional information to the weights applied to the extrapolations and the model forecasts. Considering differences of normalized intensities showed promising results and helped give more realistic intensities in blended forecasts. In-

stead of adjusting values in model forecasts based on linear weights that vary as a function of time, salient cross-dissolve also considers intensities so that pixels with high values are retained. However, updated weights,  $w_s$ , do not reflect actual data from the extrapolation nor model forecasts. Future study should consider the processes of adjusting those weights considering the real data and past performance. Additionally, frequently updating extrapolation, in other words, adding latest observational information at 15 minute time intervals should be utilized. Finally, it is also possible to consider weights as two separate variables instead of using  $w_s$  and  $1 - w_s$ . The independent variables can be adjusted using machine learning. Updated techniques using such methods are planned for future applications.

## 467 Acknowledgments

468 Funding for the authors was provided by NOAA/National Weather Service (NWS) Next  
469 Generation Air Transportation System (NextGen) weather program to the NOAA/Office of  
470 Oceanic and Atmospheric Research under NOAA-OU Cooperative Agreement NA11OAR4320072,  
471 U.S. Department of Commerce. The techniques described in this paper have been imple-  
472 mented within the Warning Decision Support System – Integrated Information (WDSS–  
473 II Lakshmanan and Smith (2007)) as the tool w2morphtrack. We thank Eric P. Jacobsen for  
474 help proof reading, Jason Levit and Mark Miller for useful suggestions on the tools developed  
475 herein, and Steven Lack at the NOAA Aviation Weather Center for information on aviation  
476 forecasting needs of the FAA.

## REFERENCES

- 479 Brown, J. M., et al., 2011: Improvement and testing of WRF physics options for application  
 480 to Rapid Refresh and High Resolution Rapid Refresh. *15th Conference on Aviation, Range,*  
 481 *and Aerospace Meteorology.*
- 482 Chou, M.-D. and M. J. Suarez, 1994: An efficient thermal infrared radiation parameterization  
 483 for use in general circulation models. *NASA Tech. Memo, 104606 (3)*, 85.
- 484 Clark, A. J., W. A. Gallus Jr, and M. L. Weisman, 2010: Neighborhood-based verification  
 485 of precipitation forecasts from convection-allowing ncar wrf model simulations and the  
 486 operational nam. *Weather and Forecasting, 25 (5)*, 1495–1509.
- 487 Clark, A. J., et al., 2011: Probabilistic precipitation forecast skill as a function of ensemble  
 488 size and spatial scale in a convection-allowing ensemble. *Monthly Weather Review, 139 (5)*,  
 489 1410–1418.
- 490 Dixon, M. and G. Wiener, 1993: TITAN: Thunderstorm Identification, Tracking, Analysis,  
 491 and Nowcasting-A radar-based methodology. *Journal of Atmospheric and Oceanic Tech-*  
 492 *nology, 10 (6)*, 785–797.
- 493 Dupree, W., et al., 2009: The advanced storm prediction for aviation forecast demonstration.  
 494 *WMO Symposium on Nowcasting.*
- 495 Fahey, T. H., T. Phaneuf, W. S. Leber, M. Huberdeau, D. P. Morin, and D. Sims, 2006:  
 496 Forecasts of convection for air traffic management strategic decision making: Compari-  
 497 son of user needs and available products. *12th Conf. on Aviation, Range, and Aerospace*  
 498 *Meteorology.*

- 499 Germann, U. and I. Zawadzki, 2002: Scale-dependence of the predictability of precipitation  
500 from continental radar images. Part I: Description of the methodology. *Monthly Weather  
501 Review*, **130** (12), 2859–2873.
- 502 Germann, U. and I. Zawadzki, 2004: Scale dependence of the predictability of precipitation  
503 from continental radar images. Part II: Probability forecasts. *Journal of Applied Meteorol-  
504 ogy*, **43** (1), 74–89.
- 505 Grundland, M., R. Vohra, G. P. Williams, and N. A. Dodgson, 2006: Cross dissolve without  
506 cross fade: Preserving contrast, color and salience in image compositing. *EUROGRAPH-  
507 ICS Annual Conference*, 577–586.
- 508 Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts.  
509 *Weather and Forecasting*, **14** (2), 155–167.
- 510 Huang, X.-Y. and P. Lynch, 1993: Diabatic digital-filtering initialization: Application to the  
511 HIRLAM model. *Monthly weather review*, **121** (2), 589–603.
- 512 Jenkner, J., C. Frei, and C. Schwierz, 2008: Quantile-based short-range qpf evaluation over  
513 switzerland. *Meteorologische Zeitschrift*, **17** (6), 827–848.
- 514 Lakshmanan, V., K. Hondl, C. K. Potvin, and D. Preignitz, 2013: An improved method for  
515 estimating radar echo-top height. *Weather and Forecasting*, **28** (2), 481–488.
- 516 Lakshmanan, V., K. Hondl, and R. Rabin, 2009: An efficient, general-purpose technique for  
517 identifying storm cells in geospatial images. *Journal of Atmospheric and Oceanic Technol-  
518 ogy*, **26** (3), 523–537.
- 519 Lakshmanan, V., R. Rabin, and V. DeBrunner, 2003: Multiscale storm identification and  
520 forecast. *J. Atmos. Res.*, **67-68**, 367–380.
- 521 Lakshmanan, V. and T. Smith, 2007: The warning decision support system - integrated  
522 information. *Wea. Forecasting*, **22**, 596–612.

- 523 Lakshmanan, V. and T. Smith, 2010: An objective method of evaluating and devising storm-  
524 tracking algorithms. *Weather and Forecasting*, **25** (2), 701–709.
- 525 Lakshmanan, V., T. Smith, K. Hondl, G. J. Stumpf, and A. Witt, 2006: A real-time,  
526 three-dimensional, rapidly updating, heterogeneous radar merger technique for reflectivity,  
527 velocity, and derived products. *Weather and Forecasting*, **21** (5), 802–823.
- 528 Li, L., W. Schmid, and J. Joss, 1995: Nowcasting of motion and growth of precipitation with  
529 radar over a complex orography. *Journal of applied meteorology*, **34** (6), 1286–1300.
- 530 Lin, C., S. Vasić, A. Kilambi, B. Turner, and I. Zawadzki, 2005: Precipitation forecast skill  
531 of numerical weather prediction models and radar nowcasts. *Geophysical research letters*,  
532 **32** (14).
- 533 Mandapaka, P. V., U. Germann, L. Panziera, and A. Hering, 2012: Can lagrangian extrapo-  
534 lation of radar fields be used for precipitation nowcasting over complex alpine orography?  
535 *Weather and Forecasting*, **27** (1), 28–49.
- 536 Matthews, M. and R. DeLaura, 2010: Assessment and interpretation of en route weather  
537 avoidance fields from the convective weather avoidance model. *10th AIAA Aviation Tech-*  
538 *nology, Integration, and Operations (ATIO) Conference, Fort Worth, TX.*
- 539 Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radia-  
540 tive transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for  
541 the longwave. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **102** (D14),  
542 16 663–16 682.
- 543 Nakanishi, M. and H. Niino, 2004: An improved Mellor-Yamada level-3 model with conden-  
544 sation physics: Its design and verification. *Boundary-layer meteorology*, **112** (1), 1–31.
- 545 Pinto, J., W. Dupree, S. Weygandt, M. Wolfson, S. Benjamin, and M. Steiner, 2010: Ad-

- 546 vances in the Consolidated Storm Prediction for Aviation (CoSPA). *14th Conference on*  
547 *Aviation, Range and Aerospace Meteorology.*
- 548 Robinson, M., W. Moser, and J. Evans, 2008: Measuring the utilization of available aviation  
549 system capacity in convective weather. *13th Conference on Aviation, Range, and Aerospace*  
550 *Meteorology.*
- 551 Sasaki, Y., 1958: An objective analysis based on the variational method. *J. Meteor. Soc.*  
552 *Japan*, **36 (3)**, 77–88.
- 553 Sasaki, Y., 1970: Some basic formalisms in numerical variational analysis. *Monthly Weather*  
554 *Review*, **98 (12)**, 875–883.
- 555 Sheth, K., T. Amis, S. Gutierrez-Nolasco, B. Spridhar, and D. Mulfinger, 2013: Develop-  
556 ment of a probabilistic convective weather forecast threshold parameter for flight-routing  
557 decisions. *Wea. Forecasting*, **28**, 1175–1187.
- 558 Skamarock, W., J. Klemp, J. Dudhia, D. Gill, and D. Barker, 2008: A description of the  
559 Advanced Research WRF Version 3. Tech. rep., NCAR Technical Note: NCAR/TN-475+  
560 STR.
- 561 Smirnova, T., J. Brown, and S. Benjamin, 1997: Evolution of soil moisture and temperature  
562 in the MAPS/RUC assimilation cycle. *Preprints, 13th Conf. on Hydrology, Long Beach,*  
563 *CA, Amer. Meteor. Soc*, 172–175.
- 564 Song, L., C. Wanke, D. Greenbaum, S. Zobell, and C. Jackson, 2008: Methodologies for  
565 estimating the impact of severe weather on airspace capacity. *Proc. 26th Int. Congress of*  
566 *the Aeronautical Sciences.*
- 567 Stratman, D. R., M. C. Coniglio, S. E. Koch, and M. Xue, 2013: Use of multiple verification  
568 methods to evaluate forecasts of convection from hot-and cold-start convection-allowing  
569 models. *Weather and Forecasting*, **28 (1)**, 119–138.

- 570 Sun, J., et al., 2014: Use of NWP for nowcasting convective precipitation: Recent progress  
571 and challenges. *Bulletin of the American Meteorological Society*, **95** (3), 409–426.
- 572 Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of  
573 winter precipitation using an improved bulk microphysics scheme. Part II: Implementation  
574 of a new snow parameterization. *Monthly Weather Review*, **136** (12), 5095–5115.
- 575 Weygandt, S. S., et al., 2011: Evaluation of the National Centers for Environmental Prediction  
576 (NCEP) implementation version of the Rapid Refresh and its skill in providing short-  
577 term guidance for aviation hazards. *15th Conference on Aviation, Range, and Aerospace  
578 Meteorology*.
- 579 Wilks, D. S., 2011: *Statistical methods in the atmospheric sciences*, Vol. 100. Access Online  
580 via Elsevier.
- 581 Wolfson, M. M., W. J. Dupree, R. M. Rasmussen, M. Steiner, S. G. Benjamin, and S. S.  
582 Weygandt, 2008: Consolidated storm prediction for aviation (CoSPA). *Integrated Com-  
583 munications, Navigation and Surveillance Conference, 2008. ICNS 2008*, IEEE, 1–19.

## 584 List of Figures

- 585 1 a) Linear cross-dissolve weights for extrapolation. Note that weights start at  
586 1 regardless of intensity at 0 h and steadily decrease to 0 at maximum forecast  
587 length. b) Linear cross-dissolve weights for HRRR model. Note that weights  
588 start at 0 regardless of intensity at 0 h and steadily increase to 1 at maximum  
589 forecast length. c) Salient cross-dissolve weights ( $w_s$ ) for extrapolation. Note  
590 that the weight of high intensity pixels (ranked saliency,  $r=y$  axis, close to  
591 1.0) remain high throughout the time period where low intensity pixels are  
592 dampened more quickly d) Salient cross-dissolve weights ( $w_s$ ) for model. The  
593 high intensity pixels ( $r$  is close to 0) remain high throughout the time period. 29
- 594 2 Illustration of the different forecast methods for an idealized sequence of fore-  
595 casts and observations at a line of discrete storm cells. (a) - (f) Observed 18  
596 dBZ echo top heights at forecast hour 0 - 6 respectively. (g) - (l), (m) - (r),  
597 (s) - (x), and (y) - (dd), same as (a) - (f) except idealized forecasts from ex-  
598 trapulation, HRRR, linear cross-dissolve and salient cross-dissolve, respectively. 30

599        3        a) Schematic of the neighborhood method. Filled green dots show locations of  
600                  observed 18 dBZ echo top height exceeding 9 km. The green circles of 20 km  
601                  radius (from filled green dots) determine hits if locations from 18 dBZ echo  
602                  top height exceed 9 km based on forecasts (extrapolation, the HRRR, linear  
603                  cross-dissolve, and salient cross dissolve) inside of green circles as shown as  
604                  filled red dots. On the contrary, unfilled dots depict missing locations either  
605                  in observation (unfilled green dots) or in forecast (unfilled red dots). Dotted  
606                  circles of 20 km radius from unfilled green and red dots are ranges where  
607                  there should be locations if 18 dBZ echo top height exceeding 9 km either  
608                  from observation or from forecast to be positively scored. b) Schematic of  
609                  the route-based segments method. The circles mean the same thing as those  
610                  in the neighborhood method, but only locations inside the route segments  
611                  (rectangular regions in black contour with width of 20 km) estimated using  
612                  two waypoints (filled black dots) are evaluated. Note that no locations with  
613                  18 dBZ echo top height exceeding 9 km inside the route segments mean that  
614                  segment is opened as indicated by the blue line on the bottom right panel.

31

- 615        4      a) Waypoints (37,736 points) in CONUS are shown as black dots. The red  
 616      contour shows the coverage of WSR-88Ds. b) Open routes are shown as blue  
 617      lines. The route segments are closed if an observed 18 dBZ echo top height  
 618      exceeded 9 km (OBS; yellow filled-contours, data from 2200 UTC 8 June 8  
 619      2013) within 10 km (or half the width) of the route. Note that only 20 %  
 620      of opened route segments are represented. c) Same as b) but 18 dBZ echo  
 621      top heights exceeding 9km are based on extrapolation (EXT; black filled-  
 622      contours). d) Same as b) but 18 dBZ echo top heights exceeding 9km are  
 623      based on model forecast (HRRR; blue filled-contours). e) Same as b) but  
 624      18 dBZ echo top heights exceeding 9km are based on linear cross-dissolve  
 625      (Lin CD; green filled-contours). f) Same as b) but 18 dBZ echo top heights  
 626      exceeding 9km are based on salient cross-dissolve (Sal CD; red filled-contours). 32
- 627        5      a) The route-based segments (20 % of 6,981) in CONUS within the WSR-  
 628      88D coverage. b) Histogram of normalized occurrence of the lengths of the  
 629      segments. c) Normalized concentration of route-based segments. d) Locations  
 630      of dense-routes are represented as red dots (normalized distribution  $\geq 0.5$ ). 33
- 631        6      Examples of observed echo top heights (OBS, (a)-(d)), extrapolations (EXT,  
 632      (e)-(h)), model forecast (HRRR, (i)-(l)), linear cross-dissolve (Lin CD, (m)-  
 633      (p)) and salient cross-dissolve (Sal CD, (q)-(t)) are shown at selected forecast  
 634      times of 2, 4, 6 and 8 h using the data on 8 June 2013. 34
- 635        7      a) POD computed using the 20 km radius neighborhood method from EXT  
 636      (black line), HRRR (blue line), linear cross-dissolve (Lin CD; green line) and  
 637      salient cross-dissolve (Sal CD; red line) for forecasts over 24 days from mid-  
 638      May to mid-June 2013. b) Same as (a) but FAR is shown. c) Same as (a) but  
 639      BIAS is shown. d) Same as (a) but ETS is shown. 35

- 640 8 As in Fig. 7, except using the 10 km route-based segments method. b) Same  
641 as (a) but FAR is shown. c) Same as (a) but BIAS is shown. d) Same as (a)  
642 but ETS is shown. 36

643 9 a) Adjusted BIAS computed using the 20 km radius neighborhood (NE)  
644 method of EXT (black line), HRRR (blue line), and Sal CD (red line) com-  
645 puted for 24 days from mid-May to mid-June 2013. b) Same as (a) but using  
646 the route-based segments (RO-seg) method. c) Bias-adjusted ETS using the  
647 NE method. d) Same as (c) but using the route-based segments (RO-seg)  
648 method. e) Percentiles of EXT, HRRR and Sal CD to correct BIAS based on  
649 areal coverage of 9 km OBS using the NE method. f) Same as (e) but using  
650 the route-based segments (RO-seg) method. 37

651 10 a) Bias-adjusted ETS computed using the 20 km radius neighborhood method  
652 of EXT (black line) and Sal CD (red line) with error bars of 2.5 and 97.5  
653 percentiles of ETS differences (10,000 resamples) ETS of Sal CD (reference)  
654 from that of EXT. b) Same as (a) but differences from EXT using the route-  
655 based segments method. c) Same as (a) but differences from HRRR. d) Same  
656 as (b) but differences from HRRR. e) Same as (a) but differences of ETS (not-  
657 bias-adjusted ETS) computed from Lin CD. f) Same as (e) but differences from  
658 Lin CD using the route-based segments method. 38

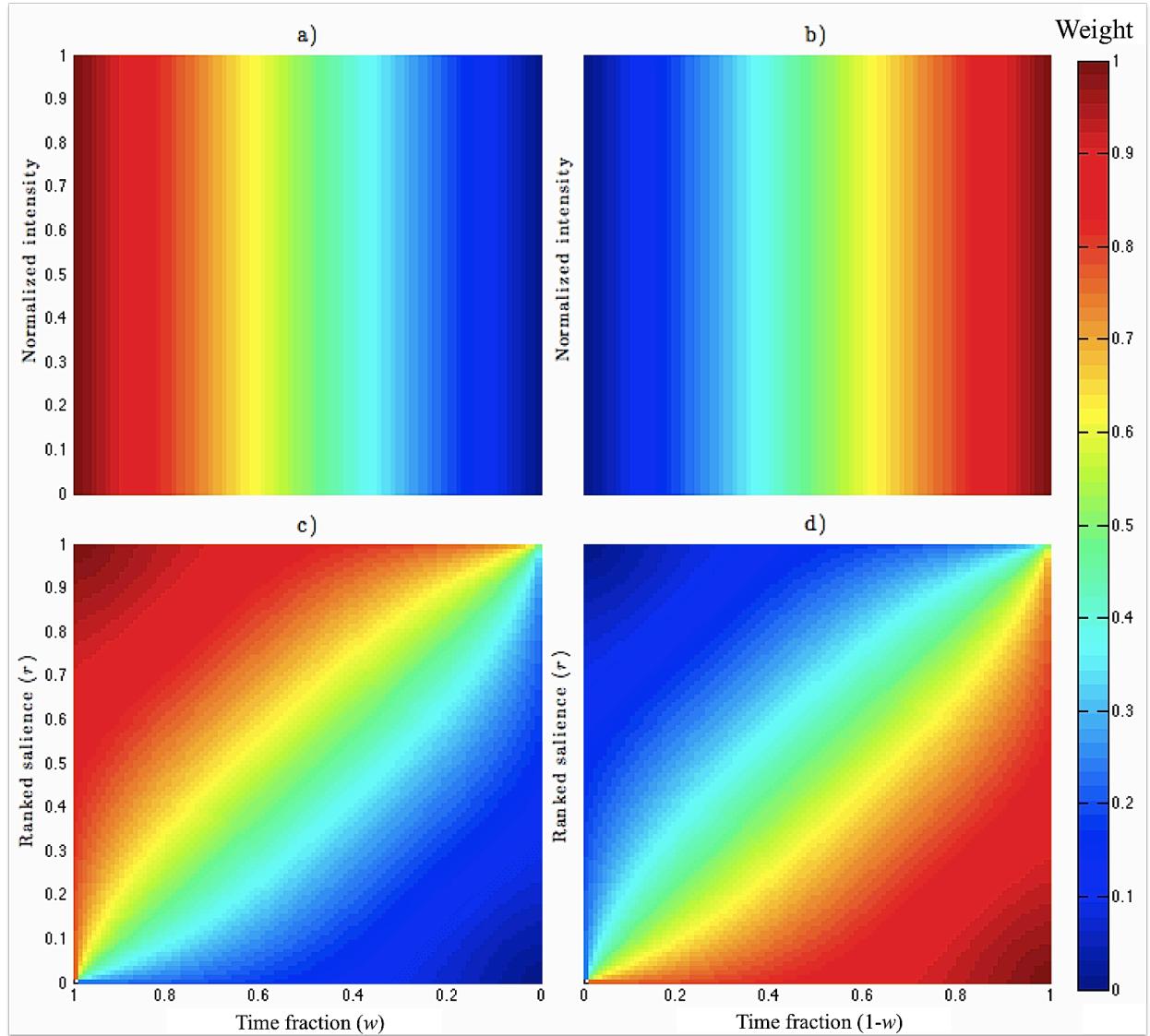


FIG. 1. a) Linear cross-dissolve weights for extrapolation. Note that weights start at 1 regardless of intensity at 0 h and steadily decrease to 0 at maximum forecast length. b) Linear cross-dissolve weights for HRRR model. Note that weights start at 0 regardless of intensity at 0 h and steadily increase to 1 at maximum forecast length. c) Salient cross-dissolve weights ( $w_s$ ) for extrapolation. Note that the weight of high intensity pixels (ranked saliency,  $r=y$  axis, close to 1.0) remain high throughout the time period where low intensity pixels are dampened more quickly d) Salient cross-dissolve weights ( $w_s$ ) for model. The high intensity pixels ( $r$  is close to 0) remain high throughout the time period.

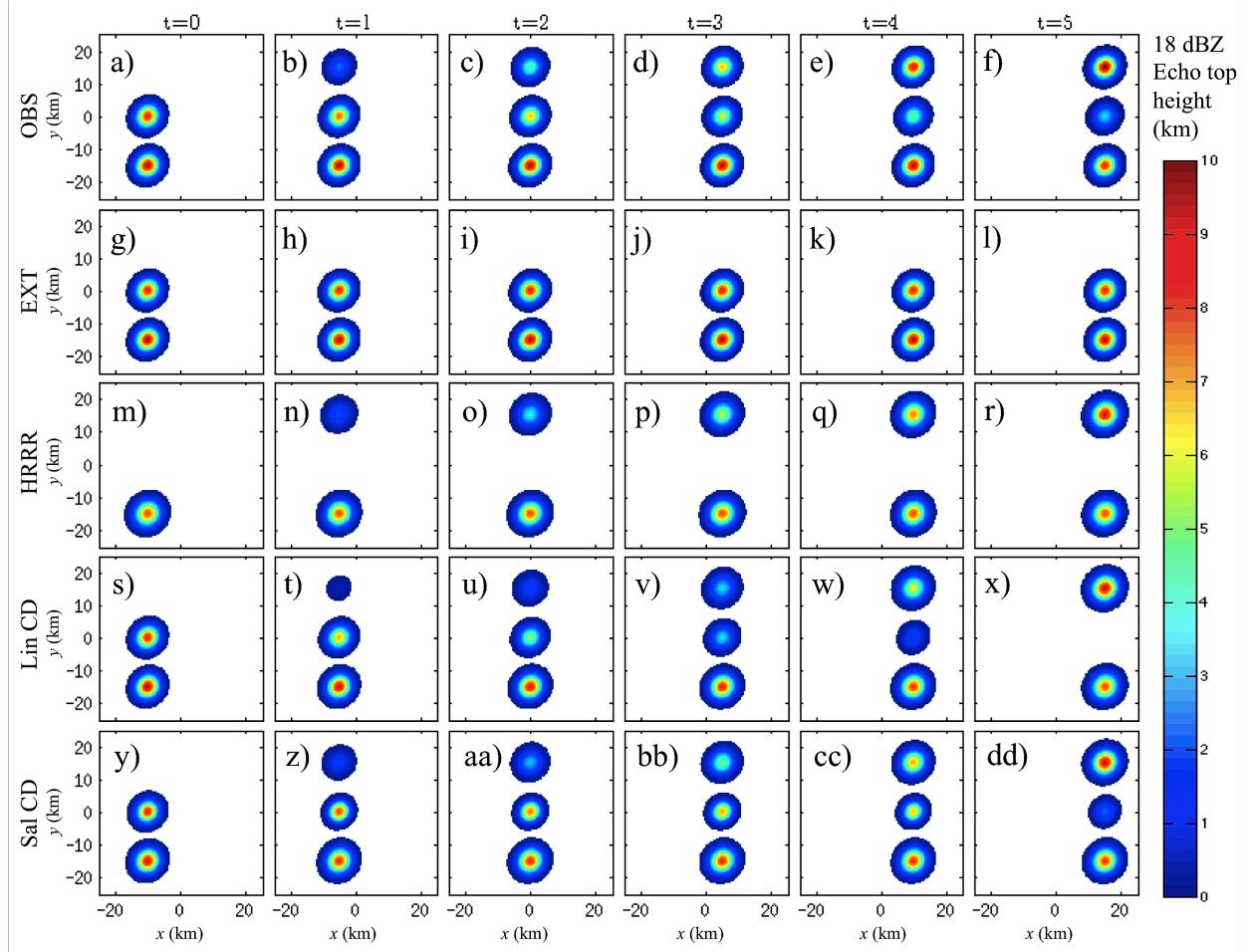


FIG. 2. Illustration of the different forecast methods for an idealized sequence of forecasts and observations at a line of discrete storm cells. (a) - (f) Observed 18 dBZ echo top heights at forecast hour 0 - 6 respectively. (g) - (l), (m) - (r), (s) - (x), and (y) - (dd), same as (a) - (f) except idealized forecasts from extrapolation, HRRR, linear cross-dissolve and salient cross-dissolve, respectively.

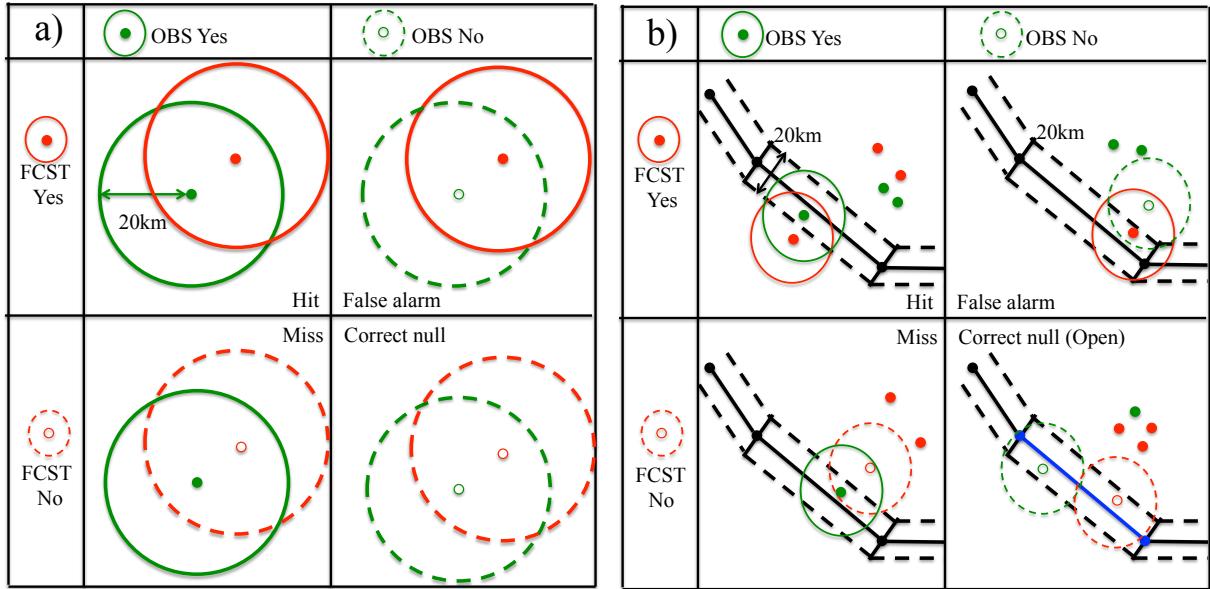


FIG. 3. a) Schematic of the neighborhood method. Filled green dots show locations of observed 18 dBZ echo top height exceeding 9 km. The green circles of 20 km radius (from filled green dots) determine hits if locations from 18 dBZ echo top height exceed 9 km based on forecasts (extrapolation, the HRRR, linear cross-dissolve, and salient cross dissolve) inside of green circles as shown as filled red dots. On the contrary, unfilled dots depict missing locations either in observation (unfilled green dots) or in forecast (unfilled red dots). Dotted circles of 20 km radius from unfilled green and red dots are ranges where there should be locations if 18 dBZ echo top height exceeding 9 km either from observation or from forecast to be positively scored. b) Schematic of the route-based segments method. The circles mean the same thing as those in the neighborhood method, but only locations inside the route segments (rectangular regions in black contour with width of 20 km) estimated using two waypoints (filled black dots) are evaluated. Note that no locations with 18 dBZ echo top height exceeding 9 km inside the route segments mean that segment is opened as indicated by the blue line on the bottom right panel.

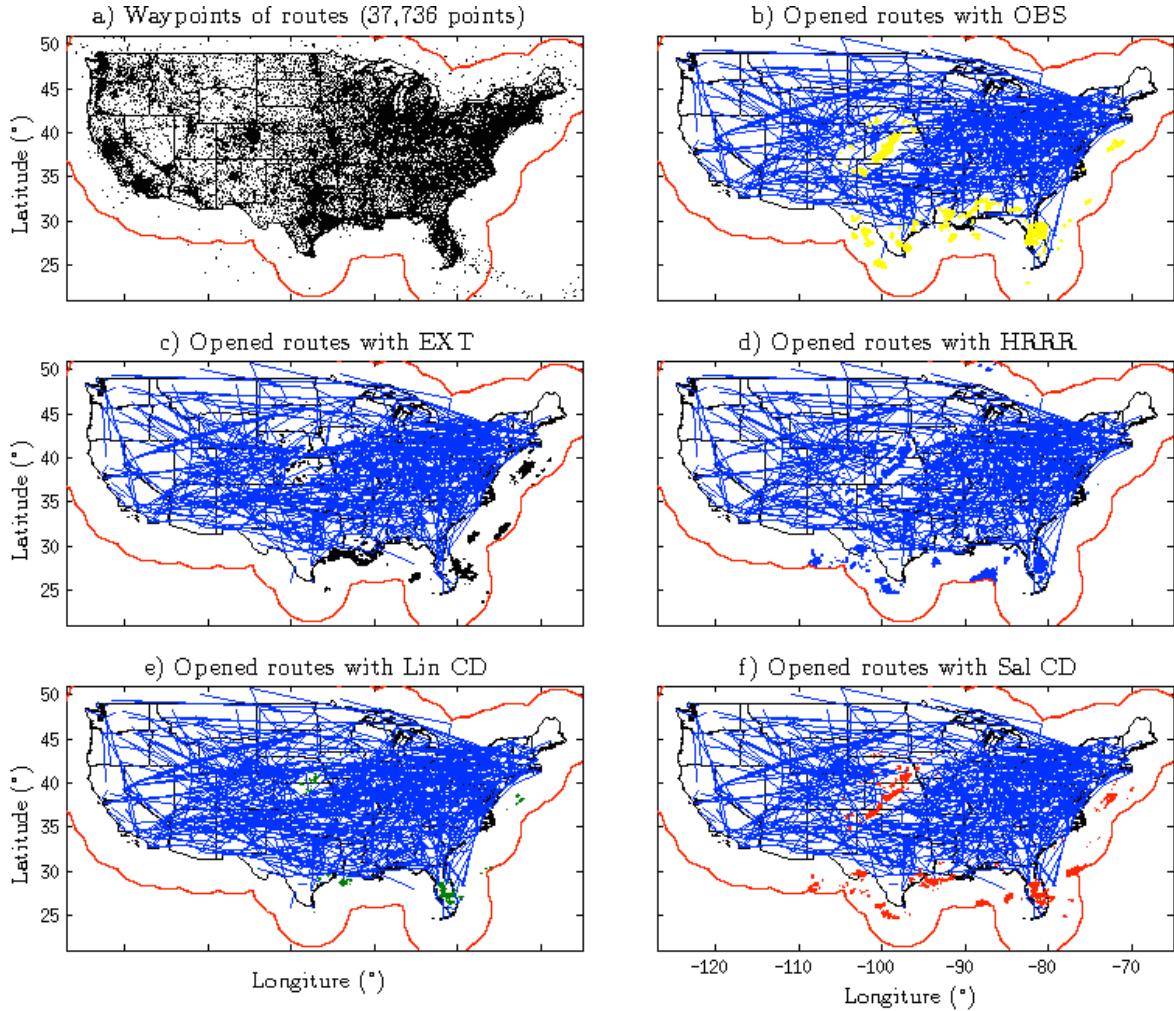


FIG. 4. a) Waypoints (37,736 points) in CONUS are shown as black dots. The red contour shows the coverage of WSR-88Ds. b) Open routes are shown as blue lines. The route segments are closed if an observed 18 dBZ echo top height exceeded 9 km (OBS; yellow filled-contours, data from 2200 UTC 8 June 8 2013) within 10 km (or half the width) of the route. Note that only 20 % of opened route segments are represented. c) Same as b) but 18 dBZ echo top heights exceeding 9km are based on extrapolation (EXT; black filled-contours). d) Same as b) but 18 dBZ echo top heights exceeding 9km are based on model forecast (HRRR; blue filled-contours). e) Same as b) but 18 dBZ echo top heights exceeding 9km are based on linear cross-dissolve (Lin CD; green filled-contours). f) Same as b) but 18 dBZ echo top heights exceeding 9km are based on salient cross-dissolve (Sal CD; red filled-contours).

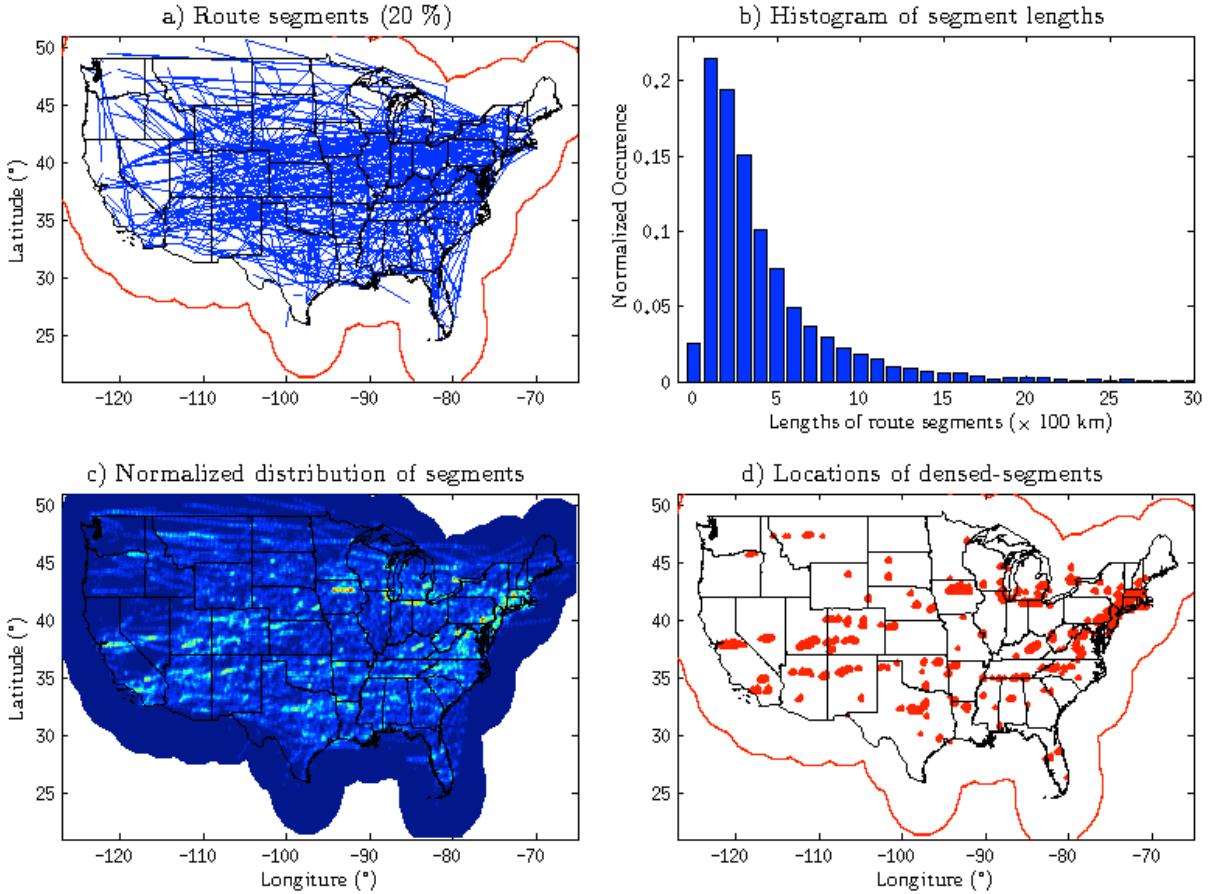


FIG. 5. a) The route-based segments (20 % of 6,981) in CONUS within the WSR-88D coverage. b) Histogram of normalized occurrence of the lengths of the segments. c) Normalized concentration of route-based segments. d) Locations of dense-routes are represented as red dots (normalized distribution  $\geq 0.5$ ).

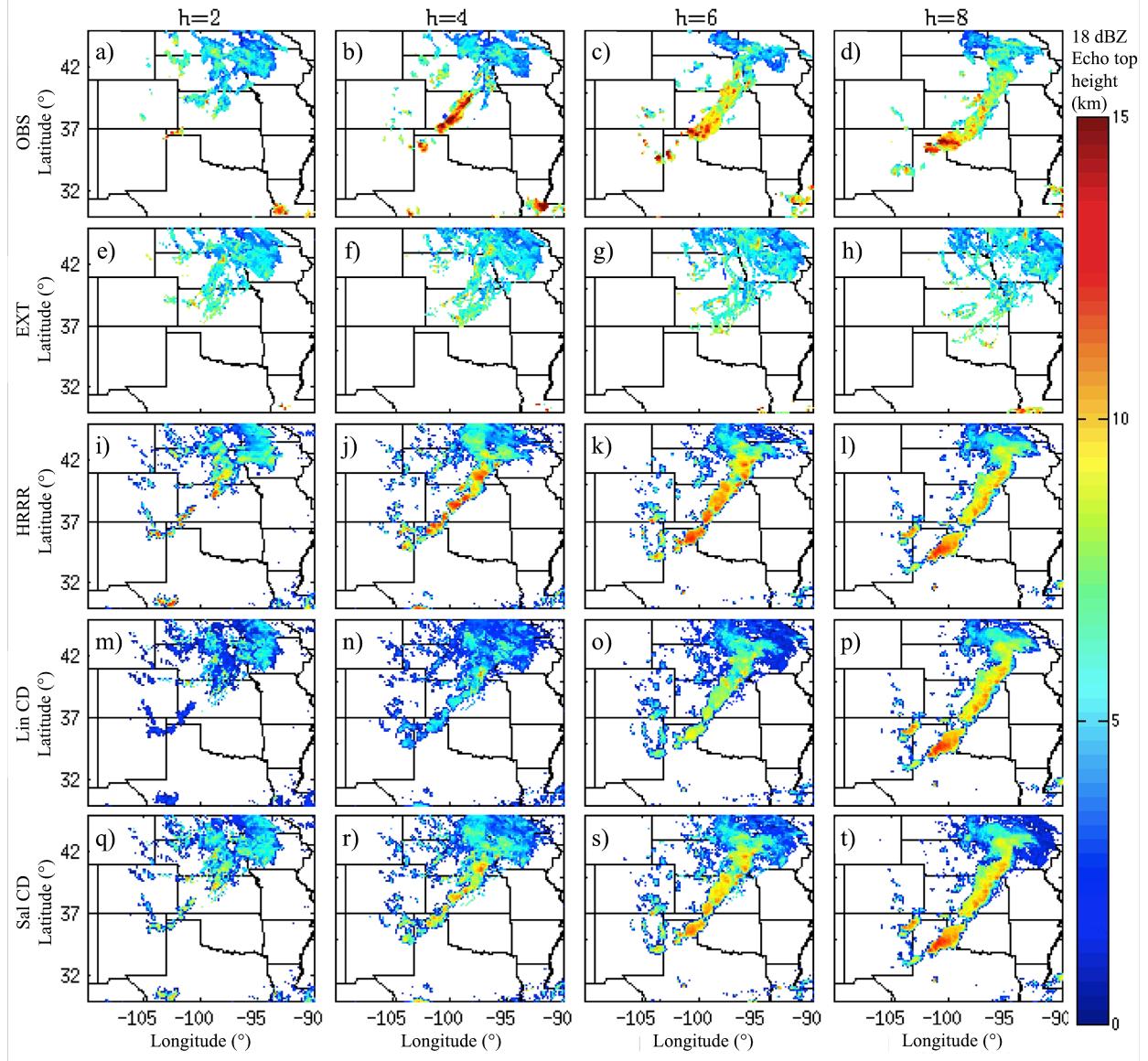


FIG. 6. Examples of observed echo top heights (OBS, (a)-(d)), extrapolations (EXT, (e)-(h)), model forecast (HRRR, (i)-(l)), linear cross-dissolve (Lin CD, (m)-(p)) and salient cross-dissolve (Sal CD, (q)-(t)) are shown at selected forecast times of 2, 4, 6 and 8 h using the data on 8 June 2013.

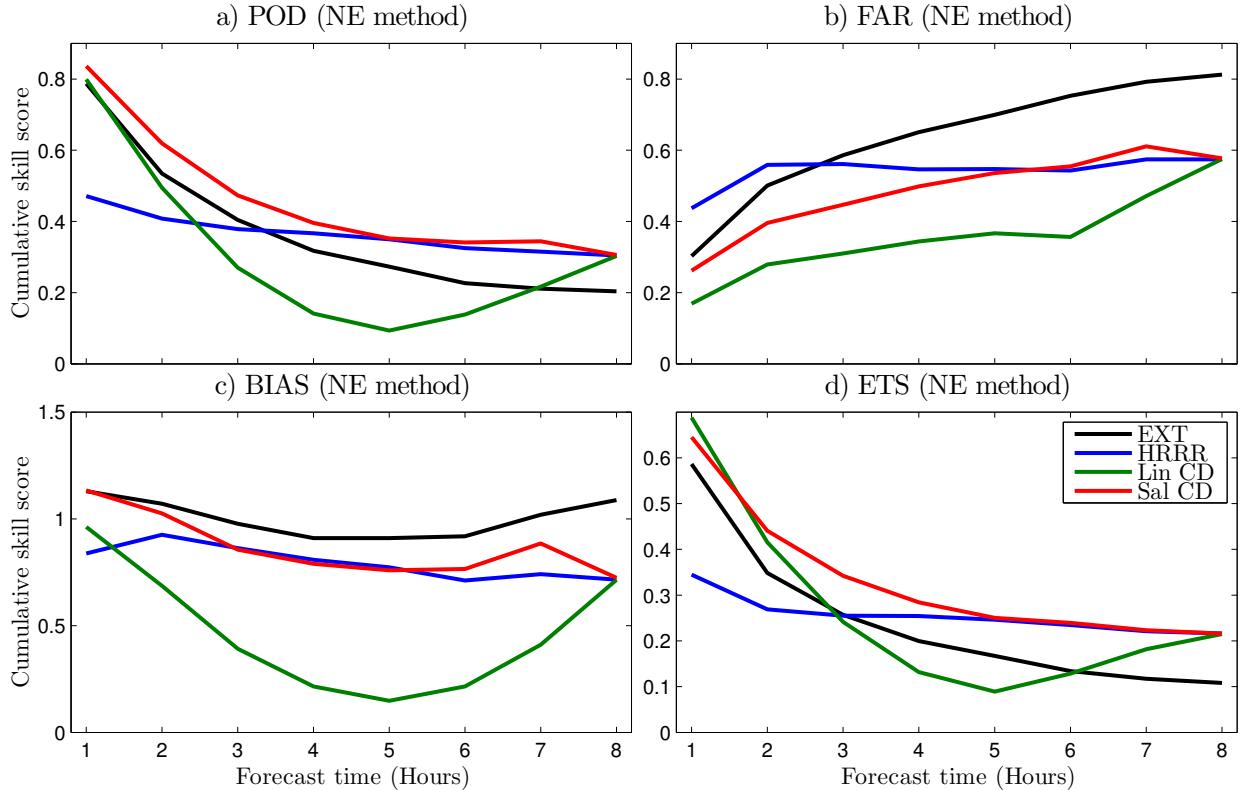


FIG. 7. a) POD computed using the 20 km radius neighborhood method from EXT (black line), HRRR (blue line), linear cross-dissolve (Lin CD; green line) and salient cross-dissolve (Sal CD; red line) for forecasts over 24 days from mid-May to mid-June 2013. b) Same as (a) but FAR is shown. c) Same as (a) but BIAS is shown. d) Same as (a) but ETS is shown.

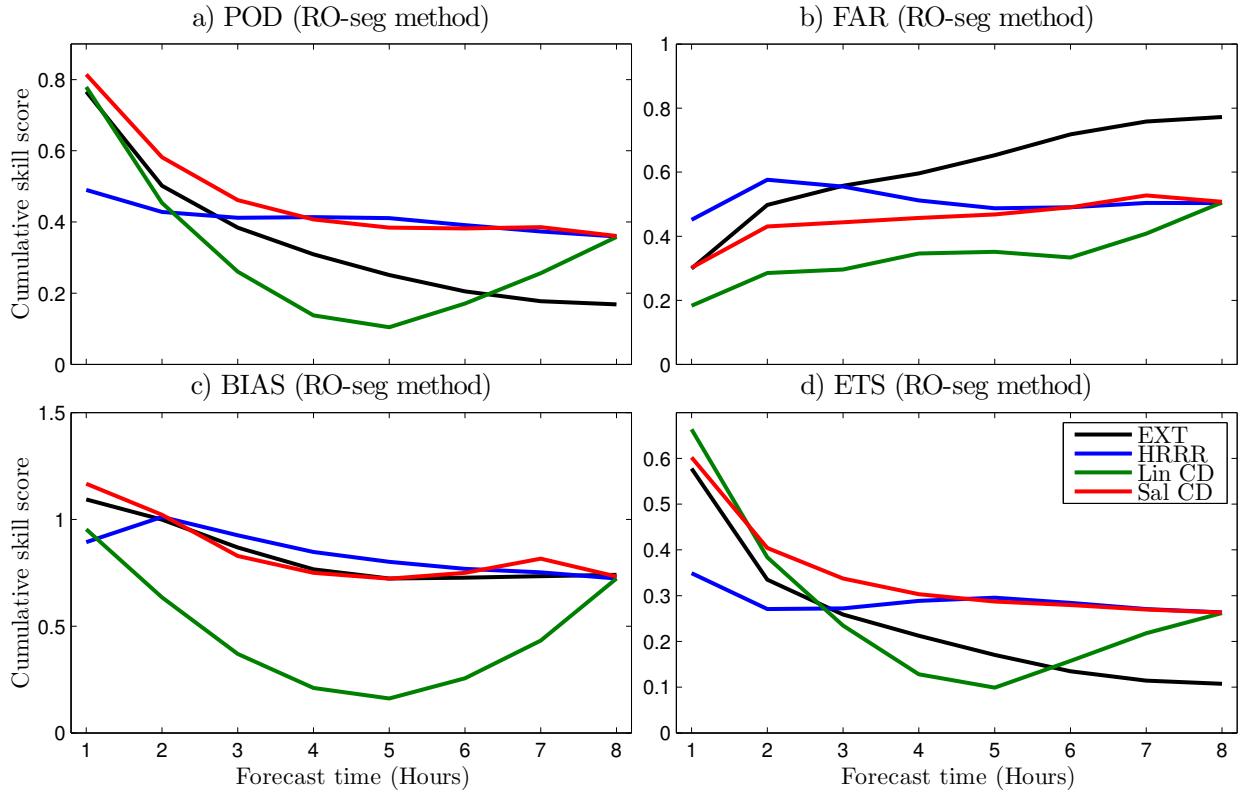


FIG. 8. As in Fig. 7, except using the 10 km route-based segments method. b) Same as (a) but FAR is shown. c) Same as (a) but BIAS is shown. d) Same as (a) but ETS is shown.

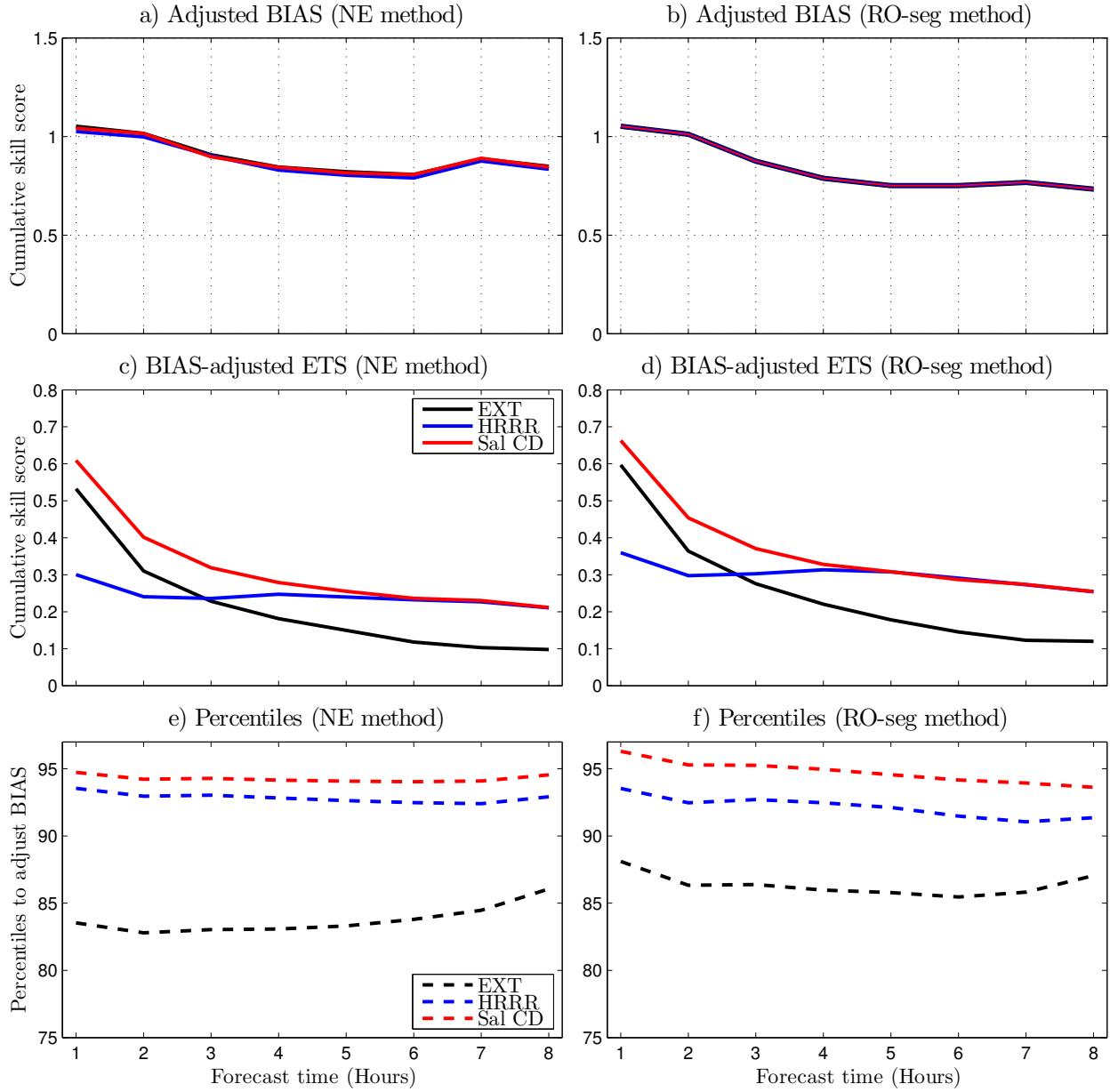


FIG. 9. a) Adjusted BIAS computed using the 20 km radius neighborhood (NE) method of EXT (black line), HRRR (blue line), and Sal CD (red line) computed for 24 days from mid-May to mid-June 2013. b) Same as (a) but using the route-based segments (RO-seg) method. c) Bias-adjusted ETS using the NE method. d) Same as (c) but using the route-based segments (RO-seg) method. e) Percentiles of EXT, HRRR and Sal CD to correct BIAS based on areal coverage of 9 km OBS using the NE method. f) Same as (e) but using the route-based segments (RO-seg) method.

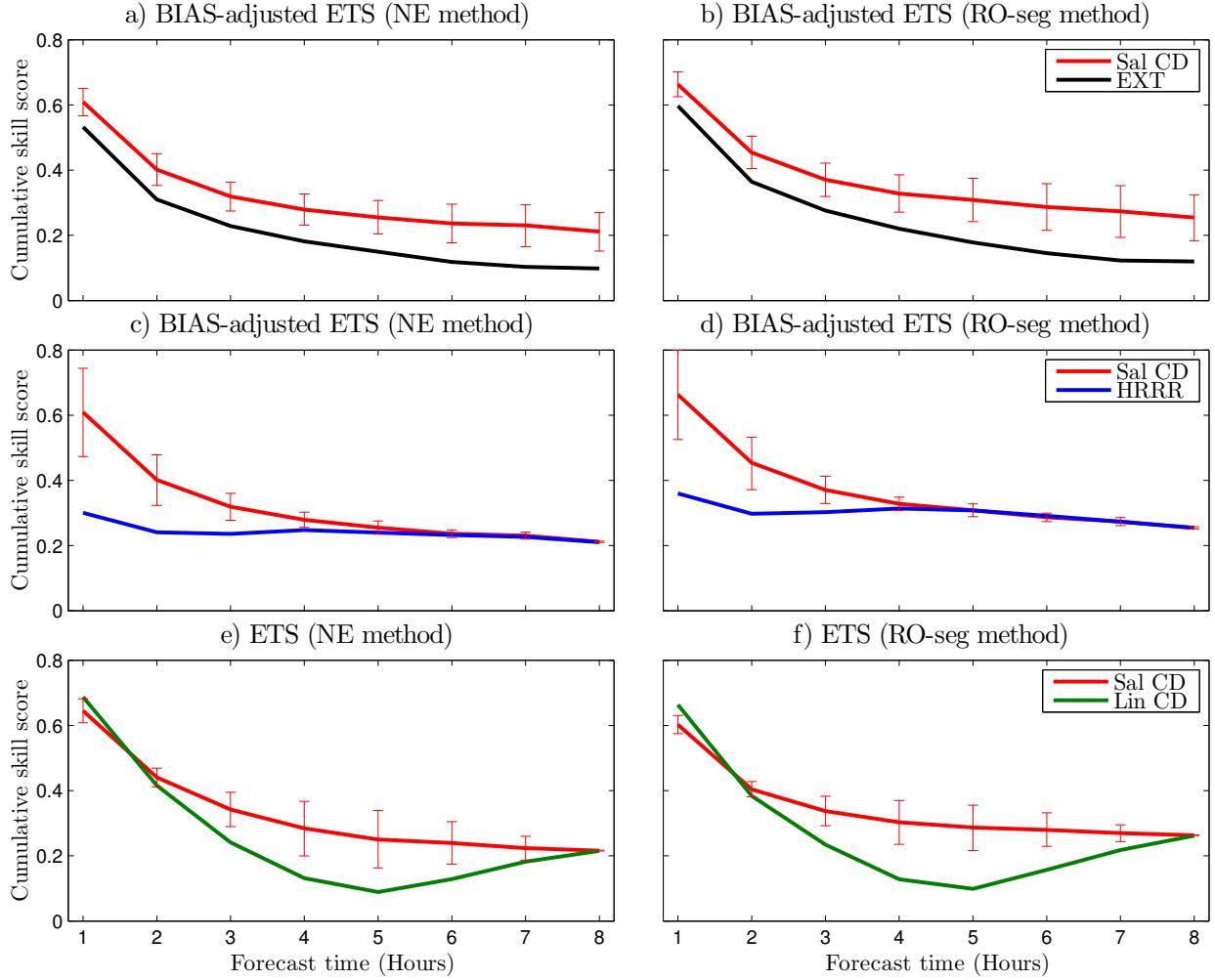


FIG. 10. a) Bias-adjusted ETS computed using the 20 km radius neighborhood method of EXT (black line) and Sal CD (red line) with error bars of 2.5 and 97.5 percentiles of ETS differences (10,000 resamples) ETS of Sal CD (reference) from that of EXT. b) Same as (a) but differences from EXT using the route-based segments method. c) Same as (a) but differences from HRRR. d) Same as (b) but differences from HRRR. e) Same as (a) but differences of ETS (not-bias-adjusted ETS) computed from Lin CD. f) Same as (e) but differences from Lin CD using the route-based segments method.