

Convective variability in real mid-latitude weather: Which processes are important?

December 21, 2016

Contents

1	Introduction	2
1.1	The cumulus parameterization problem	2
1.2	The Craig and Cohen (2006) theory	2
1.3	Research question / Aim of this study	3
2	Numerical experiments and case studies	4
2.1	Model description and set-up	4
2.2	The PSPturb turbulence scheme	4
2.3	Simulation period	5
2.4	Are the simulations realistic?	5
2.5	Are the large scales sufficiently similar and the convective scales sufficiently displaced?	5
3	Computation of M and Q statistics	6
3.1	Mass flux M	6
3.2	Heating rate Q	7
3.3	Calculation of ensemble means and variances	7
4	Simulation results	8
4.1	Relationship between variability and mean	8
4.2	Testing the assumptions of CC06 and PC08	9
5	Appendix A: Sensitivity to cloud separation algorithm	10
6	Appendix B: Sensitivity to ensemble size	11
7	Appendix C: Sensitivity to analysis height	11
8	My conclusions and questions	11
8.1	The key results and why they are relevant	11
8.2	Open questions from my side	12

1 Introduction

1.1 The cumulus parameterization problem

Convection in the atmosphere is an important process which modifies its environment by redistributing heat, moisture and momentum. It is, however, also a process which is small in scale: typical cloud sizes range from 1 km for shallow cumulus clouds to 10 km for deep convective clouds. Current global climate and weather models with grid spacings of 100 km and 10 km, respectively, are therefore unable to explicitly represent convection and rely on parameterizations to describe the effect of the unresolved convection on the resolved grid-scale. Most state-of-the-art models use a mass-flux approach (see Fig. 1) where a cloud base mass flux is used as an input for a one-dimensional cloud model which describes the mixing of updraft air with its environment, and thereby the temperature and moisture transfer. The effect of all subgrid clouds can either be represented by one single cloud in a bulk approach, or by a representative sample of clouds in what is called a spectral cloud model. The cloud base mass flux is determined by the closure assumption which itself is a function of the resolved scales, and can for instance be based on a removal of CAPE within a certain time scale. In most current operational models, parameterizations are deterministic in nature. This means that for a given large-scale state the parameterization will always produce the same output. If a model grid box contains many cloud features this approximation might be reasonable. As grid spacing become smaller, however, this determinism breaks down and fluctuations can be of the same order of magnitude as the mean response. Ignoring this subgrid variability can lead to systematic biases, under-representations of extremes and underdispersion in ensemble prediction systems. The last point has been recognized early and several ad hoc methods have been proposed to combat this lack of spread, the most prominent being SPPT. Here the parameterized tendencies are multiplied by a random factor which varies spatially and temporally:

$$\dot{\Phi} = (1 + \eta)\dot{\Phi}_{\text{param}}, \quad (1)$$

where Φ represents any model variable and η is a random variable with mean zero. The intrinsic assumption of this approach is that the standard deviation of the tendencies is proportional to the mean:

$$\langle (\delta\dot{\Phi})^2 \rangle^{1/2} = b\langle \dot{\Phi} \rangle. \quad (2)$$

The factor b depends on the distribution of the random number η . SPPT is applied to all parameterizations with the same random field. This approach has been very successful in making weather models more reliable. It has been increasingly recognized, however, that stochastic parameterizations should be process specific. This means that they should represent the true uncertainty of each physical process separately using a model of the inherent uncertainty.

1.2 The Craig and Cohen (2006) theory

One theoretical model for convective variability based on statistical mechanics has been developed by Craig and Cohen (2006)(CC06). In their work, the aim is to quantify the

fluctuations of mass flux given a certain mean mass flux $\langle M \rangle$, which is a function of the large-scale atmospheric state. The distributions of N and m are given by

$$P(m) = \frac{1}{\langle m \rangle} e^{-m/\langle m \rangle} \quad (3)$$

and

$$P(N) = \frac{\langle N \rangle^N}{N!} e^{-\langle N \rangle}. \quad (4)$$

Combining these, the distribution of the total mass flux M is given by

$$P(M) = \left(\frac{\langle N \rangle}{\langle m \rangle} \right)^{1/2} e^{-\langle N \rangle} M^{-1/2} e^{-M/\langle m \rangle} I_1 \left[2 \left(\frac{\langle N \rangle}{\langle m \rangle} M \right)^{1/2} \right], \quad (5)$$

where $I_1(x)$ is the modified Bessel function of order 1. For large (small) values of $\langle N \rangle$ the shape of this function resembles a Gaussian (Poisson) distribution. The normalized variance is given by

$$\frac{\langle (\delta M)^2 \rangle}{\langle M \rangle^2} = \frac{2}{\langle N \rangle}. \quad (6)$$

Note that $\langle M \rangle = \langle N \rangle \langle m \rangle$. Equivalently,

$$\langle (\delta M)^2 \rangle^{1/2} = \sqrt{2 \langle M \rangle \langle m \rangle}. \quad (7)$$

Note that in contrast to Eq. 2, the standard deviation here is proportional to the square root of the mean. This theory has been tested in a radiative-convective-equilibrium simulation by Cohen and Craig (2006). The simulated variance scaling agrees well with the theoretical prediction (with an error of 10%). Davies (2008) and Davoudi et al. (2010) included a diurnal cycle in their simulations. In both studies, clustering (defined as the increased probability of cloud occurrence near already existing clouds) acts to modify the normalized variability. In particular Davies (2008) found a strong diurnal cycle with increased clustering and variability in the morning and evening hours.

The CC06 theory was then used by Plant and Craig (2008)(PC08) to create a stochastic parameterization. In their approach, clouds are randomly drawn from an exponential distribution described by Eq. 3 consistent with an $\langle M \rangle$ given by the closure assumption. $\langle m \rangle = 2 \times 10^7 \text{ kg s}^{-1}$ is assumed to be a constant. This assumption is motivated by RCE simulations (e.g. Cohen and Craig, 2006). From this stochastic cloud ensemble the large-scale tendencies are then computed as the sum of the cloud model output for each plume in a spectral approach. The theoretical prediction for the variance of M (Eq. 6) is not explicitly used in PC08, but comes from the exponential m distribution combined with the random initiation of new clouds. The cloud life time is set to 45 minutes for all clouds.

1.3 Research question / Aim of this study

The primary goal of this study is to investigate the convective variability in real mid-latitude weather by using a novel approach based on a stochastic boundary-layer perturbation scheme. The research questions in particular are: (i) How well do the variance scalings proposed by CC06 and SPPT work in real weather. (ii) Which processes are the most important for convective variability? This will allow us to identify key processes which have to be included to improve stochastic convection schemes.

2 Numerical experiments and case studies

The aim is to create an ensemble of simulations of real mid-latitude weather in which the large-scale conditions are sufficiently similar, but the convective clouds have been completely displaced. The first condition is achieved by using the same initial and boundary condition for each ensemble member. To fulfill the second condition a stochastic boundary layer perturbation scheme is used. The model and the boundary layer scheme will be described now.

2.1 Model description and set-up

The model used is the COSMO model (Baldauf et al., 2011) with a horizontal grid spacing $\Delta x = 2.8$ km. The setting are the same as in the operational COSMO-DE setup with one exception, the stochastic boundary-layer scheme which will be described below. The model does not parameterize deep convection, but a parameterization for shallow convection is included. The domain size is 357 grid points in either direction with the domain centered at 10E and 50N. For the analysis a 256 by 256 grid point domain (roughly 717km) at the center of the simulation domain is considered (see Fig. 2). The 50 grid point gap to the boundary ensures that boundary spin-up effects are minimal.

Initial and boundary conditions are taken from the operational COSMO-EU (7km) deterministic forecast with a boundary condition update frequency of 1 h. All runs are started at 00UTC with a lead time of 24 h. A 50 member ensemble is created by setting a different random number seed in the stochastic boundary-layer scheme for each member. The output frequency is 30 minutes.

2.2 The PSPturb turbulence scheme

The physically-based stochastic perturbation boundary-layer scheme (PSPturb) is described and tested in Kober and Craig (2016)(KC16). A brief outline is given here now. The PSPturb scheme is an additive perturbation scheme:

$$\dot{\Phi}_{\text{total}} = \dot{\Phi}_{\text{parameterized}} + \eta \sigma_{\dot{\Phi}_{\text{parameterized}}} \quad (8)$$

These perturbations (last term) are process-specific, so for each parameterized process the perturbations have to be calculated separately. The last term in the equation above contains a random number $\eta = N(0, 1)$ and the standard deviation σ of the parameterized tendencies. The random number field has a horizontal correlation length of $5\Delta x$, the effective resolution of a numerical model (Skamarock, 2004), and is held constant for 10 minutes and then drawn again from scratch. This represents a typical eddy turnover time in the boundary layer. In KC16 the standard deviation term is approximated by

$$\sigma_{\dot{\Phi}_{\text{parameterized}}} = \alpha_{\text{const}, \Phi} \frac{l_\infty}{5\Delta x} \frac{1}{dt} \sigma_\Phi, \quad (9)$$

where $l_\infty = 150$ m is the mixing length describing the average size of an eddy. The term σ_Φ is the sub-grid scale standard deviation. For the turbulence perturbations the considered variables are vertical velocity w , potential temperature θ and humidity q . The standard deviations are calculated in the turbulence parameterization (see KC16 for details). The

factor $\frac{l_\infty}{5\Delta x} \propto 1/\sqrt{N_{\text{eddy}}}$ scales the variability according to number of unresolved eddies similar to Eq. 6. The factor $\frac{1}{dt}$ converts the term into a tendency term dependent on the time step. Finally, a scaling factor $\alpha_{\text{const},\Phi}$ is included for tuning purposes and should be of order one. It is set to 2 for these experiments.

2.3 Simulation period

The simulations were run for a continuous 12 day period from 28 May – 8 June 2016 which was characterized by strong convective rainfall over Central Europe (ask Christian for a review of this period). For a large portion of this period a low pressure system was stationed over the Central Alpine region causing South-Easterly advection over Germany. The precipitation largely followed a typical diurnal cycle (Fig. 6), along with a build up of convective available potential energy (CAPE) in the morning and a growing boundary layer (using as a measure the perturbed boundary layer height model output). The convective adjustment time scale τ_c , a measure of the synoptic forcing, shows intermediate values of around 5 h indicating moderate synoptic forcing, typical for the scattered convection seen in most of the simulated days (for an introduction to τ_c see Done et al. (2006), for a recent paper describing the calculation method used here see Flack et al. (2016)).

2.4 Are the simulations realistic?

To test whether the numerical simulations can represent convection realistically, the precipitation fields are compared to Radar-derived precipitation observations. A visual impression of the similarity between model output and observations can be gained from Fig. 2. To get a quantitative picture a precipitation histogram is used to check whether the simulations produce a realistic distribution of rain (Fig. 3). Overall the agreement is good. The model seems to produce slightly more very light rain and very heavy rain. To test whether the spatial distribution of clouds is comparable we show the radial distribution function (Fig. 4). This is particularly important since clustering will turn out to be an important factor in modulating convective variability. Both, simulations and observations, show a diurnal cycle with a clustering minimum during the convective maximum at around 15 UTC and increased clustering towards the evening hours. In general the clustering in the model appears to be sharper, meaning that the peak is larger, but then drops off faster. This implies that the model likes to produce clouds directly next to existing clouds (partly this can be an artifact of the cloud separation method), but is less likely to produce larger clusters, a known problem in convection-permitting models. With this in mind, we can still conclude that the model simulations are a realistic representation of convection.

2.5 Are the large scales sufficiently similar and the convective scales sufficiently displaced?

The premise of this study is that the convective scales are completely displaced, while the large-scale conditions are reasonably similar. A visual, qualitative test of this assumption is shown in Fig. 2. On the large scales the members agree on the location of the

precipitation, but zooming in on the convection itself reveals no perceptible correlation. A more quantitative approach is given by the ratio of the difference to the background energy spectra of kinetic energy and precipitation (Fig. 5). A complete displacement at a given scale would result in a difference energy which is twice the background energy. The precipitation, representative of the convective features, shows some error growth early in the simulations. By 06 UTC, the displacement for scales up to 50 km is mostly complete. We, therefore, choose 06 UTC to start our analysis. To see how similar the large-scale environment is we look at the difference kinetic energy spectrum. Here the largest scales seem to agree very well (with a ratio of around 10%) even at later times after the errors have grown. The small scales show a strong upscale error growth, suggesting that it takes longer for the displacement to be complete in the wind field. Since we are interested in the convection, however, it is sufficient if the convective features (i.e. the precipitation) is completely displaced. We therefore conclude that the basic assumption of displaced small scales and comparable large scales is met.

3 Computation of M and Q statistics

3.1 Mass flux M

The vertical mass flux is defined as the mass of air crossing a certain horizontal area per unit time. Since our interest is in the convective mass flux, we first have to identify the convective clouds (for an illustration of the process see Fig. 7). To do this we follow Cohen and Craig (2006) and many other previous and subsequent studies by creating a binary cloud/no cloud field using a threshold for vertical velocity $w > 1 \text{ m s}^{-1}$ combined with a positive cloud water content $q_c > 0 \text{ kg kg}^{-1}$. By doing this we aim to identify convective updrafts, ignoring downdrafts. Next, contiguous areas are identified as clouds using a 4-point segmentation algorithm so that only pixels which share an edge are considered as contiguous clouds. Additionally, “overlapping” clouds are identified with the local maximum method, followed by a watershed algorithm to find the extent of each separated cloud. The impact of this cloud separation is discussed in Appendix A.

For each identified cloud $k = 1, \dots, N_{\text{cld},i}$ in each ensemble member $i = 1, \dots, N_{\text{ens}}$ a cloud size σ_k is determined as

$$\sigma_k = N_{\text{px}} \Delta x^2, \quad (10)$$

where N_{px} is the number of pixels for each cloud k . The mass flux per cloud m_k is computed as

$$m_k = \Delta x^2 \sum_l^{N_{\text{px}}} w_l \rho_l, \quad (11)$$

where ρ is density. The resulting cloud size and cloud mass flux distributions are shown in Fig. 8. Both follow an exponential distribution which is in agreement with the assumptions of CC06. The mean cloud size is $2.1 \times 10^7 \text{ m}^2$, which corresponds to 2.7 grid boxes. The mean cloud mass flux is $4.7 \times 10^7 \text{ kg s}^{-1}$. The correlation coefficient for size and mass flux is 0.86. The RDF shows a distinct peak for the second largest search radius. This peak is partly a result of the cloud separation (see Appendix A for a discussion) but also indicates that updrafts tend to be very close together in our simulations. Since the RDF

is dominated by this distinct peak, clustering at larger radii cannot be detected using this algorithm. The diurnal variations will be discussed in more detail in Section 4.

The mass flux calculation have to be done at a certain vertical level. Ideally one would chose the cloud base mass flux, which serves as the input for many cloud models used in convective parameterizations. In our simulations, however, the height of the cloud base varies temporally and spatially. We want to make sure that we are always above the boundary layer. Another complication in our real case studies is the orography. The surface elevation varies by more than 500 m in the analysis domain (see bottom left panel in Fig. 2). The terrain-following model levels were chosen as a good compromise between a height above ground level approach close to the surface and a height above sea level approach at higher altitudes. Fig. 11 shows vertical profiles of the domain integrated mass flux at two different times (others are similar). The level at which the maximum mass flux appears changes only slightly with time and is typically located around level 30, which was chosen to conduct our analysis. This corresponds to a height of around 3000 m for a column located above sea level. Additionally, the profiles for the Northern and Southern half of the analysis domain show a good agreement. In Appendix B the sensitivity of the diagnostics to the analysis height is shown.

3.2 Heating rate Q

The heating rate used here is the temperature tendency computed by the microphysics parameterization. It is, therefore, not purely restricted to convective heating or cooling, which would be Q_1 in the typical notation. Q was summed in the vertical to represent the total heating in the column.

3.3 Calculation of ensemble means and variances

For the variance calculations, a coarse-graining is applied to create coarse boxes $j = 1, \dots, N_{\text{box},n}$ with edge lengths of $n = 256, 128, 64, 32, 16, 8$ and $4\Delta x$, where $N_{\text{box},n} = (256/n)^2$. No smaller neighborhoods are considered, since these would be significantly below the effective resolution of the model. The total mass flux per box per member $M_{i,j,n}$ is given by

$$M_{i,j,n} = \sum_{k=1}^{N_{\text{cld } i,j,n}} m_{k,i,j,n}. \quad (12)$$

To deal with clouds at the boundaries of the coarse-fields, the centers of mass for each cloud are first identified. Then the m_k is attributed to that one point in space. Therefore, the coarse box which contains the center of mass also contains the entire cloud, while the other box does not contain any of the cloud. This follows Cohen and Craig (2006). $N_{i,j,n} = N_{\text{cld } i,j,n}$ is simply the number of clouds which fall into each box. The heating rate per box $Q_{i,j,n}$ is the average of all grid points in the box.

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

$$\langle (\delta\Phi)^2 \rangle_{j,n} = \frac{1}{N_{\text{ens}} - 1} \sum_{i=1}^{N_{\text{ens}}} (\Phi_{i,j,n} - \langle \Phi \rangle_{j,n})^2, \quad (13)$$

where the ensemble mean is

$$\langle \Phi \rangle_{j,n} = \frac{1}{N_{\text{ens}}} \sum_{i=1}^{N_{\text{ens}}} \Phi_{i,j,n}. \quad (14)$$

To compute statistics for m a different approach is taken. Here the clouds in all members for each box are considered together to calculate the variance and mean. The total number of clouds over all ensemble members is denoted by $N_{\text{cldtot}} = \sum_{i=1}^{N_{\text{ens}}} N_{\text{cld},i,j,n}$.

$$\langle (\delta m)^2 \rangle_{j,n} = \frac{1}{N_{\text{cldtot}} - 1} \sum_{k=1}^{N_{\text{cldtot}}} (m_{k,j,n} - \langle m \rangle_{j,n})^2, \quad (15)$$

where the mean is

$$\langle m \rangle_{j,n} = \frac{1}{N_{\text{cldtot}}} \sum_{k=1}^{N_{\text{cldtot}}} m_{k,j,n}. \quad (16)$$

Composites were computed by averaging over the 12 days in the simulation period. For the calculation of means and standard deviations, all coarse boxes at scale n for all ensemble members were first combined, and then the means and standard deviations were calculated. This ensures that, since the number of coarse boxes with clouds will differ from case to case, every coarse box is weighted equally.

4 Simulation results

4.1 Relationship between variability and mean

Our dataset now contains values of $\langle M \rangle$, $\langle (\delta M)^2 \rangle$, $\langle Q \rangle$, $\langle (\delta Q)^2 \rangle$ for each coarsening scale n , grid box j and time t . We now proceed to test the functional dependence of the variability in terms of standard deviation on the mean values. Specifically, we would like to test Eq. 2 and Eq. 7. Fig. 12a shows a scatter plot of $\langle (\delta M)^2 \rangle^{1/2}$ and $\langle M \rangle$ for all data points. There seems to be a scale below which the standard deviation drops off. This is discussed further in Appendix B. Additionally, least square fits are plotted for a linear relationship $y = bx$ (c.f. Eq. 2) and a square root relationship $y = \sqrt{(bx)}$ (c.f. Eq. 7) for each coarsening scale separately. In Fig. 12b the fitting parameter b is plotted for each scale (note that for the square root fit b was divided by a factor given in the legend to fit the scale) along with the normalized root mean squared error (NRMSE) of the fit, computed as $NRMSE = \sqrt{1/N \sum^N (\langle (\delta M)^2 \rangle^{1/2} - (b\langle M \rangle)^p / \langle (\delta M)^2 \rangle^{1/2})^2}$ with $p = 1$ for the linear fit and $p = 1/2$ for the square root fit. The two main results are: (i) the square root fit is better than the linear fit for all scales as measured by the NRMSE; (ii) the square root fit is scale adaptive. The fitting parameter b for the linear fit decreases as the coarsening scale increases, while for the square root fit b stays relatively constant with variations of around 30%. These results suggest that the CC06 scaling works reasonably well for our real-world data. In particular it is scale-aware. The linear assumption does not seem to fit the mass flux variability very well. SPPT, based on a linear assumption, is not defined for the mass flux, however, but for the output of the parameterizations, such as the heating tendency. We, therefore, want to test whether the finding for M also apply to Q .

The first step is to see how well the means and standard deviations of M and Q correlate (Fig. 12d). For all scales there seems to be a good correlation, particularly for

the variability. A scatter plot of $\langle(\delta Q)^2\rangle^{1/2}A$ and $\langle Q \rangle A$ (Fig. 12c; the factor $A = n^2$ has to be introduced since Q is an average quantity while M is a summed quantity) reveals one major difference between the two variables. Whereas M , by definition, cannot be negative, Q can. Both SPPT and CC06 are designed so that the variability goes to zero as the mean goes to zero. This offset was recognized by Shutts and Palmer (2007) who argue that for the convective heating rate (note the different methodologies) the standard deviation depends linearly on the mean, but with a offset at zero mean. To further test how well the linear CC06 theory applies to the heating rate, $\langle(\delta Q)^2\rangle^{1/2}A$ is scattered against $\langle M \rangle$ (Fig. 12e). Again least square fits are computed for Fig. 12c,e along with the NRMSE (Fig. 12f). The results mirror these of the M fits, albeit with larger errors. Overall, Q and M behave reasonably similar suggesting that for our data a square root relation between the standard deviation and mean represents a good, scale-adaptive model of convective variability. There are, however, still significant deviations of our simulation data from the fitted curves, which we will investigate now.

4.2 Testing the assumptions of CC06 and PC08

Specifically, we will test how well the predicted variance corresponds with the variance computed from the simulation data by taking the ratio of the two and then, step by step, investigating the impact of several assumptions. The starting point is the variance which follows from using the PC08 scheme, which in addition to all CC06 assumptions also assumes that $\langle m \rangle$ is constant. We take $\langle m \rangle = m_c = 4 \times 10^7 \text{ kg s}^{-1}$, which is close to the mean for our simulations. A scatter plot showing the simulation to prediction ratio along with the temporal evolution of the mean values for each scale is shown in Fig. 13a,b. On average, the predictions are good, but the fluctuations around the mean are large. In addition, there is a distinct diurnal cycle visible with an increase from 15–18UTC, indicating that the theory over-predicts the variability in the morning up to the convective maximum at around 15UTC, and subsequently under-predicts the variability as the convective activity decreases.

Next, we include the variations in $\langle m \rangle$ (Fig. 13f) in the predictions, so that they now match the original CC06 theory (Fig. 13d,e). Including variation in the mean cloud mass flux reduces the fluctuations of the predictions (as visible by the error bars). This is particularly pronounced for small n where the diurnal variations are also reduced. Overall, it seems that there are significant fluctuations in $\langle m \rangle$ for small coarsening scales, where the number of clouds is small. Including these fluctuations in the predictions tends to improve the quality of those predictions.

To proceed, we now want to test two fundamental assumptions made by CC06: (i) N follows a Poisson distribution (Eq. 4) and (ii) m follows an exponential distribution (Eq. 3). To do this the theory of random sums (Taylor and Karlin, 1998) is used to include deviations from these assumptions in the variance predictions.

$$\langle(\delta M)^2\rangle = \langle N \rangle \langle(\delta m)^2\rangle + \langle m \rangle^2 \langle(\delta N)^2\rangle \quad (17)$$

can be rewritten as

$$\langle(\delta M)^2\rangle = (\alpha + \beta) \langle M \rangle \langle m \rangle, \quad (18)$$

where $\alpha = \frac{\langle(\delta N)^2\rangle}{\langle N \rangle}$ and $\beta = \frac{\langle(\delta m)^2\rangle}{\langle m \rangle^2}$. Both should be 1 if the CC06 assumptions hold. α is a measure of how clustered the clouds are. Values smaller than one indicate a more

regular spacing between the clouds while values larger one indicate preferred clustering. This parameter shows a strong diurnal signal (Fig. 13i), particularly for larger n , with a minimum in clustering from 12–15UTC and increased values towards the evening hours. Another measure of clustering, the RDF confirms this trend (Fig. 9). Accounting for the variations of α in the predictions strongly reduces the diurnal signal and the fluctuations of the predictions (Fig. 13g,h). For all times the theory now seems to over-predict the simulated variance.

The same can be done for β , which measures if the cloud mass flux distribution is narrower or broader than expected from an exponential distribution. Surprisingly, this parameter shows a dependence on scale (Fig. 13m). Smaller n have a narrower distribution. The reason for this is not clear yet. This behavior does not seem to be a sampling issue caused by an insufficient number of ensemble members (see Appendix B). Could it be a result of clouds not being point-like? Correcting for these deviations in the predictions increases the mean for smaller n , but otherwise does not seem to affect the predictions (Fig. 13i,l).

Finally, including both, α and β , in the predictions drastically improves the predictions compared to the original PC08 or CC06 predictions, for small and medium scales up to around 180km the predictions are basically perfect in their mean with little variations. For the largest n there still is some over-prediction and a remaining diurnal cycle. The one assumption not taken into account yet is the correlation between m and N . Fig. 10 clearly shows that the two are correlated in their mean values. When there are more clouds, they also tend to be larger.

The most striking signal seems to be caused by the clustering of clouds. There is a diurnal variation of up to 50% in the predictions which is basically removed by taking α into account.

5 Appendix A: Sensitivity to cloud separation algorithm

As can be seen in Fig. 7, the cloud separation algorithm works very aggressively resulting in many grid point clouds. For comparison Figs. 14 and 15 show the cloud size distribution and RDF without the cloud separation algorithm (c.f. Figs. 14 and 15). The mean cloud size now is $2.7 \times 10^7 \text{ m}^2$ (compared to $2.1 \times 10^7 \text{ m}^2$ with the separation) and the mean cloud mass flux is $6.1 \times 10^7 \text{ kg s}^{-1}$ (compared to $4.7 \times 10^7 \text{ kg s}^{-1}$). The distribution now has a longer tail with the largest clouds being around three times as large. The fact that the mean only changes by around 30% (the number of clouds also differs by this amount) but the largest clouds grow by around 300% indicates that the separation algorithm acts mostly on the largest clouds. For the analyses in this study we used the separated clouds.

The RDF shows a much less pronounced peak but the diurnal cycle is largely unchanged. The location of the peak has shifted by one Δx , still suggesting that updrafts as resolved by our convection-permitting simulations tend to be close together.

6 Appendix B: Sensitivity to ensemble size

Since we are sampling a continuous distribution with a finite sized ensemble of simulations it is important to check whether the key results depend on the number of ensemble members. First, the standard deviation scatter plots for 50 and 20 ensemble members are compared (Figs. 12 and 16). Two things are noticeable: (i) the error bars (NRMSE) become much smaller, indicating that the fits show a better agreement, while the mean values remain largely unchanged (except for the smallest n which will be discussed below); (ii) The scale at which the standard deviation drops off is larger for the 20 member analysis, suggesting that this is indeed a sampling issue. The scale is approximately proportional to the inverse of the ensemble size. This indicates that there is a fixed total mass flux $\sum^{N_{ens}M}$ below which the standard deviation becomes smaller. From the data available this seems to happen at around $\sum^{N_{ens}M} = 10\langle m \rangle$. Currently, a filter is applied where data points are discarded for grid points and time steps where less than 5 members have a cloud. It would potentially make sense to filter out all data points below this drop-off scale. This mostly affects the smallest n .

Second, the dependence of the simulation versus prediction comparison is discussed (Figs. 13 and 17). There is a change in the x-location of the data points for small n . This is due to the fixed filter mentioned above. The mean values change very little for the diagnostics, only the fluctuations around the mean become slightly smaller. In particular the β parameter does not seem to depend on the ensemble size. The scale dependence does therefore not seem to be a sampling issue. The insensitivity of the key features on ensemble size suggest that the sample size is large enough to make robust statements.

7 Appendix C: Sensitivity to analysis height

As mentioned in Section 3 the choice of vertical level to compute the mass flux statistics is complicated by the temporal and spatial inhomogeneities in our real world simulations. The goal of this appendix is to assess whether the variation of the key diagnostics with analysis height are significant compared with the temporal variability at one analysis height. Figs. 17, 18 and 19 show the diagnostics at 3, 2 and 4 km height, respectively (for 20 members). For most quantities shown the change with height (in this reasonable range) is much smaller than the diurnal variation (see for example α). The only exception is the β parameter which increases with height. Despite the difficulties in choosing an exact analysis height the results seem to be relatively insensitive from 2–4 km height, giving us confidence that our results are robust.

8 My conclusions and questions

8.1 The key results and why they are relevant

If I had to summarize the key points of this study I would say:

1. We used a novel approach – using BL perturbations to displace the convective scales while keeping the large scales the same – to investigate the variability of convection in “real” weather.

2. We show that the CC06 variance scaling also works in “real” weather situations with reasonable deviations. In particular, it seems to work better than the linear approach which SPPT is based on, even though SPPT is of course a much broader approach.
3. We show that the diurnal variation in clustering is the biggest factor modifying convective variability. This indicates that an improved model of convection needs to have some way of including this clustering. This probably requires identifying the key process causing the clustering and including memory.

8.2 Open questions from my side

Here are the main open questions I have about this study at the moment. These are maybe the sort of things I would questions as a reviewer.

1. Do I follow the right line of reasoning? Does it make sense to start from the parameterization view? Are the research questions well defined? Are the conclusions I am drawing (see above) reasonable?
2. Given the dataset available, am I looking at the right diagnostics to answer the research questions? This one I am particularly unsure about. The dataset spans many orders of magnitude and dimensions which caused me problems deciding how to display it best. Is there information which in the dataset which I did not consider?
3. I discussed several sensitivities in the Appendices. Do these discussion make sense? Are there any other sensitivities I did not look at, which could be important? All in all, is the analysis “bulletproof”, at least to the extent where we can say that the key results are sound?
4. Once Julia is back we will try to use her model to create improved predictions. If this work we could also include this as another part in this study. Whether this works or not, is the current analysis enough to “stand on its own legs”? Is there enough new information in there to make it publishable?

References

- 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. *Monthly Weather Review*, **139** (12), 3887–3905, doi:10.1175/MWR-D-10-05013.1, URL: <http://journals.ametsoc.org/doi/abs/10.1175/MWR-D-10-05013.1>.
- Cohen, B. G. and G. C. Craig, 2006: Fluctuations in an Equilibrium Convective Ensemble. Part II: Numerical Experiments. *Journal of the Atmospheric Sciences*, **63** (8), 2005–2015, doi: 10.1175/JAS3710.1, URL: <http://journals.ametsoc.org/doi/abs/10.1175/JAS3710.1>.
- Craig, G. C. and B. G. Cohen, 2006: Fluctuations in an Equilibrium Convective Ensemble. Part I: Theoretical Formulation. *Journal of the Atmospheric Sciences*, **63** (8), 1996–2004, doi: 10.1175/JAS3709.1, URL: <http://journals.ametsoc.org/doi/abs/10.1175/JAS3709.1>.
- Davies, L., 2008: Self-organisation of convection as a mechanism for memory. Ph.D. thesis.
- Davoudi, J., N. A. McFarlane, and T. Birner, 2010: Fluctuation of Mass Flux in a Cloud-Resolving Simulation with Interactive Radiation. *Journal of the Atmospheric Sciences*, **67** (2), 400–418, doi:10.1175/2009JAS3215.1, URL: <http://journals.ametsoc.org/doi/abs/10.1175/2009JAS3215.1>.
- Done, J. M., G. C. Craig, S. L. Gray, P. a. Clark, and M. E. B. Gray, 2006: Mesoscale simulations of organized convection: Importance of convective equilibrium. 737–756, doi:10.1256/qj.04.84, URL: <http://centaur.reading.ac.uk/5286/>.
- Flack, D. L. A., R. S. Plant, S. L. Gray, H. W. Lean, C. Keil, and G. C. Craig, 2016: Characterisation of Convective Regimes over the British Isles. *Quarterly Journal of the Royal Meteorological Society*, n/a–n/a, doi:10.1002/qj.2758, URL: <http://doi.wiley.com/10.1002/qj.2758>.
- Kober, K. and G. C. Craig, 2016: Physically Based Stochastic Perturbations (PSP) in the Boundary Layer to Represent Uncertainty in Convective Initiation. *Journal of the Atmospheric Sciences*, **73** (7), 2893–2911, doi:10.1175/JAS-D-15-0144.1, URL: <http://journals.ametsoc.org/doi/10.1175/JAS-D-15-0144.1>.
- Plant, R. S. and G. C. Craig, 2008: A Stochastic Parameterization for Deep Convection Based on Equilibrium Statistics. *Journal of the Atmospheric Sciences*, **65** (1), 87–105, doi:10.1175/2007JAS2263.1, URL: <http://journals.ametsoc.org/doi/abs/10.1175/2007JAS2263.1>.
- Selz, T. and G. C. Craig, 2015: Simulation of upscale error growth with a stochastic convection scheme. *Geophysical Research Letters*, **42** (8), 3056–3062, doi:10.1002/2015GL063525, URL: <http://doi.wiley.com/10.1002/2015GL063525>.
- Shutts, G. J. and T. N. Palmer, 2007: Convective Forcing Fluctuations in a Cloud-Resolving Model: Relevance to the Stochastic Parameterization Problem. *Journal of Climate*, **20** (2), 187–202, doi:10.1175/JCLI3954.1, URL: <http://journals.ametsoc.org/doi/abs/10.1175/JCLI3954.1>.
- Skamarock, W. C., 2004: Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra. *Monthly Weather Review*, **132** (12), 3019–3032, doi:10.1175/MWR2830.1, URL: <http://journals.ametsoc.org/doi/abs/10.1175/MWR2830.1>.

Taylor, H. and S. Karlin, 1998: *An introduction to stochastic modeling*. URL:
<https://books.google.de/books?hl=en&lr=&id=ppHiBQAAQBAJ&oi=fnd&pg=PP1&dq=An+Introduction+to+Stochastic+Modeling+by+Howard+Taylor+and+Stanislaw+Karlin>

9 Figures

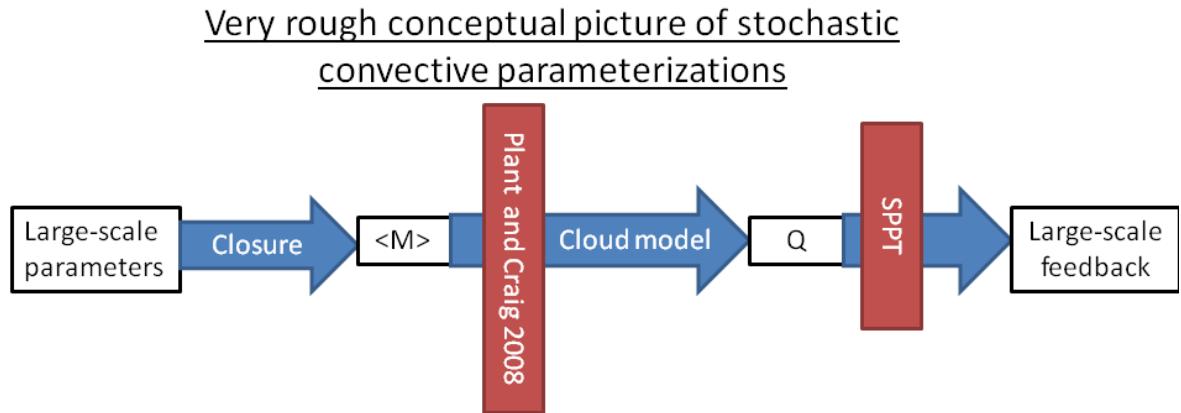


Figure 1: Conceptual picture of stochastic convection parameterizations. Red areas indicate where the stochastic parts are active in PC08 and SPPT.

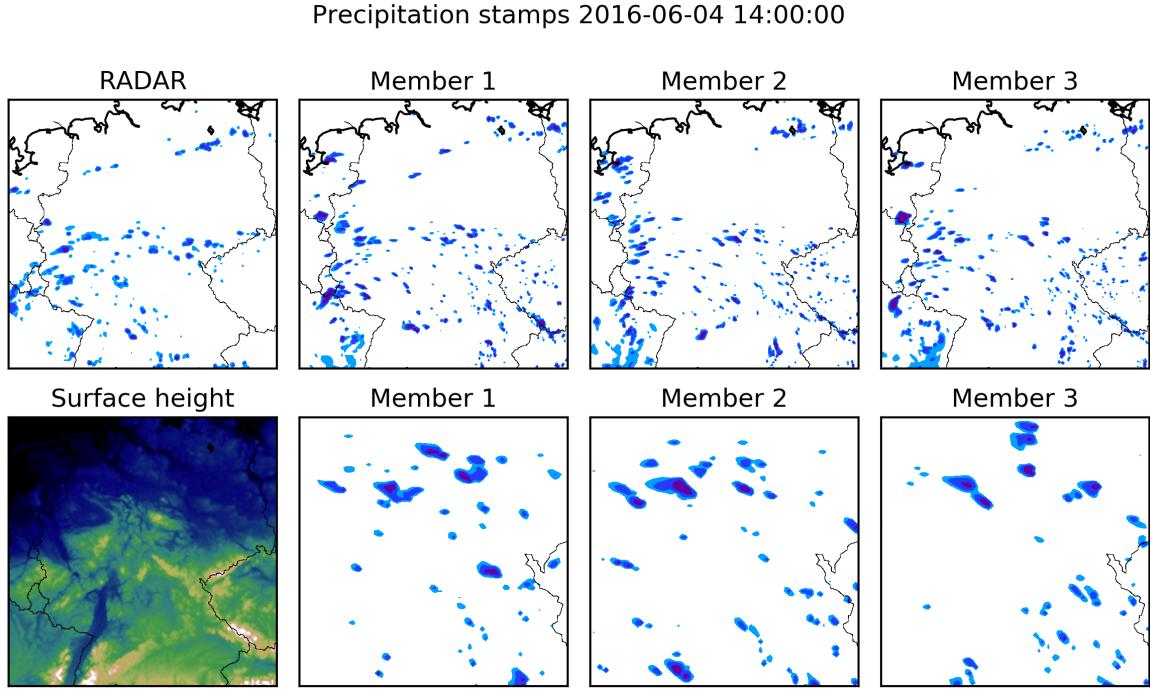


Figure 2: (top right) Radar observations, (top row) Three ensemble members, (bottom row) zoom of plot on top and (bottom left) surface elevation

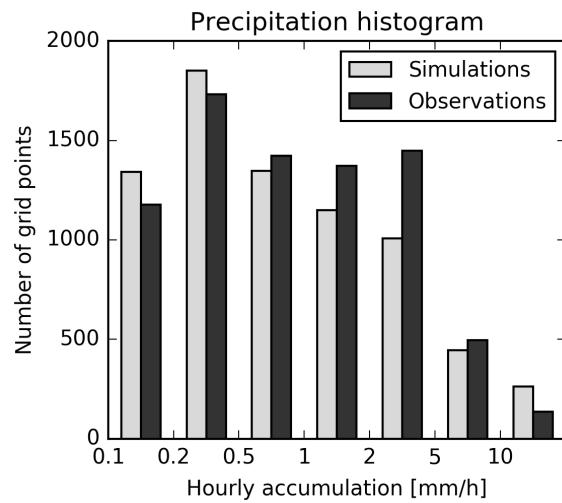


Figure 3: Precipitation histogram showing the number of grid cells within a specified hourly precipitation range. Values are averaged for all analysis times and days. Additionally the simulation numbers are ensemble averages. The “no rain” bin (0–0.1 mm/h) is not shown and accounts for the differences in the total number of shown values.

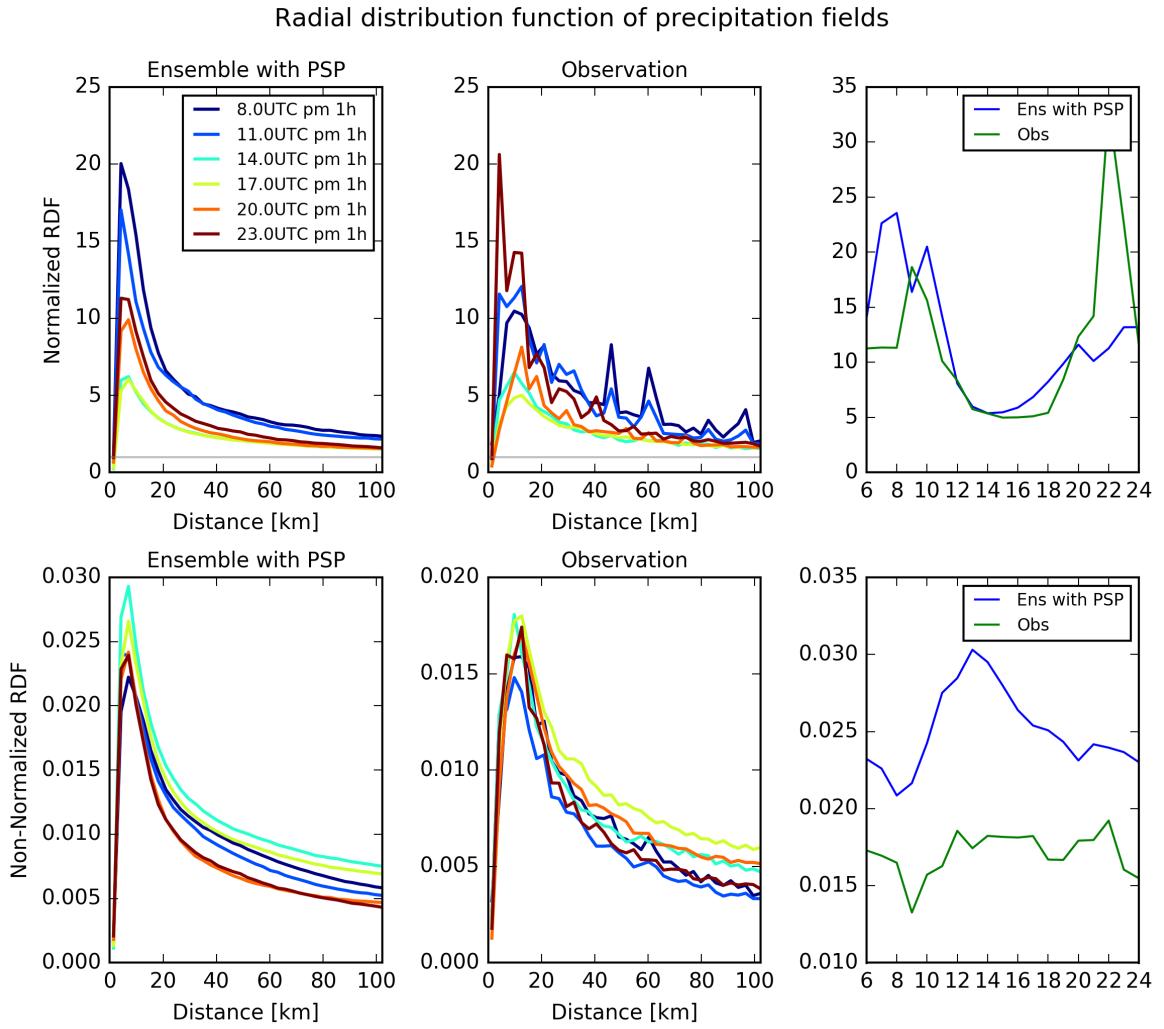


Figure 4: Normalized radial distribution function (RDF) of the hourly precipitation fields for simulations and observations. Every line shown represent a three-hourly average. The maximum search radius is $36\Delta x$ (approximately 100 km) a resolution of one Δx . To identify cloud cloud objects a rain threshold of 1 mm h^{-1} was used, followed by a separation using the same method described for mass flux objects described in Section ??.

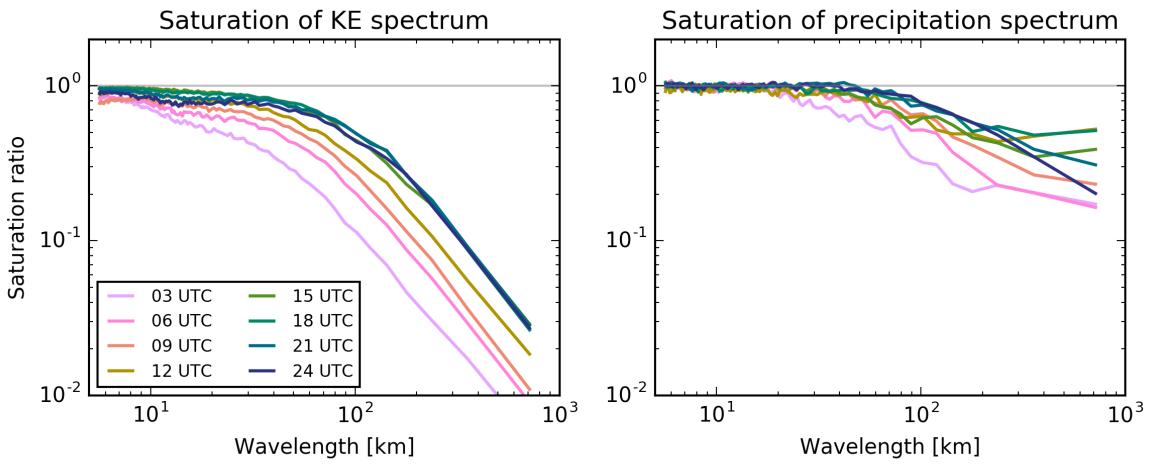


Figure 5: Saturation ratio is defined as the difference energy spectrum divided by twice the background energy spectrum for (left) precipitation and (right) kinetic energy (KE). The KE spectrum was averaged from the lowest model level to level 15 (approximately 10 km height). All spectra are instantaneous and ensemble means, where a reduced 5 member ensemble has been used. For details on the calculation of energy spectra, refer to Selz and Craig (2015)

Temporal evolution of domain averaged quantities

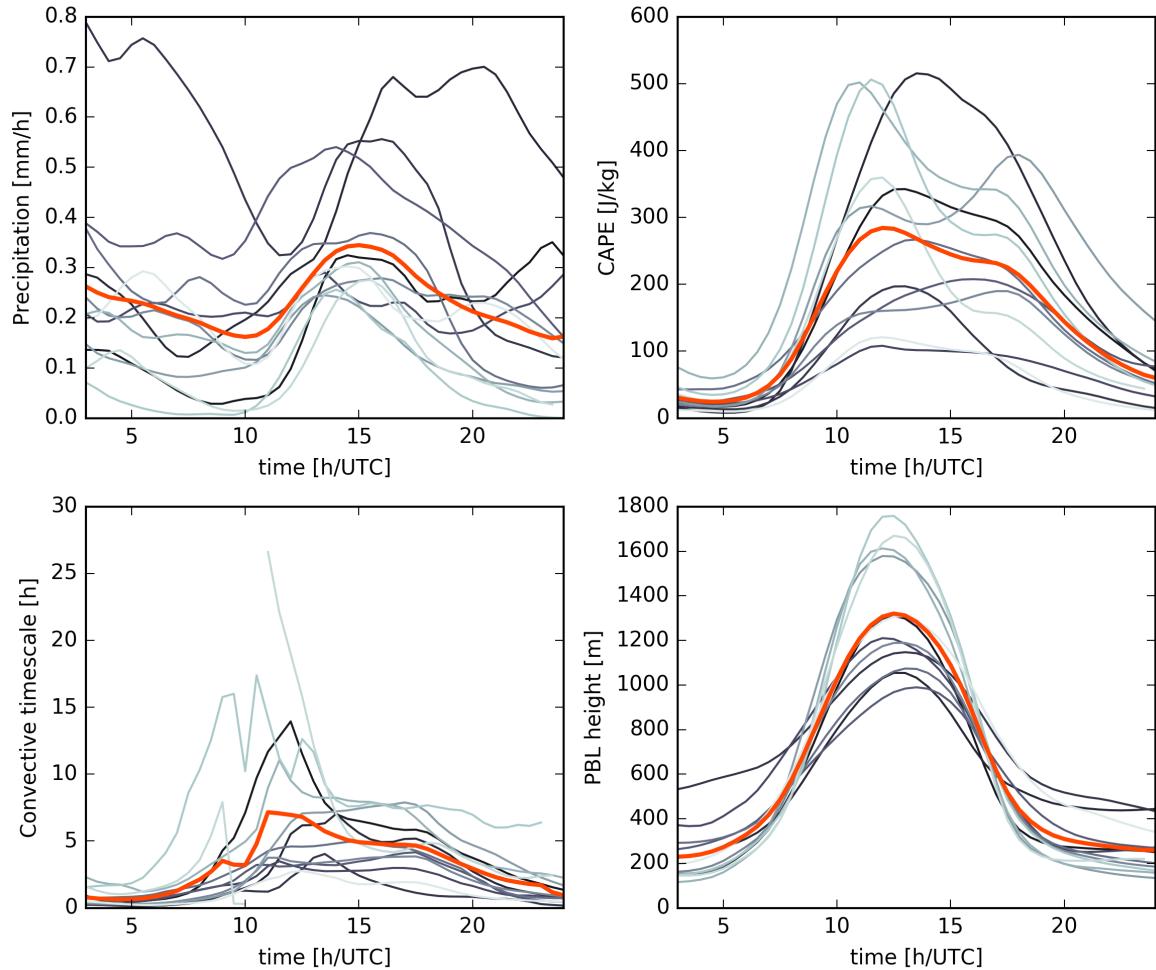


Figure 6: Time series of (a) hourly rainfall in mm/h, (b) CAPE in J/kg, (c) the convective adjustment time scale in h and (d) the perturbed boundary layer height in m. All values are domain and ensemble averages. Each gray line represents one simulation day. The red line is the mean over all simulation days.

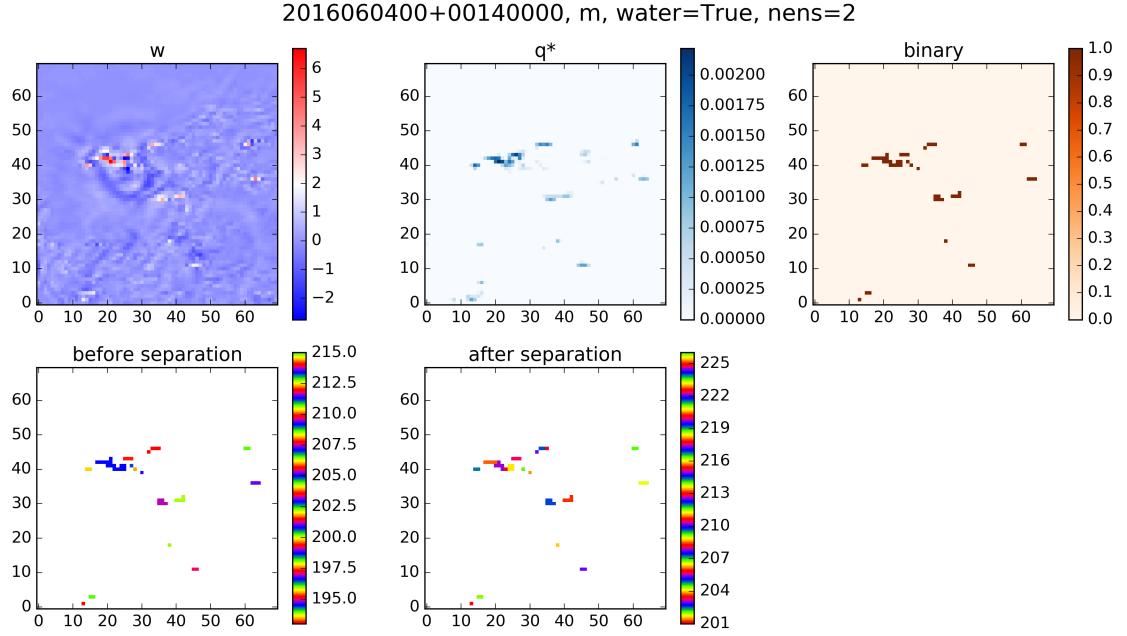


Figure 7: Cloud identification algorithm. First, a binary field (c) is created by applying thresholds to the vertical velocity (a) and cloud water (b) fields. Then, the contiguous clouds are identified (d), followed by a separation using a local maximum and watershed method (e)

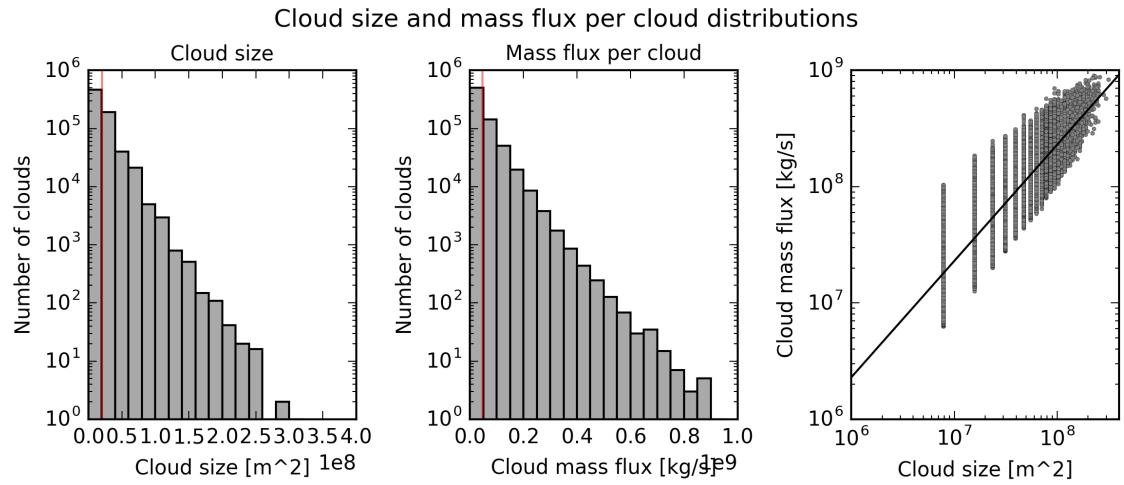


Figure 8: Distribution of (left) cloud size and (middle) cloud mass flux on a log-linear plot. (right) Scatter plot of cloud size against cloud mass flux.

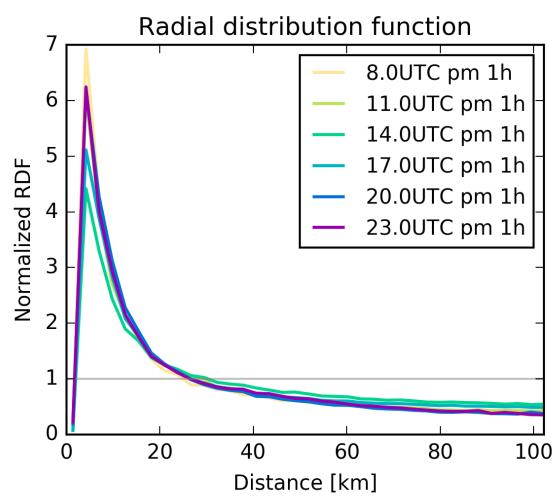


Figure 9: Radial distribution function for the center of mass of each identified cloud. The methodology is the same as used for Fig. 4.

Temporal evolution of domain averaged quantities

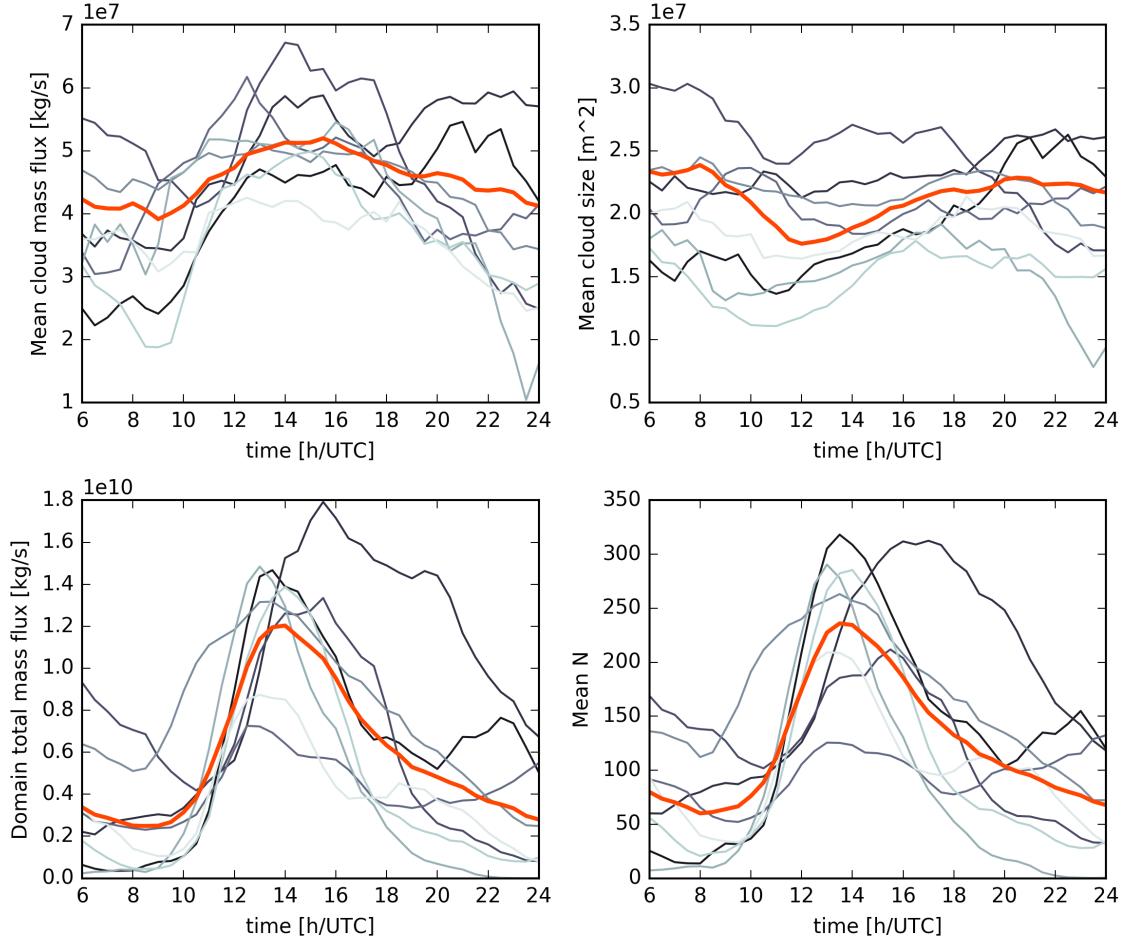


Figure 10: Temporal evolution of ensemble mean quantities. (top left) $\langle m \rangle$, (top right) $\langle \sigma \rangle$, (bottom left) domain total $\langle M \rangle$ and (bottom right) domain total $\langle N \rangle$. Each gray line represents one simulation day, the red line represents the composite mean.

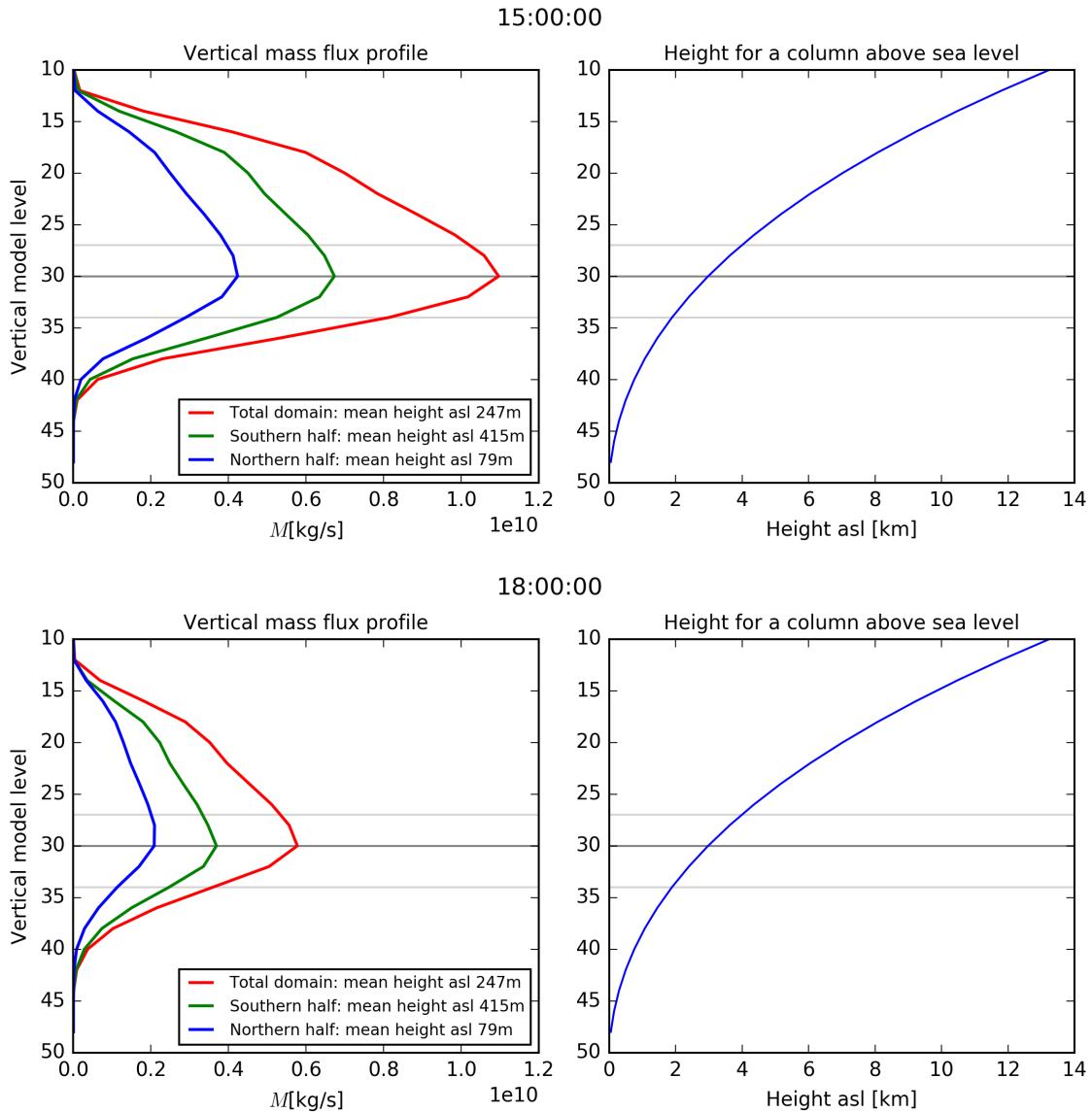


Figure 11: (Left) Vertical profiles of the domain integrated mass flux for the entire simulation domain (red), and the Southern (green) and Northern (blue) half of the domain for 15UTC (top) and 18UTC (bottom). (Right) Height above sea level for a column located over the ocean.

Scaling of standard deviation with mean

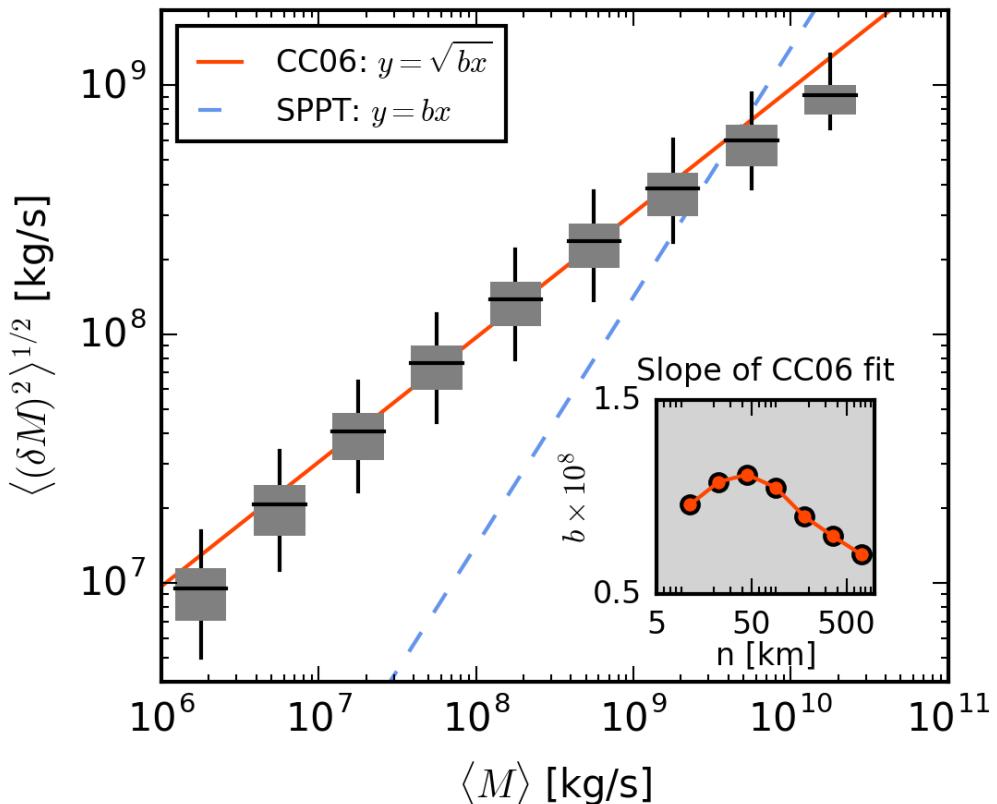


Figure 12: (left column) Scatter plots showing the relation of standard deviations to mean values. Each point represents one time and one box. Lines represent least square fits for $y = bx$ (SPPT, dashed) and $y = (by)^{1/2}$ (CC06, dash-dot). (right column top and bottom) Fit parameter b is shown with dots (SPPT) and crosses (CC06, multiplied by a factor), error bars represent the normalized RMS error of the fits.

Comparison of simulation results and predictions
of increasing complexity

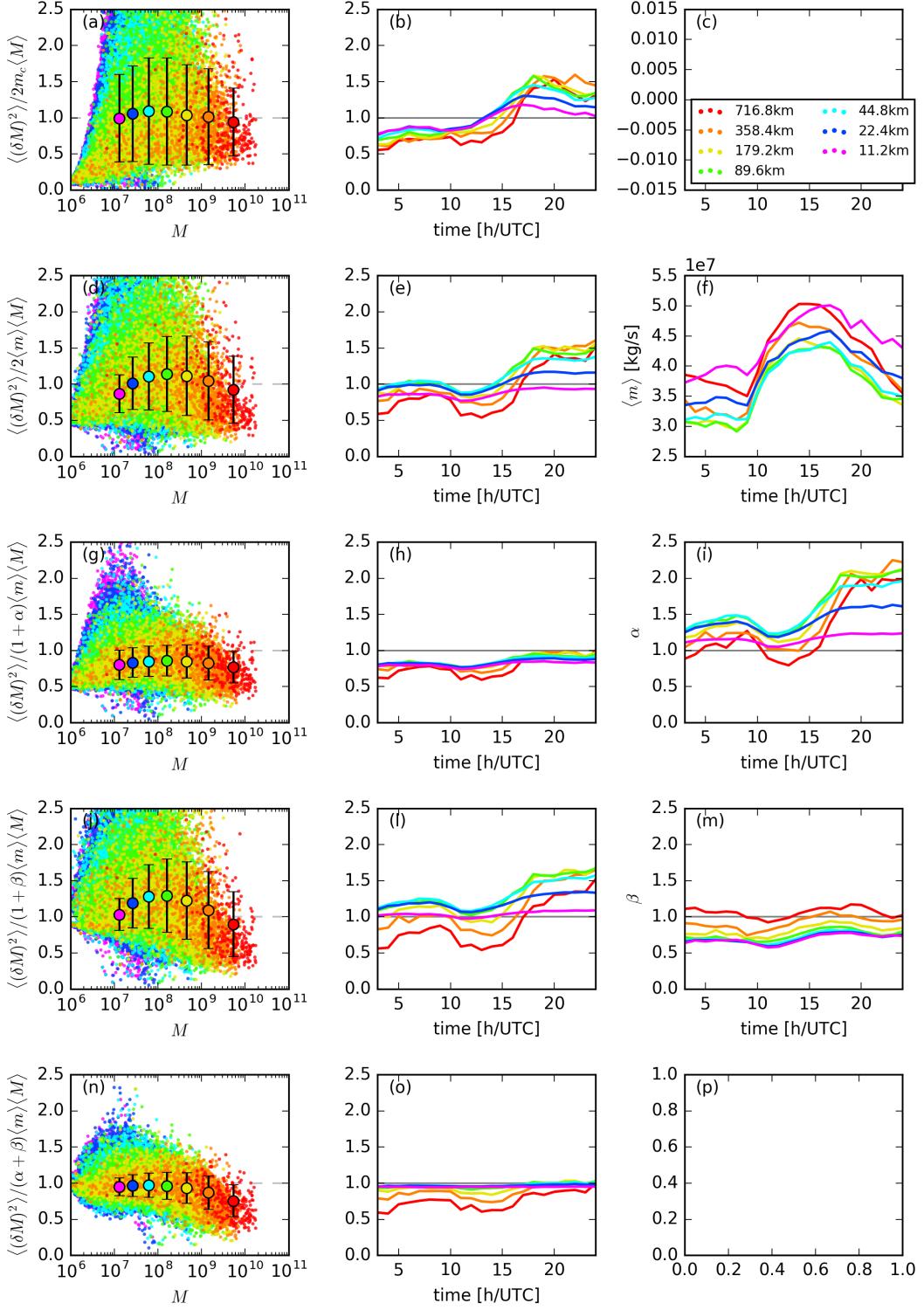


Figure 13: (left column) Ratio of simulated variance to prediction (see label) Each dot represent one time and one coarse box. Large dots represent mean values. Error bars indicate plus/minus one standard deviation. (middle column) Temporal evolution of mean ratio (corresponds to large dots in left column for one time) (right column) Temporal mean evolution of parameter used to refine prediction.

Appendix A

Cloud size and mass flux per cloud distributions

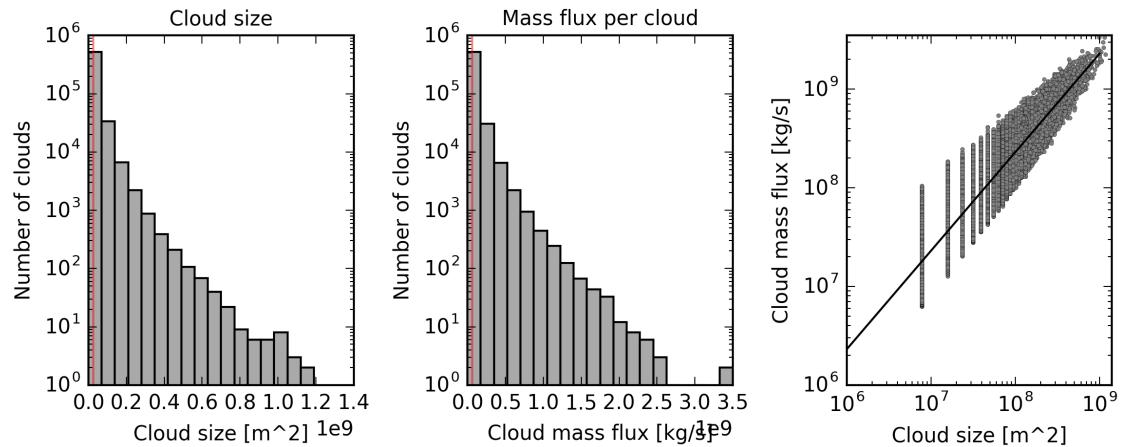


Figure 14: Same as Fig. 8 but without cloud separation. Notice the different scales of the axes.

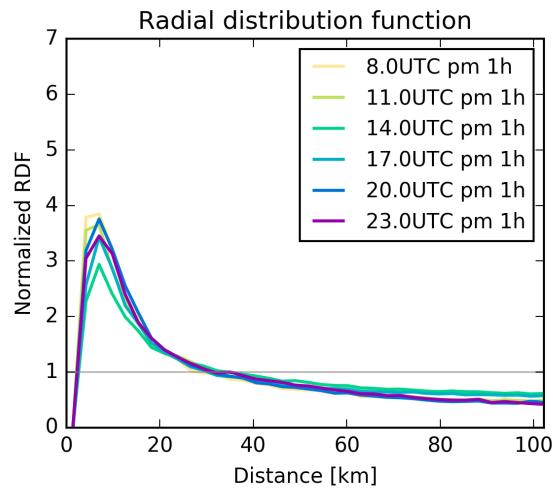


Figure 15: Same as Fig. 9 but without cloud separation.

Appendix B

composite/std_v_mean/std_v_mean_composite_ana-coarse_wat-True_lev-3000_nens-20.png

Figure 16: As Fig. 12 but for 20 ensemble members

Comparison of simulation results and predictions
of increasing complexity

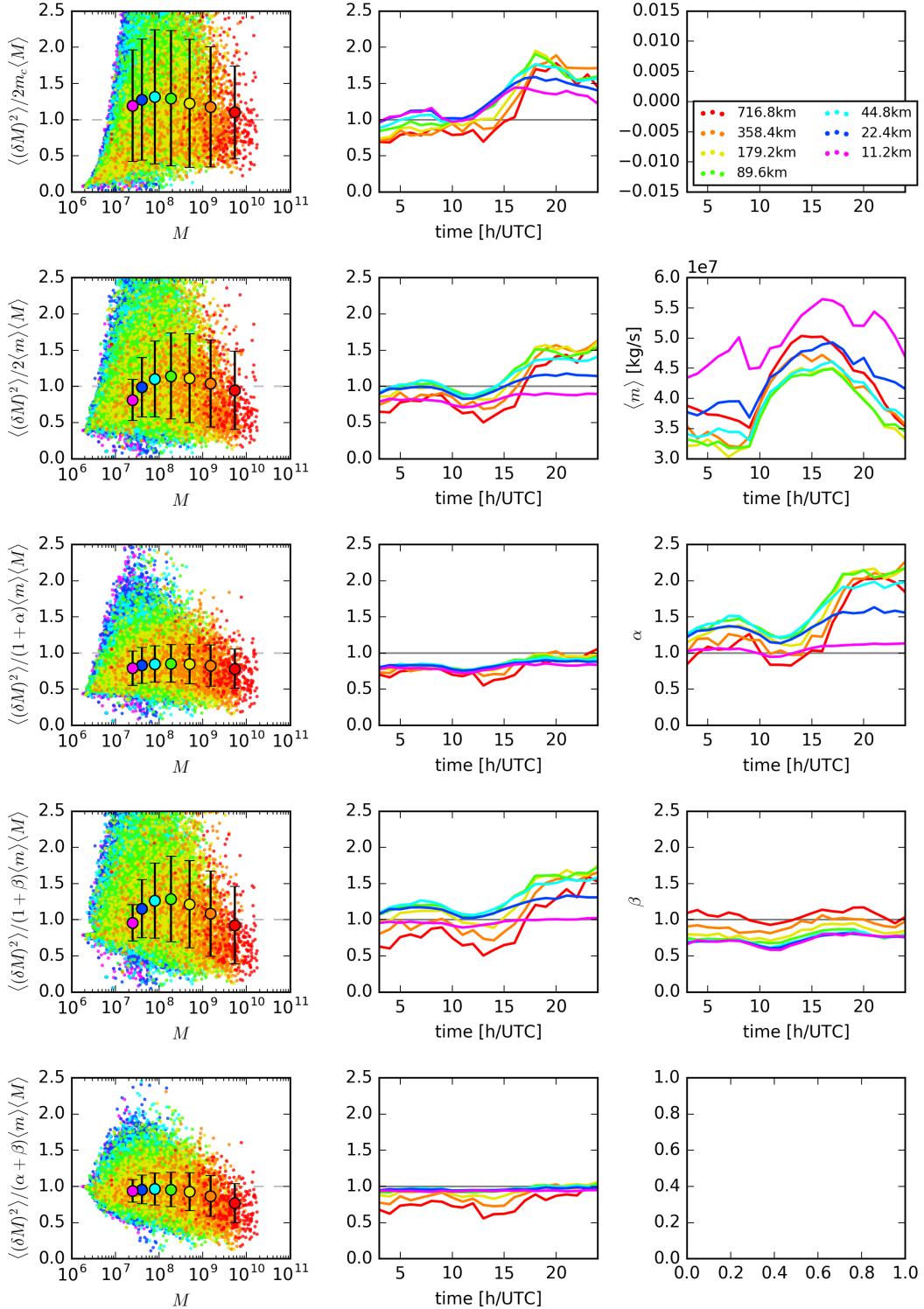


Figure 17: As Fig. 13 but for 20 ensemble members

Appendix C

Comparison of simulation results and predictions of increasing complexity

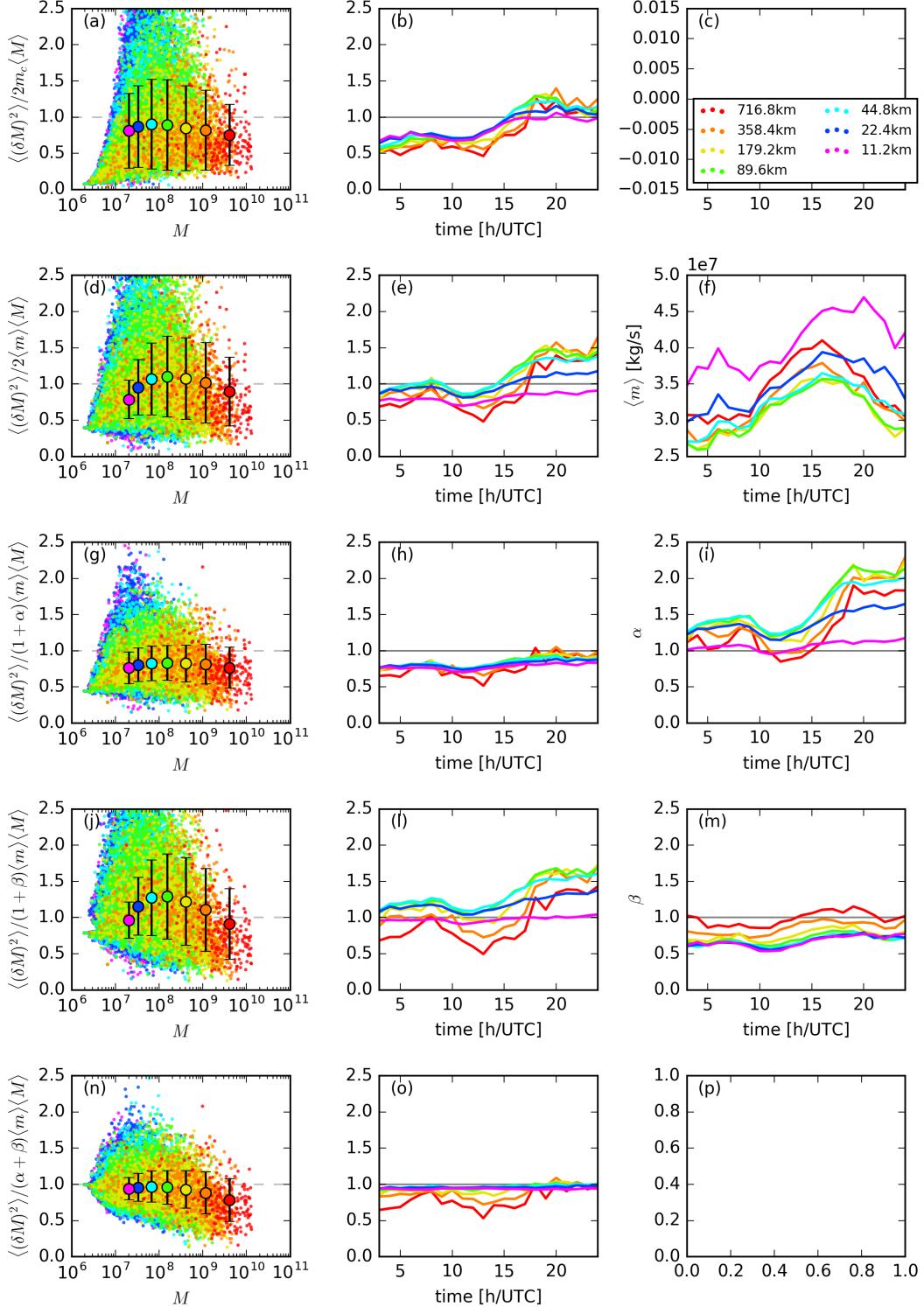


Figure 18: As Fig. 17 but at 2 km analysis height for 20 Members

Comparison of simulation results and predictions
of increasing complexity

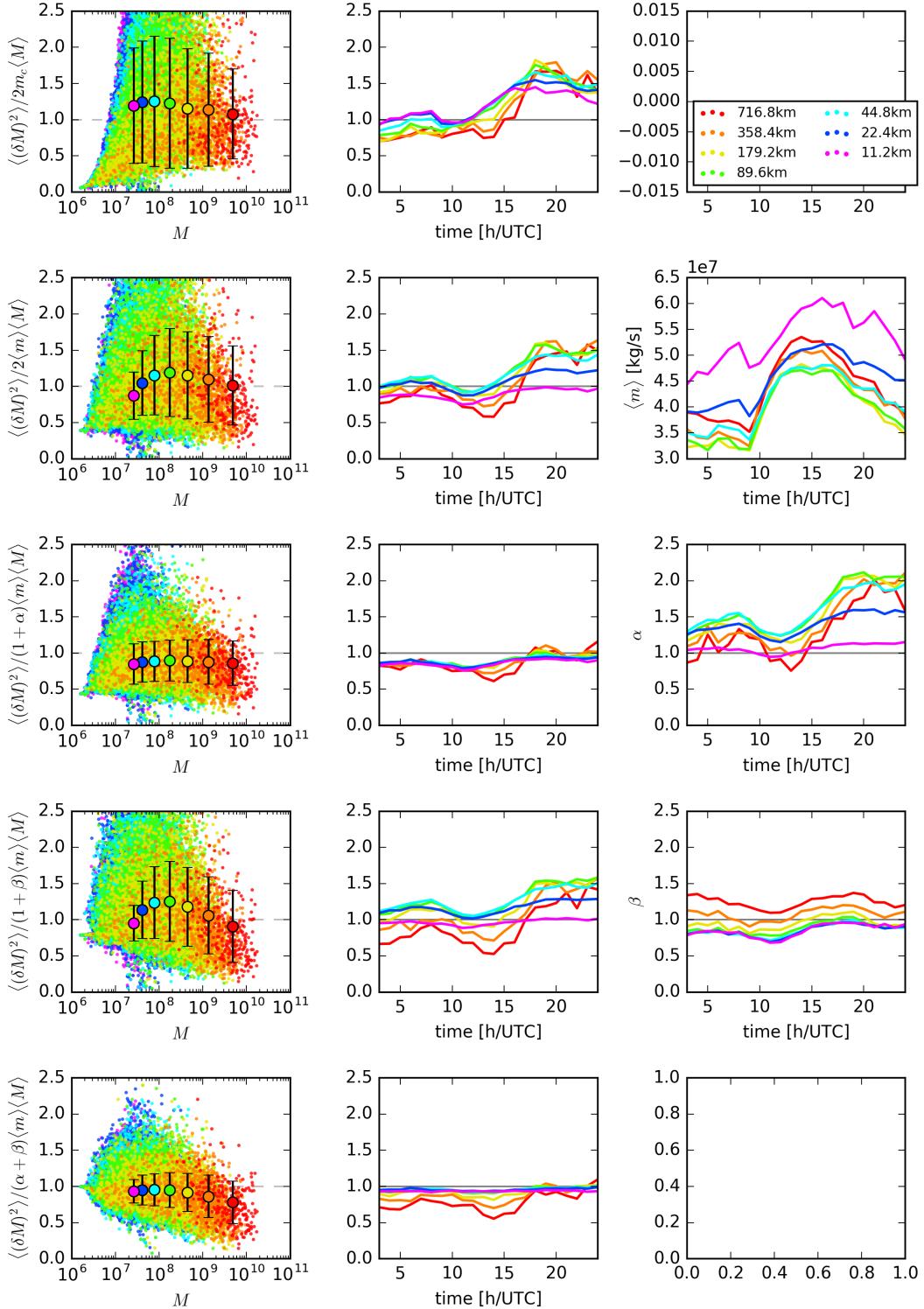


Figure 19: As Fig. 17 but at 4 km analysis height or 20 Members