Predictability of Deep Convection in Idealized and Operational Forecasts: Effects of Radar Data Assimilation, Orography, and Synoptic Weather Regime[✉]

KEVIN BACHMANN, CHRISTIAN KEIL, AND GEORGE C. CRAIG

Meteorologisches Institut, Ludwig-Maximilians-Universität, Munich, Germany

MARTIN WEISSMANN

*Hans-Ertel Centre for Weather Research, Deutscher Wetterdienst, Munich, Germany, and
Institut für Meteorologie und Geophysik, Universität Wien, Vienna, Austria*

CHRISTIAN A. WELZBACHER

Deutscher Wetterdienst, Offenbach, Germany

(Manuscript received 26 February 2019, in final form 5 September 2019)

ABSTRACT

We investigate the practical predictability limits of deep convection in a state-of-the-art, high-resolution, limited-area ensemble prediction system. A combination of sophisticated predictability measures, namely, believable and decorrelation scale, are applied to determine the predictable scales of short-term forecasts in a hierarchy of model configurations. First, we consider an idealized perfect model setup that includes both small-scale and synoptic-scale perturbations. We find increased predictability in the presence of orography and a strongly beneficial impact of radar data assimilation, which extends the forecast horizon by up to 6 h. Second, we examine realistic COSMO-KENDA simulations, including assimilation of radar and conventional data and a representation of model errors, for a convectively active two-week summer period over Germany. The results confirm increased predictability in orographic regions. We find that both latent heat nudging and ensemble Kalman filter assimilation of radar data lead to increased forecast skill, but the impact is smaller than in the idealized experiments. This highlights the need to assimilate spatially and temporally dense data, but also indicates room for further improvement. Finally, the examination of operational COSMO-DE-EPS ensemble forecasts for three summer periods confirms the beneficial impact of orography in a statistical sense and also reveals increased predictability in weather regimes controlled by synoptic forcing, as defined by the convective adjustment time scale.

1. Introduction

Convection-permitting numerical weather prediction (NWP) models underpin a step change for operational forecasting centers in their struggle to predict thunderstorms and convective precipitation (Clark et al. 2016) as they allow some key issues to be addressed. First, the intrinsically limited predictability of the small scales, including convection, necessitates the use of ensembles

to generate probabilistic forecasts and assess their confidence (Lorenz 1969; Slingo and Palmer 2011). Second, those ensembles require refined initial conditions, which can only be obtained by data assimilation (DA) of spatially dense observations on kilometer scales (Johnson and Wang 2016). And third, novel techniques are necessary to verify the forecasts and assess their skill with observations of high spatial and temporal resolution (Cintineo and Stensrud 2013).

Estimates of the forecast horizon of storm-scale features remain rather pessimistic, being on the order of only a few hours (Hohenegger and Schär 2007a,b; Zhang et al. 2015, 2016). However, there are steady Earth surface features such as orography or transient dynamical forcing patterns such as weather regimes that potentially

[✉] Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/MWR-D-19-0045.s1>.

Corresponding author: Kevin Bachmann, kevin.bachmann@lmu.de

provide the means to extend those predictability estimates (Anthes 1986). The prevailing synoptic weather regime exerts a decisive influence on the predictability of convective precipitation. In case studies, Hanley et al. (2011, 2013) and Barrett et al. (2015) showed how the larger-scale environment impacts the predictability of convective precipitation. Kühnlein et al. (2014) found systematic differences in the predictability of convection across Germany in situations of local versus synoptic forcing in summer 2011 and showed a connection to the orography in case of dominant local forcing.

The present study focuses on the effects of orography and radar data assimilation on practical predictability, where practical predictability is defined as the “ability to predict based on the procedures currently available” (Lorenz 1969; Melhauser and Zhang 2012). Similar to Surcel et al. (2015), we further distinguish the predictability of the model state, measured by ensemble dispersion, and the model predictability of the atmospheric state, incorporating a comparison to observations. The latter therefore includes model and observing system errors and represents a classic practical predictability estimate, while the predictability of the model state describes how atmospheric processes influence the model dispersion in model space independent of model deficiencies.

To reduce the complexity of the problem, Bachmann et al. (2019, B19 hereafter) performed numerical experiments with an operational convective-scale forecasting system in an idealized configuration and showed that orography increases the predictability of the model state of deep convection. Furthermore, the assimilation of radar observations can yield a noticeable beneficial impact, extending the forecast horizon by several hours. In the present study, we add synoptic-scale uncertainties to the idealized model configuration to refine the predictability estimates for deep convection. In an attempt to bridge the gap between this idealized approach and the real world, we complement the results from the idealized simulations with forecasts of real events using the same NWP system. In particular, we perform experiments with the COSMO-KENDA system using different observational data in the assimilation process for a 14-day period with intense convection across Germany in summer 2016. We apply spatial metrics, namely believable and decorrelation scale, to assess predictability. Finally, the findings regarding the effects of different weather situations are reviewed using operational forecasts of the COSMO-DE-EPS ensemble prediction system, which provides a much larger dataset, albeit with a simpler data assimilation method.

Using a state-of-the-art ensemble-based data assimilation and forecasting system at convective scale

combined with sophisticated verification metrics, we specifically address the following questions:

- What is the effect of synoptic-scale uncertainty on the predictability of deep convection in an idealized environment?
- What is the impact of radar data assimilation on the practical predictability of convective precipitation?
- Does orography increase the practical predictability of deep convection over Germany in a realistic forecasting configuration?
- What is the effect of the prevailing synoptic-scale weather regime in combination with orography?

The outline of the article is as follows: section 2 describes the ensemble data assimilation and forecasting systems, the setup, and the observations. Section 3 briefly introduces measures and scores used to evaluate the experiments. Section 4 presents the results with a focus on predictable scales in NWP model configurations with different levels of realism. Concluding remarks and a comparison to previous studies is provided in section 5.

2. Model description and experiment design

The impact of various uncertainties on the predictability of convective precipitation is examined using different configurations of the operational convection-permitting COSMO model (Consortium for Small-Scale Modeling, Baldauf et al. 2011) coupled with the kilometer-scale ensemble data assimilation system (KENDA; Schraff et al. 2016), which is based on a local ensemble transform Kalman filter (LETKF; Hunt et al. 2007).

First, the idealized setup of COSMO-KENDA including the assimilation of *synthetic* radar observations is used to investigate the potential impact of radar DA and orography in a perfect model approach excluding model errors (as in B19). New in this setup is the additional consideration of synoptic-scale uncertainty typically represented in lateral boundary condition perturbations in limited-area modeling.

Second, we investigate forecasts over Germany in a realistic forecasting configuration, which allows us to test our findings under real conditions. The state-of-the-art COSMO-KENDA system (operational since March 2017)¹ is run in three different configurations for a 14-day period with recurring, heavy convective precipitation. This period at the end of May and beginning of June 2016 features a convection-favoring weather

¹ In May 2018 the domain size as well as the horizontal and vertical resolution were increased.

TABLE 1. Summary of abbreviations used to describe the model configurations and experiments.

Acronym	Experiment	Description
IDEAL IDEAL-DA	<i>Idealized</i> COSMO-KENDA forecasts with and without radar DA	
	<i>flat</i>	Flat experiment without orography
	<i>oro</i>	Orography experiment with Gaussian-shaped mountain of 1000 m height
	IC	Only small-scale IC errors
DE-DA	IBC	IC errors and additional boundary condition perturbations
	<i>Realistic</i> COSMO-KENDA forecasts DE-DA with different DA setups	
	conv	DA of conventional observations using KENDA
	LHN	KENDA of conventional data and latent heat nudging of radar reflectivity
DE-EPS	3DRad	KENDA of both conventional data and 3D radar reflectivities and radial velocity
	<i>Operational</i> COSMO-DE-EPS based on nudging technique	
		Operational COSMO-DE-EPS ensemble forecasts for the summer periods 2014–16

situation characterized by low thermal stability and low midtropospheric winds that lead to locally extreme rainfall sums exceeding 100 mm day^{-1} (details of the synoptic situation can be found, for example, in [Piper et al. 2016](#); [Zeng et al. 2018](#); [Keil et al. 2019](#)). While the reference DA experiment is purely based on the assimilation of conventional data using KENDA, two other DA experiments make use of information from radar measurements. These two approaches comprise the latent heat nudging technique (LHN; [Stephan et al. 2008](#)) and the direct assimilation of 3D radar data using the same ensemble-based DA system (KENDA) as for the conventional observations.

Finally, the numerical experiments are complemented by operational COSMO-DE-EPS forecasts, covering three summer periods in 2014 to 2016 (403 days in total). Comparison of this big dataset to radar-derived observations allows for an examination dependent on the weather regime in a statistical sense. All model configurations and their specifications are summarized in [Table 1](#) and will be introduced in more detail below.

a. Idealized COSMO-KENDA

In the idealized setup of COSMO-KENDA, observations of *synthetic* radar reflectivity and one wind component are assimilated directly, which is in contrast to the current operational setting. The system was used in a similar setup in previous studies, and the reader is therefore referred to [Lange and Craig \(2014\)](#) and [B19](#) for a more detailed description. The most relevant settings are discussed in the following.

A COSMO (version 5.3) ensemble with 40 ensemble members is computed with a convection-permitting horizontal grid spacing of 2 km. The domain encompasses $256 \times 256 \times 50$ grid points with periodic lateral boundaries resulting in a total domain size of $512 \text{ km} \times 512 \text{ km} \times 22 \text{ km}$. A terrain-following Gal–Chen coordinate with a layer thickness ranging from $\sim 100 \text{ m}$

near the surface to $\sim 800 \text{ m}$ at the domain top is used in the vertical. We apply the standard COSMO parameterizations: the interactive soil model TERRA with surface friction; a single-moment bulk microphysics scheme accounting for cloud water, rainwater, cloud ice, snow, and graupel ([Lin et al. 1983](#); [Reinhardt and Seifert 2006](#)); as well as a two-stream radiation scheme ([Ritter and Geleyn 1992](#)). The asymptotic vertical mixing length of the boundary layer turbulence scheme is set to 500 m.

All idealized experiments are initialized with horizontally homogeneous initial conditions (IC), based on a sounding from Payerne in Switzerland (CH, Radiosonde 06610; [Fig. 1](#)), observed at 1200 UTC 30 July 2007 [[Lange and Craig \(2014\)](#), [B19](#)]. The high convective available potential energy (CAPE) of $\sim 2200 \text{ J kg}^{-1}$ and the vertical wind shear allow for strong and long-lived convection. A mean wind from the southwest of about 5 m s^{-1} at 1500 m above ground persists throughout the simulations. The temperature (vertical velocity) field is initially perturbed by white noise with an amplitude of 0.02 K (0.02 m s^{-1}) within the lowest 100 hPa of the model to break the symmetry in the ICs.

One novelty of this work is the consideration of synoptic-scale uncertainty in a second suite of ensemble experiments. Perturbations in wind, temperature and humidity are added to the initial sounding to represent synoptic-scale variations. Since this uncertainty is typically provided by lateral boundary perturbations in limited-area modeling, we will refer to these as boundary condition (BC) perturbations, although they are not strictly implemented at the boundaries. The BC perturbations consist of height-correlated noise (0.25 K for temperature, 0.25 m s^{-1} for horizontal wind, and 2% for relative humidity) added to the observed sounding ([Fig. 1a](#)). Perturbations on the relative humidity are limited to avoid supersaturation. [Figure 1b](#) provides a scatterplot of CAPE and convective inhibition (CIN) of the individual ensemble

FIG. 1. (a) Initial sounding with the added boundary condition errors. The envelope of the ensemble temperature is shown in orange and the dewpoint temperature in green. The wind barbs show the unperturbed wind. (b) Variability of CAPE and CIN in the initial sounding of the IDEAL IBC experiments for lead times within the first hour. (c) Hodographs of IDEAL IBC flat ensemble below 550 hPa (color coded).

members within the first 60 min lead time and visualizes the differences due to those perturbations. CAPE ranges from 1700 to 2800 J kg^{-1} and CIN from -80 to -25 J kg^{-1} . Both remain within these limits for the first forecast hour, showing that the sounding perturbations, although unbalanced, persist and provide realistic variability. Hodographs of the perturbed ensemble (IDEAL IBC) are shown in Fig. 1c to provide information on the variability of convective organization. These variations lead to more realistic variability in the resulting convection and its timing, visible in Fig. 3 and further discussed in section 4.

A set of experiments (called IDEAL) is started at 0400 UTC in the morning and integrated for 10 h with either only initial condition perturbations (referred to as IC) or initial and boundary condition perturbations (referred to as IBC). Every experiment is performed twice, once without any orography in the domain (referred to as *flat*), and once with a single Gaussian mountain (1000 m height and a half-width of 10 km located at $x, y = (64, 64 \text{ km})$; referred to as *oro*).

A deterministic Nature Run, representing the truth, provides *synthetic* radar observations of reflectivity and the u component of the wind, similar to the radial wind a weather radar is measuring for assimilation and validation. These observations are assimilated every 15 min for a period of 1 h in the idealized framework (resulting in IDEAL-DA). The assimilation period, between 0700 and 0800 UTC, follows after a 3 h free forecast and takes place around the onset of convective precipitation. To account for observational errors, Gaussian noise with a standard deviation of 5 dBZ on reflectivity and 1 m s^{-1} on velocity is added to the observations. Wind observations are limited to regions where reflectivity is

larger than 5 dBZ while zero-reflectivity values (no precipitation) are also assimilated to suppress spurious convection (sensitivities to the observation error are discussed in B19). A constant multiplicative covariance inflation of 1.2 is used to preserve spread. For the assimilation, the observations are averaged spatially to superobservations of $8 \text{ km} \times 8 \text{ km}$ (which reduces the observation error by the square root of the number of observations). The horizontal localization is set to 32 km following Lange and Craig (2014) and Bick et al. (2016), while the vertical localization varies between 0.075 and 0.5 (Schraff et al. 2016). The 40-member ensemble forecasts are initialized from analyses at 0800 UTC and integrated for 6 h.

b. COSMO-KENDA-DE-DA

To simulate convection in a realistic forecasting environment, COSMO (version 5.4h) was used to produce forecasts of a convectively active period in a domain centered over Germany (50°N , 10°E , Fig. 2) encompassing 461×421 grid points with a horizontal grid spacing of 0.025° or 2.8 km. The vertical coordinates are identical to the idealized setup, and the same parameterizations as in the idealized setup are applied. We use an asymptotic vertical mixing length of the boundary layer turbulence scheme of 500 m, as in the idealized configuration, but in contrast to the current operational value of 150 m. This facilitates not only direct comparisons to the idealized setup, but also removes potential biases in temperature and surface pressure (Hanley et al. 2015; Necker et al. 2018; Hirt et al. 2019).

The model setup chosen is configured as similar to the idealized COSMO-KENDA configuration as possible. The main differences comprise the model domain,

FIG. 2. Map of the investigated COSMO-KENDA and the DE-EPS domain centered over Germany and covering parts of central Europe: the altitude above sea level is shown in grayscale and the blue (red) box indicates the subdomain north (south); political borders are displayed in dotted lines for orientation.

the natural variability of weather and the DA cycling. Here DA cycling refers to the procedure of alternating short-term forecasts and performing DA to obtain subsequent analyses from which the next set of forecasts is issued. Starting from 0000 UTC 27 May 2016, the system is cycled with 40 ensemble members using 1 h assimilation windows until 10 June 2016, similar to the operational KENDA system at Deutscher Wetterdienst (DWD). The lateral BCs are provided by the ICON-EU ensemble. A combination of inflation methods [i.e., adaptive inflation, multiplicative covariance inflation, and relaxation to prior perturbations (RTPP; Zhang et al. 2004; Schraff et al. 2016)] is applied. Additionally, model errors are added to the analysis ensemble members, using the climatological error covariance matrix $\mathbf{B}$ of the global 3D-Var system (Zeng et al. 2019, manuscript submitted to *Mon. Wea. Rev.*).

In the present study, ensemble forecasts with 20 members each are initialized hourly between 1000 and 1800 UTC every day from 27 May to 9 June 2016 and run for 6 h. Ensemble analyses are provided by three independent DA experiments using different observational data with different techniques, resulting in 126 ensembles and 2520 individual forecasts each.

In a reference experiment (called DE-DA-conv), only conventional data are assimilated using KENDA. Conventional observations comprise data from surface stations, airplanes (including MODE-S data), wind profilers and radiosondes. Two other experiments are additionally performed also utilizing radar observations

via LHN (DE-DA-LHN) or radar reflectivity and radial velocity with KENDA (DE-DA-3DRad).

1) LATENT HEAT NUDGING (DE-DA-LHN)

In DE-DA-LHN observed radar information is used with LHN, while conventional observations are assimilated using KENDA. In this method, surface precipitation rates derived from radar reflectivity observations are assimilated employing the nudging technique (Stephan et al. 2008). LHN adds a heating or cooling term to the prognostic equations in case the forecast precipitation does not match the observed precipitation at any specific location. If precipitation is underestimated, the term will be positive, thereby representing latent heat release in a convective cell and vice versa. LHN is applied in the analysis cycle in every ensemble member at every time step, using interpolated radar data from the precipitation scan (available every 5 min), while the radar wind observations are discarded. Since information provided by the Operational Programme for the Exchange of Weather Radar Information (OPERA) network is used, LHN can be applied in the entire model domain beyond the coverage of the German radar network. LHN has been used operationally in COSMO at DWD since 2007.

2) 3D RADAR DA (DE-DA-3DRAD)

In this setup, conventional as well as radar observations are directly assimilated using KENDA. Model equivalents of 3D radar data as measured by the German radar network are obtained from model states using the Efficient Modular Volume scanning Radar Operator (EMVORADO; Zeng et al. 2014, 2016) every hour. The assimilation of reflectivity and the radial wind is performed following Bick et al. (2016). Additionally, warm bubbles can be triggered in each ensemble member independently every 15 min during the assimilation to account for observed convective cells missing in the model representation (Zeng et al. 2019, manuscript submitted to *Mon. Wea. Rev.*). Superobservations for the radar measurements and the model equivalents are obtained for each elevation of the used volume scan independently, using a $10 \text{ km} \times 10 \text{ km}$ Cartesian grid. The horizontal localization length scale is set to 16 km, while the vertical localization is identical to the idealized configuration and varies between 0.075 and 0.5 (Schraff et al. 2016).

c. COSMO-DE-EPS

Finally, COSMO-DE-EPS represents the convective scale ensemble forecasting system at DWD operational since 2012. Before the introduction of KENDA,

providing an ensemble of analyses from 2017 onward, the analysis of COSMO-DE-EPS relied on the deterministic COSMO-DE analyses employing LHN. The ensemble was generated by introducing uncertainties in the initial and the lateral boundary conditions as well as a combination of physics parameter perturbations to mimic model error (for details see Gebhardt et al. 2011; Peralta et al. 2012; Ben Bouallègue and Theis 2014; Kühnlein et al. 2014). Here we exploit 20-member ensemble forecasts initialized at 0000 UTC for a lead time of 27 h in the summer months May to September over the years 2014 to 2016. The horizontal grid spacing of 2.8 km allows representing deep convection explicitly, while shallow convection is parameterized. The COSMO-DE-EPS domain is identical with the DE-DA domain (Fig. 2), and the same parameterizations of physical processes as in the other configurations are applied.

d. Observations

We compare the forecasts to the so-called EY product of DWD, which provides radar-derived accumulated precipitation over central Europe with a horizontal resolution of 1 km $\times$ 1 km every 5 min. The EY product was interpolated to the COSMO grid and the precipitation was accumulated hourly to facilitate comparisons to the forecasts. It covers the entire investigation domain of this study.

3. Metrics and regime classification

The present study focuses on the practical predictability of convective precipitation and applies elaborate neighborhood and spectral verification methods to deal with issues as double penalties or the fact that a forecast of a slightly displaced convective cell might still be considered valuable by a forecaster. Those metrics and the convective adjustment time scale used to categorize the weather regime are introduced in the following section.

a. Predictability metrics

The decorrelation and believable scale are scale-separation methods, which both allow for displacements in the forecasts and scale-dependent skill evaluation (Casati et al. 2004). They complement each other as the decorrelation scale is solely based on the ensemble forecast, while the believable scale incorporates a comparison to observations. In other words, the decorrelation scale is a measure for the spatial spread and the predictability of the model state. The believable scale is a score and therefore assesses the model predictability of the atmospheric state. Both methods

are expressed in [km], which eases comparisons and interpretation.

The decorrelation scale is based on the variance of the ensemble forecasts in the following way (Surcel et al. (2015), B19): If a variable Ψ_k is fully decorrelated in all ensemble members, then the covariance $\text{Cov}(\Psi_k, \Psi_l) = 0$ and the following ratio would equal 1:

$$R(\lambda) = \frac{\sum_{k=1}^N \text{Var}_k(\Psi_k)}{\text{Var}_k\left(\sum_{k=1}^N \Psi_k\right)}, \quad (1)$$

where k, l represent ensemble members. Since the variance can be calculated by the discrete cosine transform (DCT), a scale dependence is introduced (Denis et al. 2002) and the ratio $R(\lambda)$ depends on the wavelength λ . We call the scale at which $R(\lambda)$ drops below 0.95 the decorrelation scale and assume that all smaller scales are unpredictable. A visualization of this method is a low-pass filter with a cutoff at the decorrelation scale, which yields a smoothed field containing only the information within the ensemble.

The believable scale (Dey et al. 2014, 2016) is an extension to the Fractions Skill Score (FSS, Roberts and Lean 2008; Mittermaier and Roberts 2010; Mittermaier et al. 2013; Schwartz and Sobash 2017), which ranges from 0 to 1, where 1 is a perfect forecast. It is defined as the scale at which $\text{FSS} = 0.5 + f_0/2$, where f_0 is the observed precipitation coverage. The FSS based on a percentile threshold diminishes the influence of different model biases or situations with different precipitation amounts. According to Roberts and Lean (2008), forecasts are reasonably skillful and useful above the believable scale.

The treatment of lateral boundaries poses a challenge to all neighborhood methods as window sizes must be constrained to a fraction of the maximum extent of the domain and boundary effects are introduced. For the idealized experiments in the present study, we calculate both scores on the periodic domain without further assumptions. In the realistic experiments, the decorrelation scale is computed with the DCT, which provides aliasing; for the believable scale, we pad the fields with zeros, as proposed in Roberts and Lean (2008).

b. Regime classification

The convective adjustment time scale τ_c is an indicator of the prevailing weather regime based on how fast conditional instability (measured by CAPE) is removed from the atmosphere by moist convection (measured by

FIG. 3. Horizontal cross sections of five representative ensemble members of (a) the IDEAL IC oro and (b) the IDEAL IBC oro experiments at 1 h lead time (0900 UTC); the column-maximum reflectivity is color coded, the horizontal wind at 1500 m is indicated by gray streamlines and the black circles indicate the orography.

precipitation). A convective time scale that is much smaller than the time scale on which the synoptic environment develops indicates an equilibrium state. Deep convection is therefore predominantly controlled by the synoptic-scale forcing. In contrast, if τ_c is large, no such synoptic forcing is present and the convective instability depends on local triggering mechanisms to release convection. The metric was successfully applied as an indicator of the weather regime in previous studies (e.g., Done et al. 2006; Zimmer et al. 2011; Surcel et al. 2017). The domain average of the convective time scale τ_c was computed for the hourly model output in the following way (Done et al. 2006):

$$\tau_c = \left(\frac{c_p \rho_0 T_0}{2L_v g} \right) \frac{\text{CAPE}}{P}. \quad (2)$$

All quantities within the bracket are constants (the specific heat of air at constant pressure c_p , the reference density ρ_0 and temperature T_0 , the latent heat of vaporization L_v and the gravity acceleration g). Only CAPE (J kg^{-1}) and the precipitation rate P [$\text{kg (m}^2 \text{s)}^{-1}$] vary. In the present study, we use a threshold of 6 h on the daily area averaged maximum of τ_c to distinguish between local and synoptic forcing situations (Keil et al. 2019). However, our results are robust to reasonable variations (3–12 h) of this choice.

4. Results

This section presents results on the practical predictability of convective precipitation in the presence of

orography obtained from the range of COSMO configurations introduced above. Furthermore, we investigate the impact of radar DA, as well as the interconnection of orography and the prevailing weather regime.

In the idealized setup, we focus on the effect of IC and IBC perturbations, as well as on the potential impact of radar DA on the forecast in the presence of orography in a perfect model environment. In a step toward a more realistic forecasting setting, we applied the same metrics to a 14-day period that includes strong convection over Germany in summer 2016. For this period, we investigate the impact of orography, the importance of the dynamical control and the model performance using different DA schemes. Finally, we complement our findings regarding the impact of orography and the synoptic weather regime with results based on operational ensemble forecasts of DE-EPS for three summer periods from 2014 to 2016.

a. Idealized COSMO-KENDA experiments

Before we quantify the impact of synoptic-scale uncertainty on the predictability of convection, we provide an impression of its effect in the presence of orography. Figure 3 shows the development of IDEAL oro with the column-maximum reflectivity valid at 0900 UTC (1 h lead time). Figure 3a depicts five representative ensemble members started from IC perturbations only. In all five IDEAL IC oro members, convective cells of similar size can be seen throughout the domain, but the members differ on the exact location of the convection. In contrast, Fig. 3b, depicting

five IDEAL IBC *oro* members with additional boundary condition errors, shows a range of convective scenarios. Those range from evenly distributed, predominantly small or large cells to members, where the strong convection is concentrated in the lee of the mountain (gray streamlines show the southwesterly wind). In addition to the spatial variability, the IDEAL IBC experiments (*oro* and *flat*; not shown) show higher temporal variability, with the onset of convection shifted by up to 2 h. Hence, additional synoptic scale uncertainties add variability and realism to the ensemble forecasts that was lacking in B19.

Figure 4 shows the temporal development of the decorrelation scale for the experiments performed in the idealized setup. The results of the IC experiments have already been published in B19, but are included to provide a frame of reference in the current study. We will briefly summarize the most relevant results and refer the reader to B19 for further details, metrics and experiments.

Without DA, the decorrelation scale decreases from the domain size of 512 km (IDEAL IC *flat*) to values of 10 km (~ 60 km) at initial time (at 3–6 h lead time) for IDEAL IC *oro*, indicating a higher predictability of deep convection in the presence of orography. DA of *synthetic* radar observations decreases the decorrelation scale further to 4 km at initial time (IDEAL-DA IC). The decorrelation scale gradually increases with lead time to about 30 km, which translates to an increased forecast horizon of several hours. Herein, forecast horizon differences denote differences in lead times at which different forecasts attain the same value of a verification metric. Note that there is no significant performance difference between the flat and the mountain case in the experiments with DA, since DA of radar information strongly constrains the initial states (IDEAL-DA IC *oro* vs IDEAL-DA IC *flat*).

The believable scale provides a different perspective as (*synthetic*) observations are included in its computation. It is calculated for a threshold of 19 dBZ of the column-maximum reflectivity, which translates to a rain rate of roughly 1 mm h^{-1} (Fig. 5). The believable scale of the IDEAL IC experiments increases from roughly 40 km at 0 h lead time to 100 km at 6 h lead time for IDEAL IC *oro* (to 125 km for IDEAL IC *flat*). Hence, the steady lower boundary forcing of the orography increasingly dominates with time compared to the IC uncertainty. Radar DA reduces the believable scale throughout the forecast by about 50 km in IDEAL-DA IC, again eliminating differences of *flat* and *oro*. Therefore, the radar DA increases the forecast horizon measured by the believable scale by more than 6 h in these experiments.

FIG. 4. Time series of the decorrelation scale of the idealized experiments with only IC perturbations, previously published in B19, and additional experiments with IBC perturbations.

In the absence of model errors in the idealized setup, the decorrelation and believable scale both address the predictability of the model and atmospheric state. In a perfect model, the ensemble can always reproduce the truth, in contrast to the real atmosphere. This is a crucial difference to the DE-DA experiments, which will be discussed later on.

1) EFFECT OF SYNOPTIC-SCALE UNCERTAINTY

The impact of additional synoptic-scale uncertainty on the predictability of deep convection is investigated comparing the IC to IBC experiments (Fig. 4). In IDEAL IBC *flat*, additional perturbations have no impact on the decorrelation scale as the simulations are already decorrelated to the domain scale. With values of 40 km, the decorrelation scale of IDEAL IBC *oro* is higher than that of IDEAL IC *oro* (reddish thick lines in Fig. 4). This difference persists throughout the forecast and amounts to 50 km after 6 h (50 km IDEAL IC *oro* vs 100 km IDEAL IBC *oro*), confirming our visual impression (Fig. 3). The differences highlight the detrimental impact of synoptic-scale uncertainty on the predictability of convection.

The effect of the synoptic-scale uncertainties measured by the believable scale can be seen in Fig. 5b. As expected, both IDEAL IBC experiments (*oro* and *flat*) are degraded in comparison to the IDEAL IC reference. Their believable scale remains between around 100 and 125 km for the entire lead time with a standard deviation of the ensemble members of 50 km. The fact that the mean and the standard deviation exhibit no trend over time indicates that the believable scale approaches its limit for the displacement of convection in the given weather situation. Its value saturates to a climatological value determined by the average initial and boundary conditions. This interpretation is in line with the

FIG. 5. Time series of the believable scale for a threshold of 19 dBZ of the idealized experiments: (a) experiments with only IC perturbations, previously published in B19 and (b) experiments with IBC perturbations. The shading represents the $\pm\sigma$ intervals of the distributions.

small trend of the decorrelation scale of IDEAL IBC *oro* in Fig. 4.

We find that synoptic-scale uncertainties render the idealized experiments more realistic and degrade the forecasts without radar DA. Additionally, the synoptic-scale uncertainties seem to reduce the positive impact of the orography on the predictability of precipitation (cf. IDEAL IC *oro*/IDEAL IC *flat* to IDEAL IBC *oro*/IDEAL IBC *flat* in Fig. 5). This result is in agreement with Nuss and Miller (2001) and Picard and Mass (2017), who found that orographically influenced local precipitation patterns are sensitive to perturbations of the synoptic-scale flow direction.

2) RADAR DA AND SYNOPTIC-SCALE UNCERTAINTY

All *oro* and *flat* DA experiments (both IC and IBC) perform equally well in decorrelation and believable scale indicating the dominant impact of radar DA (thin lines in Figs. 4 and 5). The direct DA of high-quality radar observations in our idealized configuration (i.e., *synthetic* observations) outweigh the orographic impact and drastically improve the predictability limits. Therefore, we do not distinguish between *oro* and *flat* experiments in the remainder of this subsection.

Radar DA shows a profound beneficial impact on the forecast skill in the idealized experiments with only IC uncertainties (Fig. 4). Similarly, only the smallest resolved scales of IDEAL-DA IBC including synoptic-scale uncertainty become decorrelated and therefore unpredictable at initial time. The difference between IDEAL-DA IC and IDEAL-DA IBC remains less than 10 km for the entire forecast duration of 6 h, which translates to a reduced forecast horizon of about 1 h. The experiments with DA clearly outperform the experiments without DA and increase the forecast horizon by 6 h.

The application of the believable scale confirms the strong impact of radar DA. Radar DA improves the skill from a climatological level to scales of $\mathcal{O}(10)$ km at 0 h lead time and $\mathcal{O}(100)$ km at 6 h lead time (Fig. 5). Comparing the IDEAL-DA IC and IDEAL-DA IBC experiments, the believable scale is almost identical at initial time, but grows faster with increasing time and exhibits a larger standard deviation within the ensemble. After 6 h lead time, the believable scale of IDEAL-DA IBC attains values of 100 km, while IDEAL-DA IC still remains below 50 km. Believable and decorrelation scale show a similar behavior, but the believable scale provides additional information on the intra-ensemble variability.

The temporal extrapolation of the believable scale curves of IDEAL IBC and IDEAL-DA IBC suggest a convergence after about 6.5 h. Thus, the impact of radar DA lasts that long in the idealized configuration including synoptic-scale uncertainty, which agrees reasonably well with earlier realistic forecasting studies (Vié et al. 2011; Kühnlein et al. 2014). After this time, the beneficial effect of improved ICs is lost and the remaining forecast skill is a climatological one in the sense that it stems from the boundary conditions and forcings.

Our study shows that direct radar DA provides a promising approach to increase the practical predictability of deep convection on lead times of several hours. The idealized experiments including DA outperform the free forecasts by about 40 km in case solely initial uncertainties are present and by up to 80 km if synoptic-scale uncertainties are added. The KENDA system proved to be effective in correcting the position of deep convection and is also able to correct the detrimental impact of uncertain boundary conditions that a limited-area model inevitably inherits from the driving larger-scale model. In an idealized setup without model errors, COSMO then propagates and develops the

FIG. 6. Time series of (a) the decorrelation scale and (b) the believable scale, both for the 75th percentile of the precipitation as threshold for the DE-DA experiments (between 27 May and 10 Jun 2016).

convection properly. Synoptic-scale uncertainties reduce the predictability limits by about 1 h in the idealized COSMO-KENDA setup, but, provided radar DA is applied, still yield remarkably valuable forecasts.

b. Realistic COSMO-KENDA experiments

COSMO-KENDA data assimilation and forecasting experiments for real weather situations are performed for the convectively active high impact weather (HIW) period between 27 May and 10 June 2016 to put the findings of the idealized experiments in the context of a realistic forecasting scenario. In contrast to the previous section, the metrics are applied to the hourly aggregated precipitation fields. The time aggregation will impact the scores slightly positively as it compensates for small timing errors. Apart from that, scores for hourly precipitation instead of instantaneous column-maximum reflectivity showed comparable results (Stratman et al. 2013). We compare two assimilation approaches, namely LHN using radar-observed surface rain rates and direct assimilation of the 3D radar reflectivity and radial wind in KENDA, to a reference experiment using only conventional observations in KENDA.

1) OROGRAPHIC IMPACT ON PREDICTABILITY

To assess the orographic influence on convection we divide Germany in a southern and a northern part to represent regions that are strongly and weakly influenced

by mountains, respectively [similar to Seifert et al. (2012) and Kühnlein et al. (2014), Fig. 2].

Figure 6a depicts the development of the decorrelation scale with lead time in the DE-DA experiments. As the forecasts are initialized hourly every day of the HIW period between 1000 and 1800 UTC and run for 6 h, the Figure depicts averages over lead times that are valid for different times of the day. Forecasts initialized later tend to perform slightly better as more radar information is used (not shown). For the DE-DA-conv experiment, the decorrelation scale grows continuously from approximately 10 km (25 km) at 1 h lead time to around 50 km (70 km) after 6 h in the southern (northern) part. In our study, the more orographically influenced south shows significantly lower values in all forecasts and the difference is increasing with lead time. We interpret this as an indicator for increased predictability of the model state of deep convection in southern Germany due to orography.

This line of argumentation is corroborated by the fact that a significant portion of deep convection in other midlatitude regions of the world can be linked to orographic features (Lin et al. 2001). Carbone and Tuttle (2008) found that about 60% of midsummer rainfall between the Rocky Mountains and the Appalachians are caused by propagating rain systems triggered by elevated terrain. Levizzani et al. (2010) showed similar results for the Mediterranean. Convective precipitation over orography is, in a probabilistic sense,

more constrained in those regions favored by the interaction of orography and, for example, synoptic-scale flow or radiation, which increases the predictability.

While the decorrelation scale only assesses intra-ensemble variability and involves no comparison to observations, the believable scale offers additional information about the practical predictability, more specifically the model predictability of the atmospheric state (Fig. 6b). It is computed for the 75th percentile precipitation, a variable threshold to account for spatial and temporal differences in rainfall intensities during the 14-day period in 2016.

Similar to the decorrelation scale, the believable scale initially grows, but seems to start saturating after only 2–3 h lead time for the DE-DA-conv experiment. However, the believable scale is larger (by a factor of 2–3) than the decorrelation scale due to model errors (revealed by comparison to observations) that pose additional challenges to the forecasting system. The difference between the orographic south and the comparatively flat north is still visible and amounts to more than 100 km.

The believable scale highlights that the skill of short-range forecasts and the predictability of the atmospheric state over southern Germany is increased by at least 6 h due to orography. It also shows that the skill of state-of-the-art thunderstorm forecasts in central Europe converges toward a regime-dependent climatological value governed by synoptic-scale dynamical control and the geographical region within only a couple of hours.

2) IMPACT OF RADAR DA

To estimate the impact of the assimilation of radar observations in realistic conditions and to assess the impact of a specific DA method, we compare a reference without radar DA (but including the KENDA assimilation of conventional observations, DE-DA-conv) to the operational LHN technique (DE-DA-LHN) and the assimilation of 3D radar data in KENDA in a realistic yet experimental setup (DE-DA-3DRad).

In both experiments using radar data, the position of convection, measured by the decorrelation scale (Fig. 6a), is more constrained than in DE-DA-conv for both northern and southern Germany. The impact of radar DA persists throughout the lead time of 6 h and is stronger in the north. This might be due to the presence of orography in the south, which already constrains the position of convection, thereby limiting the positive impact of radar DA. Furthermore, radar observations over mountainous terrain are affected by orographic blocking, thus deteriorating the quality of the assimilated observations.

FIG. 7. Comparison of the total frequency of observed and forecasted reflectivities in IDEAL-DA-IBC and DE-DA-3DRad experiments. Counts for dBZ ≤ 0.3 are not shown.

The believable scale shows a similar beneficial impact of radar DA in comparison to observations. Radar DA improves the believable scale by about 50 km after 1 h lead time to 50 km in the south and 170 km in the north. The believable scale of DE-DA-LHN and DE-DA-3DRad approach the value of DE-DA-conv, resulting in comparable skill after 6 h. For early lead times, however, the operational radar DA (DE-DA-LHN) provides a forecast horizon increase of 2–3 h (around 1 h for DE-DA-3DRad) compared to DE-DA-conv. The comparably simple LHN approach, which does not incorporate wind observations, outperforms the direct radar data assimilation indicating that there is still room for improvement. Furthermore, the forecast improvement with both approaches is smaller than in the idealized setup.

A direct comparison of the findings in the realistic setup, using real observations, to the idealized configuration shown in Fig. 5b (IDEAL-DA IBC *flat/oro*), utilizing *synthetic* observations for assimilation and verification, has to be performed with caution. In contrast to the ideal case, we find a significant difference in the believable scale of the north (cf. *flat*) and the south (cf. *oro*), independent of the usage of radar data. However, in the DE-DA experiments, a variety of transient meteorological situations occur across a geographical region in central Europe with complex orography. The believable scale in the idealized configuration amounts to $\mathcal{O}(10\text{ km})$ while Fig. 6b shows scales of $\mathcal{O}(50\text{ km})/\mathcal{O}(100\text{ km})$ in the south/north in the realistic DE-DA experiments. This gap may be attributed to the design of the idealized experiments including more dense observations, no model and no forward operator error, horizontally homogeneous atmospheric conditions and a single Gaussian mountain.

To shed some light on potential reasons for the differences of the IDEAL-DA and DE-DA experiments,

Fig. 7 displays the frequency of observed and forecast reflectivities in both configurations. In the DE-DA experiments, model equivalents for reflectivities are obtained from the model forecast using the EMVORADO forward operator. In the idealized experiments, the reflectivity is computed from mixing ratios of graupel (QG), rain (QR), and snow (QS), using the simpler approach of [Done et al. \(2004\)](#) at each model grid point (as in [Lange and Craig 2014](#); [Bachmann et al. 2019](#)).

Figure 7 shows that observations and model equivalents in IDEAL-DA are in good agreement, due to the perfect-model approach. The small discrepancies at about 60 dBZ are caused by the added observation errors. In contrast, the DE-DA experiments exhibit differences, particularly in the occurrence of higher reflectivities. One component of the model that limits the forecasted reflectivity is the single-moment microphysics scheme. Nonetheless, these cases are comparatively rare during the investigated 14-day HIW period in the DE-DA experiments. Although statistics for both sets of experiments are shown in Fig. 7, a direct comparison is difficult as the computations reflect a different level of realism. However, we hypothesize that major differences regarding radar DA in the idealized setup and the realistic forecast configuration are the absence of systematic deficiencies in the model, approximations in the observation operator, and potentially correlated observation errors that are not accounted for in KENDA and that are not represented in the idealized experiments.

The differences between the three realistic forecast experiments is illustrated in Fig. 8 for a representative case based on radar observations of the German radar network and model equivalents obtained by applying the EMVORADO operator to the forecasts. We focus on lead times of 1, 2, and 3 h (first to third row) for the experiments DE-DA-conv (2nd column), DE-DA-LHN (third column) and DE-DA-3DRad (fourth column) issued from the corresponding analyses at 1100 UTC 1 June 2016, complemented by radar observations (1st column) at valid time. In the fourth row, the probability of the ensemble forecasts to exceed a threshold of 20 dBZ in the 3 h forecast is displayed.

For all forecast lead times, the positive impact of either LHN or 3D radar data KENDA is apparent. Both perform visually similar for a lead time of 1 h, representing the prominent feature over central Germany. This feature is also present in DE-DA-conv, but the agreement with the actual observations is poor. Notably, for this lead time LHN leads to false positive signals in the East of the Netherlands. For 2 h lead time, the simulated DE-DA-3DRad reflectivities are more comparable with radar observation over Germany,

while the high reflectivities over northern Germany are captured better by DE-DA-LHN. After 3 h, DE-DA-3DRad still predicts the features over central Germany more accurately compared to the other forecasts, while the isolated cells over northern Germany appear better in DE-DA-LHN.

This different behavior can be attributed to the different nature of both radar DA approaches. LHN uses a forcing term in the prognostic equations that can also produce precipitation and convection in places where none is observed. In contrast, 3D radar DA in KENDA lacks the potential to create new cells as the analysis is a locally weighted linear combination of the ensemble members. As a result, ensemble methods have a disadvantage in terms of creating new cells without specific inflation techniques, such as the warm bubbles.

However, we find that the DE-DA-LHN ensemble forecasts are more constrained than DE-DA-3DRad forecasts (fourth row of Fig. 8), displaying larger maximum probabilities. The use of LHN in combination with ensemble methods for other observations can lead to a suboptimal spread-skill-ratio, especially for forecasts of precipitation, which are often under-dispersive. In this case study, DE-DA-3DRad gives a better representation of the overall structures, whereas DE-DA-LHN predicts isolated cells more accurately, although some false positives are apparent.

c. Impact of weather regime

The practical predictability of convective precipitation and the impact of the lower boundary condition like orography on the initiation of convection considerably depend on the prevailing synoptic-scale weather regime ([Keil et al. 2014](#); [Kühnlein et al. 2014](#); [Baur et al. 2018](#)). Application of the convective adjustment time scale allows a weather regime dependent evaluation ([Keil et al. 2019](#)). The diagnostic identifies 5 days within the 2-week period to be dominated by locally forced conditions: 28 May and 4–7 June.

Figure 9 depicts the believable scale of the DE-DA-3DRad experiments for the north and south of Germany for these days. A comparison to the believable scale of the entire HIW period (Fig. 6b) clearly shows higher values (i.e., lower predictability) during locally forced conditions across both domains at all forecast lead times (e.g., 210 km vs 120 km at 3 h lead time in the south). At short lead times of less than 3 h, there is a beneficial effect of orography on the predictability of convection during locally forced weather regimes indicated by the believable scale amounting to 80 km after 1 h lead time in the south. Across the northern flat region, the believable scale reveals no skill (values larger than 250 km).

FIG. 8. Radar observations (first column) and model equivalents (second column) for the experiments DE-DA-conv, (third column) DE-DA-LHN, and (fourth column) DE-DA-3DRad started at 1100 UTC 1 Jun 2016. (top three rows) Lead times of 1–3 h. (fourth row) Probabilistic 3 h ensemble forecast. The model equivalents translate to the 0.5° elevation (lowest elevation in volume scan).

Additionally, we extend our weather regime-dependent investigation on a larger dataset to put our findings on more solid statistical grounds. Unfortunately, the only large dataset of COSMO ensemble forecasts currently available builds on a different DA method (see [section 2c](#)).

Based on the convective adjustment time scale, around 30% of the 403 days are classified as locally forced cases.

[Figure 10a](#) depicts the diurnal cycle of the decorrelation scale averaged over the complete period. As forecasts are started at 0000 UTC, the lead time and the time

FIG. 9. As in Fig. 6b, but only for the 5 days with locally forced weather regime (28 May, 4–7 Jun) of the DE-DA-3DRad experiments (for details see text).

of the day are identical. Within the first 12 h lead time, before local noon, the decorrelation scale remains below 30 km for both synoptic regimes. After that, the decorrelation scale grows rapidly. We attribute the growth of forecast errors between 1200 and 1800 UTC to the more frequent occurrence of deep convection following the diurnal cycle that contributes strongly to error growth in the atmosphere (Zhang et al. 2007; Selz and Craig 2015; Sun and Zhang 2016). Especially in situations characterized by local forcing, the decorrelation scale grows rapidly and reaches about 60 km at the end of the forecast. The predictable scales become three times as large in the absence of synoptic forcing organizing convection. Keil et al. (2014) and Kühnlein et al. (2014) also described this higher practical predictability during synoptic control.

The increased predictability of the model state in the mountainous south caused by orography also becomes evident in this dataset. While the decorrelation scale is very similar at early lead times, the differences grow

with the development of convection. For both synoptic situations, the south depicts a decorrelation scale that is roughly 20 km smaller than the one in the north, which means those scales remain predictable for a longer time in the mountainous south.

The believable scale highlights different aspects than the decorrelation scale (Fig. 10b). The comparison of forecasts with observations reveals a rapid loss of predictability of the atmospheric state within the first 3 h, a period during which the decorrelation scale suggests a high agreement within the ensemble. The small decorrelation scales indicate a too small ensemble spread that is not representative of the uncertainty of the situation and does not capture the observed precipitation. After the initial rapid growth period from 50 to 100 km in the synoptic regime (170 to 200 km in the local forcing regime), the believable scale approaches an upper limit that represents the climatological predictability of deep convection in summer in central Europe.

In contrast to the DE-DA experiment for the HIW period, this limit is dominated by the weather regime in the operational DE-EPS forecasts, while the orography plays only a minor role in both regimes (only significant differences until 1000 UTC in Fig. 10b determined by bootstrapping during synoptic forcing). The above-mentioned upper limit for local forcing regimes is insensitive to the percentiles the FSS believable scale is based on. For the synoptic regime, the choice of the threshold shifts the saturation level, but the characteristics remain unchanged (see online supplemental material).

Finally, we want to highlight a similarity between the realistic and idealized experiments. Note that the sounding we based the idealized simulations on represents a typical locally forced situation. Interestingly, the

FIG. 10. Mean time series of (a) the decorrelation scale and (b) the believable scale for the 75th percentile of precipitation as threshold of the operational COSMO-DE-EPS forecasts in summer 2014–16. As forecasts are started at 0000 UTC, the lead time and the time of the day are identical. The shading represents the 95% confidence intervals determined by bootstrapping ($n = 10\,000$). Only cases with more than 100 precipitating grid boxes are considered in (a),(b).

DE-EPS forecasts share similar features in decorrelation scale² during local forcing conditions with IDEAL-DA IBC *oro* forecasts exhibiting a similar growth rate that amounts to roughly 50 km within 6 h (1200–1800 UTC, Fig. 4 vs Fig. 10a) during the convective active period.

5. Conclusions

The examination of the predictability of convective precipitation using convective-scale ensemble prediction systems represents a complex challenge. The various components of such modern NWP systems must address the intermittent, highly variable nature of convection influenced by persistent terrain features and the changes in large scale conditions. The efficient assimilation of spatiotemporally highly resolved observations to provide a sound estimate of the initial state as well as scale-dependent quality measures are crucial for investigating predictability limits. Nowadays, computational resources and numerical tools are available allowing to quantify hypotheses that the presence of orography and the prevailing weather regime can potentially extend the predictability limits (see, e.g., Anthes 1986). In the present paper, we investigate the predictability of deep convection using the COSMO model with KENDA at different levels of complexity and try to bridge the gap from an idealized case study to a quasi-operational NWP configuration.

First, we extend the study of B19 by adding synoptic-scale uncertainties of temperature, relative humidity and wind to represent significant error sources in convective-scale NWP (Gustafsson et al. 2018) in our idealized COSMO-KENDA setup. These vertically correlated perturbations represent the lateral boundary condition uncertainty in operational ensemble systems and lead to a decrease in spatial predictability to a scale of about 100 km (in FSS believable scale). Although the general convective environment is correctly predicted, it is impossible to predict the location of individual convective cells without DA in the idealized framework. Including radar DA improves the forecasts in the idealized setup immensely and throughout the lead time of 6 h. The presence of synoptic-scale uncertainties reduces the forecast horizon by only one hour compared to DA experiments with only IC uncertainty when using the decorrelation scale, a metric to

evaluate the ensemble dispersion (30 km at 5 h lead time, Fig. 4). The assimilation of radar data effectively reduces the displacement errors of convective cells at the initial time and outweighs existing synoptic-scale uncertainties. This emphasizes the potential of the direct assimilation of high-quality radar observations using an LETKF data assimilation system in a perfect-model approach that neglects representativeness, systematic and correlated observation errors, as well as approximations in the numerical forecast model and the forward operator.

To relate the findings of the idealized experiments to real-world predictions, we performed COSMO-KENDA experiments (called DE-DA) in a domain covering Germany for a 14-day period that featured several extreme convective precipitation events. As those experiments are evaluated with radar-derived observations, observation uncertainty and model error are present. Both metrics, the decorrelation scale and the believable scale, indicate increased predictability over the orographic south compared to the comparatively flat north of Germany. We attribute this increase to the presence of orography, which is in agreement with studies based on radar climatologies (Duda and Gallus 2013; Isotta et al. 2014; Kovacs and Kirshbaum 2016). The believable scale, taking observations into account, highlights increased predictability of the atmospheric state across the mountainous south. The fact that the decorrelation scale is smaller than the believable scale by a factor of 2–3 (cf. Fig. 6 and supplemental material) is a strong indicator that the ensemble is persistently underdispersive, a typical finding for precipitation forecasts (e.g., Romine et al. 2014).

The assimilation of radar data also reveals a clear beneficial impact in the real-world experiments. The full 3D radar data assimilation (DE-DA-3DRad) outperforms a reference simulation using conventional observations only by 1–3 h in forecast horizon. A case study shows that DE-DA-3DRad produces superior mesoscale structures, wind fields and probabilistic precipitation forecasts. However, the impact is considerably more short lived than in the idealized environment.

These results are contrasted with forecasts using the currently operational KENDA setup at DWD (direct assimilation of conventional observations plus the LHN assimilation of radar precipitation, DE-DA-LHN). Both systems show an overall comparable forecast quality. Given that DE-DA-3DRad is currently in an experimental status, while the LHN technique has been operational for almost ten years, we see this as an encouraging result. In principle, the full 3D ensemble assimilation utilizes a more sophisticated approach than

²Note that the comparison of the believable scale of idealized and realistic forecasts obscures this similarity due to the absence of a model error in the idealized setup, whereas there are systematic differences between observation and COSMO in DE-EPS.

LHN and assimilates more reflectivity and additional wind observations indicating room for further improvement in the direct assimilation of radar observations.

The gap in the predictability limits between idealized and the real world DA experiments is related to the major differences between the idealized and the realistic system. These include systematic deficiencies in the model, approximations in the observation operator, and uncertainty in the observations. For example, the effect of the parameterization of microphysical processes is a characteristic model error, and simplifications of the radar forward operator used here represent error sources in the observation operator. Mitigation of such deficiencies is of utmost importance for the assimilation of radar observations in real systems.

Finally, our results confirm the importance of the prevailing weather regime. Below-average predictability is found during locally forced weather situations in both regions at all forecast lead times within the 14-day HIW period in 2016. There is a beneficial effect of orography on the predictability of convection in the absence of strong dynamical forcing especially for short lead times of less than 3 h. We find a decisive influence of the weather regime on precipitation predictability exploiting 403 days of multiyear summertime operational 20-member COSMO-DE-EPS forecasts. Situations with synoptic forcing show a believable scale of $\mathcal{O}(100\text{ km})$, which is of the same order as found in our idealized experiments and comparable to previous studies (Schwartz et al. 2009; Clark et al. 2010; Mittermaier et al. 2013), based on FSS and other neighborhood methods. Done et al. (2006, 2012) found that in those situations, area-averaged precipitation tends to be well-predicted, but the location of individual storms is not. However, southern Germany exhibits consistently increased predictability, as measured with the decorrelation scale, compared to northern Germany in our experiments, which indicates that even in situations of low location predictability, orography can act as a trigger providing increased local probability of convection.

This study concurs with previous literature that synoptically forced situations exhibit increased predictability of convection compared to locally forced situations (Done et al. 2006; Keil and Craig 2011; Done et al. 2012), demonstrating that dynamical regimes comprise a source of predictability as hypothesized by Anthes (1986). The largest beneficial impact of orography was found in the late afternoon during the local forcing regime. Local forcing implies a convectively active situation that is primarily dependent on surface triggers to initiate convection, which orography can provide (Fig. 10a).

The idealized and the short-range DE-DA experiments in the HIW period agree on the general scales that remain predictable at certain lead times ($\mathcal{O}(100\text{ km})$ after 6 h). The forecast horizon after which predictability is lost amounts to 5–6 h in the idealized and 2–3 h in the DE-DA experiments. In addition, in the idealized experiments, DA is able to compensate the positive impact of orography on the predictability for several hours and is able to outweigh the uncertainty of the synoptic conditions. We find a smaller, albeit beneficial impact of radar DA for the realistic COSMO-KENDA experiments covering a 14-day period. The fact that both metrics in the different subdomains do not converge to each other suggests further potential to improve the assimilation of radar observations, and subsequent precipitation forecasts. This is especially true for northern Germany, where less orography and therefore fewer triggers of convection are present.

To conclude, we identified a positive impact of orography on the predictability of deep convection in a range of COSMO-KENDA configurations using two different scale-dependent metrics. We showed the importance of the synoptic weather regime on the predictability of deep convection and demonstrated the considerable beneficial potential of radar DA providing guidance on spatial scales, on which forecasts should be considered.

Acknowledgments. The research leading to these results was supported by the Hans-Ertel-Centre for Weather Research (Weissmann et al. 2014; Simmer et al. 2016). This German research network of universities, research institutes, and Deutscher Wetterdienst is funded by the BMVI (Federal Ministry of Transport and Digital Infrastructure). It was also supported within the Transregional Collaborative Research Center SFB/TRR 165 “Waves to Weather” funded by the German Research Foundation (DFG). C. A. Welzbacher was supported by the Deutscher Wetterdienst research program Innovation Programme for applied Researches and Developments (IAFE) in the course of the SINFONY project. Furthermore, the authors want to acknowledge the continuous advice provided by Ulrich Blahak and colleagues at DWD as well as our colleague Josef Schrötte at LMU. We are thankful to three anonymous reviewers for their constructive and insightful comments on an earlier version of the article.

REFERENCES

- Anthes, R. A., 1986: The general question of predictability. *Mesoscale Meteorology and Forecasting*, P. S. Ray, Ed., Vol. II, Amer. Meteor. Soc., 636–656.

- Bachmann, K., C. Keil, and M. Weissmann, 2019: Impact of radar data assimilation and orography on predictability of deep convection. *Quart. J. Roy. Meteor. Soc.*, **145**, 117–130, <https://doi.org/10.1002/qj.3412>.
- Baldauf, M., A. Seifert, J. Förstner, D. Majewski, M. Raschendorfer, and T. Reinhardt, 2011: Operational convective-scale numerical weather prediction with the COSMO Model: Description and sensitivities. *Mon. Wea. Rev.*, **139**, 3887–3905, <https://doi.org/10.1175/MWR-D-10-05013.1>.
- Barrett, A. I., S. L. Gray, D. J. Kirshbaum, N. M. Roberts, D. M. Schultz, and J. G. Fairman, 2015: Synoptic versus orographic control on stationary convective banding. *Quart. J. Roy. Meteor. Soc.*, **141**, 1101–1113, <https://doi.org/10.1002/qj.2409>.
- Baur, F., C. Keil, and G. C. Craig, 2018: Soil moisture–precipitation coupling over Central Europe: Interactions between surface anomalies at different scales and the dynamical implication. *Quart. J. Roy. Meteor. Soc.*, **144**, 2863–2875, <https://doi.org/10.1002/qj.3415>.
- Ben Bouallègue, Z., and S. E. Theis, 2014: Spatial techniques applied to precipitation ensemble forecasts: From verification results to probabilistic products. *Meteor. Appl.*, **21**, 922–929, <https://doi.org/10.1002/met.1435>.
- Bick, T., and Coauthors, 2016: Assimilation of 3D radar reflectivities with an ensemble Kalman filter on the convective scale. *Quart. J. Roy. Meteor. Soc.*, **142**, 1490–1504, <https://doi.org/10.1002/qj.2751>.
- Carbone, R. E., and J. D. Tuttle, 2008: Rainfall occurrence in the U.S. warm season: The diurnal cycle. *J. Climate*, **21**, 4132–4146, <https://doi.org/10.1175/2008JCLI2275.1>.
- Casati, B., G. Ross, and D. B. Stephenson, 2004: A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteor. Appl.*, **11**, 141–154, <https://doi.org/10.1017/S1350482704001239>.
- Cintineo, R. M., and D. J. Stensrud, 2013: On the predictability of supercell thunderstorm evolution. *J. Atmos. Sci.*, **70**, 1993–2011, <https://doi.org/10.1175/JAS-D-12-0166.1>.
- Clark, A. J., W. A. Gallus, and M. L. Weisman, 2010: Neighborhood-based verification of precipitation forecasts from convection-allowing NCAR WRF Model simulations and the operational NAM. *Wea. Forecasting*, **25**, 1495–1509, <https://doi.org/10.1175/2010WAF2222404.1>.
- Clark, P., N. M. Roberts, H. W. Lean, S. P. Ballard, and C. Charlton-Perez, 2016: Convection-permitting models: A step-change in rainfall forecasting. *Meteor. Appl.*, **23**, 165–181, <https://doi.org/10.1002/met.1538>.
- Denis, B., J. Côté, and R. Laprise, 2002: Spectral decomposition of two-dimensional atmospheric fields on limited-area domains using the discrete cosine transform (DCT). *Mon. Wea. Rev.*, **130**, 1812–1829, [https://doi.org/10.1175/1520-0493\(2002\)130<1812:SDOTDA>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<1812:SDOTDA>2.0.CO;2).
- 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, <https://doi.org/10.1175/MWR-D-14-00172.1>.
- , N. M. Roberts, R. S. Plant, and S. Migliorini, 2016: A new method for the characterization and verification of local spatial predictability for convective-scale ensembles. *Quart. J. Roy. Meteor. Soc.*, **142**, 1982–1996, <https://doi.org/10.1002/qj.2792>.
- Done, J., C. A. Davis, and M. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the Weather Research and Forecasting (WRF) model. *Atmos. Sci. Lett.*, **5**, 110–117, <https://doi.org/10.1002/asl.72>.
- , 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, <https://doi.org/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, <https://doi.org/10.1002/qj.943>.
- Duda, J. D., and W. A. Gallus, 2013: The impact of large-scale forcing on skill of simulated convective initiation and upscale evolution with convection-allowing grid spacings in the WRF. *Wea. Forecasting*, **28**, 994–1018, <https://doi.org/10.1175/WAF-D-13-00005.1>.
- Gebhardt, C., S. E. Theis, M. Paulat, and Z. Ben Bouallègue, 2011: Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. *Atmos. Res.*, **100**, 168–177, <https://doi.org/10.1016/j.atmosres.2010.12.008>.
- Gustafsson, N., and Coauthors, 2018: Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Quart. J. Roy. Meteor. Soc.*, **144**, 1218–1256, <https://doi.org/10.1002/qj.3179>.
- Hanley, K. E., D. J. Kirshbaum, S. E. Belcher, N. M. Roberts, and G. Leoncini, 2011: Ensemble predictability of an isolated mountain thunderstorm in a high-resolution model. *Quart. J. Roy. Meteor. Soc.*, **137**, 2124–2137, <https://doi.org/10.1002/qj.877>.
- , —, N. M. Roberts, and G. Leoncini, 2013: Sensitivities of a squall line over central Europe in a convective-scale ensemble. *Mon. Wea. Rev.*, **141**, 112–133, <https://doi.org/10.1175/MWR-D-12-00013.1>.
- , R. S. Plant, T. H. Stein, R. J. Hogan, J. C. Nicol, H. W. Lean, C. Halliwell, and P. A. Clark, 2015: Mixing-length controls on high-resolution simulations of convective storms. *Quart. J. Roy. Meteor. Soc.*, **141**, 272–284, <https://doi.org/10.1002/qj.2356>.
- Hirt, M., S. Rasp, U. Blahak, and G. C. Craig, 2019: Stochastic parameterization of processes leading to convection initiation in kilometer-scale models. *Mon. Wea. Rev.*, **147**, 3917–3934, <https://doi.org/10.1175/MWR-D-19-0060.1>.
- Hohenegger, C., and C. Schär, 2007a: Atmospheric predictability at synoptic versus cloud-resolving scales. *Bull. Amer. Meteor. Soc.*, **88**, 1783–1793, <https://doi.org/10.1175/BAMS-88-11-1783>.
- , and —, 2007b: Predictability and error growth dynamics in cloud-resolving models. *J. Atmos. Sci.*, **64**, 4467–4478, <https://doi.org/10.1175/2007JAS2143.1>.
- Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, **230**, 112–126, <https://doi.org/10.1016/j.physd.2006.11.008>.
- Isotta, F. A., and Coauthors, 2014: The climate of daily precipitation in the Alps: Development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data. *Int. J. Climatol.*, **34**, 1657–1675, <https://doi.org/10.1002/joc.3794>.
- Johnson, A., and X. Wang, 2016: A study of multiscale initial condition perturbation methods for convection-permitting ensemble forecasts. *Mon. Wea. Rev.*, **144**, 2579–2604, <https://doi.org/10.1175/MWR-D-16-0056.1>.
- Keil, C., and G. C. Craig, 2011: Regime-dependent forecast uncertainty of convective precipitation. *Meteor. Z.*, **20**, 145–151, <https://doi.org/10.1127/0941-2948/2011/0219>.
- , 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, <https://doi.org/10.1002/qj.2143>.
- , F. Baur, K. Bachmann, S. Rasp, L. Schneider, and C. Barthlott, 2019: Relative contribution of soil moisture, boundary layer and microphysical perturbations on convective predictability in different weather regimes. *Quart. J. Roy. Meteor. Soc.*, **145**, 3102–3115, <https://doi.org/10.1002/qj.3607>.
- Kovacs, M., and D. J. Kirshbaum, 2016: Topographic impacts on the spatial distribution of deep convection over southern Quebec. *J. Appl. Meteor. Climatol.*, **55**, 743–762, <https://doi.org/10.1175/JAMC-D-15-0239.1>.
- Kühnlein, C., C. Keil, G. C. Craig, and C. Gebhardt, 2014: The impact of downscaled initial condition perturbations on convective-scale ensemble forecasts of precipitation. *Quart. J. Roy. Meteor. Soc.*, **140**, 1552–1562, <https://doi.org/10.1002/qj.2238>.
- Lange, H., and G. C. Craig, 2014: The impact of data assimilation length scales on analysis and prediction of convective storms. *Mon. Wea. Rev.*, **142**, 3781–3808, <https://doi.org/10.1175/MWR-D-13-00304.1>.
- Levizzani, V., F. Pinelli, M. Pasqui, S. Melani, A. G. Laing, and R. E. Carbone, 2010: A 10-year climatology of warm-season cloud patterns over Europe and the Mediterranean from Meteosat IR observations. *Atmos. Res.*, **97**, 555–576, <https://doi.org/10.1016/j.atmosres.2010.05.014>.
- Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climate Appl. Meteor.*, **22**, 1065–1092, [https://doi.org/10.1175/1520-0450\(1983\)022<1065:BPOTSF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2).
- , S. Chiao, T.-A. Wang, M. L. Kaplan, and R. P. Weglarz, 2001: Some common ingredients for heavy orographic rainfall. *Wea. Forecasting*, **16**, 633–660, [https://doi.org/10.1175/1520-0434\(2001\)016<0633:SCIFHO>2.0.CO;2](https://doi.org/10.1175/1520-0434(2001)016<0633:SCIFHO>2.0.CO;2).
- Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307, <https://doi.org/10.3402/tellusa.v21i3.10086>.
- Melhauser, C., and F. Zhang, 2012: Practical and intrinsic predictability of severe and convective weather at the mesoscales. *J. Atmos. Sci.*, **69**, 3350–3371, <https://doi.org/10.1175/JAS-D-11-0315.1>.
- Mittermaier, M., and N. M. Roberts, 2010: Intercomparison of spatial forecast verification methods: Identifying skillful spatial scales using the fractions skill score. *Wea. Forecasting*, **25**, 343–354, <https://doi.org/10.1175/2009WAF2222260.1>.
- , —, and S. A. Thompson, 2013: A long-term assessment of precipitation forecast skill using the Fractions Skill Score. *Meteor. Appl.*, **20**, 176–186, <https://doi.org/10.1002/met.296>.
- Necker, T., M. Weissmann, and M. Sommer, 2018: The importance of appropriate verification metrics for the assessment of observation impact in a convection-permitting modelling system. *Quart. J. Roy. Meteor. Soc.*, **144**, 1667–1680, <https://doi.org/10.1002/qj.3390>.
- Nuss, W. A., and D. K. Miller, 2001: Mesoscale predictability under various synoptic regimes. *Nonlinear Processes Geophys.*, **8**, 429–438, <https://doi.org/10.5194/npg-8-429-2001>.
- Peralta, C., Z. Ben Bouallègue, S. E. Theis, C. Gebhardt, and M. Buchhold, 2012: Accounting for initial condition uncertainties in COSMO-DE-EPS. *J. Geophys. Res.*, **117**, D07108, <https://doi.org/10.1029/2011JD016581>.
- Picard, L., and C. F. Mass, 2017: The sensitivity of orographic precipitation to flow direction: An idealized modeling approach. *J. Hydrometeor.*, **18**, 1673–1688, <https://doi.org/10.1175/JHM-D-16-0209.1>.
- Piper, D., M. Kunz, F. Ehmele, S. Mohr, B. Mühr, A. Kron, and J. Daniell, 2016: Exceptional sequence of severe thunderstorms and related flash floods in May and June 2016 in Germany—Part 1: Meteorological background. *Nat. Hazards Earth Syst. Sci.*, **16**, 2835–2850, <https://doi.org/10.5194/nhess-16-2835-2016>.
- Reinhardt, T., and A. Seifert, 2006: A three-category ice scheme for LMK. COSMO Newsletter, No. 6, Consortium for Small-Scale Modeling, Offenbach, Germany, 115–120.
- Ritter, B., and J.-F. Geleyn, 1992: A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations. *Mon. Wea. Rev.*, **120**, 303–325, [https://doi.org/10.1175/1520-0493\(1992\)120<0303:ACRSFN>2.0.CO;2](https://doi.org/10.1175/1520-0493(1992)120<0303:ACRSFN>2.0.CO;2).
- 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, <https://doi.org/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, <https://doi.org/10.1175/MWR-D-14-00100.1>.
- Schraff, C., H. Reich, A. Rhodin, A. Schomburg, K. Stephan, A. Perriáñez, and R. Potthast, 2016: Kilometre-scale ensemble data assimilation for the COSMO model (KENDA). *Quart. J. Roy. Meteor. Soc.*, **142**, 1453–1472, <https://doi.org/10.1002/qj.2748>.
- Schwartz, C. S., and Coauthors, 2009: Next-day convection-allowing WRF Model guidance: A second look at 2-km versus 4-km grid spacing. *Mon. Wea. Rev.*, **137**, 3351–3372, <https://doi.org/10.1175/2009MWR2924.1>.
- , and R. A. Sobash, 2017: Generating probabilistic forecasts from convection-allowing ensembles using neighborhood approaches: A review and recommendations. *Mon. Wea. Rev.*, **145**, 3397–3418, <https://doi.org/10.1175/MWR-D-16-0400.1>.
- Seifert, A., C. Köhler, and K. D. Beheng, 2012: Aerosol-cloud-precipitation effects over Germany as simulated by a convective-scale numerical weather prediction model. *Atmos. Chem. Phys.*, **12**, 709–725, <https://doi.org/10.5194/acp-12-709-2012>.
- Selz, T., and G. C. Craig, 2015: Upscale error growth in a high-resolution simulation of a summertime weather event over Europe. *Mon. Wea. Rev.*, **143**, 813–827, <https://doi.org/10.1175/MWR-D-14-00140.1>.
- Simmer, C., and Coauthors, 2016: HERZ: The German Hans-Ertel Centre for Weather Research. *Bull. Amer. Meteor. Soc.*, **97**, 1057–1068, <https://doi.org/10.1175/BAMS-D-13-00227.1>.
- Slingo, J., and T. Palmer, 2011: Uncertainty in weather and climate prediction. *Philos. Trans. Roy. Soc. A*, **369**, 4751–4767, <https://doi.org/10.1098/rsta.2011.0161>.
- Stephan, K., S. Klink, and C. Schraff, 2008: Assimilation of radar-derived rain rates into the convective-scale model COSMO-DE at DWD. *Quart. J. Roy. Meteor. Soc.*, **134**, 1315–1326, <https://doi.org/10.1002/qj.269>.
- Stratman, D. R., M. C. Coniglio, S. E. Koch, and M. Xue, 2013: Use of multiple verification methods to evaluate forecasts of convection from hot- and cold-start convection-allowing models. *Wea. Forecasting*, **28**, 119–138, <https://doi.org/10.1175/WAF-D-12-00022.1>.
- Sun, Y. Q., and F. Zhang, 2016: Intrinsic versus practical limits of atmospheric predictability and the significance of the butterfly effect. *J. Atmos. Sci.*, **73**, 1419–1438, <https://doi.org/10.1175/JAS-D-15-0142.1>.

- Surcel, M., 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, <https://doi.org/10.1175/JAS-D-14-0071.1>.
- , —, —, M. Xue, and F. Kong, 2017: More on the scale dependence of the predictability of precipitation patterns: Extension to the 2009–13 CAPS spring experiment ensemble forecasts. *Mon. Wea. Rev.*, **145**, 3625–3646, <https://doi.org/10.1175/MWR-D-16-0362.1>.
- 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, <https://doi.org/10.1175/2010MWR3487.1>.
- Weissmann, M., and Coauthors, 2014: Initial phase of the Hans-Ertel Centre for Weather Research—A virtual centre at the interface of basic and applied weather and climate research. *Meteor. Z.*, **23**, 193–208, <https://doi.org/10.1127/0941-2948/2014/0558>.
- Zeng, Y., U. Blahak, M. Neuper, and D. Jerger, 2014: Radar beam tracing methods based on atmospheric refractive index. *J. Atmos. Oceanic Technol.*, **31**, 2650–2670, <https://doi.org/10.1175/JTECH-D-13-00152.1>.
- , —, and D. Jerger, 2016: An efficient modular volume-scanning radar forward operator for NWP models: Description and coupling to the COSMO model. *Quart. J. Roy. Meteor. Soc.*, **142**, 3234–3256, <https://doi.org/10.1002/qj.2904>.
- , T. Janjić, A. De Lozar, U. Blahak, H. Reich, C. Keil, and A. Seifert, 2018: Representation of model error in convective-scale data assimilation: Additive noise, relaxation methods, and combinations. *J. Adv. Model. Earth Syst.*, **10**, 2889–2911, <https://doi.org/10.1029/2018MS001375>.
- Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253, [https://doi.org/10.1175/1520-0493\(2004\)132<1238:IOIEAO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2).
- , N. Bei, R. Rotunno, C. Snyder, and C. C. Epifanio, 2007: Mesoscale predictability of moist baroclinic waves: Convection-permitting experiments and multistage error growth dynamics. *J. Atmos. Sci.*, **64**, 3579–3594, <https://doi.org/10.1175/JAS4028.1>.
- Zhang, Y., F. Zhang, D. J. Stensrud, and Z. Meng, 2015: Practical predictability of the 20 May 2013 tornadic thunderstorm event in Oklahoma: Sensitivity to synoptic timing and topographical influence. *Mon. Wea. Rev.*, **143**, 2973–2997, <https://doi.org/10.1175/MWR-D-14-00394.1>.
- , —, —, and —, 2016: Intrinsic predictability of the 20 May 2013 tornadic thunderstorm event in Oklahoma at storm scales. *Mon. Wea. Rev.*, **144**, 1273–1298, <https://doi.org/10.1175/MWR-D-15-0105.1>.
- 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, <https://doi.org/10.1029/2010GL046199>.