Variability and Clustering of Midlatitude Summertime Convection: Testing the Craig and Cohen Theory in a Convection-Permitting Ensemble with Stochastic Boundary Layer Perturbations

STEPHAN RASP, TOBIAS SELZ, AND GEORGE C. CRAIG
Meteorological Institute, Ludwig-Maximilians-Universität München, Munich, Germany

(Manuscript received 31 August 2017, in final form 23 November 2017)

ABSTRACT

The statistical theory of convective variability developed by Craig and Cohen in 2006 has provided a promising foundation for the design of stochastic parameterizations. The simplifying assumptions of this theory, however, were made with tropical equilibrium convection in mind. This study investigates the predictions of the statistical theory in real-weather case studies of nonequilibrium summertime convection over land. For this purpose, a convection-permitting ensemble is used in which all members share the same large-scale weather conditions but the convection is displaced using stochastic boundary layer perturbations. The results show that the standard deviation of the domain-integrated mass flux is proportional to the square root of its mean over a wide range of scales. This confirms the general applicability and scale adaptivity of the Craig and Cohen theory for complex weather. However, clouds tend to cluster on scales of around 100 km, particularly in the morning and evening. This strongly impacts the theoretical predictions of the variability, which does not include clustering. Furthermore, the mass flux per cloud closely follows an exponential distribution if all clouds are considered together and if overlapping cloud objects are separated. The nonseparated cloud mass flux distribution resembles a power law. These findings support the use of the theory for stochastic parameterizations but also highlight areas for improvement.

1. Introduction

Stochastic parameterizations have the potential to increase forecast skill and decrease model biases by capturing the inherently turbulent nature of many subgrid processes [for a comprehensive overview, see Berner et al. (2016)]. In the case of atmospheric deep convection, the fluctuations around the mean state within a grid box become significant for model grid spacing less than 100 km (Jones and Randall 2011). This subgrid noise can feed back onto the resolved scales, impacting tropical oscillations (Wang et al. 2016; Christensen et al. 2017) and the upscale growth of forecast errors (Selz and Craig 2015b). Designing a stochastic parameterization requires some model of the subgrid-scale variability. Simple approaches include perturbing parameterized model tendencies in a multiplicative way (Buizza et al. 1999) or perturbing parameters in the parameterizations (Ollinaho et al. 2017). More complex schemes have been designed based on subgrid Markov chains (Dorrestijn et al. 2013) and cellular automata (Bengtsson et al. 2013).

In this study, we focus on a theory of convective variability based on statistical physics developed by Craig and Cohen (2006; the theory is hereafter abbreviated CC06). Its application in a stochastic parameterization framework (Plant and Craig 2008) proved beneficial in a number of ways: it produces scale-aware fluctuations in a mesoscale model (Keane et al. 2014); forecast errors grow upscale realistically, as opposed to deterministic parameterizations in which errors grow too slowly (Selz and Craig 2015a); and in global climate simulations, precipitation variability and tropical wave activity, such as the Madden–Julian oscillation, are improved (Wang et al. 2016; Wang and Zhang 2016). Recently, Sakradzija et al. (2015) extended the approach to shallow convection, which improves coupling the resolved
model dynamics (Sakradzija et al. 2016). For all its benefits, the CC06 theory is based on strongly simplifying assumptions about how convection behaves. The main purpose of this study is to test the applicability of the CC06 theory outside of its comfort zone.

a. The CC06 theory

The aim of the CC06 theory is to derive a minimally simple model of convective variability. Under the quasi-equilibrium assumption, the large-scale state prescribes the mean mass flux in a certain domain \( \langle M \rangle \) — through a closure assumption in a parameterization. The mean mass flux of an individual cloud \( \langle m \rangle \) is determined solely by local properties such as boundary layer turbulence or entrainment and is therefore independent of \( \langle M \rangle \). The average number of clouds is then \( \langle N \rangle = \langle M \rangle / \langle m \rangle \). The individual clouds are assumed to be pointlike, randomly distributed in space, and noninteracting. The most likely state of such a cloud ensemble is characterized by an exponential distribution of the cloud mass flux \( m \):

\[
p(m) = \frac{1}{\langle m \rangle} e^{-m/\langle m \rangle};
\]  

(1)

and a Poisson distribution of the cloud number in the domain

\[
p(N) = \frac{\langle N \rangle^n}{n!} e^{-\langle N \rangle} \quad \text{for} \quad n = 0, 1, \ldots
\]  

(2)

Combining these equations yields a distribution of the domain-total mass flux \( p(M) \) as a function of its mean \( \langle M \rangle \) and the mean cloud mass flux \( \langle m \rangle \). In principle, it is possible and interesting to investigate the full distribution function or several higher-order moments. For this study, however, we focus on the second moment. This is done for two main reasons: first, higher moments require a larger sample size to yield statistically significant results, and second, deviations in the variances allow for clear and physical interpretations. The second moment can be expressed in terms of the normalized variance

\[
\langle (\delta M)^2 \rangle = \frac{2}{\langle N \rangle}
\]  

(3)

or in terms of the unnormalized standard deviation

\[
\langle (\delta M)^2 \rangle^{1/2} = \sqrt{2\langle M \rangle / \langle m \rangle}.
\]  

(4)

\footnote{Note that the angle brackets used throughout the text describe ensemble means.}

b. Previous tests of the CC06 theory

So far, few studies have directly tested the assumptions and predictions of CC06. Cohen and Craig (2006) used a convection-permitting model in a radiative–convective equilibrium setup with different large-scale forcing and vertical wind shear strengths and found that \( p(m) \) was well approximated by an exponential distribution for all settings but that \( p(N) \) was broader than predicted by Eq. (2) because of cloud clustering. Despite this, the simulated mass flux variability was up to 20% lower than predicted, which they attributed to compensating errors. Davoudi et al. (2010) confirmed this result in their simulations with a diurnal cycle of radiation. Scheufele (2014) tested the sensitivity of the Cohen and Craig (2006) results to model resolution and found that clouds are more strongly clustered at higher resolutions. Furthermore, to reproduce the exponential cloud mass flux distributions, a separation of connected clouds into individual updrafts becomes necessary. Davies (2008) focused on the variation of the CC06 predictions in an idealized diurnal cycle setup. Cloud clustering and convective variability were strongest shortly after convective initiation and during the decline of convective activity. Studying the variability scaling in trade wind shallow cumulus, Sakradzija et al. (2015) saw a drastic increase in cloud organization once precipitation started to form. Consequently, the mass flux variability was many times larger than the predictions of the statistical theory.

c. Motivation and aims of this paper

The studies mentioned above have two things in common: first, they show that the cloud field is organized in contradiction with the CC06 assumptions, and second, all the studies were conducted with simplified, idealized setups. These setups generally aim to represent tropical equilibrium convection or a diurnal cycle over land in the absence of any large-scale changes in forcing or other complications such as land surface variations or orography. Stochastic parameterizations in general circulation models, however, must be able to cope with a wide variety of convection around the globe. It remains unclear if, and to what extent, the CC06 theory and the stochastic parameterizations based on it are useful for representing complex real-world weather situations.

In this study, we aim to test the assumptions and predictions of CC06 in simulations of midlatitude summertime convection over land. In particular, we ask the following research questions:

- How well do the CC06 predictions hold up for complex, nonequilibrium convection?
- Are there systematic deviations, and can we explain them?
- What role does convective organization play?
To answer these questions, we set up ensemble simulations (section 2) that allow us to quantitatively measure convective variability in real-world case studies (section 3). The analysis details are described in section 4. In section 5, we compare the predictions of the CC06 theory to our simulation results and identify systematic deviations. We then focus on cloud organization (section 6) and the cloud mass flux distribution (section 7). A discussion of the implications for stochastic parameterizations follows in section 8 before a summary in section 9.

2. Numerical simulations, observations, and computational reproducibility

a. COSMO model and ensemble setup

The numerical simulations are done with the Consortium for Small-Scale Modeling (COSMO) model (Baldauf et al. 2011). The horizontal grid spacing $\Delta x$ is 0.025°, roughly 2.8 km, with 50 levels in the vertical. There is no parameterization of deep convection, but shallow convection is parameterized using the Tiedtke scheme (Doms et al. 2011). The domain spans 357 grid points in either horizontal direction centered over Germany at 50°N, 10°E. For the analysis, a $256 \times 256$ gridpoint subdomain, roughly $717 \text{ km} \times 717 \text{ km}$, at the center of the simulation domain is considered to avoid boundary spinup effects. The 50-member ensemble simulations are started at 0000 UTC on each of the 12 consecutive days (see section 3) with a simulation time of 24 h. For initial and boundary conditions, we use hourly interpolated deterministic COSMO European version (COSMO-EU) analyses, which have a horizontal resolution of 7 km. Each ensemble member differs only in the random seed used for the stochastic perturbation scheme, described below, which has the effect of randomly shuffling the convective cells. Additionally, one deterministic simulation is run without the stochastic perturbation scheme. The output frequency is 60 min. The model name lists are saved in the online repository accompanying this paper (see section 2c).

Given unlimited computational resources, it would be desirable to run the model at a higher resolution to actually resolve the cloud features. With our computational constraints, however, a trade-off between resolution, ensemble size, and length of the simulation period is necessary. To ensure the simulations create realistic cloud features, we compare the model to radar observations (see section 2c). We will also discuss the potential impact of resolution on our results in section 9.

b. Stochastic boundary layer perturbations

The physically based stochastic perturbation (PSP) scheme was first proposed by Kober and Craig (2016). It represents a general framework for adding process-specific perturbations to reintroduce missing variability on the grid scale of convection-permitting models that is associated with unresolved processes. The process considered here is boundary layer turbulence. In the appendix, we present an updated formulation of the scheme to clarify the physical rationale. In this study, the stochastic scheme enables us to obtain many different realizations of the convective cloud fields for the same large-scale flow. One limitation of this approach is the fact that only subgrid turbulence is considered in the stochastic perturbations. Other subgrid processes such as orography might preferentially trigger convection in particular locations, thereby violating the CC06 assumptions.

c. Radar-derived precipitation observations

To validate our simulations we use the Radar Online Aneichung (RADOLAN) radar-weighted (RW) product provided by the German Weather Service (DWD; DWD 2017). These data contain estimates of 1-hourly precipitation accumulations based on radar reflectivities adjusted using rain gauges. Even though the precipitation values are not direct observations, we use the term observation for the RADOLAN RW product in the rest of the text. The original data, which have a spatial resolution of 1 km, are adapted to the COSMO model grid. For all model-observation comparisons, only grid points with observation data are used. If daily time series are computed, a joint mask is used only including grid points for which observation data are available at all times throughout the day. The hourly RADOLAN RW products are valid at 10 min to the full hour, whereas our model precipitation accumulations are written to the full hour. For the purposes of this study, we think it is reasonable to neglect this difference. Last, because of the automatic adjustment of the radar observations with rain gauge data, some unrealistically high precipitation values occur, sometimes exceeding $300 \text{ mm h}^{-1}$ (K. Stephan 2017, personal communication). Since we are not computing forecast scores in this study, an ad hoc measure of removing all grid points with hourly precipitation accumulations larger than $100 \text{ mm h}^{-1}$ turned out to be sufficient to remove any significant artifacts.

d. Displacement growth of stochastic perturbations

Figure 1a shows precipitation snapshots of the observations and two ensemble members. The large-scale
precipitation pattern is very similar, but the individual convective elements appear to be completely uncorrelated. To quantitatively assess the displacement at different scales, we look at the saturation of the ensemble difference spectra of kinetic energy and precipitation [Figs. 1b,c; for a detailed description of the calculation of the spectra, see Selz and Craig (2015b)]. The kinetic energy spectrum shows typical signs for upscale error growth: small scales saturate first, followed by saturation at increasingly larger scales. At the scale of large precipitation patterns, around 300 km, the difference between the ensemble members is below 20%. This confirms that the large-scale forcing is similar between all ensemble members. In contrast, the precipitation spectrum at the scale of individual convective elements, approximately 50 km, is almost fully decorrelated after 6 h. The combination of similar large-scale conditions and displaced convection allows us to investigate convective variability in real-weather case studies.

e. Computational details and reproducibility

This subsection closely follows the guidelines on publishing computational results proposed by Irving (2016). The analysis and plotting of model and observation data were done using Python. The Python libraries Numerical Python (NumPy; van der Walt et al. 2011) and Scientific Computing Tools for Python (SciPy; Jones et al. 2001) were used heavily. The raw data were read with the Python module cosmo_utils (code available upon request). The figures were plotted using the Python module Matplotlib (Hunter 2007). Plotting colors were chosen according to the hue–chroma–luminance color space (Stauffer et al. 2015). Some plots were postprocessed using the vector graphics program Inkscape.

To enable reproducibility of the results, this paper is accompanied by a version-controlled code repository (https://github.com/raspstephan/convective_variability_analysis) and a Figshare repository (Rasp 2017), which contains a snapshot of the code repository at the time of submission and supplementary log files for each figure. These log files contain information about the computational steps taken from the raw data to the generation of the plots. While the model code and initial data are not openly available, a detailed technical description of the model simulations can be found in the cosmo_runs scripts directory of the code repository. The Jupyter notebooks (Kluyver et al. 2016) mentioned in the text are stored in the directory jupyter_notebooks of the repository. Links to noninteractive versions of the notebooks can be found on the front page of the Github repository; rendered PDF versions are also added to the supplement of this paper.

![Fig. 1. (a) Hourly precipitation accumulations at 1400 UTC 4 Jun 2016 for the observations and two ensemble members. (top) Entire analysis domain; (bottom) zoom in on a smaller region. (b),(c) Ratio of difference to twice the difference spectrum of (b) horizontal kinetic energy in the troposphere, omitting the top 15 model levels, and (c) hourly precipitation for different times of day. For details on the calculation, see Selz and Craig (2015b). Values of one indicate a full displacement at this scale. Lines represent composites over all days.](image-url)
3. Weather situation and precipitation in model and observations

a. Synoptic situation and convective regime

The period from 26 May to 9 June 2016 was characterized by extraordinary extreme weather over central Europe and, in particular, Germany (Piper et al. 2016). Heavy precipitation, exceeding a 200-yr return period in some regions of southern Germany, caused flash floods that, together with hail measuring up to 5 cm in diameter and 12 confirmed tornadoes, resulted in damages of over EUR 5 billion. A persistent heavy-precipitation period of similar length is unprecedented in a 55-yr climatology. For this study, we selected 12 contiguous days from 28 May to 8 June.

The period can be roughly divided into two phases (Fig. 2). In the first, from 28 May to approximately 3 June, an upper-level trough dominated European weather, subsequently developing into a cutoff low. This upper-level feature caused strong synoptic lifting and several weak surface low pressure systems. These were accompanied by cyclonic circulation centered over the Alpine region and southeasterly advection of moist air over central Europe. This lead to a destabilization of the atmosphere, particularly over southern Germany. In the second phase, from 4 to 8 June, the cutoff gave way to a stationary upper-level ridge, typical for an omega-blocking situation. This caused very persistent weather with large instability building up over southern Germany.

The synoptic instability, combined with strong surface heating, provided a favorable environment for the development of deep convection. Precipitation followed a diurnal pattern on most days (Fig. 3) but was modulated by synoptic lifting on several days. This is most noticeable in the night of 29–30 May, in which a mesoscale convective system in association with large-scale ascent covered most of southern Germany (domain-mean precipitation plots for each individual day can be found in the supplement). The precipitation lags the convective available potential energy (CAPE) by around 4 h. The chosen period represents a variety of nonequilibrium convection over land and is therefore well suited to test the CC06 theory outside of the regime for which it was originally designed.

b. Precipitation in simulations and observations

Finally, we want to compare the mean precipitation in our simulations with the PSP scheme (ens), without the PSP scheme (det), and in the observation (obs) in Fig. 3. The PSP scheme causes an earlier onset and a higher maximum in precipitation on several days, which is the expected systematic effect (Kober and Craig 2016). The maximum precipitation increase caused by the stochastic perturbations is 10%, indicating that the model behavior is not drastically altered.

![Fig. 2. Synoptic charts at (a) 0000 UTC 30 May and (b) 0000 UTC 5 Jun 2016. White lines represent mean sea level pressure (hPa). Colors represent 500-hPa geopotential (dam). Figures created from ECMWF analyses.](image)

![Fig. 3. Composite of domain-averaged precipitation for observations (black), deterministic (green), and ensemble simulations (blue). The blue shaded region represents plus and minus one standard deviation of the ensemble. Additionally, ensemble-mean domain-averaged CAPE (J kg⁻¹) is shown in red. The gray vertical line indicates the time at which the analysis starts.](image)
Larger differences can be seen between the model simulations and the observations. It is important to note that we are not interested in obtaining simulations that are as close to the observations as possible. Rather, our aim is to test whether the general weather situation is reproduced and if there are any systematic differences. On most days, the precipitation amounts in the simulations and observations match well. There is, however, a systematic early decline of precipitation in the late afternoon. Inspecting the precipitation stamps (available in the supplement) suggests that, in some situations, the model is not able to reproduce lasting organized convection when the forcing becomes weaker (an example for this can be seen in Fig. 8). While we are confident that the general characteristics of the convection are well captured in our simulations, it is important to keep this systematic deviation in mind for the subsequent analysis. Furthermore, there is a spinup peak in the first few forecast hours caused by the stochastic perturbations, which is typical when starting from a downscaled analysis. To avoid this effect and to allow the perturbations to develop, we start all our subsequent analyses from 0600 UTC.

4. Cloud identification and computation of statistics

To test the CC06 predictions and assumptions, we need to identify the individual clouds and compute statistics of the modeled updraft mass flux field. The process of identifying cloud objects and computing the radial distribution function is illustrated in an accompanying Jupyter notebook (called cloud_identification_and_rdf.ipynb). We identify convective updrafts by using a vertical velocity threshold $w > 1 \text{ m s}^{-1}$ combined with a positive cloud water content at model level 30. This corresponds to a height above ground level of around 3000 m in the terrain-following COSMO grid. Ideally, one would choose the cloud base since this is what many convection parameterizations use, but this is difficult to determine. Previous studies have shown that the mass flux statistics are relatively insensitive to changes in the analysis height in a reasonable range around the chosen level [see, e.g., Fig. 13 of Davoudi et al. (2010) or Fig. 5.6 in Davies (2008)]. From the resulting binary field, objects are classified as pixels that share an edge. Visual inspection suggested that many “objects” are in fact conglomerates of several touching updrafts [this has also been found by Scheufele (2014)]. We therefore separate the cloud objects using a local maximum filter in combination with a watershed algorithm (Beucher and Meyer 1992). For the local maximum filter, we use a search footprint of $3 \times 3$ grid points. All subsequently presented analysis is done using the separated objects unless otherwise stated. For further information and sensitivity tests of the cloud separation algorithm, see the aforementioned Jupyter notebook cloud_identification_and_rdf.ipynb.

Computation of statistics

For each identified cloud $k = 1, \ldots, N_{\text{cld},i}$ in each ensemble member $i = 1, \ldots, N_{\text{ens}}$, the cloud size, defined as the horizontal area, $\sigma_k$ is computed:

$$\sigma_k = N_{\text{px}} \Delta x^2,$$

where $N_{\text{px}}$ is the number of pixels for each cloud $k$. The mass flux per cloud $m_k$ is given by

$$m_k = \Delta x^2 \sum_i N_{\text{px}} w_i \rho,$$

where $\rho$ is density.

To compute the domain statistics, a coarse graining is applied to create coarse boxes $j = 1, \ldots, N_{\text{box},n}$ with edge lengths $n \in \{256, 128, 64, 32, 16, 8\} \Delta x$, where the number of coarse boxes for each analysis time step is $N_{\text{box},n} = (256/n)^2$. No smaller neighborhoods are considered, since these would be below the effective resolution of the model and the sample size of clouds within the coarse boxes becomes too small. The total mass flux per box $j$ per member $i$, denoted $M_{i,j,n}$, is given by

$$M_{i,j,n} = \sum_{k=1}^{N_{\text{cld},i,n}} N_{\text{px}} m_{k,i,j,n},$$

where $N_{\text{cld},i,n}$ is simply the number of clouds that fall into each box. To avoid splitting clouds at the boundaries of the coarse fields, the centers of mass for each cloud are first identified. Then $m_k$ is attributed to that one point in space. Therefore, the coarse box that contains the center of mass also contains the entire cloud, while the other box does not contain any of the cloud (Cohen and Craig 2006).

Additionally, we compute statistics for the mean heating rate $Q$ for each coarse box $j$ at the same model level. Note that, unlike $M$, $Q$ can be negative. The variables $M$ and $Q$ are well correlated, with a correlation coefficient of 0.8 across all scales.

Ensemble statistics of $\Phi \in \{M, N, Q\}$ are then calculated for each box $j$. The sample variance is computed as

$$\langle (\Phi)^2 \rangle_{j,n} = \frac{1}{N_{\text{ens}} - 1} \sum_{i=1}^{N_{\text{ens}}} \left( \Phi_{i,j,n} - \langle \Phi \rangle_{j,n} \right)^2,$$

where the ensemble mean is given by
\[
\langle \Phi \rangle_{j,n} = \frac{1}{N_{\text{ens}}} \sum_{i=1}^{N_{\text{ens}}} \Phi_{i,j,n},
\]

To compute statistics for \( m \), a different approach is taken. Here, the clouds in all ensemble members for each box are considered together to calculate the variance and mean:

\[
\langle (\delta m)^2 \rangle_{j,n} = \frac{1}{N_{\text{cldtot}} - 1} \sum_{k=1}^{N_{\text{cldtot}}} (m_{k,j,n} - \langle m \rangle_{j,n})^2,
\]

where the mean is given by

\[
\langle m \rangle_{j,n} = \frac{1}{N_{\text{cldtot}}} \sum_{k=1}^{N_{\text{cldtot}}} m_{k,j,n},
\]

and \( N_{\text{cldtot}} = \sum_{n=1}^{N_{\text{ens}}} N_{\text{cld},i,j,n} \) is the total number of clouds in all ensemble members. If \( N_{\text{cldtot}} \) becomes too small, the statistics are severely affected by sampling issues. After inspecting numerical tests, we decided to drop all data where the total number of clouds across all ensemble members is less than 11 (see supplementary Jupyter notebook beta_sample_size_dependency.ipynb).

5. Comparison of CC06 predictions with simulations

a. Scaling of standard deviation with the mean

Our first research question is whether the CC06 theory is applicable for complex real-weather situations. Assuming that the variation of the mean cloud mass flux \( \langle m \rangle \) is small, an assumption we will revisit later, the CC06 theory states that the domain-total mass flux standard deviation increases with the square root of its mean [Eq. (4)]. The combined data, including all coarse boxes for all scales \( n \) for all days and time steps, show that, over more than three orders of magnitude in \( \langle M \rangle \) and almost two orders of magnitude in horizontal scale, the mass flux standard deviation is well described by the proposed square root relation (Fig. 4a). The fit parameter \( b \) is 9.62 \( \times \) 10\(^7\) kg s\(^{-1}\). The variability drops off at both ends, which is reflected in a change in \( b \) if the curve is fitted to each scale individually—smaller scales have a steeper slope.

The heating rate \( Q \), being a horizontally averaged quantity, needs to be multiplied by the area of the coarse box to test for the scaling; \( Q \times A \) behaves similarly to \( M \) (Fig. 4b). The square root scaling applies over an even larger range of scales, but the uncertainty increases toward small values of \( Q \times A \). In contrast to the updraft mass flux, which is by definition positive, the heating rate can also be negative. Therefore, the heating rate standard deviation is not constrained to go to zero as its mean goes to zero, which is a potential cause for the deviations at smaller scales.

Additionally, we fit a linear relation, which corresponds to multiplicative noise as in the stochastically perturbed parameterization tendencies (SPPT) scheme (Sherts and Palmer 2007). Note that changing \( b \) does not change the apparent slope on a log–log plot but only displaces the line. The linear relation is not able to capture the standard deviation scaling of \( \langle M \rangle \) or \( Q \times A \) in our simulations. This is in line with the findings of Sherts and Pallarès (2014), who argue that, for convection, the standard deviation is better described by a square root scaling.

This first general test of the main prediction of the CC06 theory confirms its applicability even in the complex situations in our simulations. One key feature of the theory is its scale awareness. This can be seen in the slope of the fit \( b \), which varies by a factor of 2 as the mean mass flux increases by more than three orders of magnitude (see insert of Fig. 4a). A small variation of the slope indicates that the CC06 scaling describes the variability for a range of different coarsening resolutions without the need for retuning of the parameters. This resolution independence is a desirable trait for stochastic parameterizations, particularly as the gray zone is approached. Multiplicative perturbations, in contrast, appear to be inherently resolution dependent. We discuss the implications of our findings for SPPT in section 8.

b. Deviations from the CC06 theory

While the overall variability scaling is reasonably described by the CC06 theory, the analysis above implies differences between the coarsening scales. In this section, we focus on the systematic deviations from the CC06 theory as a function of the coarsening scale and time of day—our second research question. Since most of the 12 simulation days show similar characteristics (see supplement), we concentrate on composites of all days. To test the CC06 variance predictions, the ratio of simulated to predicted variance is defined as

\[
R_V = \frac{1}{N_{\text{box},n}} \sum_{j=1}^{N_{\text{box},n}} \frac{\langle (\delta M)^2 \rangle_j}{\langle M \rangle_j^2}.
\]

This metric is similar to the one used by Davies (2008) and Davoudi et al. (2010) and describes whether the simulated variance is less (values smaller than one) or more (values larger than one) compared to the variance predicted by the CC06 theory given the same values of \( \langle M \rangle \) and \( \langle m \rangle \). In Fig. 5a \( R_V \) is shown for three representative scales. Two trends jump out. First, there is a diurnal cycle with lower variability in the early
afternoon, from 1200 to 1500 UTC, and increased variability in the morning and, particularly, the evening. The trend becomes stronger with scale. Second, the mean RV is largest for scales around 100 km. Our subsequent analysis aims to identify the causes of these systematic deviations.

Recall that the CC06 variance prediction arises from two distributions: a Poisson distribution of the cloud number $N$ and an exponential distribution of the cloud mass flux $m$. To test these distributions in our simulations, we define parameters describing the width of the...
distributions with respect to their mean. Starting with
the cloud number distribution,
\[ \alpha = \frac{1}{N_{\text{box},n}} \sum_j \langle (\delta N)^2 \rangle_j / \langle N \rangle_j \] (13)
describes whether clouds are more clustered, \( \alpha > 1 \), or
more regularly spaced, \( \alpha < 1 \), compared to a completely
random distribution in space [see appendix A of Davoudi
et al. (2010)]. Figure 5b indicates that the cloud clustering
varies significantly throughout the diurnal cycle and also
depends on the coarsening scale. We will investigate
these aspects further in section 6. The changes in \( \alpha \)
closely resemble the behavior of \( R_V \). In fact, we can
remove the deviations in \( R_V \) caused by the deviations in
\( \alpha \) in each coarse grid box \( j \). The CC06 theory states that
the factor 2 in Eq. (3) comes in equal parts from the
normalized variance of \( N \) and \( m \). Therefore, we can
correct this factor by taking into account the simulated
variance of one of the two (or later both) distributions:
\[ \alpha - \text{adjusted} \ R_V = \frac{1}{N_{\text{box},n}} \sum_j \langle (\delta N)^2 \rangle_j / \langle N \rangle_j (1 + \alpha_j \langle M \rangle_j / \langle m \rangle_j) \] (14)

Doing so halves the amplitude of the diurnal variation
for the medium and large scales (Fig. 5e). In other
words, changes in cloud clustering in the convective life
cycle are responsible for around 50% of the deviations
from the CC06 variance predictions. The small and
medium scales are now closely aligned, which suggests
that cloud organization constitutes the main difference
between these two scales (discussed further in section 6).

Similarly, for the cloud mass flux distribution,
\[ \beta = \frac{1}{N_{\text{box},n}} \sum_j \langle (\delta M)^2 \rangle_j / \langle M \rangle_j^2 \] (15)
indicates whether the distribution is narrower or
broader compared to an exponential distribution with
the same mean. Figure 5c primarily shows a strong
scale dependence but also a weaker diurnal cycle.
Small and medium scales consistently have narrower
distributions than large scales, for which \( \beta \) is around
one. Again, we can account for the \( \beta \) deviations in \( R_V \)
(Fig. 5f):
\[ \beta - \text{adjusted} \ R_V = \frac{1}{N_{\text{box},n}} \sum_j \langle (\delta M)^2 \rangle_j / \langle M \rangle_j (1 + \beta_j \langle m \rangle_j) \] (16)

This shifts the means of the small and medium scales
upward but also slightly decreases the diurnal amplitu
The large scales are hardly affected.

Finally, we remove both \( \alpha \) and \( \beta \) deviations in \( R_V \)
(Fig. 5d):
\[ \alpha \text{} \beta \text{} \text{adjusted} \ R_V = \frac{1}{N_{\text{box},n}} \sum_j \langle (\delta N)^2 \rangle_j / \langle N \rangle_j \left( \alpha_j + \beta_j \langle M \rangle_j / \langle m \rangle_j \right) \] (17)

According to the theory of random sums (Taylor and
Karlin 1998, p. 72), the variance of \( M \) should be fully
described by the variances of its underlying distributions
\( N \) and \( m \), assuming that the distribution of \( m \), conditionally
averaged on \( N \), does not depend on \( N \). This applies well to the small and medium scales, where \( R_V \)
fluctuates around one after adjusting for differences in
\( \alpha \) and \( \beta \). For the large scales, however, a significant devi-

The analysis above highlights some systematic de-

6. Cloud clustering

The parameter \( \alpha \) indicates that cloud clustering
strongly affects convective variability. Now we
introduce another, independent metric of cloud clustering, the radial distribution function (RDF), which allows us to better understand how clouds organize. The RDF measures how many times more likely than random a cloud is located at a certain distance \( r \) to another cloud. A completely random distribution would give RDF \( 1 \) for all \( r \). The mathematical and computational details of the algorithm are explained in the accompanying Jupyter notebook cloud_identification_and_rdf.ipynb. To enable comparison with observation data, we use the hourly precipitation field for the RDF analysis. The results for the mass flux field are qualitatively similar. The RDF shows some variability between simulation days (see supplement). In particular, the first, synoptically dominated phase shows relatively constant RDFs, while the second, locally forced phase shows a stronger diurnal variation. This diurnal cycle also shows up in the composite in Fig. 7. 

Figure 7a shows the RDF as a function of the radius for two times, 1400 and 2100 UTC. The simulations with and without stochastic perturbations are very similar, confirming that the PSP scheme does not drastically alter the behavior of clouds. The observations have their peak RDF value at larger distances. This agrees with the visual impression that clouds in the model are more intermittent than in the observations, a characteristic of cloud-resolving models also observed in other studies (Hanley et al. 2015; Nguyen et al. 2017). In general, the RDF indicates that cloud clusters have a scale of around 50–100 km (25–50 km is the radius at which the RDFs drop to half their peak value) and that the clustering is stronger in the evening.

Since the RDF changes mostly in amplitude, not in shape, we use the maximum RDF value as a proxy of clustering strengths. Figure 7b therefore illustrates the diurnal variation in clustering. The results strongly correspond to the evolution of \( \alpha \); clouds are more clustered in the morning and evening. Similar results were found by Davies (2008) in her idealized diurnal cycle experiments. The scale dependence of \( \alpha \) may be explained by the typical cluster size of around 50–100 km. The small scales are close to the typical cloud separation distance, around 10 km. Consequently, the clustering does not fully impact these scales. The medium scales correspond to the typical clustering size and, therefore, experience the strongest impact resulting in larger values of \( \alpha \). Finally, the large scales can contain many cloud clusters. During the day, these clusters appear to be more regularly spaced, leading to values of \( \alpha \) smaller than one. This could be related to orography, the land surface, or synoptic variations within the domain. Toward the evening, the convective activity drops off rapidly. The few remaining active clusters are less constrained by the large-scale forcing and rely more on internal processes to maintain convection. Supporting evidence for this argument can be found in the ensemble precipitation variability (Fig. 3), which stays approximately constant even as the total precipitation amount rapidly declines.

To understand the nature of cloud clustering, it is important to note that the RDF measures the cloud clustering relative to all clouds in the domain. In fact, the absolute cloud number density around existing clouds changes little in the diurnal cycle. The variation in the RDF stems primarily from the increased isolation of the existing clusters in the morning and evening. Figure 8 illustrates a typical sequence of these events. In the morning, convection is clustered because of two processes: larger precipitation patches from the previous day in the south and spatially confined regions of early convective cells in the north. In the early afternoon, when convection is strongest, convective cells are
distributed over most of the domain. Local clustering still occurs, but the clusters are simply more numerous. In the evening, as the forcing subsides, only strong, congregated convection survives. On this particular day, the formation of a squall line over western Germany is missed by all of the simulations. As mentioned in section 3, this failure of the model to produce larger organized features is apparent on several days.

7. Cloud mass flux distribution

The parameter $\beta$ shows a substantial dependence on the coarse-graining scale and a weaker systematic diurnal cycle. In this section, we further investigate the cloud mass flux distribution. Specifically, we inspect the variations of the mean mass flux, the geographical variation of $h_m$, and the impact of the cloud separation algorithm.

a. Temporal and spatial variations of $\langle m \rangle$

Figure 9 shows how the mean cloud mass flux and size varies with the time of day. Note that $\langle m \rangle$ increases as the total convective activity picks up, but its peak is delayed around 3 h behind $\langle M \rangle$. Also $\langle r \rangle$ is smallest during the increase of convection around 1200 UTC and then increases toward the evening. The difference between the separated and nonseparated $\langle m \rangle$ are also strongest at that time; $\langle m \rangle$ fluctuates by about ±20% throughout the diurnal cycle, while $\langle M \rangle$ varies by a factor of 4. This is in line with previous findings of how the mean cloud mass flux changes with the forcing (Cohen and Craig 2006; Scheufele 2014): the primary effect of increasing the large-scale cooling is an increase in the number of clouds; the changes in the cloud properties are secondary. For this reason, $\langle m \rangle$ is often a prescribed constant, for example, in the Plant and Craig (2008) parameterization. To assess the impact of using a fixed $\langle m \rangle$, we compute $R_V$ using the overall mean cloud mass flux of $5.07 \times 10^7 \text{kg s}^{-1}$ (Fig. 6a).

This increases the diurnal variation for the small and medium scales. Additionally, the spread of these scales is increased. These changes indicate that the variations in $\langle m \rangle$ impact the CC06 variance predictions, but the magnitude is secondary compared to other systematic biases.

In section 5b, we found that small and medium scales seem to have a narrower than expected cloud mass flux distribution. The precipitation snapshots (Figs. 1 and 8 and supplemental material) lead us to hypothesize that $h_m$ varies geographically. There appear to be regions with mostly larger clouds and other regions with mostly smaller clouds. This could result from differences in the synoptic situation, orography, cloud–cloud interactions, or land surface variations leading to changes in the Bowen ratio, which was found to be important for the shallow cumulus mass flux distribution (Sakradzija and Hohenegger 2017). On the domain scale, regions with different mean cloud sizes are included, potentially increasing the width of cloud mass flux distribution.

b. Overall cloud mass flux distribution—The impact of cloud separation

Next, we take a closer look at the overall cloud size distribution for all dates, times, and ensemble members (Fig. 10). If the clouds are separated using the local maximum method, the distribution is close to exponential. Clouds collect near the grid scale, an observation also made by Scheufele (2014). Without the cloud separation, the distribution is closer to a power law. Windmiller (2017) argues that a power-law cloud size distribution is the result of cloud clustering when individual clouds drawn from an exponential distribution overlap. It is curious, however, that this effect already shows up at resolutions of 2.8 km. Previous kilometer-scale studies in idealized setups (Cohen and Craig 2006; Scheufele 2014) found that cloud overlap only became significant at much higher resolutions. This
seems to suggest that in our real-weather case studies, clouds tend to congregate more. The deterministic run and the stochastically perturbed ensemble agree well for smaller mass fluxes. The differences for the larger bins are most likely caused by the much smaller sample size of the deterministic run.

8. Implications for (stochastic) parameterizations of convection

Much of the interest in a theory for convective variability such as CC06 comes from the application to stochastic convective parameterization. The theory provides guidance for how the stochastic variability should change with meteorological situation and model resolution. The evaluation of the theory presented here has both positive and negative implications for a convection scheme, like that of Plant and Craig (2008), which is based on the CC06 theory. Most importantly, the basic square root scaling of mass flux standard deviation with the total mass flux in any finite region holds to a reasonable degree of approximation. Although the convective cases considered here are much more complex than the radiative–convective equilibrium environment for which the theory was first developed and

FIG. 8. Hourly precipitation snapshots at 1100, 1500, and 2000 UTC 5 Jun 2016 for radar observations, deterministic, and two ensemble simulations. Contours in the radar snapshots indicate radar coverage.
tested, the additional complexity does not completely alter the behavior. Since it is the variability scaling that relates the amplitude of convective variability to both the convective closure and the model resolution, its use in stochastic parameterization is supported. In contrast, the most commonly used stochastic parameterization method in numerical weather prediction, SPPT, scales the standard deviation linearly proportional to the convective amount, which is not supported by our results. This does not imply that SPPT should not be used since it is not a convection parameterization but an “all inclusive” method to account for model errors and primarily increase ensemble spread. Furthermore, SPPT must also represent other processes that may scale differently. However, to the extent that small-scale variability is associated with convective clouds, the incorrect scaling implies that the scheme will need to be retuned whenever factors such as model resolution are changed.

However, the significant deviations from CC06 theory found in this study should be accounted for in parameterization. This is not simple, since the dominant deviation was related to convective clustering on a characteristic scale of order 100 km, varying in intensity throughout the diurnal cycle. Unlike the simple model of unorganized convection, the clustering couples different grid columns in the large-scale model—an effect that can be difficult to implement. The model dynamics can only represent organization on scales larger than the effective resolution, approximately 5Δx (Skamarock 2004). Furthermore, even within a grid box, convective clustering may impact other parameterized processes including radiation and surface fluxes, which would need to be accounted for in the relevant schemes. The changes in 〈m〉 are technically easier to implement in a stochastic convection scheme, but a good understanding of what determines 〈m〉 is still lacking. Additionally, we note that it is important to consider the individual updrafts rather than overlapping features to construct distributions that agree with the theory.

Finally, it should be emphasized that we focused on spatial variability in this study and did not address the important question of temporal structures or convective memory, both of which are important for parameterization. The Plant and Craig (2008) parameterization has some memory by giving each random cloud a fixed cloud lifetime larger than the convective time step. Sakradzija et al. (2016) further related the cloud lifetime to the mass flux. No attempt has been made, however, to include the effect of organization on convective memory.

9. Conclusions

In this study, we tested a minimally simple theory of convective variability developed by Craig and Cohen (2006) in complex summertime weather over land.

**Fig. 9.** Evolution of mean cloud mass flux (red) and mean cloud size (blue) for separated (solid) and unseparated (dashed) clouds as a function of time as a composite over all days. The gray line indicates the evolution of 〈M〉.

**Fig. 10.** Histogram of cloud mass flux 〈m〉 for all dates and times: (a) separated and (b) nonseparated. An exponential and a power-law curve are fitted using a least squares algorithm. The bin widths are 1.67 × 10^7 kg s^-1 in (a) and 5.93 × 10^6 kg s^-1 in (b).
The theory’s core building blocks, a random distribution of clouds in space and an exponential mass flux distribution, were mainly developed with tropical maritime convection in mind. We chose 12 consecutive, high-impact weather days over Germany to test the theory’s prediction outside of its comfort zone. To quantify the variability of convection in real-weather case studies, we set up a 50-member ensemble in which all members had the same large-scale conditions but the convective cells were displaced using stochastic boundary layer perturbations.

In general, the mass flux standard deviation scales with the square root of its mean, in accordance with the theory. We did find systematic deviations, however. First, clouds tend to be clustered, violating the no cloud–cloud interaction assumption, more strongly so in the morning and evening. The typical cluster size is around 50–100 km. Second, the mean mass flux per cloud varies geographically—clouds of a certain size appear to congregate—and temporally, by around ±20% in the diurnal cycle. This indicates that the cloud properties are not entirely independent of the large-scale conditions, as assumed by the theory.

Our findings support the applicability of the CC06 theory for stochastic parameterizations in global models but also highlight areas for improvement. Particularly, the organization of clouds, most likely caused by cold pool dynamics, remains an outstanding issue. The incorporation of cloud organization in convection parameterization requires either a diagnostic grid-scale indicator of subgrid clustering or a prognostic subgrid variable.

While the 12-day period in our study presents a variety of nonequilibrium convections, they still only represent a narrow selection of all possible convective situations. In particular, it would be interesting to investigate how well the theory holds up for larger organized systems such as mesoscale convective systems or even tropical cyclones. Furthermore, the realism of our simulations is somewhat limited by our grid spacing. As shown in comparison with radar observations, our kilometer-scale model produces a large number of gridcell storms and is, on several days, not able to create lasting organization into the night. A more detailed investigation of the resolution dependence of cloud statistics would certainly be desirable, particularly as recent studies (Craig and Dörnbrack 2008; Scheufele 2014; Hanley et al. 2015; Heath et al. 2017) show a lack of convergence even at horizontal grid spacings of around 100 m.

Acknowledgments. The research leading to these results has been done within the subproject “Representing forecast uncertainty using stochastic physical parameterizations” of the Transregional Collaborative Research Center SFB/TRR 165 “Waves to Weather” funded by the German Research Foundation (DFG). We thank Robert Redl for technical support and the German Weather Service (DWD) for providing the COSMO model and the initial and boundary conditions, as well as ECMWF for providing the data for our synoptic charts. Last, we would like to thank the three anonymous reviewers for their help in improving this paper.

APPENDIX

Rationale and Formulation of the Stochastic Boundary Layer Perturbations

In the atmosphere, the boundary layer is characterized by turbulent eddies, which occur on a wide range of scales from millimeters to approximately the height of the boundary layer, typically around 1 km. Convection-permitting models with horizontal grid spacings of order 1 km are not able to represent these eddies. Therefore, parameterizations are necessary to describe the effect of the unresolved turbulence on the grid scale. Usually, these boundary layer parameterizations are assumed to represent the average effect of many eddies and are deterministic in nature: given a certain grid-scale condition, they always produce the same mean response. The only variance thus comes from the resolved boundary layer circulation. On a scale equivalent to the grid length of a kilometer-scale model, however, the turbulent response can vary significantly from realization to realization. Therefore, the variability in the model, expressed by the joint probability density function (PDF) of boundary layer quantities, can be much smaller than the corresponding variability in nature. While the mean boundary evolution might still be adequately represented, this lack of variability can drastically alter the grid-scale behavior if nonlinear convection occurs.

On typical summer days, turbulence is driven by surface heating, leading to a growth of the boundary layer after sunrise until the capping inversion is reached. At this point, only parcels with enough momentum as well as positive humidity and temperature perturbations can break through the inversion layer to eventually trigger deep convection. Parcels that have these properties originate from the extreme end of the joint boundary layer PDF. Reduced variability accordingly reduces the probability of such parcels existing. This can lead to systematic biases in model behavior. In other words, there is a disparity between the convection, which is assumed to be resolved, and the process responsible for triggering it, namely, boundary layer turbulence, which is not resolved. Note that operational models such as the COSMO
model, which is specifically designed for precipitation forecasts, are often tuned to produce the correct diurnal cycle of precipitation at the expense of other biases.

The PSP scheme aims to reintroduce the missing variability on the smallest model-resolved scale, around $\Delta x_{eff} = 5\Delta x$ (Bierdel et al. 2012). In other words, perturbations with a scale equal to the effective model resolution $\Delta x_{eff}$ and with an amplitude proportional to the subgrid standard deviation $\sqrt{\Phi^2}$ of each variable $\Phi \in \{T, q_n, w\}$ are added to resolved flow. The perturbations are introduced as a forcing term in the model equations that persists over a representative eddy lifetime $\tau_{eddy}$. Mathematically, this can be expressed as

$$\frac{\partial \Phi}{\partial t}_{PSP} = \alpha_{tuning} \frac{1}{\tau_{eddy}} \frac{I_{eddy}}{\Delta x_{eff}} \sqrt{\Phi^2}, \quad (A1)$$

The subgrid standard deviation $\sqrt{\Phi^2}$ is taken directly from the turbulence scheme. In the case of the COSMO model, this is a 1.5-order closure (Raschendorfer 2001) based on level 2.5 of Mellor and Yamada (1982). Here, the second moments are diagnostically computed based on the turbulence kinetic energy and the vertical gradient of the variable in question. The factor $I_{eddy}/\Delta x_{eff}$ scales the amplitude of the perturbations to the grid length; $\tau_{eddy} = 1\text{ km}$ is the typical size of an eddy spanning the daytime convective boundary layer. Therefore, this ratio is equal to $1/N_{eddy}$, where $N_{eddy}$ is the number of eddies in a square with edge length $\Delta x_{eff}$. This follows the CC06 theory for convective variability by assuming that the variability of the domain total depends on the number of elements in question. A larger domain relative to the eddy size contains more eddies and the variability is, therefore, reduced. We define $\eta$ as a two-dimensional random field with mean zero and standard deviation one, which is horizontally correlated using a Gaussian kernel with half width $2.5\Delta x$. The random field is kept constant for $\tau_{eddy} = 10\text{ min}$, after which a completely new random field is drawn. Finally, $\alpha_{tuning}$ is a tuning factor, which is set to 7.2. Note that the value of the tuning factor is different than in Kober and Craig (2016) because of the changes in the formulation. The tuning factor was chosen so that the effects of the PSP scheme were noticeable but reasonable (Kober and Craig 2016).

REFERENCES


