

# Convective variability in real mid-latitude weather: Which model best explains it?

October 3, 2016

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	The cumulus parameterization problem . . . . .	2
1.2	The Craig and Cohen (2006) theory and its application in Plant and Craig (2008) . . . . .	2
1.2.1	Deviations from theory in other studies . . . . .	3
1.3	SPPT . . . . .	4
1.4	Other approaches . . . . .	4
1.5	Research question / Aim of this study . . . . .	5
<b>2</b>	<b>Numerical experiments and case studies</b>	<b>5</b>
2.1	Model description and set-up . . . . .	5
2.2	The PSPturb turbulence scheme . . . . .	5
2.3	Simulation period . . . . .	6
2.4	Are the simulations realistic? . . . . .	6
2.5	Are the large scales sufficiently similar and the convective scales sufficiently displaced? . . . . .	7
<b>3</b>	<b>Computation of <math>M</math> and <math>Q</math> statistics</b>	<b>7</b>
3.1	Mass flux $M$ . . . . .	7
3.2	Heating rate $Q$ . . . . .	8
3.3	Calculation of ensemble means and variances . . . . .	8
<b>4</b>	<b>Comparing two models for convective variability</b>	<b>9</b>
<b>5</b>	<b>Which assumptions are how wrong?</b>	<b>10</b>
<b>6</b>	<b>Conclusion</b>	<b>11</b>
<b>7</b>	<b>Appendix A: Sensitivity to ensemble size</b>	<b>11</b>
<b>8</b>	<b>Appendix B: Sensitivity to analysis height</b>	<b>12</b>

# 1 Introduction

## 1.1 The cumulus parameterization problem

Effect of unresolved scales on the resolved scales is described by finding a simple model of what the unresolved scales do given some resolved parameters. In mass flux schemes typically a closure assumption is used to determine the mean mass flux, which is then used in a cloud model to determine the heating rate, etc.

If many small scale features are in the LS box, this approximation can be done almost deterministically, but as the grid sizes get smaller, the sampling issue and therefore the fluctuations around the mean state become significant (when is mean eq std?). In these cases deterministic parameterizations can lead to systematic biases and under-representations of extremes. Stochastic parameterizations aim to randomly chose one small scale state compatible with the large scale. The goal is to find a model for the variability around the mean response.

Schematic diagram.

## 1.2 The Craig and Cohen (2006) theory and its application in Plant and Craig (2008)

The Craig and Cohen (2006)(CC06) theory aims to quantify the mass flux fluctuations of a cloud field in convective equilibrium. Convective equilibrium implies that the average properties of the convection are determined by the large-scale forcing. In more detail, the average total mass flux  $\langle M \rangle$  is a function of the large-scales. Other assumptions are: (a) the mean mass flux per cloud  $\langle m \rangle$  does not depend on the large-scale forcing, only the mean number of clouds  $\langle N \rangle$  does; (b) non-interacting clouds: Cloud are spatially separated (no clustering) and do not influence each other. (c) Equal a priori probabilities: This statistical equilibrium assumption implies that “that clouds are equally likely to occur in any location and with any mass flux”. Using these arguments as a basis, a statistical theory is constructed for the distributions of  $N$  and  $m$ :

$$P(N) = \frac{\langle N \rangle^N}{N!} e^{-\langle N \rangle} \quad (1)$$

$$P(m) = \frac{1}{\langle m \rangle} e^{-m/\langle m \rangle} \quad (2)$$

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

$$P(M) = \left( \frac{\langle N \rangle}{\langle m \rangle} \right)^{1/2} e^{-\langle N \rangle} M^{-1/2} e^{-M/\langle m \rangle} I_1 \left[ 2 \left( \frac{\langle N \rangle}{\langle m \rangle} M \right)^{1/2} \right], \quad (3)$$

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. It is also possible to derive an equation for the normalized variance of  $M$ :

$$\mu_2 = \frac{\langle (\delta M)^2 \rangle}{\langle M \rangle^2} = \frac{2}{\langle N \rangle} \quad (4)$$

Always note that  $\langle M \rangle = \langle N \rangle \langle m \rangle$ . Eq. 4 can be derived directly from Eq. 3 or from the theory of random sums (Taylor and Karlin, 1998, p.70ff): Assume  $X = \xi_1 + \dots + \xi_N$  where  $\xi_k$  and  $N$  have the finite moments  $E[\xi_k] = \mu$ ,  $Var[\xi_k] = \sigma^2$  and  $E[N] = \nu$ ,  $Var[N] = \tau^2$ . Then the first and second moment of  $X$  are  $E[X] = \mu\nu$ ,  $Var[X] = \nu\sigma^2 + \mu^2\tau^2$ .

The theoretical predictions above were tested against numerical simulations in radiative-convective equilibrium (RCE) by Cohen and Craig (2006). The results of these simulations agreed well with the theory. The error in  $\mu_2$  is around 10%, with  $\mu_2 \langle N \rangle \approx 1.6$ . Other studies introduced time-varying forcings and looked at the differences in mass flux statistics as described below.

In the Plant and Craig (2008)(PC08) stochastic parameterization approach, the exponential  $m$  distribution (Eq. 2) is used to create a random population of plumes for each grid-box consistent with a large scale  $\langle M \rangle$ . From this distribution the large-scale tendencies are then computed as the sum of the cloud model output for each plume.  $\langle m \rangle = 2 \times 10^7 \text{ kg s}^{-1}$  is assumed to be a constant. This assumption is motivated by RCE simulations (e.g. Cohen and Craig, 2006). The theoretical prediction for the variance of  $M$  (Eq. 4) 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.

The PC08 scheme has been tested in a GCM study with some success, improving the precipitation patterns and equatorial waves (Wang et al., 2016).

### 1.2.1 Deviations from theory in other studies

Two studies looked at the deviations from the CC06 predictions in their simulations of convection with a time varying forcing: Davies (2008) and Davoudi et al. (2010). A quick definition of clustering for this text: Clustering describes the increased probability of occurrence of clouds near already existing clouds. So basically a spike in an RDF.

**Davies (2008)** She used a model with 1km resolution, a prescribed radiative cooling and time-varying surface fluxes or temperature. The domain size was 64 km by 64 km. For the reference RCE simulation she found  $\mu_2 \langle N \rangle \approx 1.5$  at 3 km, a deviation of 10% in  $\mu_2$ , which is in agreement with CC06. When looking at their time-varying simulations, they see that  $\mu_2$  is increased (about 2.2) 1h after convection is first triggered and at around 15UTC. They find that at the triggering time and at 18UTC there is strong clustering at scales from 5–20 km. At the time of maximum convection (12UTC), the RDF is almost uniform and  $\mu_2 \approx 0.7$ . She argues that the deviation from the predicted variance can be largely explained by clustering (see Fig. 1).

**Davoudi et al. (2010)** They used a similar model setup to CC06, but with a diurnal cycle through interactive radiation with fixed SST. In their Fig. 13, they show their values of  $\mu_2 \langle N \rangle$  for different heights. They find that for  $z < 8$  km, this value is less than two. Additionally, in their Fig. 12. they plot histograms of  $P(M)$  from their data. They then fit Eq. 3 with  $\langle M \rangle$  and  $\langle N \rangle$  as free parameters. When they compare these fitted values to the calculated values of  $\langle M \rangle$  and  $\langle N \rangle$  from their data, they find that  $\langle M \rangle$  is similar but the fitted  $\langle N \rangle$  is larger than the observed  $\langle N \rangle$ . They state that “Therefore, predictions of  $\mu_2$  are smaller than the corresponding normalized variance from the data. Figure 13

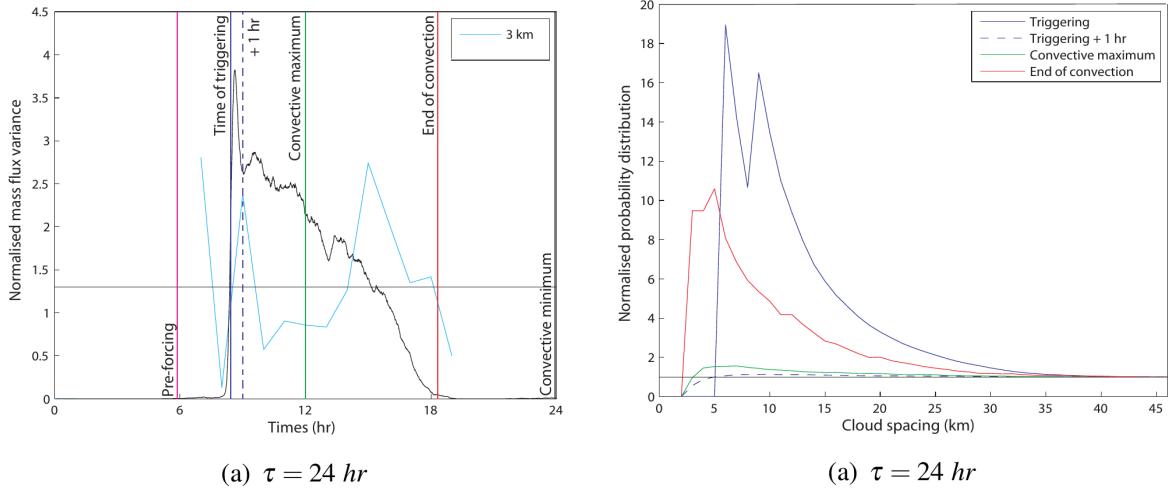


Figure 1: From Davies (2008): (left)  $\mu_2$  from their 24 h time-varying forcing simulation at a height of 3 km (blue). The black line shows the domain total mass flux at that level. (right) The corresponding RDF for the times indicated (left).

demonstrates that the variance, as well as skewness, is underestimated by the theory close to the cloud base and for the range of altitudes in  $z \in [2, 8] \text{ km}$ .” *This statement seems wrong. Shouldn’t it be the other way around? I sent them an email.*

They then look at two clustering metrics. First, the radial distribution function (their Fig. 14), where they find strong clustering for 5-10 km. Second,  $\alpha = \frac{\sigma_N^2}{\langle N \rangle}$  (their Fig. 14), where they find values of about 110% at cloud base, which is in agreement with the findings by CC06 and Davies (2008).

### 1.3 SPPT

SPPT acts on the output of all parameterizations in the model on the assumption that the standard deviation is proportional to the mean tendencies, in case of convection this would be the heating rate (and moisture?) (cite Christensen).

### 1.4 Other approaches

Several approaches to get a physical model of the underlying uncertainty of convection have been tried. Dorrestijn et al. (2015) and Gottwald et al. (2016) used conditional Markov chains to describe the transition of cloud states. Bengtsson et al. (2013) used cellular automata.

Dorrestijn et al. (2015) and Gottwald et al. (2016) used conditional Markov chains to describe the transition from one cloud-state to another for several micro-nodes in a GCM grid box. The transition probabilities are dependent on some large scale indicator, in their case the large scale vertical velocity, and the exact values are calculated from observations. This approach allows them to calculate a fraction of deep convective clouds for each grid-box, which can then be used to estimate mass flux for use in a convective parameterization. The advantage of using conditional Markov chains is that they inherently have memory.

Application in a simple GCM shows improvements in the distribution of precipitation, and some improvements for equatorial waves (Dorrestijn et al., 2016).

## 1.5 Research question / Aim of this study

The primary goal of this paper is to use a novel technique to characterize the relationship between convective variability and the mean state in real mid-latitude weather situations. How this was achieved is outlined in Section ??? . The results from these numerical experiments will then be compared to the simple models of CC06 and SPPT, and we will try to find explanations for the deviations. Based on these findings we will try to use a more appropriate model and lastly investigate how this could be used to improve parameterizations.

## 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???. The 50 grid point gap to the boundary ensures that boundary 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 was available every 30 minutes.

### 2.2 The PSPturb turbulence scheme

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

$$\left( \frac{\partial \Phi}{\partial t} \right)_{\text{total}} = \left( \frac{\partial \Phi}{\partial t} \right)_{\text{parameterized}} + \eta \sigma_{\left( \frac{\partial \Phi}{\partial t} \right)_{\text{parameterized}}} \quad (5)$$

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  $\sigma$  of the parameterized tendencies. The random number field has a horizontal correlation length of  $5\Delta x$ , the effective resolution 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

$$\sigma_{\left(\frac{\partial \Phi}{\partial t}\right)_{\text{parameterized}}} = \alpha_{\text{const}, \Phi} \frac{l_\infty}{5\Delta x} \frac{1}{dt} \sigma_\Phi, \quad (6)$$

where  $l_\infty = 150$  m is the mixing length describing the average size of an eddy. The term  $\sigma_\Phi$  is the sub-grid scale standard deviation. For the turbulence perturbations the considered variables are vertical velocity  $w$ , potential temperature  $\theta$  and humidity  $q$ . The standard deviations are calculated in the turbulence parameterization (see KC06 for details). The factor  $\frac{l_\infty}{5\Delta x} \propto \frac{1}{\sqrt{N_{\text{eddy}}}}$  scales the variability according to number of unresolved eddies similar to Eq. 4. 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. ??), 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  $\tau_c$ , 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  $\tau_c$  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???. To get a quantitative picture a precipitation histogram is used to check whether the simulations produce a realistic distribution of rain (Fig. 4). 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. 5). 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?

This is the premise upon which this study is built. A visual, qualitative test of this assumption is shown in Fig. 3. 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. 6). 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. 8). 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 then 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.

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

$$\sigma_k = N_{px} \Delta x^2, \quad (7)$$

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

$$m_k = \Delta x^2 \sum_l^{N_{\text{px}}} w_l \rho_l, \quad (8)$$

where  $\rho$  is density.

A vertical level has to be chosen for this analysis. Ideally, one would use the cloud base at the top of the boundary layer, since many convective parameterizations use the cloud base mass flux  $M_b$  as the input for their cloud models. In our simulations, however, the cloud base varies spatially and temporally. A dynamic detection of the cloud base level would be complex and prone to errors. FIG??? shows the vertical mass flux profiles at two times, during the convective maximum and during the evening hours.

There are, however, some complications. First, shallow convective clouds also occur at this level. These are typically ignored in parameterizations of deep convection, which is the focus of this study. Therefore, it makes sense to do the analysis above the shallow convective layer. Second, in our simulations of real weather, the height of the boundary layer, and therefore also the cloud base, varies temporally and spatially in the domain. Thus, our reasoning is to make sure that our chosen height is always above the boundary layer. Third, our studies have orography meaning that the height above sea level is variable. Due to this we chose to do our analysis on the terrain following model levels of the COSMO model (?), which in the lower troposphere closely follows the distance to the ground, but is basically parallel to the sea level height at the tropopause and above. The vertical profile of the domain average convective updraft mass flux is shown in Fig. ?? for the entire analysis domain, but also separately for Northern and Southern Germany at the time where the boundary layer is highest. Based on the reasoning laid out in this paragraph we chose model level 30 for our mass flux analyses, but in the Appendix also show the dependence of certain diagnostics with height.

### 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 neighborhoods smaller are considered, since these would be significantly below the effective resolution of the model (Skamarock and Skamarock, 2004). The total mass flux per box per member  $M_{i,j,n}$  is given by

$$M_{i,j,n} = \sum_{k=1}^{N_{\text{cld } i,j,n}} m_{k,i,j,n}. \quad (9)$$

To deal with clouds at the boundaries of the coarse-fields, the centers of mass for each cloud is 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.  $N_{i,j,n} = N_{\text{cld},i,j,n}$  is simply the number of clouds which fall into each box. This follows Cohen and Craig (2006). The heating rate per box  $Q_{i,j,n}$  is simply 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

$$\langle(\delta\Phi)^2\rangle_{j,n} = \frac{1}{N_{\text{ens}} - 1} \sum_{i=1}^{N_{\text{ens}}} (\Phi_{i,j,n} - \langle\Phi\rangle_{j,n})^2, \quad (10)$$

where the ensemble mean is

$$\langle\Phi\rangle_{j,n} = \frac{1}{N_{\text{ens}}} \sum_{i=1}^{N_{\text{ens}}} \Phi_{i,j,n}. \quad (11)$$

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  $N_{\text{cldtot}} = \sum_{i=1}^{N_{\text{ens}}} N_{\text{cld},i,j,n}$ .

$$\langle(\delta m)^2\rangle_{j,n} = \frac{1}{N_{\text{cldtot}} - 1} \sum_{k=1}^{N_{\text{cldtot}}} (m_{k,j,n} - \langle m \rangle_{j,n})^2, \quad (12)$$

where the mean is

$$\langle m \rangle_{j,n} = \frac{1}{N_{\text{cldtot}}} \sum_{k=1}^{N_{\text{cldtot}}} m_{k,j,n}. \quad (13)$$

Since we are sampling a distribution with a limited number of data points  $N_{\text{ens}}$ , sampling issues arise when  $\langle N \rangle$  becomes small ( $\approx \frac{1}{N_{\text{ens}}}$ ). In particular, if only one member contains a cloud chances are that the real  $\langle N \rangle < \frac{1}{N_{\text{ens}}}$  and we therefore overestimate the mean mass flux  $\langle M \rangle$ . To avoid this issue, a criterion is introduced where at least 5 out of 20 ensemble members must contain at least one cloud. *This threshold is a quick fix and there probably should be a threshold with better reasoning.*

Composites were computed by averaging over the 12 days in the simulation period. For the calculation of means (every overbar) and standard deviations (std), 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 Comparing two models for convective variability

We now have a dataset containing 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 would now like to test how the distribution of the mass flux can be described. Ideally one would like to find a simple model  $f$  which returns the PDF of convective states  $p(M)$  or  $p(Q)$  given a certain input:

$$p(M) = f(\langle M \rangle, ?) \quad (14)$$

The challenge is to find the model and the parameters which are important for the model. Since testing a complete PDF is rather difficult, we will focus on the variance as the

main measure of quality for our simple model predictions. The first moment, the mean, is assumed to be perfect by design, which corresponds to the assumption that the closure of the convective parameterization scheme is perfect. Two theories will be tested. The first one is the CC06 theory as used in the PC08 scheme which results in the following prediction for the standard deviation of the mass flux:

$$\langle(\delta M)^2\rangle^{1/2} = c\langle M\rangle^{1/2} \quad (15)$$

The constant factor  $c$  is given by twice a constant mass flux per cloud  $m_c$ . The second one is the SPPT assumption which states:

$$\langle(\delta Q)^2\rangle^{1/2} = c\langle Q\rangle \quad (16)$$

Here  $c$  depends on the standard deviation of the random number used to multiply the parameterized tendencies (Shutts and Palmer, 2007). Fig. 9a shows the relation of  $\langle(\delta M)^2\rangle^{1/2}$  and  $\langle M\rangle$  for all simulation data. Additionally, least square fits for the functions 15 and 16 are plotted for each scale separately. The fit parameter  $c$  along with the normalized root mean square error of the fit is shown in panel (b). The two main results seem to be: i) For all scales the CC06 fit is just as good or better than the SPPT fit as measured by the NRMSE. ii) The CC06 relation is scale adaptive. This means that for all scales the fit parameter is relatively constant with fluctuations of around 30%. The SPPT fit parameter, on the other hand, decreases monotonically with scale, indicating that tuning is necessary if the resolution is changed, while the CC06 theory is inherently scale-aware.

SPPT was not designed for  $M$ , but for the parameterization output, however. We will use  $Q$  as a proxy for the parameterized tendencies. To make  $Q$ , which is a mean quantity, and  $M$ , which is a summed quantity comparable for all scales, the heating rate is multiplied by the area of the coarse box  $A = n^2$ . Fig. 9c shows the relation between  $\langle(\delta Q)^2\rangle^{1/2}$  and  $\langle Q\rangle$  along with the SPPT fits.  $Q$  can be negative and there can be a non-zero variability for a zero mean  $Q$ . This is not possible in either CC06 or SPPT. To see how well the results for  $M$  apply to  $Q$ , the correlation coefficient for the means and standard deviations of the two variables is shown in panel (d). With values of at least 0.75 for the means and at least 0.9 for the standard deviation the correlation is good. To test how well the CC06 theory applies to the heating rate the input  $M$  is related to the output of the parameterization  $Q$  (Fig. 9e with CC06 fits). This assumes a perfect cloud model. In panel (f), analogous to (b), the fit parameters and NRMSE is shown. The general conclusions for  $Q$  mirror the ones for  $M$  with larger errors. These results show the applicability of the CC06 theory in real mid-latitude weather and the advantages over the SPPT theory. There are however, still large deviations of the prediction from the simulation results.

## 5 Which assumptions are how wrong?

The goal of this section is to go through the assumption made by CC06 and PC08 and quantitatively assess how big the error from a violation of these assumptions is. Assumption one is that the mean cloud mass flux is constant  $\langle m\rangle = m_c$ . This is an assumption made by PC08 when construction their parameterization. To make such a prediction we

take  $m_c = 4 \times 10^7 \text{ kgs}^{-1}$  which is close to the mean value for the entire simulation. The ratio of predicted variance to simulated variance is shown in Fig. 10a and b for each scale. One can clearly see that the mean predictions are relatively close to the model output but the fluctuations are large. A large diurnal cycle is also apparent. Panel (f) now shows the temporal evolution of  $\langle m \rangle$ . The differences in scale come from the filtering criteria and the non-weighted averaging. There also is a clear diurnal cycle with larger clouds during the convective maximum. The “corrected” prediction (which now is the pure CC06 prediction) is shown in panels (d) and (e). The fluctuations around the mean value have clearly decreased, particularly for small  $n$ .

The next step is to look at the assumption of spatially randomly distributed clouds, which results in a Poisson distribution of  $N$ . We define a parameter  $\alpha = \frac{\langle (\delta N)^2 \rangle}{\langle N \rangle}$  to see whether the clouds are more clustered or more regularly distributed. In panel (i) one can see that there is a very strong diurnal variation of this clustering parameter with increased clustering in the evening hours. The diurnal variations are strongest for medium and large  $n$ . Accounting for  $\alpha$  in the prediction (panel h) removes almost all of the diurnal variations and greatly reduces the fluctuations (panel g). The adjusted theory now generally overpredicts the variance by 20–30%.

Another assumption is that the cloud mass flux is exponentially distributed. Here we look at the parameter  $\beta = \frac{\langle (\delta m)^2 \rangle^2}{\langle m \rangle}$  to see whether the distribution is broader or narrower than exponential. Panel (l) shows this parameter. It is striking that there is a strong dependence on scale. For large  $n$ ,  $\beta = 1$  with little variation, while for smaller  $n$  the distribution seems to be narrower. The reason for this is not clear yet. It does not seem to be a sampling issue (see Appendix B). Could it be an effect of finite sized clouds? Accounting for  $\beta$  in the predictions seems to mostly affect the mean, particularly for small and medium  $n$ . The fluctuations of the prediction to simulation ratio does not seem to be altered greatly.

Accounting for both  $\alpha$  and  $\beta$  (bottom row) yields greatly improved prediction compared to the pure CC06 prediction. For small and medium scales the predictions are almost perfect with little fluctuations. For large  $n$  a general overprediction remains. The remaining assumption is that the cloud number and mass flux are uncorrelated. From the mean plots (FIG???) this is obviously not true. For very large scales then (300 km and up) this covariance seems to play an important role, while for scales smaller 100 km the effect is negligible.

Overall the greatest deviations seem to come from clustering as measured by  $\alpha$ . In particular there is a strong diurnal signal which can impact the variance predictions by up to a factor of two.

## 6 Conclusion

## 7 Appendix A: 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. In particular the scale dependence of the  $\beta$  parameter we suspected to be a sampling issue. Figs ??? and ??? show the diagnostics evaluated for 50 and 20 ensemble

members respectively. The main change is in the x-axis, the total mass flux  $M$ . This is an artifact of the filtering mentioned in Section ??? which puts a lower threshold on the number of ensemble members which have clouds. This threshold is 5 for both the 20 and 50 member analysis, thereby effectively creating a different minimum  $M$ . Apart from this analysis artifact, the most values are largely unchanged, indicating that the sampling is robust. The fluctuations about the mean are slightly smaller in the 50 member analysis. In particular the  $\beta$  parameter does not seem to depend on the ensemble size. The scale dependence is therefore not simply an issue of sample size.

## 8 Appendix B: Sensitivity to analysis height

As mentioned in Section ??? 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 ??? to ??? show the diagnostics at 2, 3 and 4 km height. For most quantities shown the change with height (in this reasonable range) is much smaller than the diurnal variation (see for example  $\alpha$ ). The only exception is the  $\beta$  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 the results are robust.

## 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>.
- Bengtsson, L., M. Steinheimer, P. Bechtold, and J.-F. Geleyn, 2013: A stochastic parametrization for deep convection using cellular automata. *Quarterly Journal of the Royal Meteorological Society*, **139** (675), 1533–1543, doi:10.1002/qj.2108, URL: <http://doi.wiley.com/10.1002/qj.2108>.
- 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/>.
- Dorrestijn, J., et al., 2015: Stochastic Parameterization of Convective Area Fractions with a Multicloud Model Inferred from Observational Data. *Journal of the Atmospheric Sciences*, **72** (2), 854–869, doi:10.1175/JAS-D-14-0110.1, URL: <http://journals.ametsoc.org/doi/10.1175/JAS-D-14-0110.1>.
- Dorrestijn, J., et al., 2016: Stochastic Convection Parameterization with Markov Chains in an Intermediate-Complexity GCM. *Journal of the Atmospheric Sciences*, **73** (3), 1367–1382, doi:10.1175/JAS-D-15-0244.1, URL: <http://journals.ametsoc.org/doi/10.1175/JAS-D-15-0244.1>.
- 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>.
- Gottwald, G. A., K. Peters, and L. Davies, 2016: A data-driven method for the stochastic parametrisation of subgrid-scale tropical convective area fraction. *Quarterly Journal of the Royal Meteorological Society*, **142** (694), 349–359, doi:10.1002/qj.2655, URL: <http://doi.wiley.com/10.1002/qj.2655>.

- Kober, K., G. C. Craig, K. Kober, 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. and W. C. Skamarock, 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>
- Wang, Y., G. J. Zhang, and G. C. Craig, 2016: Stochastic convective parameterization improving the simulation of tropical precipitation variability in the NCAR CAM5. *Geophysical Research Letters*, **43** (12), 6612–6619, doi:10.1002/2016GL069818, URL: <http://doi.wiley.com/10.1002/2016GL069818>.

## 9 Figures

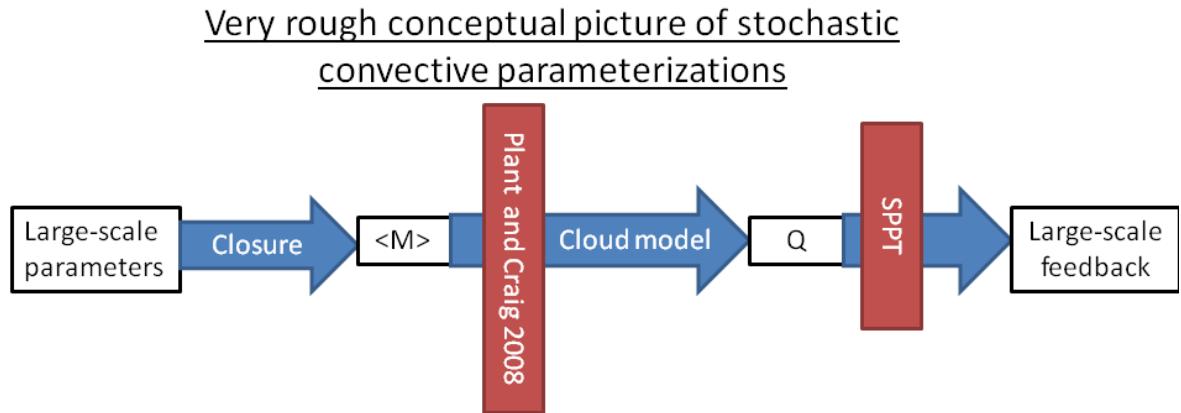


Figure 2: Conceptual picture of stochastic convection parameterizations. Red areas indicate where the stochastic parts are active in PC08 and SPPT.

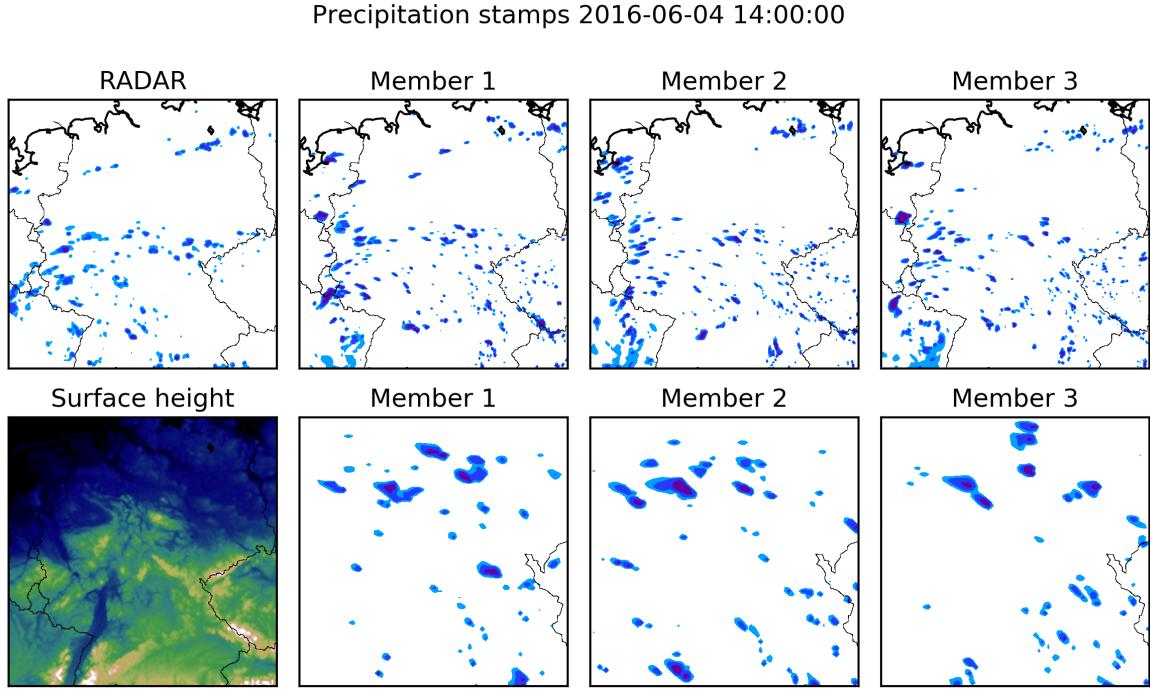


Figure 3: (top right) Radar observations, (top row) Three ensemble members, (bottom row) zoom of plot on top and (bottom left) surface elevation

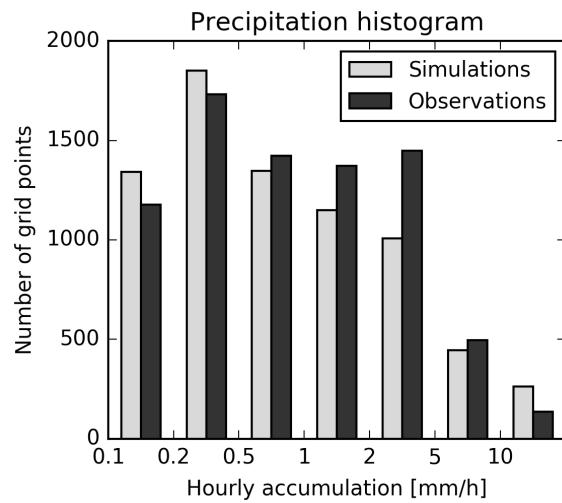


Figure 4: Precipitation histogram showing the number of grid cells within a specified hourly precipitation range. Values are averaged for all analysis times and days. Additionally the simulation numbers are ensemble averages. The “no rain” bin (0–0.1 mm/h) is not shown and accounts for the differences in the total number of shown values.

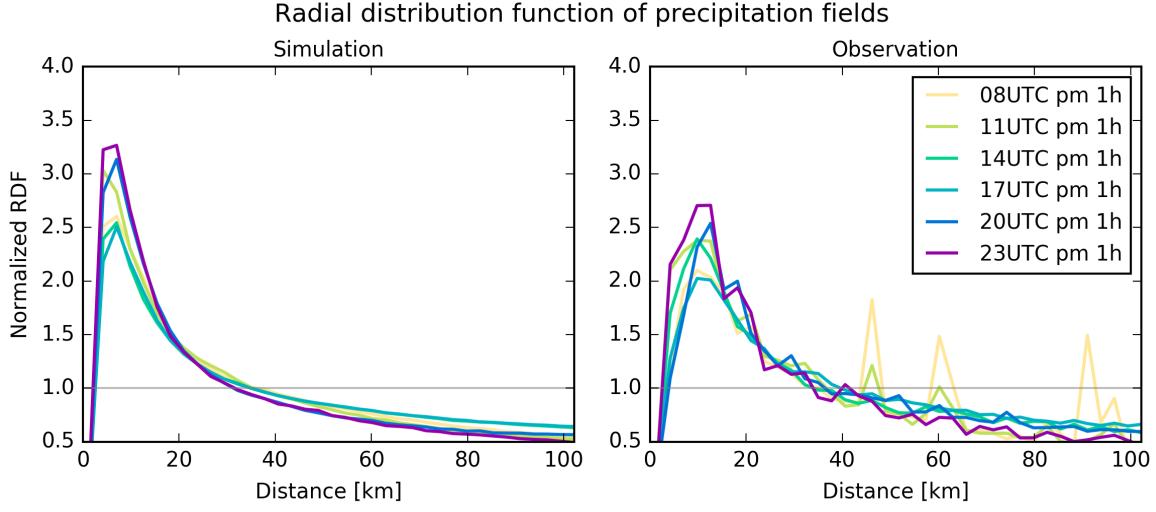


Figure 5: Normalized radial distribution function (RDF) of the hourly precipitation fields for simulations and observations. Every line shown represent an average over three time steps. The maximum search radius is  $36\Delta x$  (approximately 100 km) a resolution of one  $\Delta x$ . To identify cloud cloud objects a rain threshold of  $1 \text{ mm h}^{-1}$  was used, followed by a separation using the same method described for mass flux objects described in Section ??.

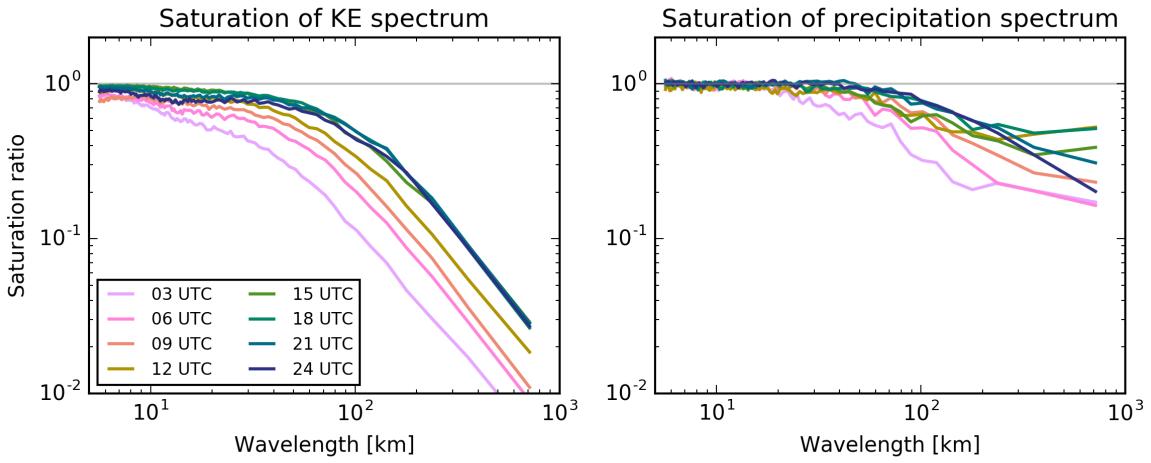


Figure 6: Saturation ratio is defined as the difference energy spectrum divided by twice the background energy spectrum for (left) precipitation and (right) kinetic energy (KE). The KE spectrum was averaged from the lowest model level to level 15 (approximately 10 km height). All spectra are instantaneous 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

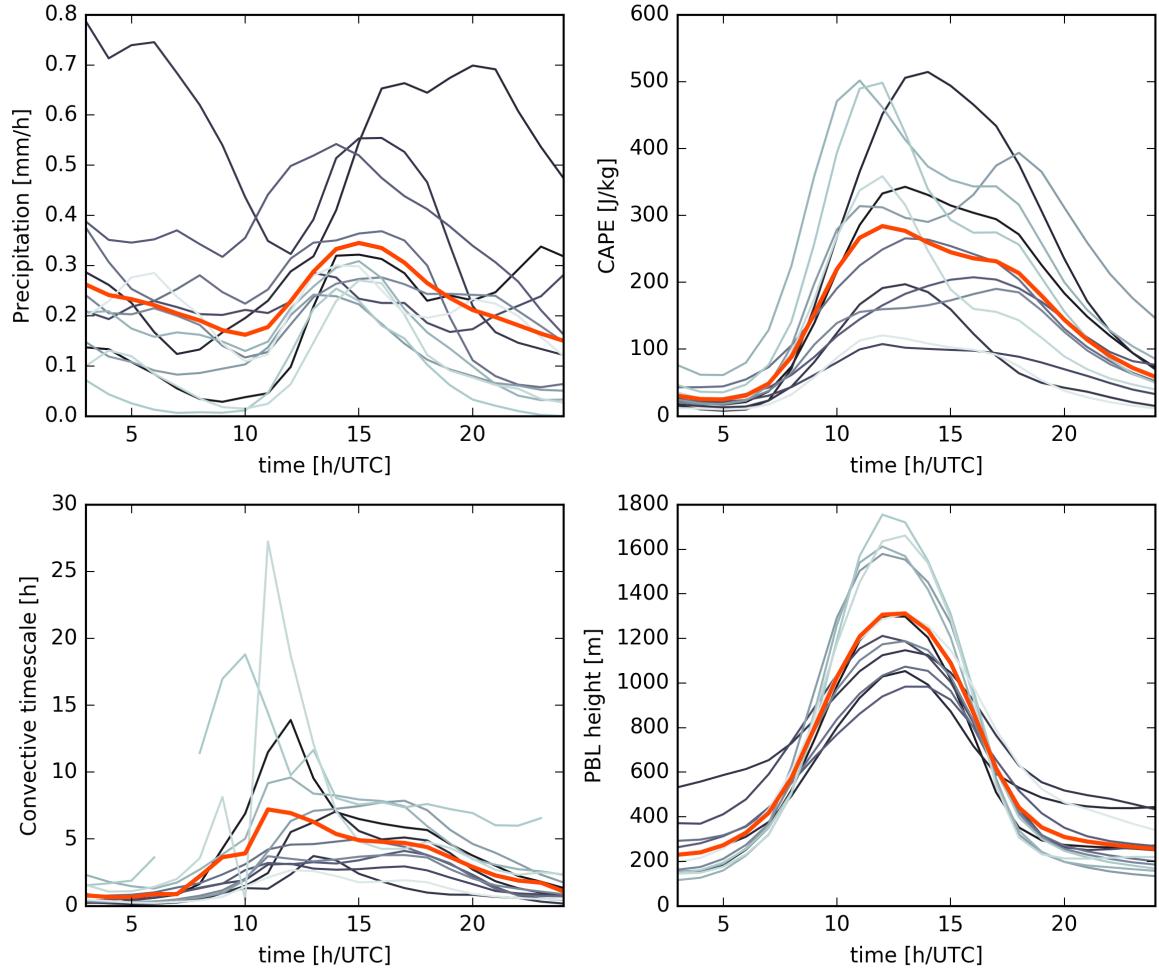


Figure 7: Time series of (a) hourly rainfall in mm/h, (b) CAPE in J/kg, (c) the convective adjustment time scale in h and (d) the perturbed boundary layer height in m. All values are domain and ensemble averages. Each gray line represents one simulation day. The red line is the mean over all simulation days.

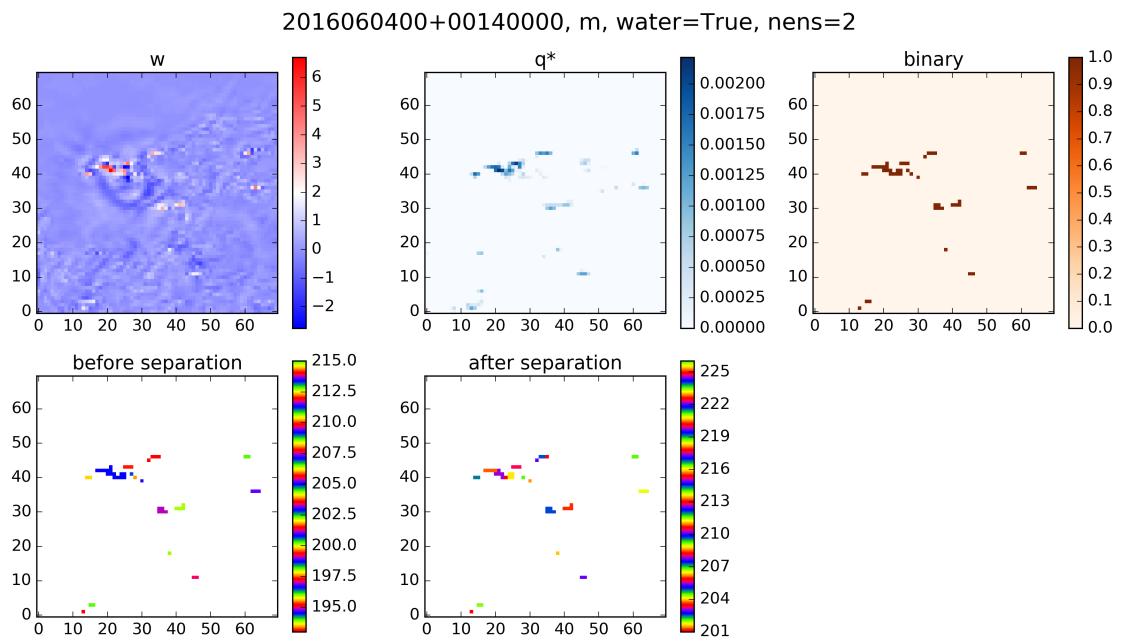


Figure 8: Cloud identification algorithm. First, a binary field (c) is created by applying thresholds to the vertical velocity (a) and cloud water (b) fields. Then, the contiguous clouds are identified (d), followed by a separation using a local maximum and watershed method (e)

composite  
coarse, water=True, lev= 3000, nens=50  
CC06 variance scaling fits data reasonably well

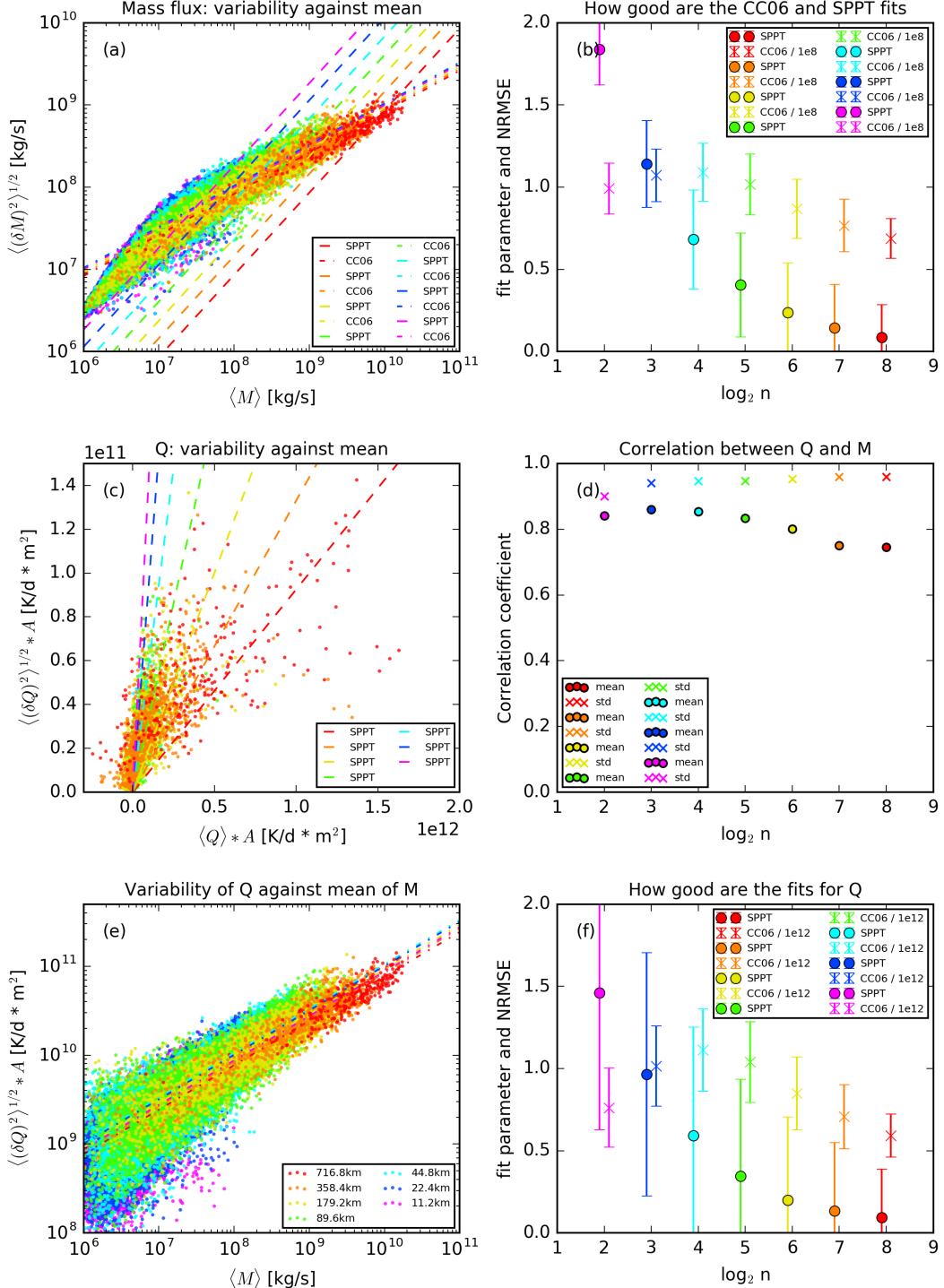


Figure 9: (left column) Scatter plots showing the relation of standard deviations to mean values. Each point represents one time and one box. Lines represent least square fits for  $y = bx$  (SPPT, dashed) and  $y = (by)^{1/2}$  (CC06, dash-dot). (right column top and bottom) Fit parameter  $b$  is shown 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

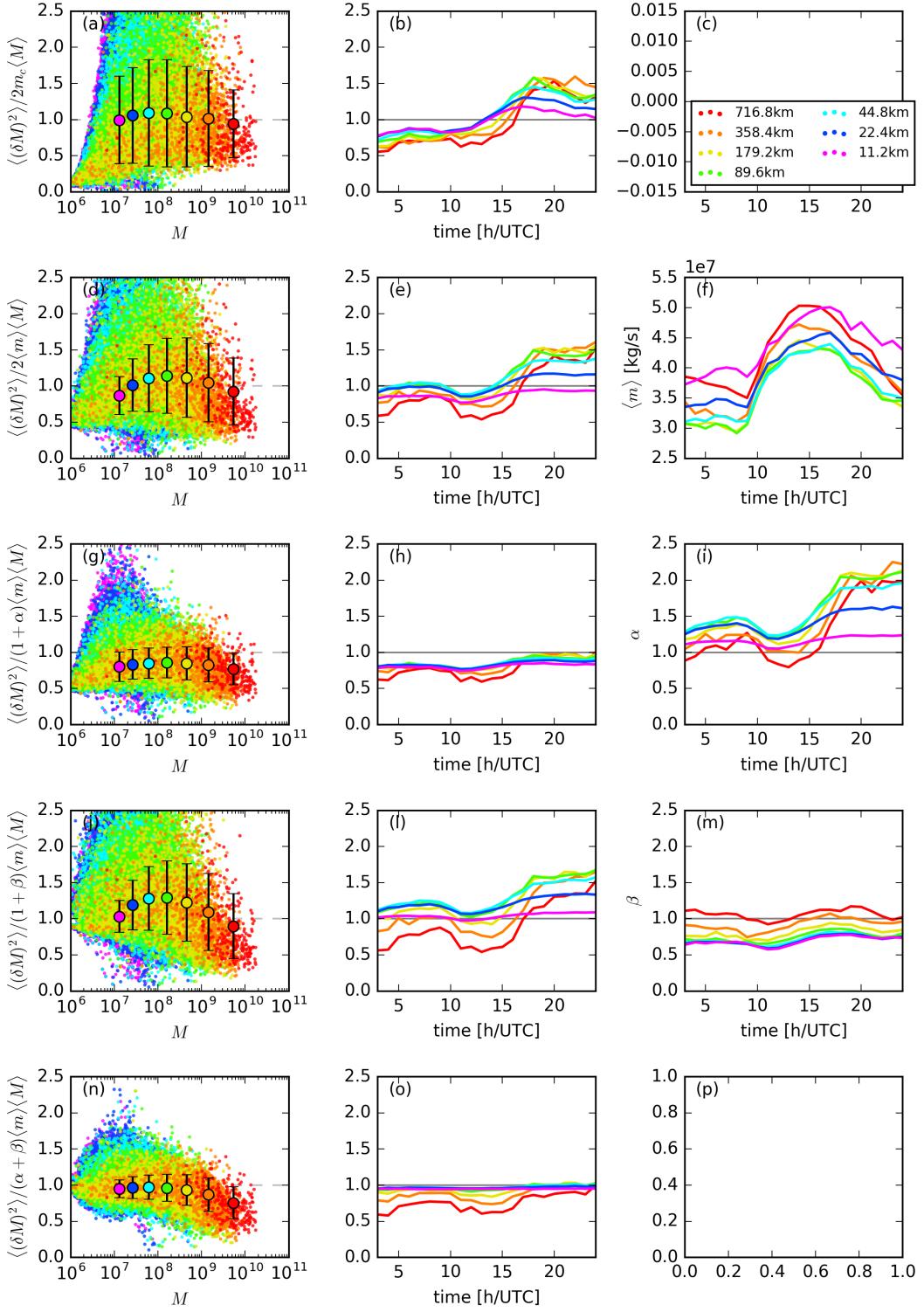


Figure 10: (left column) Ratio of simulated variance to prediction (see label) Each dot represent one time and one coarse box. Large dots represent mean values. Error bars indicate plus/minus one standard deviation. (middle column) Temporal evolution of mean ratio (corresponds to large dots in left column for one time) (right column) Temporal mean evolution of parameter used to refine prediction.

Comparison of simulation results and predictions  
of increasing complexity

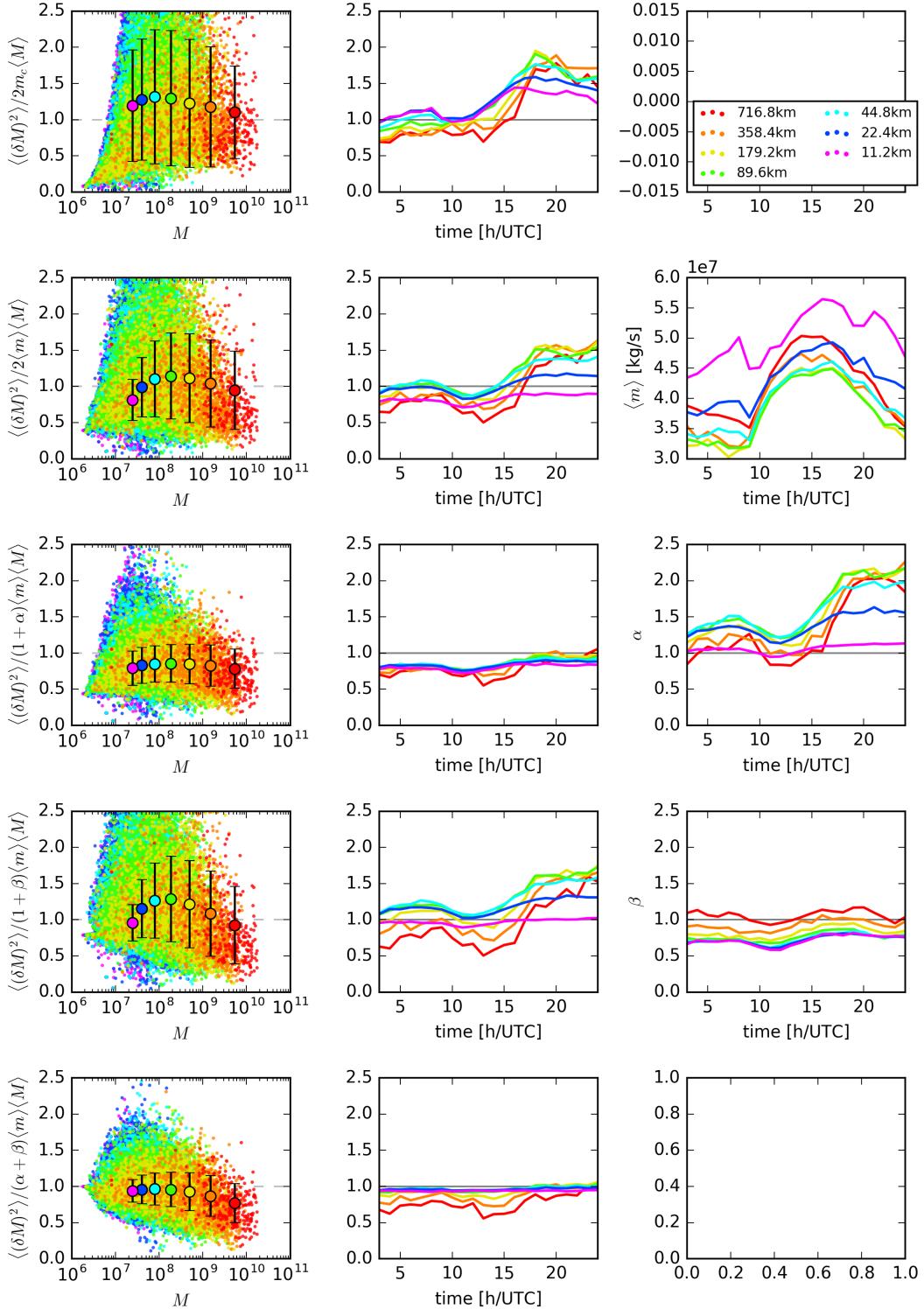


Figure 11: As above but for 20 ensemble members

Comparison of simulation results and predictions  
of increasing complexity

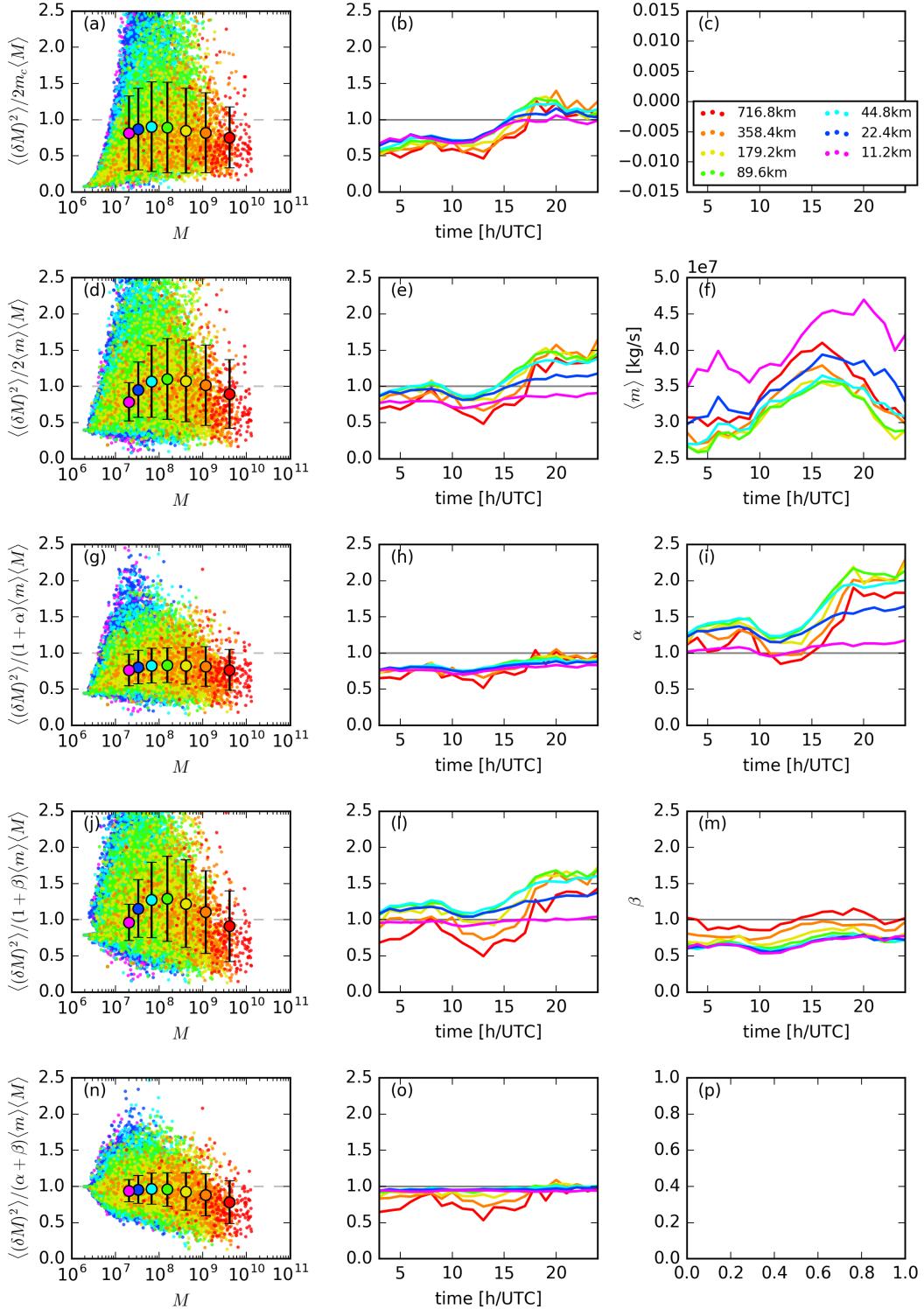


Figure 12: As above but at 2 km analysis height for 20 Members

Comparison of simulation results and predictions  
of increasing complexity

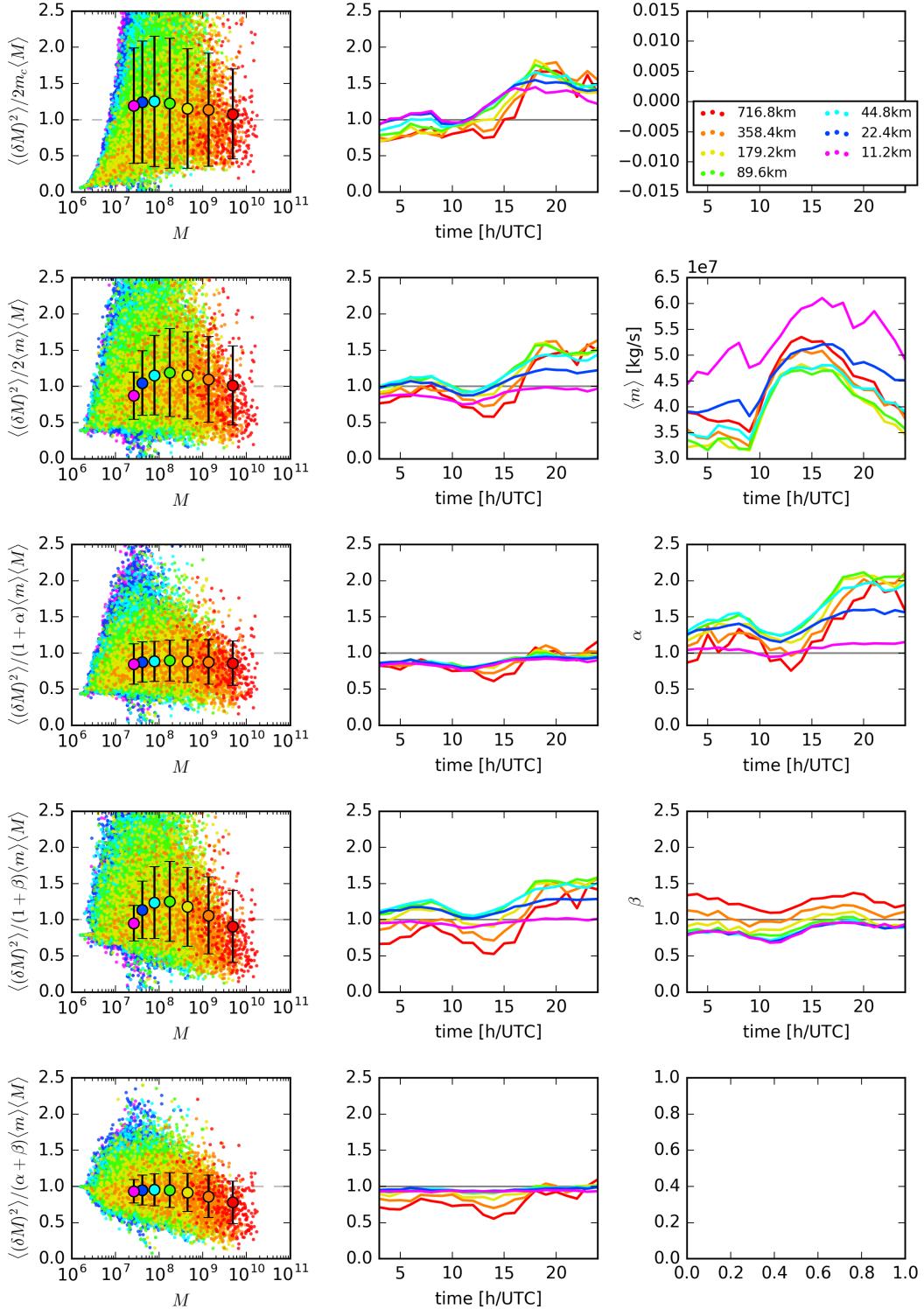


Figure 13: As above but at 4 km analysis height or 20 Members