

More on the Scale Dependence of the Predictability of Precipitation Patterns: Extension to the 2009–13 CAPS Spring Experiment Ensemble Forecasts

MADALINA SURCEL, ISZTAR ZAWADZKI, AND M. K. YAU

McGill University, Montreal, Quebec, Canada

MING XUE

School of Meteorology and Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma

FANYOU KONG

Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma

(Manuscript received 19 September 2016, in final form 10 May 2017)

ABSTRACT

This paper analyzes the scale and case dependence of the predictability of precipitation in the Storm-Scale Ensemble Forecast (SSEF) system run by the Center for Analysis and Prediction of Storms (CAPS) during the NOAA Hazardous Weather Testbed Spring Experiments of 2008–13. The effect of different types of ensemble perturbation methodologies is quantified as a function of spatial scale. It is found that uncertainties in the large-scale initial and boundary conditions and in the model microphysical parameterization scheme can result in the loss of predictability at scales smaller than 200 km after 24 h. Also, these uncertainties account for most of the forecast error. Other types of ensemble perturbation methodologies were not found to be as important for the quantitative precipitation forecasts (QPFs). The case dependences of predictability and of the sensitivity to the ensemble perturbation methodology were also analyzed. Events were characterized in terms of the extent of the precipitation coverage and of the convective-adjustment time scale τ_c , an indicator of whether convection is in equilibrium with the large-scale forcing. It was found that events characterized by widespread precipitation and small τ_c values (representative of quasi-equilibrium convection) were usually more predictable than nonequilibrium cases. No significant statistical relationship was found between the relative role of different perturbation methodologies and precipitation coverage or τ_c .

1. Introduction

Predictability limitations of deterministic convection-allowing models, related to the rapid upscale growth of initial condition errors at small scales, as well as to model design, make it necessary to develop reliable probabilistic (ensemble) forecasting systems at these scales in order to account for forecast uncertainty. Significant effort has been devoted to understanding error growth at the kilometer scale and to developing appropriate techniques for sampling initial and lateral boundary condition (IC/LBC) uncertainties, as well as uncertainties in physical parameterizations (e.g., Zhang

et al. 2006; Walser et al. 2004; Romine et al. 2014). However, the best way of designing convection-allowing ensembles is still not established. While the current perturbation approaches (such as large-scale IC/LBC perturbations, convective-scale IC perturbations, mixed physics, stochastic perturbations in the physical schemes, etc.) can be used together to increase ensemble spread, achieving sufficient spread without negatively affecting the skill of perturbed members is still challenging (Clark et al. 2011; Vié et al. 2011; Romine et al. 2014; Schwartz et al. 2014). Moreover, the sensitivity to the perturbation method is both scale and case dependent (Johnson et al. 2014; Johnson and Wang 2016), thereby adding further complexity to the matter.

Surcel et al. (2015, hereafter SZY15) and Surcel et al. (2016, hereafter SZY16) characterized the scale and

Corresponding author: Madalina Surcel, madalina.surcel@mail.mcgill.ca

case dependences of precipitation predictability by a Storm-Scale Ensemble Forecast (SSEF) system with IC/LBC perturbations *and* variations in the model physical parameterization schemes (hereafter referred to as IC/LBC/PHYS perturbations). This ensemble was run by the Center for Analysis and Prediction of Storms (CAPS) as part of the 2008 NOAA Hazardous Weather Testbed (HWT) Spring Experiment. The results showed that this ensemble suffered a rapid loss of predictability of precipitation at meso- γ and meso- β scales, both in terms of the agreement among the ensemble members (increase in spread), and in terms of the comparison to observations (decrease in skill). However, while predictive skill at scales smaller than 100 km was lost during the first 12 h, the ensemble was still indicating some predictability at these scales (i.e., there was some resemblance between the ensemble members). It is unknown whether this inconsistency between skill and spread at small scales is related to the ensemble being underdispersive or to the ensemble being biased.

As mentioned in SZY15 and SZY16, these results were applicable to an ensemble with IC/LBC/PHYS perturbations, and they might differ for ensembles that use other perturbation methodologies. In fact, the sensitivity to the perturbation methodology has been reported upon in previous studies. For example, Stensrud et al. (2000) investigated the relative importance of IC and model physics (PHYS) perturbations for two cases. They showed that IC perturbations produced a more skillful ensemble for an event with strong large-scale forcing, while PHYS perturbations were more important for a weak-forcing event. On the other hand, Kong et al. (2014) found that IC/LBC perturbations derived from a regional-scale ensemble caused a much larger ensemble spread than the PHYS perturbations. They also noted that perturbing the LBCs is important in maintaining spread throughout the forecast time.

Vié et al. (2011) also compared the effects of uncertainties in convective-scale ICs to the effects of LBC uncertainty. For short lead times (less than 12 h), IC uncertainties had a dominant effect, but that effect was shown to be case dependent. For longer lead times, the LBC uncertainty was found to have the dominant effect. Romine et al. (2014) investigated the impact of adding LBC and stochastic PHYS perturbations to an ensemble that had ICs derived from a continuously cycled ensemble data assimilation system. While the additional perturbation methodologies resulted in increased spread, especially at longer forecast times, they also had a negative impact on the deterministic QPF skill of individual members. Johnson et al. (2014) analyzed the multiscale characteristics and evolution of different types of ensemble perturbations for precipitation

forecasts of the CAPS SSEF run during the spring of 2010. In particular, they investigated how small-scale IC perturbations compare to larger-scale IC/LBC perturbations obtained from the Short-Range Ensemble Forecasting (SREF) system operational at NCEP (Du et al. 2009) and to PHYS perturbations. They found that on average large-scale IC/LBC/PHYS perturbations have a dominant impact compared to the small-scale IC perturbations. However, for one case study characterized by convection organizing upscale into a mesoscale convective system, all perturbation methods generated differences with respect to the control precipitation forecast comparable to the forecast error. The sensitivity of precipitation predictability to the structure of IC perturbations was further analyzed by Johnson and Wang (2016) within the context of perfect model observing system simulation experiments (OSSEs). They found that probabilistic evaluation measures indicate an improvement in forecasting skill when multiscale IC perturbations produced by ensemble data assimilation are used rather than downscaling IC perturbations from coarser-resolution ensembles.

While the 2008 CAPS SSEF consisted of only eight IC/LBC/PHYS-perturbed members, more perturbation methodologies were used during 2009–13. These methodologies include PHYS perturbations, achieved through varying only the microphysical parameterization scheme (hereafter MP perturbations) or only the planetary boundary layer scheme (PBL perturbations), or by using a stochastic kinetic energy backscatter scheme (SKEB perturbations). In 2010, additional approaches to simulate small-scale IC uncertainty were attempted. Therefore, the ensemble precipitation forecasts produced during the 2008–13 Spring Experiments allow further study of the effects of various sources of uncertainty on the precipitation predictability by convection-allowing models.

The work presented here is an extension to the studies of SZY15 and SZY16. The scale and case dependences of precipitation predictability by an ensemble employing different perturbation methodologies is characterized. In addition, the factors affecting the relative importance of different sources of uncertainty are explored. The following issues are addressed through these analyses.

First, we investigate the influence of the perturbation method on the quantitative estimates of predictability limits. In previous predictability studies, IC uncertainty was accounted for in various ways: through time-lag forecasting techniques (Hohenegger et al. 2006; Walser et al. 2004), by adding correlated or uncorrelated random noise to the initial temperature or moisture fields (Zhang et al. 2002, 2006; Done et al. 2012), or by using an ensemble analysis system, such as the ensemble Kalman

filter (EnKF; Snook et al. 2012; Schwartz et al. 2014; Johnson and Wang 2016). While the results of such studies are qualitatively similar and in agreement with Lorenz's (1969) speculations, the actual quantitative estimates might differ. Here, we offer quantitative estimates for the loss of predictability with spatial scale and forecast lead time due to large-scale IC/BC uncertainty only (IC/LBC member), IC/LBC and PHYS uncertainty (IC/LBC/PHYS members), small-scale IC uncertainty achieved by adding uncorrelated or correlated noise to the initial temperature and humidity fields (hereafter RAND and RC members, respectively), and uncertainties in the representation of PHYS (achieved through MP, PBL, and SKEB perturbations).

Then, we analyze the relative importance of the various perturbations, and whether any of the perturbation methodologies used in the CAPS SSEF generate spread comparable to the forecast error. Furthermore, we explore whether the results are case dependent, as suggested by Johnson et al. (2014) and Stensrud et al. (2000).

This paper is organized as follows. Section 2 presents the ensemble forecasts and the verification data. Section 3 shows the effects of the different perturbations on the loss of precipitation predictability with scale and forecast time. Section 4 discusses the case dependences of predictability and of the sensitivity to different perturbation methodologies. Finally, section 5 offers a brief summary with conclusions.

2. Data description

a. CAPS SSEF ensemble forecasts

Forecasts from six Spring Experiments (2008–13) are analyzed here. The main setup of the CAPS SSEF system is similar from year to year. During April–June of each year, convection-allowing (4-km grid spacing) forecasts are initialized almost daily at 0000 UTC, and are run for at least 30 h over a domain covering most of the contiguous United States (CONUS).¹ Each year, the CAPS SSEF system uses the latest version of the Advanced Research version of the Weather Research and Forecasting Model (WRF-ARW; Skamarock et al. 2008). The ensemble consists of two control members (CN and C0), and a number of perturbed members. The background ICs and the control LBCs are provided by the North American Mesoscale Forecast System (NAM; Janjić 2003) 12-km analysis and forecasts. The control members CN and C0 have identical

model configurations, but CN has mesoscale data assimilation (including radar), performed using the ARPS 3DVAR system (Gao et al. 2004; Hu et al. 2006a,b; Xue et al. 2003), while C0 obtains the ICs directly from NAM. The perturbed members are configured exactly as for CN, except for their perturbations, as described later (see Table 2).

The IC/LBC perturbations are obtained directly from the SREF system operational at NCEP (Du et al. 2009). The SREF members are based on different dynamic cores (ETA, WRF-NMM, and WRF-ARW) and are run with a grid spacing of 32 or 45 km. Therefore, the IC/LBC perturbations do not have variability near the grid scale of the 4-km members. The IC perturbations are obtained as the difference between 3-h forecasts of SREF members and the SREF control. The perturbations of the u - and v -wind components, potential temperature, and specific humidity are rescaled to have root-mean-square values of 1 m s^{-1} , 0.5 K, and 0.02 g kg^{-1} , respectively.

The random (RAND) perturbations were added to the temperature and humidity fields and have standard deviations of 0.5 K and 5%, respectively. The correlated random (RC) perturbations have the same standard deviations, but correlation distances of 12 km in the horizontal and 3 km in the vertical.

Each year, different model physical parameterization schemes were varied between the members. In 2008 the microphysical (MP), planetary boundary layer (PBL), and shortwave radiation (SW) parameterization schemes were varied for the perturbed members, while the land surface model (LSM) and the longwave radiation (LW) scheme were kept the same, as for the control CN. In 2009, the LSM was also varied for the perturbed members. After 2010, the SW and LW schemes were kept identical between the members and PHYS variability was accounted for only through the MP, PBL, and LSM schemes. A more detailed description of the ensembles is given in Table 3.

b. Verification data

The precipitation forecasts were verified against NCEP's Stage IV multisensor precipitation product (Baldwin and Mitchell 1997). The Stage IV product consists of hourly rainfall accumulations over the CONUS, and was obtained from the NCEP website (www.emc.ncep.noaa.gov/mmb/ylin/pcpanl/). These hourly precipitation analyses are available on a 4-km grid and are developed at NCEP by mosaicking the regional hourly multisensor (radar and rain gauge) precipitation analyses produced by the NWS's 12 River Forecast Centers over the CONUS.

The entire verification was also performed for instantaneous reflectivity fields, with respect to radar

¹ Starting in 2010, the CAPS SSEF domain covered all of the CONUS. However, to keep the evaluation objective, the analysis domain has been kept the same from year to year.

TABLE 1. List of analyzed cases.

Year	No. of cases	Dates
2008	22	23, 24, 25 Apr 1, 2, 5, 6, 7, 8, 19, 20, 21, 22, 24, 27, 28, 29, 30 May 2, 4, 5, 6 Jun
2009	23	27 Apr 1, 2, 4, 5, 6, 7, 8, 11, 12, 18, 19, 20, 22, 26, 27, 28, 29 May 1, 2, 3, 4, 5 Jun
2010	34	29, 30 Apr 3, 4, 5, 6, 10, 11, 12, 13, 14, 17, 18, 19, 20, 21, 24, 25, 27, 28 May 1, 2, 3, 4, 7, 8, 9, 10, 11, 14, 15, 16, 17, 18 Jun
2011	29	2, 3, 4, 9, 11, 12, 13, 16, 17, 18, 19, 20, 22, 23, 24, 25, 26, 27, 28, 30, 31 May 1, 2, 3, 6, 7, 8, 9, 10 Jun
2012	32	26, 27 Apr 1, 2, 3, 4, 7, 8, 9, 10, 11, 14, 15, 16, 17, 18, 21, 22, 23, 24, 25, 27, 30, 31 May 1, 4, 5, 6, 7, 8, 9, 10 Jun
2013	29	30 Apr 1, 2, 3, 6, 7, 8, 9, 10, 13, 14, 15, 16, 18, 19, 20, 21, 22, 23, 24, 27, 28, 30, 31 May 3, 5, 6, 7, 12 Jun

reflectivity mosaics similar to those presented by SZY16. As the results were fully consistent for the different datasets, only the verification of forecasts of hourly precipitation accumulations against the Stage IV estimates is discussed herein.

c. Method

A total of 169 cases distributed over six Spring Experiments (2008–13) were analyzed. Table 1 enumerates the forecasts as labeled by starting dates. All runs were initialized at 0000 UTC each year. The data were remapped using a nearest-neighbor method onto the Stage IV product polar stereographic grid, which has a grid spacing of 4.7625 km at 60°N (Fig. 1). This method assigns to a pixel in the new grid the exact same value as the geographically closest pixel in the original grid [somewhat differing from the nearest-neighbor-average method of Accadia et al. (2003)].

3. The limits of the predictability of precipitation for the 2008–13 Spring Experiments

It was shown in SZY15 that ensemble perturbations result in errors that grow and propagate upscale throughout the forecast integration time, causing a total loss of the predictability of precipitation at increasingly larger scales with forecast lead time. The loss of predictability with spatial scale and forecast lead time due to IC/LBC/PHYS perturbations was quantified for the 2008 Spring Experiment in SZY15 in terms of the decorrelation scale. The decorrelation scale is defined for any two precipitation fields as the scale λ_0 below which these two fields are fully decorrelated. A complete decorrelation of two precipitation fields means that they

are as similar to each other as two random fields, and is interpreted as a complete lack of predictability (i.e., one of the fields cannot be used to predict the other).

To compute the decorrelation scale between any two precipitation fields, X and Y , we must first compute the power ratio as a function of spatial scale λ :

$$R(\lambda) = \frac{\text{Var}_X(\lambda) + \text{Var}_Y(\lambda)}{\text{Var}_{X+Y}(\lambda)}, \quad (1)$$

where $\text{Var}_X(\lambda)$ and $\text{Var}_Y(\lambda)$ represent the variances of the fields X and Y at scale λ and $\text{Var}_{X+Y}(\lambda)$ represents the variance of the field $X + Y$ at scale λ . If the value of this ratio is equal to 1, then the fields X and Y are

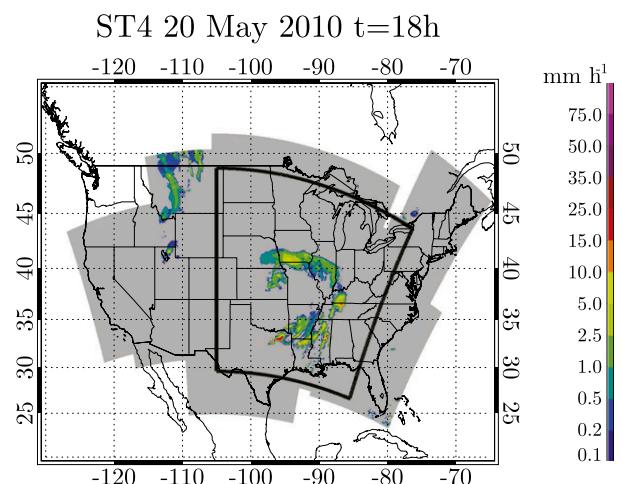


FIG. 1. Analysis domain. The gray shading represents the coverage of the Stage IV data, and the black contour indicates the analysis domain. All the data were remapped onto the analysis domain using nearest-neighbor interpolation.

decorrelated at scale λ , as the variance of a sum of two variables is equal to the sum of the variances only when the variables are not correlated.

To obtain the variance of a precipitation field at a given scale, we compute the Fourier power spectrum of the field. The values for $R(\lambda)$ vary between 1, which represents complete decorrelation between the fields, and $1/2$, which represents perfect resemblance between the fields. The largest scale at which $R(\lambda)$ has a value of 1 is λ_0 .

The decorrelation scale can only identify the scales at which there is a complete lack of predictability (or skill), without offering any measure of predictability at predictable scales. Therefore, to complement the analysis, we have also quantified the similarity between precipitation fields in terms of the normalized root-mean-square error (NRMSE; Surcel et al. 2015) and the fractions skill score (FSS; Roberts and Lean 2008).

The NRMSE is defined as

$$\text{NRMSE} = \frac{\sum_{i=1}^I \sum_{j=1}^J [X(i,j) - Y(i,j)]^2}{\sum_{i=1}^I \sum_{j=1}^J [X(i,j) + Y(i,j)]^2}, \quad (2)$$

where X and Y are two different precipitation fields of spatial dimensions I and J . Small values of NRMSE signify good agreement between the two fields, while the larger the values, the poorer the resemblance, with values larger than 1 meaning that two forecasts are as similar to each other as two random fields. To obtain information on predictability at different spatial scales, the NRMSE is computed for different scale components of the precipitation fields, rather than the original fields. In this case, X and Y are bandpass- or low-pass-filtered precipitation fields. In this paper, similarly to Johnson et al. (2014), a Haar wavelet transform is used to obtain different scale components of the precipitation fields. The normalization in the NRMSE is done to eliminate the dependence of the metric on the variance of the precipitation fields and, thus, makes the evaluation results less sensitive to forecast biases and to the changes in precipitation variance due to the diurnal cycle.

The FSS is a neighborhood verification method often used for the evaluation of high-resolution precipitation forecasts (e.g., Roberts and Lean 2008; Schwartz 2016; Dey et al. 2014). The FSS is computed as a function of precipitation threshold and spatial scale. Given the two precipitation fields X and Y mentioned above, a precipitation threshold p , and a neighborhood size r , two fraction fields are obtained corresponding to X and Y , f_X and f_Y , having the same dimensions as the initial precipitation fields. A fractional value for the grid point (i, j) in f_X is determined by the proportion of points

within r km of (i, j) in X with precipitation accumulations greater than p , and similarly for f_Y . Then, the FSS between X and Y is defined as

$$\text{FSS} = 1 - \frac{\sum_{i=1}^I \sum_{j=1}^J [f_X(i,j) - f_Y(i,j)]^2}{\sum_{i=1}^I \sum_{j=1}^J f_X^2(i,j) + \sum_{i=1}^I \sum_{j=1}^J f_Y^2(i,j)}. \quad (3)$$

In this paper, p is a percentile threshold (e.g., the highest 0.5% of precipitation amounts) and the absolute thresholds corresponding to p are determined independently for each separate precipitation field. The neighborhood around each grid point (i, j) is considered a square of size r km.

As in SZY15, to characterize the model predictability of the atmospheric state, values of λ_0 , NRMSE, and FSS are computed for Stage IV–member pairs of hourly rainfall accumulation fields for each of the cases listed in Table 1. Similarly, to characterize the predictability of the model state, λ_0 , NRMSE and FSS are computed for member–CN pairs of precipitation fields. The evolution of $\lambda_0(t)$ corresponding to different model configurations shows the effects of different perturbation methodologies.

Figure 2 shows $\lambda_0(t)$ for each member–CN pair and for Stage IV–CN, averaged over all the cases of each year, for the years 2008–13. The color coding indicates the differently perturbed members, as described in the legend. The number of members corresponding to each perturbation methodology changed from year to year. For example, in 2008, there were seven IC/LBC/PHYS members, and hence there are seven dark green lines in Fig. 2a.

For the IC/LBC and IC/LBC/PHYS members, $\lambda_0(t)$ with respect to CN increases with forecast lead time following a power law, reaching 200 km after 18 h (see the evolution of the green lines in Fig. 2). SZY15 showed that λ_0 increases faster with time for the IC/LBC/PHYS members relative to the IC/LBC members at the beginning of the forecast, whereas after the first 12 h they become similar for most cases. This is not evident in the average $\lambda_0(t)$ curves in Fig. 2a, but will be investigated further in the next section. We have also computed the average values of λ_0 for all Stage IV–member pairs and they were similar between all members, independent of their configuration (not shown). Furthermore, the value of λ_0 for member–CN becomes similar to the Stage IV–CN λ_0 only at the end of the forecast period (i.e., $t = 30$ h). The results relative to the IC/LBC/PHYS members are consistent for all years. Note that while the evolution of λ_0 is similar for all IC/LBC/PHYS members, in 2008, 2010, 2011, and 2013, there are some IC/LBC/PHYS members that show consistently lower λ_0 values with respect to CN relative

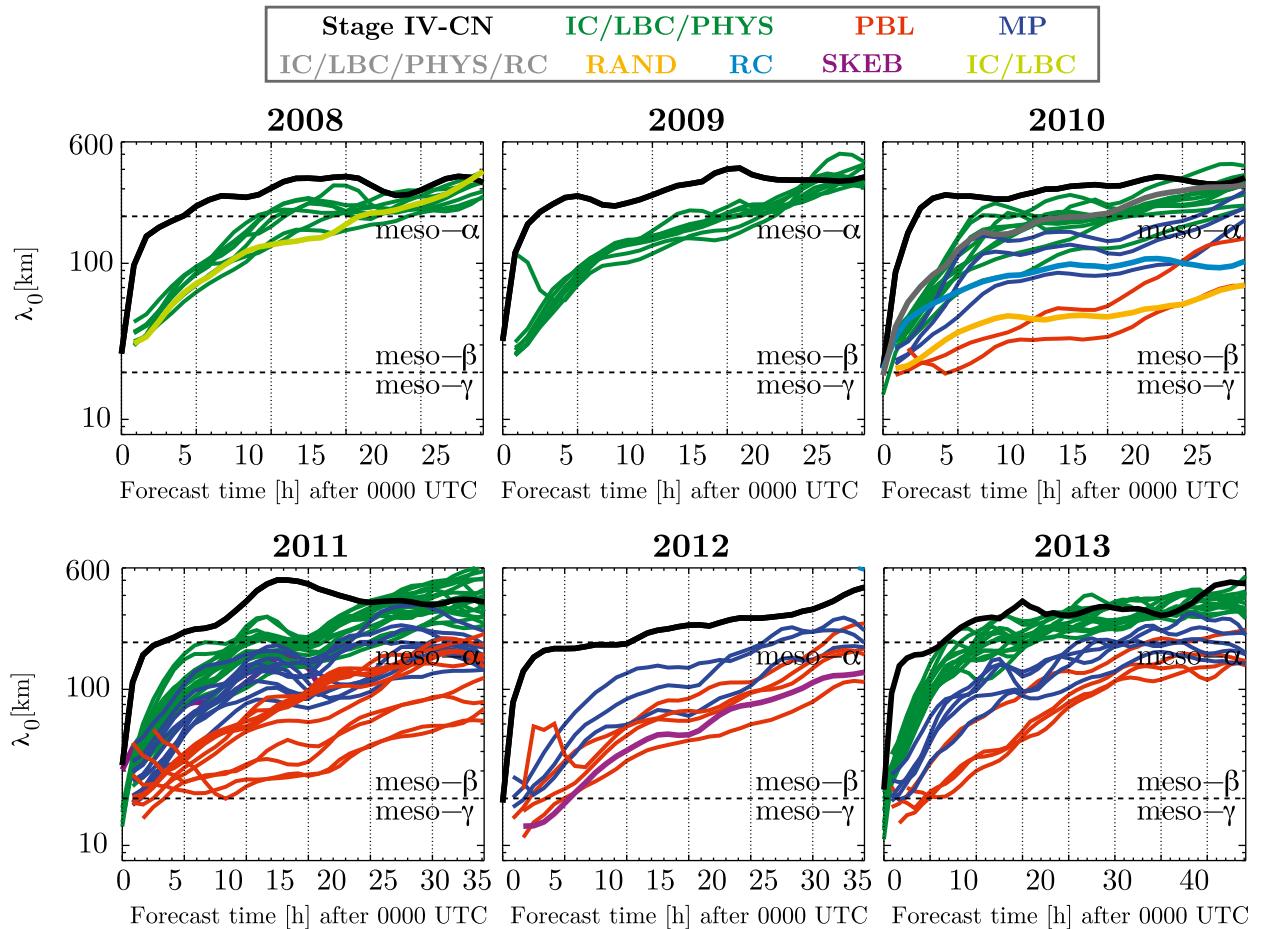


FIG. 2. The decorrelation scale λ_0 as a function of forecast time for members-CN averaged over all events of each year for the years 2008–13. The black line, denoted Stage IV-CN, corresponds to λ_0 between Stage IV estimates and CN, and the color lines correspond to member-CN. The different-colored lines correspond to different types of members, as indicated in the legend: IC/LBC/PHYS, members with both SREF-derived IC/LBC perturbations and mixed PHYS; PBL, members identical to CN except for the planetary boundary layer parameterization; MP, members identical to CN except for the microphysics parameterization; IC/LBC/PHYS/RC, member with SREF-derived IC/LBC perturbations, mixed PHYS, and random correlated noise added to the ICs (see text for details); RAND, random noise added to the ICs, otherwise identical to CN; RC, random correlated noise added to the ICs; IC/LBC, only SREF-derived IC/LBC perturbations; and SKEB, a member identical to CN but with SKEB perturbations.

to the rest. These members have the same MP parameterization and LSM as CN. Similarities between members that have the same MP schemes despite IC/LBC/PHYS perturbations have been reported previously by Johnson et al. (2011) for the 2009 CAPS SSEF system.

To allow for predictability studies, members with only IC perturbations were added in 2010: RAND and RC. The RAND member was configured identically to CN, but random noise was added to the initial temperature and humidity fields. As shown in Fig. 2, the values of λ_0 with respect to CN corresponding to this member are much lower on average than for the IC/LBC/PHYS members (orange line is lower than the dark green lines), reaching only 100 km at the end of the forecast

period. When some spatial structure is imposed on the RAND perturbations (by applying a recursive filter resulting in a decorrelation distance of 12 km in the horizontal and 3 km in the vertical), as for member RC, the values of λ_0 become much larger and comparable to those corresponding to the IC/LBC/PHYS members over the first 12 h (light blue line compared to dark green lines in Fig. 2). After 12 h, the IC/LBC/PHYS perturbations result in a larger decorrelation scale. This might be due to the type of IC perturbations (large scale, derived from a regional ensemble, versus small-scale, RC perturbations), or to the effect of LBC and PHYS perturbations. Previous studies (e.g., Vié et al. 2011) noted the importance of LBC perturbations for ensemble spread, especially for later forecast times.

TABLE 2. List of ensemble perturbations discussed in the text.

Type	Description	Years
IC/LBC	SREF-derived IC/LBC perturbations	2008
IC/LBC/PHYS	SREF-derived IC/LBC perturbations and mixed PHYS	2008–11, 2013
RAND	Uncorrelated random noise added to the initial moisture and temperature fields	2010
RC	Correlated random noise added to the initial moisture and temperature fields	2010
MP	Microphysical parameterization scheme different than for CN (Thompson)	2010–13
PBL	Planetary boundary layer scheme different than for CN (MYJ)	2010–13
SKEB	Stochastic kinetic energy backscatter scheme	2012

The rather constant λ_0 values corresponding to RC–CN after the first 12 h may be due to the lack of LBC perturbations in RC. These types of small-scale RC perturbations are usually used in predictability studies (e.g., Zhang et al. 2006; Hohenegger et al. 2006) to assess intrinsic predictability limits. If the growth of such perturbations can indeed provide reliable estimates of intrinsic predictability limits, then this limit is still lower than the practical predictability limit illustrated by the IC/LBC/PHYS members, and lower still than the actual forecast performance. Therefore, as noted by Durran and Gingrich (2014), there is still room for improvement before reaching the limit of intrinsic predictability. However, as Johnson et al. (2014) showed for this same dataset (using a different analysis method), there are some cases in which the error with respect to CN obtained from RC perturbations is the same as the error due to IC/LBC/PHYS perturbations and to the error of CN. This is not the case for the decorrelation scale with respect to CN, which after the first 2 h is always smaller for the RAND member than for the IC/LBC/PHYS members. At the beginning of the forecast period the RAND perturbations result in spurious precipitation, which also has a negative effect on QPF skill (Johnson et al. 2014). Therefore, adding RC perturbations to an IC/LBC/PHYS member can improve the spread with respect to CN in the first 6 h, but also has a negative effect in QPF skill at $t = 1$ h.

So far, we have discussed the loss of precipitation predictability due to IC/LBC/PHYS errors. After 2010, PHYS members were added to the ensemble. As mentioned in section 2, these members have the same ICs/LBCs and general configuration as CN, but they have differences in their physical parameterization schemes (see Table 2).

Figure 2 shows that the PBL members have the lowest values of λ_0 with respect to CN, and all PHYS variations lead to smaller decorrelation scales with respect to CN than for the IC/LBC/PHYS members. Therefore, it seems that the errors accounted for in the representation of precipitation microphysics generate more spread on average than the errors in the PBL scheme. However, note

that the evaluation is done for precipitation fields, which are directly impacted by the MP scheme. The errors in the PBL parameterizations might appear more significant if fields other than precipitation were analyzed. Also, the evolution of the member–CN $\lambda_0(t)$ with time is different between the MP members and the other members, showing a more rapid growth in the early forecast hours.

It seems from Fig. 2 that in 2011 there is larger variability in the decorrelation scale caused by PHYS variations. Some MP members have λ_0 values with respect to CN as large as the IC/LBC/PHYS members. However, a closer examination of the difference in predictability metrics between 2010 and 2011 reveals that this apparent difference is due to more PHYS members being available in 2011 than in 2010, and to the Ferrier member in particular. There is no significant difference in the member–CN λ_0 , NRMSE, or FSS for a given MP member from year to year.

Forecasts from a member that uses the SKEB scheme as a way of perturbing PHYS were also available for 2012. In terms of the evolution of $\lambda_0(t)$ with respect to CN, the SKEB member shows similar results to the PBL members (purple line). Romine et al. (2014) showed that SKEB schemes do increase ensemble spread when added to an IC ensemble, especially at later forecast times, but that they also caused forecast biases. Here, the Stage IV–member λ_0 corresponding to the SKEB member values are not systematically different than for the other members (not shown).

It was shown that independent of the perturbation method, the values of λ_0 for member–CN do not reach the values of λ_0 for Stage IV–CN for most forecast times. From the different perturbation methodologies employed in the CAPS SSEF, the SREF-derived IC/LBC/PHYS perturbations generate the largest λ_0 values on average. The effect of ensemble perturbations was also quantified in terms of the NRMSE and the FSS. The average NRMSE and FSS results computed for different scales show values consistent with the decorrelation scale. The next section further explores the sensitivity to the type of ensemble perturbations in terms of NRMSE and FSS, as well as the case dependence of this sensitivity.

4. The case-to-case variability of the predictability of precipitation

The case dependence of the predictability of precipitation was investigated in SZY16 for the 2008 CAPS SSEF. It was found that both ensemble spread and QPF skill were case dependent. The analyzed events for which the evolution of precipitation was controlled by the diurnal cycle of solar heating were associated with larger ensemble spread and poorer QPF skill than widespread events characterized by strong large-scale forcing. On the other hand, the evolution of $\lambda_0(t)$ for the ensemble did not exhibit such case dependence. In 2008, the CAPS SSEF consisted of members with IC/LBC perturbations, derived from the regional SREF system, and mixed PHYS, while other types of ensemble perturbations were included in the following years (as mentioned in section 2). Keil et al. (2014) showed that the spread of an ensemble with only varied PHYS depended on the weather regime, being much lower for strongly forced cases than for weakly forced cases. Conversely, they found that the predictability by an ensemble with only LBC perturbations did not exhibit such case dependence. Also, Stensrud et al. (2000) found that accounting for PHYS uncertainty was more important for weakly forced cases, while IC uncertainty had a dominant effect for strongly forced cases. Moreover, Johnson et al. (2014) investigated the effect of different types of IC perturbations. They found that in general IC perturbations derived from regional ensembles have the dominant impact for precipitation predictability. However, for one case characterized by weak large-scale forcing, all perturbation methodologies produced differences with respect to CN that were comparable to the forecast error.

Therefore, in this section we investigate two different aspects of the case-to-case variability of predictability. First, we determine whether the ensemble perturbations used in 2009–13 show the same case dependence reported in SZY16 for the 2008 ensemble, by analyzing the relationships between forecast skill or ensemble spread and event type. Second, we explore whether the relative effect of different types of perturbations is case dependent. For instance, are PHYS perturbations more important for some types of events than for others? Also, for which cases, if any, are the errors due to the ensemble perturbations of similar magnitude to the forecast error?

a. Case classification

In SZY16, we have attempted to discriminate between strongly and weakly forced events in terms of

synoptic-scale forcing. However, we found no statistical relationship between the strength of the quasigeostrophic forcing for ascent and the predictability of precipitation. On the other hand, we did find that the domain-averaged, convective adjustment time scale and the areal coverage of precipitation showed a relation to predictability. The convective adjustment time scale was proposed by Done et al. (2006) as a measure of differentiating objectively between equilibrium and nonequilibrium cases. Previous studies showed that when convection is in equilibrium with the large-scale flow, the properties of convective rainfall are set by the large-scale environment and are thus more predictable. On the other hand, for nonequilibrium cases, the development and evolution of convection are controlled by less predictable, small-scale, local factors (SZY16; Done et al. 2006; Keil et al. 2014). These studies have used the convective adjustment time scale as a loose indication of whether a precipitation event is characterized by strong or weak large-scale forcing. In reality, there can be events in which large areas of instability are consumed by convection that was initiated at small scales (such as thunderstorms initialized by cold pools), and then suffered upscale growth (coalescence of small storms into larger mesoscale systems). Because these events were triggered by small-scale features, it might be inappropriate to consider them associated with large-scale forcing (even though a beneficial large-scale environment might be necessary to allow for the upscale growth in the first place). As finding the best estimate and definition of the term *large-scale forcing* is outside the scope of the current paper, and to prevent confusion, hereafter we will avoid the use of this term. However, finding a relationship between predictability and the convective adjustment time scale would still be useful, as this convective adjustment time scale can be calculated a priori from the model output. If a relationship did exist, then we could have a measure of the expected performance and uncertainty of forecasts associated with equilibrium and nonequilibrium cases.

The convective-adjustment time scale τ_c is defined as the ratio between convective available potential energy (CAPE) and the rate of change of CAPE:

$$\tau_c \sim \frac{\text{CAPE}}{d(\text{CAPE})/dt}. \quad (4)$$

The rate of change of CAPE is controlled by the removal of CAPE by supplying latent heat, and can be calculated from the precipitation rate. Thus,

$$\tau_c = \frac{1}{2} \frac{\text{CAPE}}{P} \times 49.58 \text{ mm s}^3 \text{ m}^{-2} \text{ h}^{-1}, \quad (5)$$

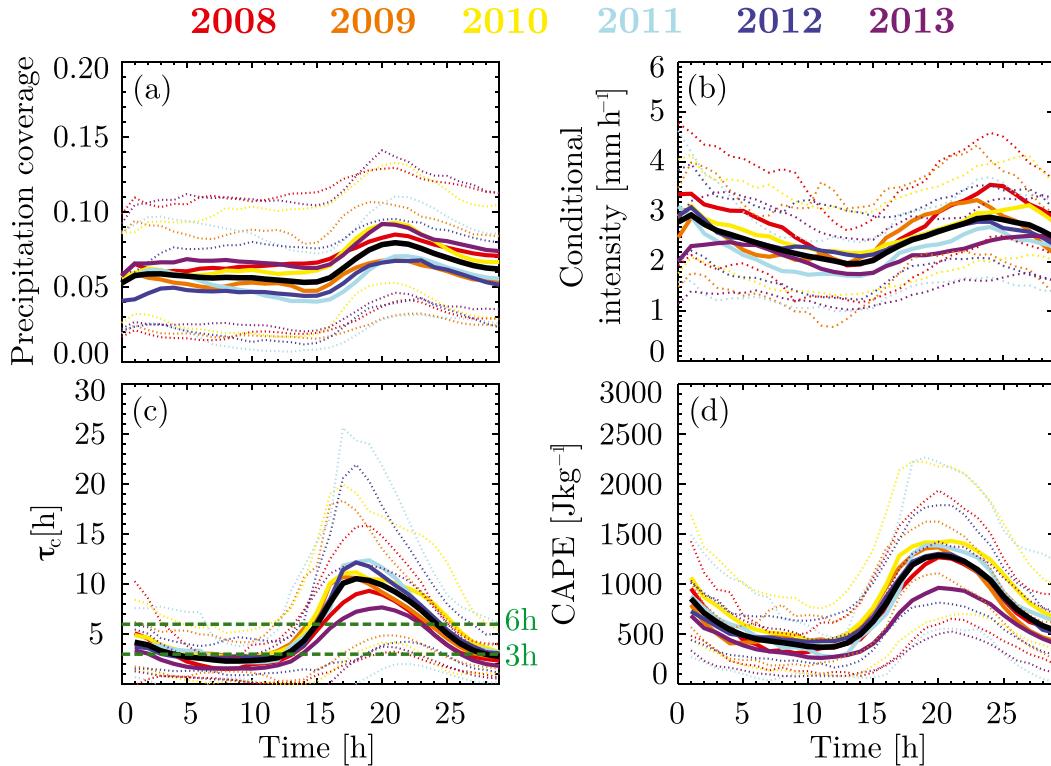


FIG. 3. (a) Precipitation coverage, (b) conditional intensity, (c) τ_c , and (d) CAPE corresponding to CN as a function of forecast lead time (forecasts initialized at 0000 UTC), for the years 2008–13. The solid colored lines represent averages for all cases of a given year according to the legend (red, 2008; orange, 2009; yellow, 2010; light blue, 2011; indigo, 2012; purple, 2013). The dotted lines represent $\pm\sigma$ for each year. The thick black line represents the average over all cases. Dashed green lines in (b) indicate the 3- and 6-h thresholds of τ_c used in previous studies to differentiate between quasi-equilibrium and nonequilibrium cases.

with units of seconds, and with the units of P and CAPE being millimeters per hour and joules per kilogram, respectively. Small values of τ_c indicate that CAPE is rapidly consumed by convection and is interpreted as convection being in equilibrium with the large scales. On the other hand, large values of τ_c are associated with nonequilibrium cases. Thresholds of 3 or 6 h were used in previous studies to differentiate between equilibrium and nonequilibrium cases (Zimmer et al. 2011; Keil et al. 2014).

Here, τ_c is calculated using the above formula as a function of forecast lead time from maps of hourly rainfall accumulations and most unstable CAPE (MUCAPE, hereafter referred to simply as CAPE) corresponding to the control member CN. As in Keil et al. (2014), the CAPE and hourly accumulations fields are smoothed using a Gaussian filter with a half-width of 60 km before computing τ_c , in order to eliminate noise in the calculations, but to keep the level of detail representative of scales at which convection occurs. For each forecast time, the spatial values of τ_c are averaged over

all the data pixels with rain rates higher than 1 mm h^{-1} to exclude dry areas over which τ_c cannot be computed.² While only the values corresponding to CN are used here for classification, note that all of the ensemble members showed very similar values of τ_c and precipitation coverage.

It was reported by SZY16 and confirmed by Dey et al. (2016) for a different forecasting system that a relationship exists between precipitation coverage and predictability. It is expected for cases with lower precipitation coverage to be inherently less predictable and more likely to suffer from displacement errors, which will result in poorer forecast performance. We will verify here whether

² This helps eliminate points over which τ_c values would be very large as a result of division by a very small number. The threshold of 1 mm h^{-1} is a subjective choice borrowed from Keil et al. (2014). Slightly modifying this threshold does not impact the results. However, using a threshold higher than 5 mm h^{-1} would cause much noisier τ_c values because of the small number of data points with precipitation values larger than 5 mm h^{-1} .

the relation between precipitation coverage and predictability is maintained for the entire dataset.

Figures 3a–d show, respectively, the temporal evolution of the mean and the standard deviation of the precipitation coverage (defined as the fraction of all data points with hourly accumulations of 0.2 mm h^{-1} or more), the conditional rainfall intensity (defined as the average precipitation intensity over raining areas), the domain-averaged τ_c , and the CAPE for all years. Each metric shows a diurnal maximum around 2000 UTC, confirming the importance of the diurnal cycle of solar heating for the evolution of convective precipitation over the CONUS (Surcel et al. 2010). In terms of precipitation coverage, the years 2009, 2011, and 2012 show both lower values and less variability (smaller standard deviations) on average than the other years. The years 2008 and 2013 show the smallest average values of τ_c , with 2013 being the year with the lowest values and the lowest variability of average CAPE as well. On the other hand, 2008 and 2013 differ in terms of average conditional intensity, with 2008 having the largest values of conditional intensity and 2013 the lowest overall totals. We note however that daily τ_c values are often larger than the threshold of 3 or 6 h used to differentiate between equilibrium and nonequilibrium cases in previous studies on the predictability of convective rainfall in Europe (Flack et al. 2016; Keil et al. 2014). Therefore, in this paper, we do not attempt a case classification based on a categorical threshold. Instead, we investigate whether a statistical relationship exists between spatially averaged τ_c values and precipitation coverage and predictability metrics.

b. The case dependence of predictability

In SZY16 it was shown that no clear relationship existed between λ_0 and precipitation coverage or τ_c . On the other hand, ensemble spread and the skill of the control member CN showed weather regime dependence. Therefore, in this paper, rather than investigating the case-to-case variability of λ_0 , we investigate the variability of the skill and spread.

The skill of the control member CN is quantified in terms of the NRMSE for scales larger than 128 km and FSS at the 100-km scale and for the 95th percentile threshold computed with respect to Stage IV precipitation accumulations. We have chosen to show the relationship between the event classifier and the skill and spread metrics at scales of roughly 100 km because, as Fig. 2 indicates, predictability is lost very rapidly at scales smaller than 100 km. Figure 4 illustrates the correlation between NRMSE (Figs. 4a,b) and FSS (Figs. 4c,d) and precipitation coverage and τ_c , respectively, as indicated in the panels' titles. Precipitation coverage is usually negatively correlated with NRMSE, and positively correlated with FSS, consistent with events of limited extent showing

on average lower skill than for widespread events. On the other hand, τ_c is generally positively correlated with NRMSE and negatively correlated with FSS, consistent with quasi-equilibrium events, characterized by small τ_c values showing better predictability than nonequilibrium events. For simplicity, the relationship between the predictability metric and the event classifier is shown in Fig. 4 either in terms of the correlation coefficient r or as $-r$, depending on the sign, such that we always show a range from 0 to 1. The thick lines in Fig. 4 indicate that the correlation coefficients are significantly different than 0 for an $\alpha = 0.05$ confidence level, while the thin lines reflect that no statistical significance is found for the $\alpha = 0.05$ level and hence the relationship is not deemed significant.

Except for 2008, during the first 15–20 h of the forecasts, about 20%–30% of the variability in QPF skill is explained by the variability in precipitation coverage, such that higher precipitation coverage values are associated with better QPF skill (lower NRMSE and higher FSS values). In other words, it seems that QPFs of widespread precipitation systems are more skillful than for events with less extensive precipitation coverage. As indicated in Fig. 4, there is generally a lower correlation between the skill and coverage in the afternoon hours, which is probably associated with the maximum in the diurnal cycle of precipitation. It was shown previously by Berenguer et al. (2012) that the models evaluated here had some difficulties in forecasting accurately the timing of the diurnal maximum.

The relationship between skill and τ_c is stronger here than was reported in SZY16 (Fig. 4). This is due to the fact that in this paper the values of τ_c are estimated differently than in SZY16, being calculated using output from the member CN, rather than using NAM data. The relationship between CN skill and τ_c is similar to that between skill and coverage, with anywhere between 20% and 30% of the variance in skill being explained by the variance in τ_c , and with weaker relationships during afternoon hours. In agreement with Keil et al. (2014), large τ_c values, representative of nonequilibrium convection, are weakly associated with poor QPF skill.

The strength of the relationship between skill and event type varies somewhat from year to year. The relationship between skill and coverage is weaker or not statistically significant for 2008, while the relationship between skill and τ_c is not significant for 2009 and 2013.

Note from Fig. 4 that the results are not sensitive to whether the skill is quantified using NRMSE or FSS at similar scales. The relationships presented in Fig. 4 hold also when NRMSE and FSS are computed for larger scales, but no longer hold when the skill is quantified for scales smaller than 100 km. This is related to all events

2008 2009 2010 2011 2012 2013

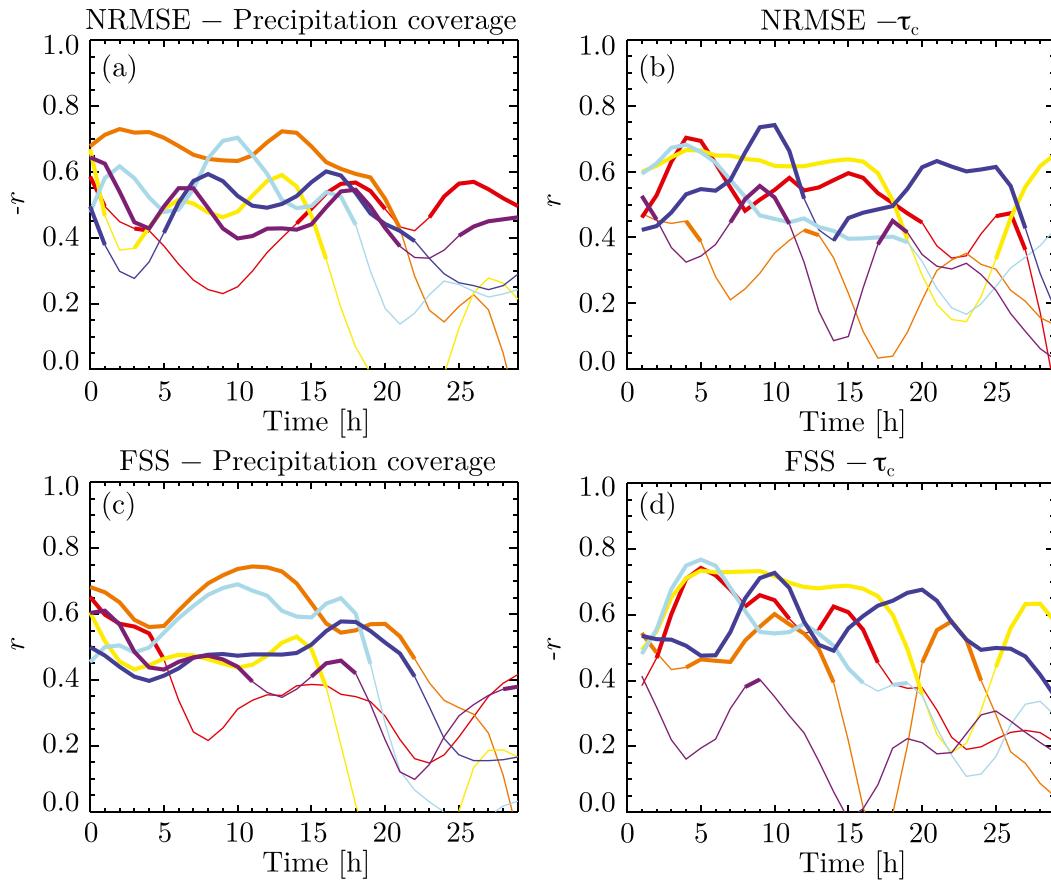


FIG. 4. The correlation coefficient r (or $-r$ when $r < 0$) between (a),(c) the skill of CN and precipitation coverage and (b),(d) τ_c . The skill is measured in terms of the Stage IV–CN NRMSE computed at scales larger than 128 km in (a) and (b), and in terms of the Stage IV–CN FSS for a scale of 100 km and a 95th percentile threshold in (c) and (d). The thick lines indicate that the correlation coefficient is statistically significantly different than 0 at the $\alpha = 0.05$ confidence level. The values of the correlation coefficient have been smoothed in time using a 3-h running mean for illustration purposes.

showing poor QPF skill (independent of the evaluation metric) at scales smaller than 100 km for most forecast hours, in agreement with the values of the decorrelation scale presented in Fig. 2. Therefore, we find that the QPF skill of the control member CN at scales larger than 100 km shows a case-by-case variability consistent with poorer performance for small-coverage, nonequilibrium events, and better performance for large-coverage, equilibrium events. However, the QPF skill at scales smaller than 100 km does not show any systematic case dependence. These results are consistent with small-scale convective events, with low precipitation coverage being inherently less predictable than widespread events. However, the fact that all events show poor skill at small scales might indicate that all events suffer from similar errors at small scales, while more extensive precipitation systems allow for verification of skill at predictable scales.

As for both metrics the scores increase with increasing scales and as FSS is computed for percentile thresholds, thus eliminating the sensitivity of the results to forecast bias, the lack of sensitivity of the results to the evaluation metric might be an indication that the main types of errors are displacement errors, which becomes less important as the verification scale increases.

Unfortunately, it is worth noting that all the statistical relationships presented here are quite weak. Therefore, it would be difficult to make quantitative use of these relationships to gain insights into the predictability of precipitation systems a priori.

As mentioned before, it was reported by Keil et al. (2014) that the sensitivity of ensemble spread to the weather regime is dependent to the type of ensemble perturbations. The case dependence of ensemble spread is also investigated here for two subensembles: IC/LBC/PHYS and

2008 2009 2010 2011 2012 2013

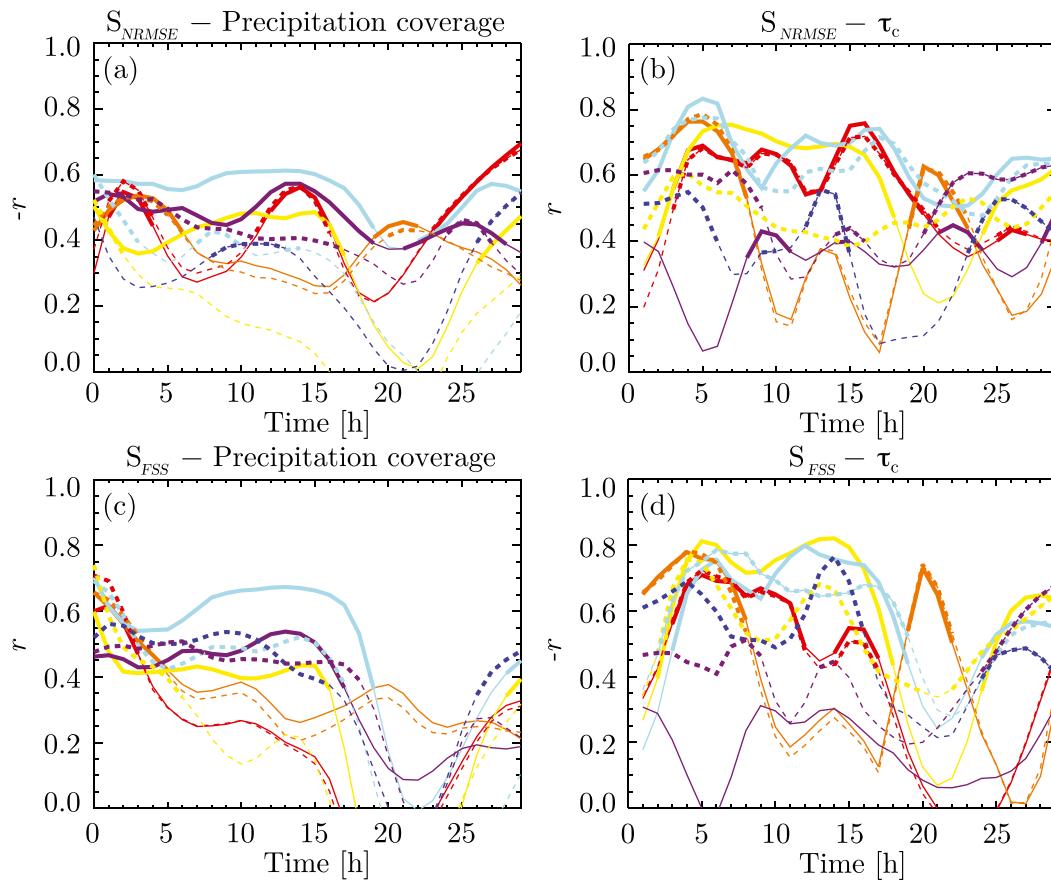


FIG. 5. The correlation coefficient r (or $-r$ when $r < 0$) (a),(c) between spread and precipitation coverage and (b),(d) between spread and τ_c . The spread is measured in terms of the member-CN NRMSE computed at scales larger than 128 km and averaged over perturbed members in (a) and (b), and in terms of the member-CN FSS for a scale of 100 km, a 95th percentile threshold, and averaged over perturbed members in (c) and (d). The solid lines correspond to the IC/LBC/PHYS subensembles, while the dashed lines correspond to the MP subensembles. The thick lines indicate that the correlation coefficient is statistically significantly different than 0 at the $\alpha = 0.05$ confidence level. The values of the correlation coefficient have been smoothed in time using a 3-h running mean for illustration purposes.

MP. To maintain consistency with the rest of the paper, and to allow an investigation of spread as a function of spatial scale, rather than computing the traditional spread metric for the two subensembles, we define here ensemble spread in terms of the average of NRMSE and FSS with respect to CN corresponding to the IC/LBC/PHYS and MP members, respectively. In particular, the spread of a subensemble with N members in terms of NRMSE is defined as

$$S_{\text{NRMSE}} = \frac{1}{N} \sum_{\text{all members}} \text{NRMSE}_{\text{member-CN}} \quad (6)$$

and similarly for the spread in terms of FSS as

$$S_{\text{FSS}} = \frac{1}{N} \sum_{\text{all members}} \text{FSS}_{\text{member-CN}} \quad (7)$$

Figure 5 shows the correlation coefficient computed between S_{NRMSE} at 128 km and S_{FSS} at 100 km and for the 95th percentile threshold and precipitation coverage and τ_c , respectively, for the two subensembles (solid lines for IC/LBC/PHYS and dashed lines for MP), for each year. Independent of the perturbation type, the relation between precipitation coverage and ensemble spread is more consistent in time than the relationship between τ_c and ensemble spread. This is probably due to τ_c being a noisier variable, as it is proportional to the ratio of two variables that exhibit themselves large spatial and temporal variability at small scales (precipitation and CAPE). This also suggests that while τ_c could have some qualitative predictive value for the predictability of a case in general, it should be used with caution in quantitative applications. For all years, Fig. 5 indicates that the

correlation between spread and both precipitation coverage and τ_c is stronger in the first 15 h of the forecast, this being especially true for 2010 and 2011. While the values of the correlation coefficient corresponding to the IC/LBC/PHYS subensemble seem higher than those for the MP subensemble, this is likely not significant considering that the number of ensemble members is higher for the IC/LBC/PHYS subensemble than for the MP. Finally, there is no relationship between precipitation coverage and τ_c and ensemble spread at scales smaller than 128 km, consistent with the results of SZY16 (not shown).

c. *The relative importance of ensemble perturbation methodologies*

We have shown in section 3 the effect of different types of perturbations on average, for each of the years. In this section, we are interested in determining whether the *relative importance of different perturbation methodologies depends on the weather regime*. For instance, does the addition of PHYS perturbations to an IC/LBC ensemble result in more variability for convective-equilibrium or for nonequilibrium cases? Therefore, the results presented in this section are consistent with, but complementary to, the results shown in section 3, as here we show the relative contributions of different types of errors as a function of precipitation coverage and τ_c .

First, we analyze the case dependence of the effect of adding PHYS perturbations to the IC/LBC perturbations, by plotting the ratio of the n2–CN NRMSE to the n1–CN NRMSE as a function of precipitation coverage and τ_c for all cases during 2008 (Fig. 6). The 2008 n2 member was the only member that had only IC/LBC perturbations. The member n1 is just one of the members run in 2008 that had IC/LBC/PHYS perturbations, using the Ferrier MP scheme, the Goddard SW scheme, and the Yonsei University (YSU) PBL scheme (see Table 3 for more details on the ensemble configuration). The NRMSE was computed for large scales (i.e., low-pass-filtered precipitation fields with a cutoff scale of 256 km) and medium scales (i.e., bandpass-filtered fields between 64 and 256 km). This ratio is more often larger than 1 than smaller than 1, meaning that in most cases adding PHYS perturbations results in increased spread, consistent with the average results illustrated in Fig. 2. This effect is more important at the beginning of the forecasts (left panels of Fig. 6), whereas after $t = 24$ h, about 30% of the cases show n2–CN NRMSE values larger than n1–CN NRMSE values (right panels). This could mean that PHYS perturbations are important in the increasing spread at early forecast times, whereas for late forecast times the ensemble spread is mostly dominated by the IC/LBC perturbations. Also, from these diagrams, it seems that the error ratio has values closer to 1 for events with larger

precipitation coverage. Moreover, the ratio values are more often larger than 1 for τ_c values greater than 3 h, especially for $t > 24$ h, hinting that PHYS perturbations are more important for nonequilibrium, small-coverage events. The error ratio at scales smaller than 64 km is around 1 for all cases (not shown), most likely because of the errors being saturated at these scales for all types of perturbations.

We also investigate in more detail the effects of the different types of ensemble perturbations used in 2010 (Table 3). Figures 7 and 8 show how the effects of other types of perturbations compare to the effects of the IC/LBC/PHYS perturbations traditionally used in the CAPS SSEF in previous years at large and medium scales in terms of the error ratio. Figure 7 shows the error ratios at large scales (larger than 256 km), while Fig. 8 shows the error ratios at medium scales (between 64 and 256 km). The ratios in Fig. 8 are less variable and closer to 1 than at large scales. This is related to the actual NRMSE values being larger and having a smaller dynamical range at medium scales. In other words, the different perturbations have a similar effect at medium scales. Otherwise, the relative effects of the different perturbations are consistent between the medium and large scales. For example, the effect of the RAND perturbations is less important than the effects of IC/LBC/PHYS perturbations both at the medium and large scales.

As shown in Figs. 7a and 7b, using simply random noise as the IC perturbation almost never generates as much variability as the IC/LBC/PHYS perturbations (the NRMSE ratios are almost always smaller than 1). If the RAND perturbations have some spatial structure, as for member RC, the upscale growth is large at first, resulting in variability comparable to that caused by IC/LBC/PHYS perturbations up to $t = 6$ h. Afterward, only in a few cases do RC perturbations have an effect comparable to the IC/LBC/PHYS perturbations (fewer data points fall around the 1 ratio line in Figs. 7c and 7d at later forecast times). Furthermore, it seems that most of the points situated above the 1 ratio line are associated with small precipitation coverage and large τ_c values. Accounting for small-scale IC uncertainty by adding RC perturbations to a member with IC/LBC/PHYS perturbations has the largest effect at lead times shorter than 18 h and for events with small precipitation coverage and large τ_c values.

Figures 7g–j and 8g–j show how the relative effects of the IC/LBC/PHYS and PHYS perturbations vary with precipitation coverage and τ_c . For the majority of events, PHYS perturbations have a much smaller effect than IC/LBC/PHYS perturbations, and no systematic dependence on precipitation coverage or τ_c is evident.

This result also remains valid for the years 2011–13, with the caveat that the MP perturbations had a larger relative effect in 2011 than during the other years (not shown).

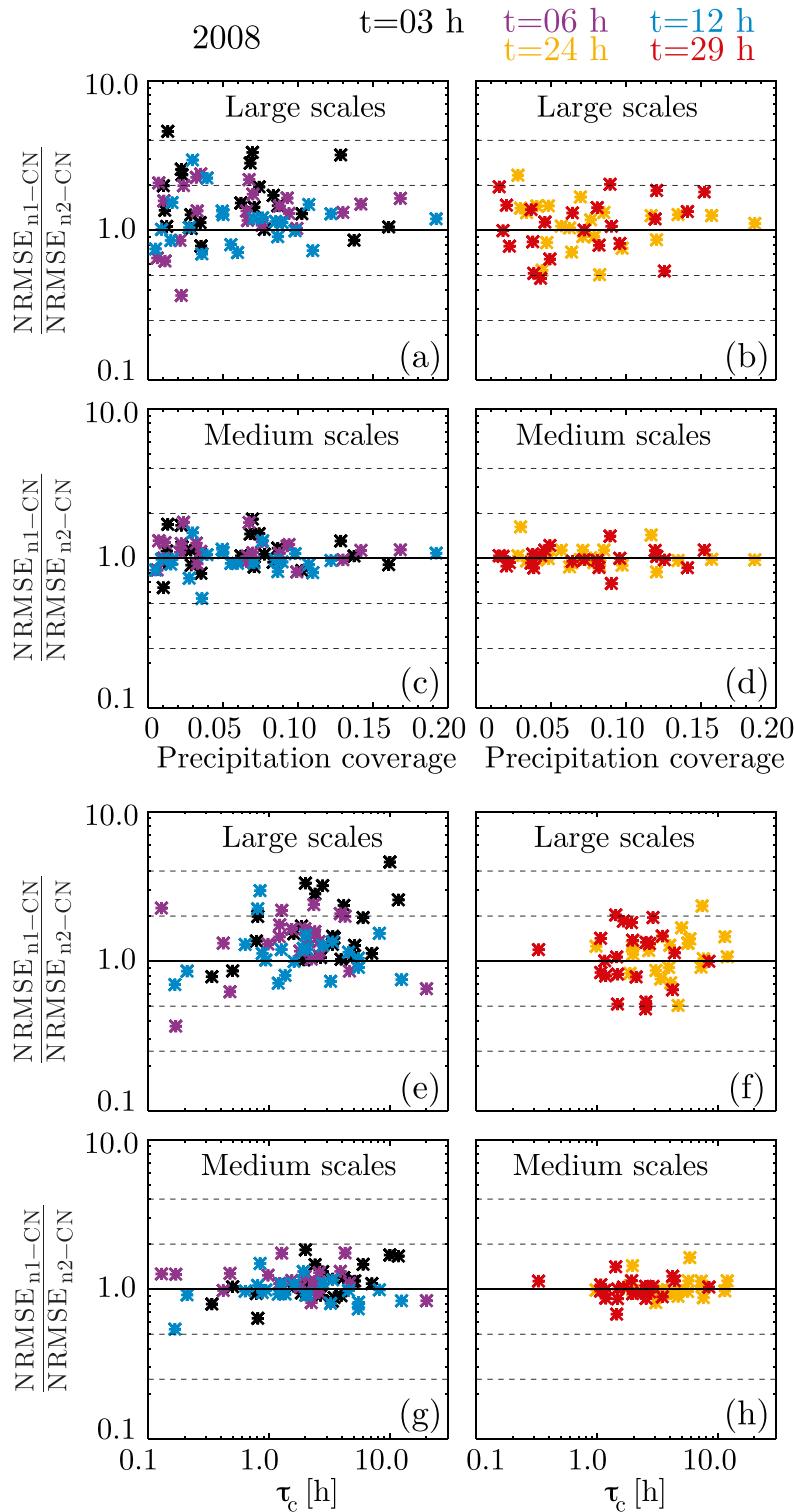


FIG. 6. Scatterplots of the ratio between n1-CN and n2-CN NRMSEs, and (a)–(d) precipitation coverage and (e)–(h) τ_c at large (larger than 256 km) and medium (between 64 and 256 km) scales, as indicated in the figure. The different colors correspond to different lead times, with early lead times in the right panels and later lead times in the left.

TABLE 3. Description of the CAPS SSEF ensembles. The first two columns indicate the year and the forecast length. Columns 3–5 indicate the number of members and describe the variable PHYS options for each subensemble (IC/LBC/PHYS, MP, PBL). The last column provides any important additional information about the entire ensemble, such as the WRF-ARW version used, PHYS configuration of the control member (CN), the PHYS options common to *all* ensemble members, etc. The following PHYS options are used in the ensemble: MP—Thompson (Thompson et al. 2008), Ferrier (Ferrier et al. 2002), WSM6 (Hong and Lim 2006), WRF double-moment 6-class (WDM6; Lim and Hong 2010), Morrison (Morrison et al. 2009), Milbrandt and Yau (M–Y; Milbrandt and Yau 2005); PBL—MYJ (Janjić 1994), YSU (Hong et al. 2006), quasi-normal scale elimination (QNSE; Sukoriansky et al. 2005), Mellor–Yamada–Nakanishi–Niino (MYNN; Nakanishi and Niino 2006), Asymmetric Cloud Model version 2 (ACM2; Pleim 2007); LSM—Noah (Tewari et al. 2004), RUC (Benjamin et al. 2004); SW—Dudhia (Dudhia 1989), Rapid Radiative Transfer Model for GCMs (RRTMG; Iacono et al. 2008), Goddard (Chou and Suarez 1999); LW—RRTM (Mlawer et al. 1997), RRTMG. All the forecasts are performed with a grid spacing of 4 km.

Year	Forecast length (h)	IC/LBC/PHYS	MP only	PBL only	Miscellaneous
2008	30	8 members. MP: Thompson, Ferrier, WSM6; SW: Dudhia, Goddard; PBL: MYJ, YSU	—	—	WRF-ARW version 2.2. CN: MP—Thompson, SW—Goddard, PBL—MYJ; alt: LW—RRTM, LSM—Noah
2009	30	8 members. MP: Thompson, Ferrier, WSM6; SW: Dudhia, Goddard; PBL: 2MYJ, YSU; LSM: Noah, RUC	—	—	WRF-ARW v3.0.1.1. CN: MP—Thompson, SW—Goddard, PBL—MYJ, LSM—Noah; alt: LW—RRTM
2010	30	9 members. MP: Thompson, Ferrier, WSM6, WDM6, Morrison; PBL: MYJ, YSU, QNSE, MYNN; LSM: Noah, RUC	3 members. MP: Thompson, WSM6, WDM6, Morrison; PBL: MYJ; LSM: Noah	2 members. PBL: MYJ, QNSE, MYNN; MP: Thompson; LSM: Noah	WRF-ARW v3.1.1. CN: MP—Thompson, PBL—MYJ, LSM—Noah; alt: LW—RRTM, SW—Goddard
2011	36	17 members. MP: Thompson, Ferrier (+), WSM6, WDM6, Morrison, M–Y; PBL: MYJ, YSU, QNSE, MYNN, ACM2; LSM: Noah, RUC	10 members. MP: Thompson (2), WSM6 (5), WDM6, Morrison, Ferrier (2), M–Y; PBL: MYJ; LSM: Noah	10 members. MP: Thompson; PBL: MYJ (4), QNSE, MYNN, ACM2 (3), YSU, YSU—Thompson; LSM: Noah	WRF-ARW v3.2.1. CN: MP—Thompson, PBL—MYJ, LSM—Noah; alt: LW—RRTM, SW—Goddard, LSM—Noah, PBL—MYJ; radar—DA-cycled member
2012	36	—	3 members. MP: WDM6, Morrison, M–Y; PBL: MYJ; LSM: Noah	4 members. MP: Thompson; PBL: QNSE, MYNN, ACM2, YSU; LSM: Noah	WRF-ARW v3.3.1. CN: MP—Thompson, PBL—MYJ, LSM—Noah; alt: LW—RRTM, SW—Goddard; one SKEB member
2013	48	13 members. MP: Thompson, WDM6, Morrison, M–Y; PBL: MYJ, YSU, QNSE, MYNN, ACM2; LSM: Noah, RUC	6 members. MP: WDM6, Morrison, M–Y, WSM, NSSL, Thompson (coupled); PBL: MYJ; LSM: Noah	4 members. MP: Thompson; PBL: QNSE, MYNN, ACM2, YSU; LSM: Noah	WRF-ARW v3.4.1. CN: MP—Thompson, PBL—MYJ, LSM—Noah; alt: LW—RRTM, SW—RRTMG; one member with coupled MP—radiation

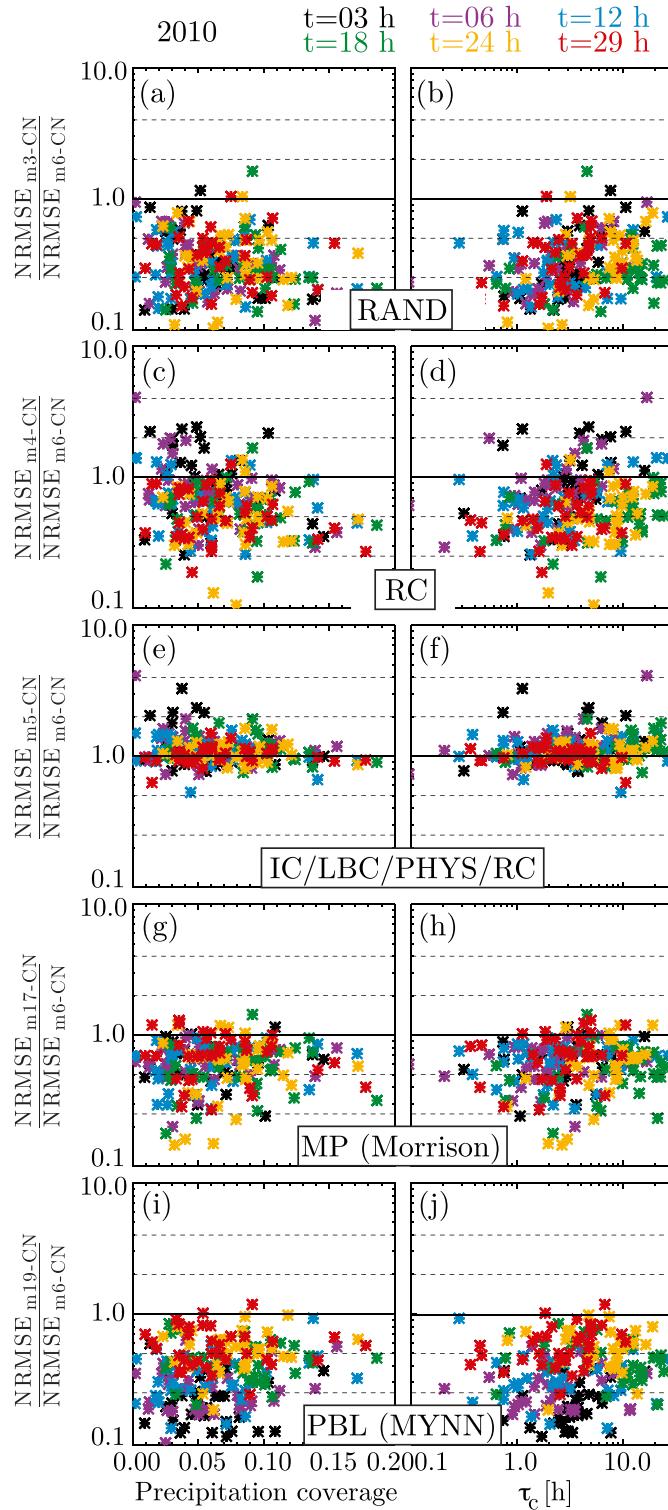


FIG. 7. The ratios of NRMSE computed between a perturbed member (as described in the different panels) and CN and the m6-CN NRMSE (m6 is a IC/LBC/PHYS member) at scales larger than 256 km plotted against (left) precipitation coverage and (right) τ_c . The different colors correspond to different forecast lead times, as shown in the legend.

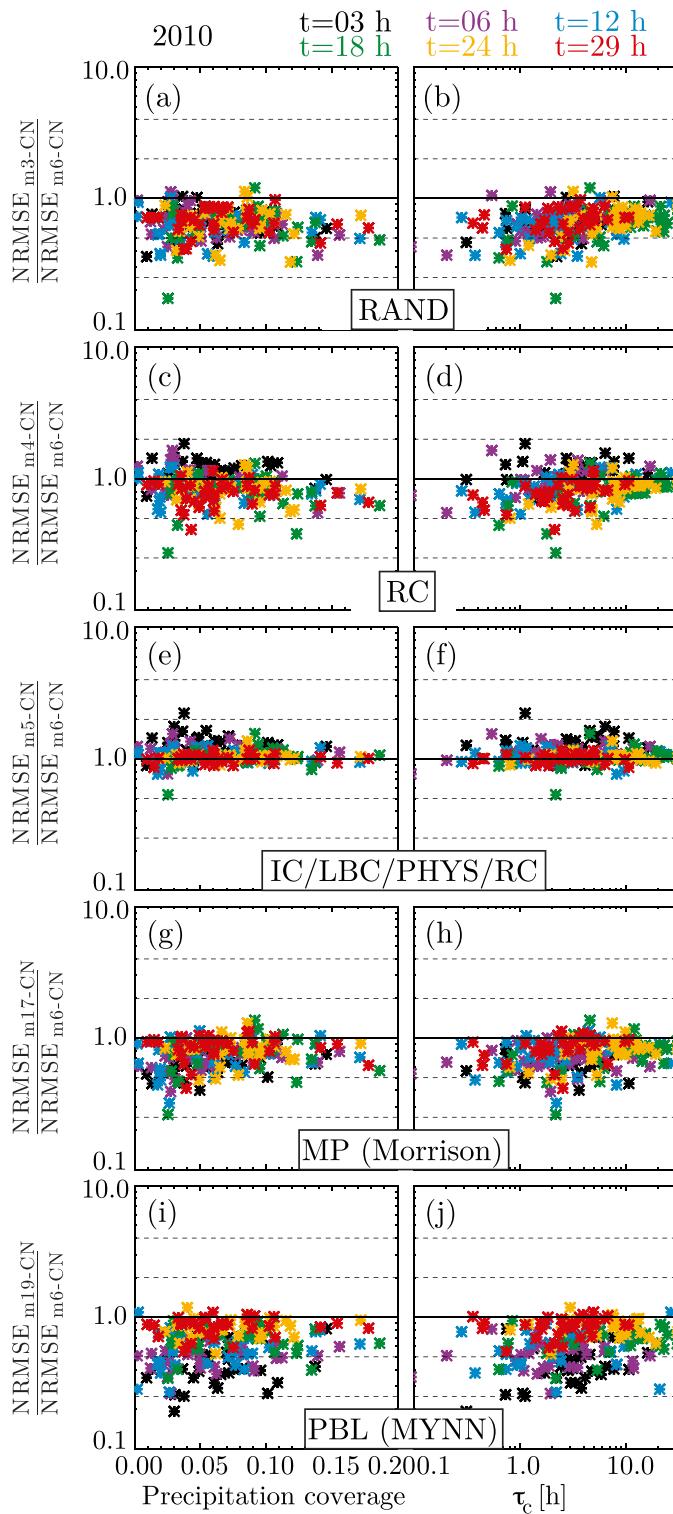


FIG. 8. As in Fig. 7, but here the error is computed for medium scales (64–256 km).

We note that Johnson et al. (2014) present a case, 20 May 2010, for which all IC perturbation methodologies resulted in differences with respect to the control member comparable to the forecast error. In fact, we find that this is the only case for which the RAND and RC perturbations have an effect comparable to the IC/LBC/PHYS perturbations. Unfortunately, this event is not characterized by very small precipitation coverage or very large τ_c values. This is consistent with the strength (or lack thereof) of the statistical relationship found between predictability and precipitation coverage and τ_c .

As it was found that the IC/LBC/PHYS/RC perturbations generate the most spread on average, we investigate whether the error attributable to these perturbations is comparable to the forecast error. Figure 9 shows a scatterplot of the ratio $\text{NRMSE}_{m5-CN}/\text{NRMSE}_{\text{StageIV-CN}}$ against precipitation coverage (Figs. 9a–d) and τ_c (Figs. 9e–g). At early forecast times (Fig. 9, left), the forecast error (Stage IV–CN NRMSE) is larger than the difference caused by the ensemble perturbations (most points lie below the 1 line). Toward the end of the forecast time (Fig. 9, right), the error due to the perturbations becomes equal to and even surpasses the forecast error both at medium and large scales. Furthermore, we note that the m5–CN NRMSE is larger than the Stage IV–CN NRMSE, especially for events with small values of precipitation coverage and τ_c values larger than 3 h. The difference between the results presented here and those of Johnson et al. (2014) may be due to the use of different analysis methods. In particular, the predictability measures used here are presented in normalized units to remove the variability due to the diurnal cycle of precipitation, in contrast to Johnson et al. (2014), who use absolute error.

We have also investigated whether the perturbations affect the deterministic QPF skill of the perturbed members. Romine et al. (2014) found that while stochastic PHYS perturbations increased ensemble spread, they also negatively affected the deterministic QPF skill of the perturbed members, noting that SKEB perturbations had the least negative impact. In addition, Johnson et al. (2014) found that adding RAND perturbations to temperature and humidity resulted in spurious precipitation at very early forecast times and, thus, decreased QPF skill. We show in Fig. 10 how the skill levels of the IC/LBC/PHYS/RC members compare to the QPF skill of the control CN at scales larger than 256 km in terms of the NRMSE and as a function of precipitation coverage and τ_c . We limit our investigation to those scales, as it was shown both here and in SZY15 that the model predictability of the atmospheric state is lost in a matter of a few hours at meso- γ and meso- β scales. While for early forecast hours the skill of the two members is similar (Fig. 10, left), at late forecast times, the IC/LBC/PHYS/RC

member shows poorer skill (larger NRMSE values) than CN for the majority of cases. This is more often the case for events with low precipitation coverage. Finally, the rest of the perturbed ensemble members do not seem to have a significantly different skill than CN (not shown).

5. Conclusions

This paper has investigated the scale and case dependence of the predictability of precipitation by the CAPS SSEF system run during NOAA's HWT Spring Experiments of 2008–13. Depending on the year, the ensemble consisted of members with IC/LBC perturbations derived from a regional ensemble, small-scale IC perturbations, and different types of PHYS perturbations.

The scale dependence of the predictability of precipitation was analyzed following the methodology of SZY15, in terms of the decorrelation scale λ_0 . In agreement with previous studies, IC/LBC perturbations derived from a regional ensemble system, representative of large-scale IC/LBC errors, were found to sample the main source of forecast uncertainty for storm-scale models. Perturbing the model physical schemes in addition to the IC/LBC perturbations leads to slightly higher values of λ_0 , but their values are not sufficient to reach those of λ_0 for Stage IV–CN until 24 h into the forecast, and for scales larger than 200 km. Adding small-scale RC perturbations to the initial temperature and humidity fields resulted in increased λ_0 values with respect to CN, but with results still smaller than the Stage IV–CN λ_0 .

From the different methodologies of perturbing PHYS, varying the MP schemes had the largest effect both in terms of the differences in the precipitation forecasts with respect to CN and on the QPF skill of the perturbed members. The uncertainty in the model microphysical parameterization scheme is sufficient to cause the loss of predictability at meso- β scales after about 1 day. The other types of PHYS perturbations (PBL, SKEB) resulted in smaller values of λ_0 with respect to CN throughout the forecast lead time. This study evaluated the effect of ensemble perturbations on precipitation forecasts only. Precipitation is the direct result of microphysical processes, and thus it is reasonable that accounting for MP errors would lead to variability in QPF. On the other hand, the effects of uncertainties in other physical schemes might be more evident in fields other than precipitation.

The case-by-case variability of the precipitation predictability of the CAPS SSEF system was also investigated following the methodology of SZY16. Both the model performance and the ensemble spread showed relationships with both precipitation coverage and τ_c . In the case of ensemble spread, it seems that the spread of an IC/LBC/PHYS subensemble showed higher correlation

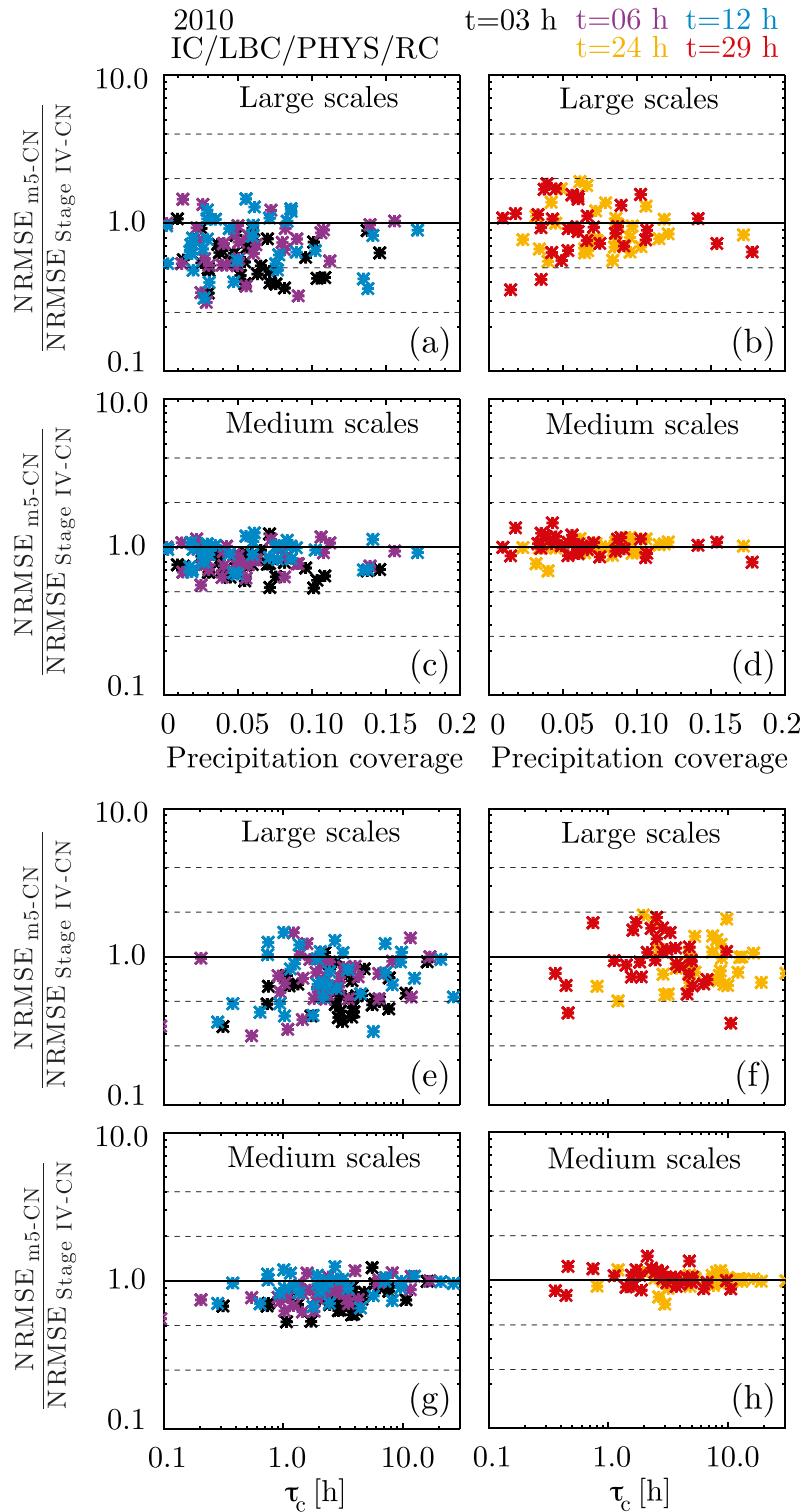


FIG. 9. Comparison of the effects of IC/LBC/PHYS/RC perturbations (member m5) quantified in terms of the NRMSE with respect to CN for the forecast error of CN (NRMSE between CN and Stage IV). The ratio between the m5-CN and Stage IV-CN NRMSEs is plotted against (a)–(d) precipitation coverage and (e)–(h) τ_c . The NRMSE is shown for large scales (larger than 256 km) and medium scales (64–256 km), as indicated in the figure. The different colors correspond to different forecast lead times, with early lead times on the left and late lead times on the right.

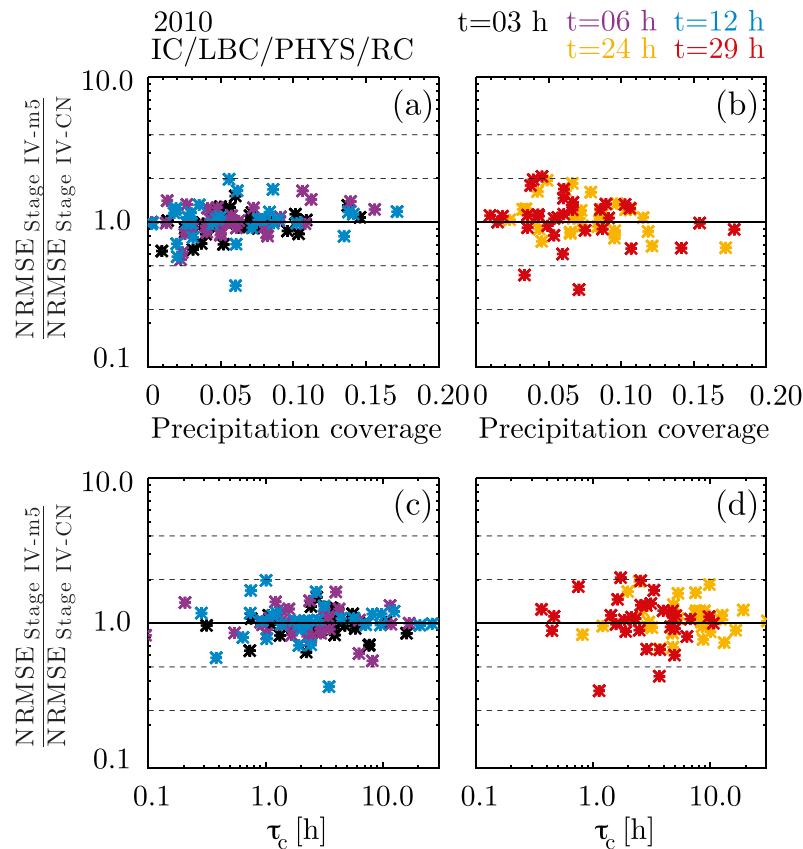


FIG. 10. The ratio between the Stage IV-m5 and Stage IV-CN NRMSEs is plotted against (a),(b) precipitation coverage and (c),(d) τ_c for scales larger than 256 km. The different colors correspond to different forecast lead times, with early lead times on the left and late lead times on the right.

with τ_c than the spread of an MP subensemble during 2010 and 2011. Moreover, the analysis highlighted the large effect of the diurnal cycle of precipitation for both model skill and ensemble spread. Given the importance of the diurnal cycle of solar heating on the evolution of severe weather over the continental United States, future research should investigate in more detail whether the models have a bias in depicting the intensity and phase of the diurnal cycle, or whether the presence of more convective instability is simply the cause for faster error growth as suggested by previous studies.

With respect to the case dependence of the relative effect of ensemble perturbations, the results were inconclusive. While the addition of PHYS perturbations to an IC/LBC ensemble seems to have a larger effect on events of small precipitation coverage and large τ_c values, this relationship is too weak to be usable in a quantitative sense.

In conclusion, in agreement with Johnson et al. (2014), we found that accounting for large-scale IC/LBC and

microphysics uncertainties is most important for quantitative precipitation forecasts over the continental United States. While other types of perturbation methodologies (such as varying the PBL scheme or SKEB perturbations) might be important for other applications, their effect is negligible for QPF. However, accounting for small-scale IC uncertainties seems important, and future work should focus on finding the best way of accounting for IC errors across scales.

Acknowledgments. The CAPS SSEF forecasts were produced mainly under the support of a grant from the NOAA CSTAR program, and the 2008 ensemble forecasts were produced at the Pittsburgh Supercomputer Center. Kevin Thomas, Jidong Gao, Keith Brewster, and Yunheng Wang of CAPS made significant contributions to the forecasting efforts. This work was funded by the Natural Science and Engineering Research Council of Canada (NSERC) and Hydro-Quebec through the IRC program. Dr. Adam Clark is kindly acknowledged for his

assistance in obtaining the data. We thank two anonymous reviewers for their comments and suggestions that helped improve the manuscript.

REFERENCES

- Accadia, C., S. Mariani, M. Casaioli, A. Lavagnini, and A. Speranza, 2003: Sensitivity of precipitation forecast skill scores to bilinear interpolation and a simple nearest-neighbor average method on high-resolution verification grids. *Wea. Forecasting*, **18**, 918–932, doi:10.1175/1520-0434(2003)018<0918:SOPFSS>2.0.CO;2.
- Baldwin, M. E., and K. E. Mitchell, 1997: The NCEP hourly multi-sensor U.S. precipitation analysis for operations and GCIP research. Preprints, *13th Conf. on Hydrology*, Long Beach, CA, Amer. Meteor. Soc., 54–55.
- Benjamin, S. G., G. A. Grell, J. M. Brown, T. G. Smirnova, and R. Bleck, 2004: Mesoscale weather prediction with the RUC hybrid isentropic–terrain-following coordinate model. *Mon. Wea. Rev.*, **132**, 473–494, doi:10.1175/1520-0493(2004)132<0473:MWPWTR>2.0.CO;2.
- Berenguer, M., M. Surcel, I. Zawadzki, M. Xue, and F. Kong, 2012: The diurnal cycle of precipitation from continental radar mosaics and numerical weather prediction models. Part II: Intercomparison among numerical models and with nowcasting. *Mon. Wea. Rev.*, **140**, 2689–2705, doi:10.1175/MWR-D-11-00181.1.
- Chou, M.-D., and M. J. Suarez, 1999: A solar radiation parameterization for atmospheric studies. NASA Tech. Memo. NASA/TM-1999-104606, NASA Tech. Rep. Series on Global Modeling and Data Assimilation, Vol. 15, 40 pp. [Available online at <https://gmao.gsfc.nasa.gov/pubs/docs/Chou136.pdf>.]
- Clark, A. J., and Coauthors, 2011: Probabilistic precipitation forecast skill as a function of ensemble size and spatial scale in a convection-allowing ensemble. *Mon. Wea. Rev.*, **139**, 1410–1418, doi:10.1175/2010MWR3624.1.
- Dey, S. R. A., G. Leoncini, N. M. Roberts, R. S. Plant, and S. Migliorini, 2014: A spatial view of ensemble spread in convection permitting ensembles. *Mon. Wea. Rev.*, **142**, 4091–4107, doi:10.1175/MWR-D-14-00172.1.
- , R. S. Plant, N. M. Roberts, and S. Migliorini, 2016: Assessing spatial precipitation uncertainties in a convective-scale ensemble. *Quart. J. Roy. Meteor. Soc.*, **142**, 2935–2948, doi:10.1002/qj.2893.
- Done, J. M., G. C. Craig, S. L. Gray, P. A. Clark, and M. E. B. Gray, 2006: Mesoscale simulations of organized convection: Importance of convective equilibrium. *Quart. J. Roy. Meteor. Soc.*, **132**, 737–756, doi:10.1256/qj.04.84.
- , —, —, and —, 2012: Case-to-case variability of predictability of deep convection in a mesoscale model. *Quart. J. Roy. Meteor. Soc.*, **138**, 638–648, doi:10.1002/qj.943.
- Du, J., G. Dimego, Z. Toth, D. Jovic, B. Zhou, J. Zhu, J. Wang, and H. Juang, 2009: NCEP Short Range Ensemble Forecast (SREF) system upgrade in 2009. *23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction*, Omaha, NE, Amer. Meteor. Soc., 4A.4. [Available online at https://ams.confex.com/ams/23WAF19NWP/techprogram/paper_153264.htm.]
- Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107, doi:10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2.
- Durran, D. R., and M. Gingrich, 2014: Atmospheric predictability: Why butterflies are not of practical importance. *J. Atmos. Sci.*, **71**, 2476–2488, doi:10.1175/JAS-D-14-0007.1.
- Ferrier, B. S., Y. Jin, Y. Lin, T. Black, E. Rogers, and G. DiMego, 2002: Implementation of a new grid-scale cloud and rainfall scheme in the NCEP Eta model. *19th Conf. on Weather Analysis and Forecasting/15th Conf. on Numerical Weather Prediction*, San Antonio, TX, Amer. Meteor. Soc., 10.1. [Available online at https://ams.confex.com/ams/SLS_WAF_NWP/techprogram/paper_47241.htm.]
- Flack, D. L. A., R. S. Plant, S. L. Gray, H. W. Lean, C. Keil, and G. C. Craig, 2016: Characterisation of convective regimes over the British Isles. *Quart. J. Roy. Meteor. Soc.*, **142**, 1541–1553, doi:10.1002/qj.2758.
- Gao, J., M. Xue, K. Brewster, and K. K. Droegemeier, 2004: A three-dimensional variational data analysis method with recursive filter for Doppler radars. *J. Atmos. Oceanic Technol.*, **21**, 457–469, doi:10.1175/1520-0426(2004)021<0457:ATVDAM>2.0.CO;2.
- Hohenegger, C., D. Lüthi, and C. Schär, 2006: Predictability mysteries in cloud-resolving models. *Mon. Wea. Rev.*, **134**, 2095–2107, doi:10.1175/MWR3176.1.
- Hong, S.-Y., and J.-O. J. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc.*, **42**, 129–151.
- , Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341, doi:10.1175/MWR3199.1.
- Hu, M., M. Xue, and K. Brewster, 2006a: 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of the Fort Worth, Texas, tornadic thunderstorms. Part I: Cloud analysis and its impact. *Mon. Wea. Rev.*, **134**, 675–698, doi:10.1175/MWR3092.1.
- , —, J. Gao, and K. Brewster, 2006b: 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of the Fort Worth, Texas, tornadic thunderstorms. Part II: Impact of radial velocity analysis via 3DVAR. *Mon. Wea. Rev.*, **134**, 699–721, doi:10.1175/MWR3093.1.
- Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, doi:10.1029/2008JD009944.
- Janjić, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945, doi:10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2.
- , 2003: A nonhydrostatic model based on a new approach. *Meteor. Atmos. Phys.*, **82**, 271–285, doi:10.1007/s00703-001-0587-6.
- Johnson, A., and X. Wang, 2016: A study of multi-scale initial condition perturbation methods for convection-permitting ensemble forecasts. *Mon. Wea. Rev.*, **144**, 2579–2604, doi:10.1175/MWR-D-16-0056.1.
- , —, M. Xue, and F. Kong, 2011: Hierarchical cluster analysis of a convection-allowing ensemble during the Hazardous Weather Testbed 2009 Spring Experiment. Part II: Ensemble clustering over the whole experiment period. *Mon. Wea. Rev.*, **139**, 3694–3710, <https://doi.org/10.1175/MWR-D-11-00016.1>.
- , and Coauthors, 2014: Multiscale characteristics and evolution of perturbations for warm season convection-allowing precipitation forecasts: Dependence on background flow and method of perturbation. *Mon. Wea. Rev.*, **142**, 1053–1073, doi:10.1175/MWR-D-13-00204.1.
- Keil, C., F. Heinlein, and G. C. Craig, 2014: The convective adjustment time-scale as indicator of predictability of convective

- precipitation. *Quart. J. Roy. Meteor. Soc.*, **140**, 480–490, doi:10.1002/qj.2143.
- Kong, F., and Coauthors, 2014: CAPS storm-scale ensemble forecasting system: impact of IC and LBC perturbations. *26th Conf. on Weather Analysis and Forecasting/22nd Conf. on Numerical Weather Prediction*, Atlanta, GA, Amer. Meteor. Soc., 119. [Available online at <https://ams.confex.com/ams/94Annual/webprogram/Paper234762.html>.]
- Lim, K.-S. S., and S.-Y. Hong, 2010: Development of an effective double-moment cloud microphysics scheme with prognostic cloud condensation nuclei (CCN) for weather and climate models. *Mon. Wea. Rev.*, **138**, 1587–1612, doi:10.1175/2009MWR2968.1.
- Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307, doi:10.3402/tellusa.v21i3.10086.
- Milbrandt, J. A., and M. K. Yau, 2005: A multimoment bulk microphysics parameterization. Part I: Analysis of the role of the spectral shape parameter. *J. Atmos. Sci.*, **62**, 3051–3064, doi:10.1175/JAS3534.1.
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682, doi:10.1029/97JD00237.
- Morrison, H., G. Thompson, and V. Tatarskii, 2009: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Mon. Wea. Rev.*, **137**, 991–1007, doi:10.1175/2008MWR2556.1.
- Nakanishi, M., and H. Niino, 2006: An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Bound.-Layer Meteor.*, **119**, 397–407, doi:10.1007/s10546-005-9030-8.
- Pleim, J. E., 2007: A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: Model description and testing. *J. Appl. Meteor. Climatol.*, **46**, 1383–1395, doi:10.1175/JAM2539.1.
- Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, doi:10.1175/2007MWR2123.1.
- Romine, G. S., C. S. Schwartz, J. Berner, K. R. Fossell, C. Snyder, J. L. Anderson, and M. L. Weisman, 2014: Representing forecast error in a convection-permitting ensemble system. *Mon. Wea. Rev.*, **142**, 4519–4541, doi:10.1175/MWR-D-14-00100.1.
- Schwartz, C. S., 2016: Improving large-domain convection-allowing forecasts with high-resolution analyses and ensemble data assimilation. *Mon. Wea. Rev.*, **144**, 1777–1803, doi:10.1175/MWR-D-15-0286.1.
- , G. S. Romine, K. R. Smith, and M. L. Weisman, 2014: Characterizing and optimizing precipitation forecasts from a convection-permitting ensemble initialized by a mesoscale ensemble Kalman filter. *Wea. Forecasting*, **29**, 1295–1318, doi:10.1175/WAF-D-13-00145.1.
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., doi:10.5065/D68S4MVH.
- Snook, N., M. Xue, and Y. Jung, 2012: Ensemble probabilistic forecasts of a tornadic mesoscale convective system from ensemble Kalman filter analyses using WSR-88D and CASA radar data. *Mon. Wea. Rev.*, **140**, 2126–2146, doi:10.1175/MWR-D-11-00117.1.
- Stensrud, D. J., J.-W. Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.*, **128**, 2077–2107, doi:10.1175/1520-0493(2000)128<2077:UICAMP>2.0.CO;2.
- Sukoriansky, S., B. Galperin, and V. Perov, 2005: Application of a new spectral theory of stably stratified turbulence to the atmospheric boundary layer over sea ice. *Bound.-Layer Meteor.*, **117**, 231–257, doi:10.1007/s10546-004-6848-4.
- Surcel, M., M. Berenguer, and I. Zawadzki, 2010: The diurnal cycle of precipitation from continental radar mosaics and numerical weather prediction models. Part I: Methodology and seasonal comparison. *Mon. Wea. Rev.*, **138**, 3084–3106, doi:10.1175/2010MWR3125.1.
- , I. Zawadzki, and M. K. Yau, 2015: A study on the scale dependence of the predictability of precipitation patterns. *J. Atmos. Sci.*, **72**, 216–235, doi:10.1175/JAS-D-14-0071.1.
- , —, and —, 2016: The case-to-case variability of the predictability of precipitation by a storm-scale ensemble forecasting system. *Mon. Wea. Rev.*, **144**, 193–212, doi:10.1175/MWR-D-15-0232.1.
- Tewari, M., and Coauthors, 2004: Implementation and verification of the unified Noah land surface model in the WRF model. *20th Conf. on Weather Analysis and Forecasting/16th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., 14.2a. [Available online at https://ams.confex.com/ams/84Annual/techprogram/paper_69061.htm.]
- Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, doi:10.1175/2008MWR2387.1.
- Vié, B., O. Nuissier, and V. Ducrocq, 2011: Cloud-resolving ensemble simulations of Mediterranean heavy precipitating events: Uncertainty on initial conditions and lateral boundary conditions. *Mon. Wea. Rev.*, **139**, 403–423, doi:10.1175/2010MWR3487.1.
- Walser, A., D. Lüthi, and C. Schär, 2004: Predictability of precipitation in a cloud-resolving model. *Mon. Wea. Rev.*, **132**, 560–577, doi:10.1175/1520-0493(2004)132<0560:POPIAC>2.0.CO;2.
- Xue, M., D. H. Wang, J. Gao, K. Brewster, and K. K. Droegemeier, 2003: The Advanced Regional Prediction System (ARPS) storm-scale numerical weather prediction and data assimilation. *Meteor. Atmos. Phys.*, **82**, 139–170, doi:10.1007/s00703-001-0595-6.
- Zhang, F., C. Snyder, and R. Rotunno, 2002: Mesoscale predictability of the “surprise” snowstorm of 24–25 January 2000. *Mon. Wea. Rev.*, **130**, 1617–1632, doi:10.1175/1520-0493(2002)130<1617:MPOTSS>2.0.CO;2.
- , A. M. Odins, and J. W. Nielsen-Gammon, 2006: Mesoscale predictability of an extreme warm-season precipitation event. *Wea. Forecasting*, **21**, 149–166, doi:10.1175/WAF909.1.
- Zimmer, M., G. C. Craig, C. Keil, and H. Wernli, 2011: Classification of precipitation events with a convective response timescale and their forecasting characteristics. *Geophys. Res. Lett.*, **38**, L05802, doi:10.1029/2010GL046199.