RMetS
Royal Meteorological Society

Metrics for the evaluation of warm convective cloud fields in a large-eddy simulation with Meteosat images

Sebastian Bley,[a]* Hartwig Deneke,[a] Fabian Senf[a] and Leonhard Scheck[b]
[a]*Leibniz Institute for Tropospheric Research, Leipzig, Germany*
[b]*Meteorological Institute Munich, Ludwig-Maximilian University of Munich, Germany*

*Correspondence to: S. Bley, TROPOS, Permoserstrasse 15, 04318 Leipzig, Germany.
E-mail: sebastian.bley@tropos.de

The representation of warm convective clouds in atmospheric models and satellite observations can considerably deviate from each other partly due to different spatial resolutions. This study aims to establish appropriate metrics to evaluate high-resolution simulations of convective clouds by the ICON Large-Eddy Model (ICON-LEM) with observations from Meteosat SEVIRI over Germany. The time series and frequency distributions of convective cloud fraction and liquid water path (LWP) are analyzed. Furthermore, the study focuses on size distributions and decorrelation scales of warm convective cloud fields. The investigated metrics possess a pronounced sensitivity to the apparent spatial resolution. At the fine spatial scale, the simulations show higher occurrence frequencies of large LWP values and a factor of two to four smaller convective cloud fractions. Coarse-graining of simulated fields to the optical resolution of Meteosat essentially removes the differences between the observed and simulated metrics. The distribution of simulated cloud sizes compares well with the observations and can be represented by a power law, with a moderate resolution sensitivity. A lower limit of cloud sizes is identified, which is 8–10 times the native grid resolution of the model. This likely marks the effective model resolution beyond which the scaling behaviour of considered metrics is not reliable, implying that a further increase in spatial resolution would be desirable to better resolve cloud processes below 1 km. It is finally shown that ICON-LEM is consistent with spatio-temporal decorrelation scales observed with Meteosat having values of 30 min and 7 km, if transferred to the true optical satellite resolution. However, the simulated Lagrangian decorrelation times drop to 10 min at 1 km resolution, a scale covered by the upcoming generation of geostationary satellites.

Key Words: warm convective cloud fields; model evaluation; geostationary satellite remote sensing; cloud field metrics; resolution sensitivity; large-eddy simulation

Received 12 December 2016; Revised 10 March 2017; Accepted 3 May 2017; Published online in Wiley Online Library 5 June 2017

1. Introduction

The realistic representation of clouds in general circulation models (GCMs) in general, and convective clouds in particular, remains a fundamental challenge of climate research (Bony *et al.*, 2015). This difficulty arises in part from the coarse horizontal resolution (∼100 km) of GCMs, which does not allow one to resolve individual clouds nor the underlying turbulent, microphysical and convective processes. Instead, subgrid-scale processes below the model resolution have to be parametrized. To evaluate and improve these parametrizations, and to advance our understanding of the effects of such small-scale processes, the characterization of the spatio-temporal characteristics of convective clouds across all relevant scales based on observations and high-resolution models is essential. Towards this goal, suitable metrics for comparing observations and model results are required, and their sensitivity to the spatial resolution of the underlying data needs to be quantified.

Weisman *et al.* (1997) reported a strong sensitivity of the temporal evolution of convective cloud systems to the grid resolution of the atmospheric model. They also demonstrated that a resolution below 1 km is necessary to resolve the cellular structure of cumulus clouds. The cloud size distribution and the morphology of simulated clouds has also been found to be sensitive to the horizontal resolution (Nagasawa *et al.*, 2006). Brown (1999) showed that the cloud size distribution is clearly

shifted towards smaller clouds if the grid resolution of their large-eddy simulation (LES) model is increased up to 20 m. They also found that the higher frequency of smaller clouds is compensated by a reduction of larger clouds, resulting in no significant change of the total cloud fraction (CF). In contrast, Neggers *et al.* (2003) demonstrated that the sizes of the dominating clouds are robust in their LES model considering a horizontal grid spacing from 25 to 100 m. Heus and Seifert (2013) tracked shallow cumulus clouds in a LES with 25 m grid resolution covering a relatively large domain, and investigated the resulting cloud size distributions. Applying different liquid water path (LWP) thresholds for cloud masking, a power law dependency was observed, with exponents ranging from −2.2 to −2.9. Moreover, they found the cloud size distributions for different resolution set-ups start to converge if only clouds larger than the effective model resolution are considered. The latter was assumed to be six times the grid resolution in their study. Dorrestijn *et al.* 2012) pointed out that a spatial resolution of at least 100 m is required to realistically resolve convective dynamics of small-scale clouds in numerical models. At the same time, large-domain calculations are needed to obtain meaningful cloud statistics (Stevens *et al.*, 2002). Based on 30 m resolution Landsat observations, Koren *et al.* (2008) showed that, as the spatial resolution of the satellite sensor decreases, an increasing fraction of small clouds are missed. Cloudy pixels also increasingly contain significant clear-sky contributions from the cloud-free surface. Overall, the CF was found to be strongly overestimated due to the resolution sensitivity. The described effects will result in signficant and resolution-dependent uncertainties in derived cloud properties, and highlight the sensitivity of satellite products to sensor resolution.

Due to the increasing capacity of supercomputers, large-domain simulations at a cloud-resolving scale are becoming feasible, and show, for example, a better timing of the diurnal cycle of convection (Hohenegger *et al.*, 2008). Within the High Definition Clouds and Precipitation for Climate Prediction (HD(CP)^2) project, a number of LESs were conducted with the ICOsahedral Non-hydrostatic (ICON) atmospheric model (Dipankar *et al.*, 2015) covering a large domain over Germany. These ICON-LEM runs were carried out on a 156 m grid, and thus partially resolve turbulence and large eddies, and allow study of their influence on shallow cumulus convection. A comprehensive evaluation of the ability of ICON simulations to reproduce the general thermodynamic and cloud-precipitation characteristics was performed by Heinze *et al.* (2016). This resulted in the overall agreement that high-resolution ICON-LEM simulations enable an improved understanding of cloud and precipitation processes, and can serve as reference for the development of novel cloud parametrizations in GCMs. Our study extends the efforts of Heinze *et al.* (2016) and assesses the representation of warm convective cloud fields including their spatio-temporal variability in more detail.

One challenge for the evaluation of high-resolution and large-domain simulations is the lack of suitable reference observations. Polar-orbiting satellite instruments like MODIS provide global datasets with high spatial resolution (∼250–1000 m), but do not allow to evaluate the temporal evolution of convective cloud fields. In contrast, the geostationary Meteosat satellites observe Central Europe with a 5 min repeat cycle, but only at a nadir resolution of ∼3 km, which is relatively coarse in comparison to the typical size of warm convective clouds. Nevertheless, Meteosat offers the unique opportunity to characterize both the spatial and the temporal variability of small-scale convective cloud fields, and to use these characteristics for model evaluation. Towards this goal, Bley *et al.* (2016) introduced metrics to characterize the spatio-temporal evolution of convective cloud fields. Considering the fields of the retrieved liquid water path, a characteristic decorrelation time-scale of 30 min in a Lagrangian reference frame, and a decorrelation length-scale of 7 km was identified. However, these scales are only valid at the coarse horizontal resolution of Meteosat (∼3.1 × 6 km^2), and a decrease of the decorrelation time when considering the high-resolution visible (HRV) channel with 1.2 × 2.1 km^2 horizontal resolution was also observed. This suggests a strong resolution dependency of the decorrelation scales on spatial resolution. Typical lifetimes of cumulus clouds simulated by LES lie between 20 and 40 min, but small clouds dissipate much faster after a few minutes (Jiang *et al.*, 2006).

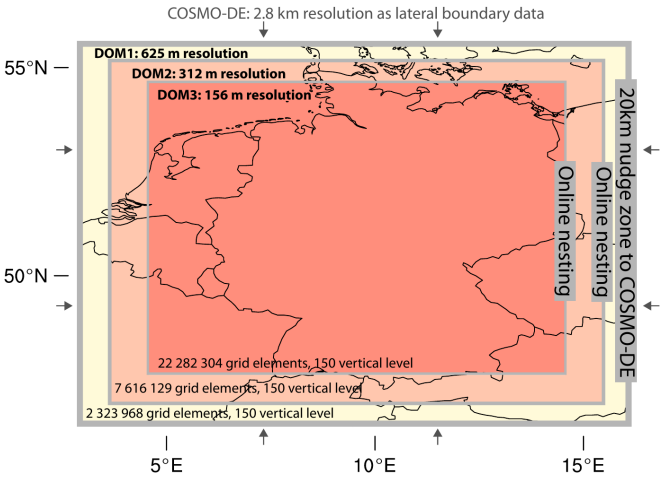

Figure 1. ICON-LEM domain and its two nests with grid refinement. The open lateral boundaries are relaxed towards the COMSO-DE analysis (Heinze *et al.*, 2016). [Colour figure can be viewed at wileyonlinelibrary.com].

The central goal of the present study is to establish appropriate metrics to evaluate the representation of convective cloud fields in ICON-LEM with observations from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) on board Meteosat, and to investigate the effects of model and sensor resolution. The main research questions can be summarized as:

(i) Does ICON-LEM realistically simulate the CF frequency distribition of LWP, as well as the horizontal structure and temporal evolution of convective cloud fields?
(ii) How sensitive is the comparison of satellite observations and model results to the spatial resolution of the satellite instrument and model grid?
(iii) Can we understand some limitations of the Meteosat observations using the much higher-resolved ICON-LEM results?

For this study, ICON-LEM simulations are carried out with three horizontal grid resolutions of 156, 312 and 625 m. This allows us to analyze differences in the cloud fields arising from different model resolutions, and to separate model-inherent resolution effects from those resulting from a coarse-graining of the model outputs.

The ICON-LEM, COSMO-DE and Meteosat data used in this study are briefly described in section 2. Results are presented and discussed in section 3. The main conclusions and an outlook are given in section 4.

2. Data and methods

2.1. Simulations

The ICON-LEM model is based on the unified modelling system for climate and numerical weather forecast ICON (ICOsahedral Non-hydrostic), which has been extended to a LES that partially resolves turbulence and convection within the HD(CP)^2 project. For the unresolved scales, a new three-dimensional turbulence scheme based on the classical Smagorinsky scheme has been implemented on a triangular grid (Dipankar *et al.*, 2015).

Within the project, four ICON-LEM runs have been performed for 24–26 April and 2 May 2013 with a very high spatial resolution of 156 m over a large domain covering Germany (Figure 1). The four simulated days comprise a range of typical weather conditions

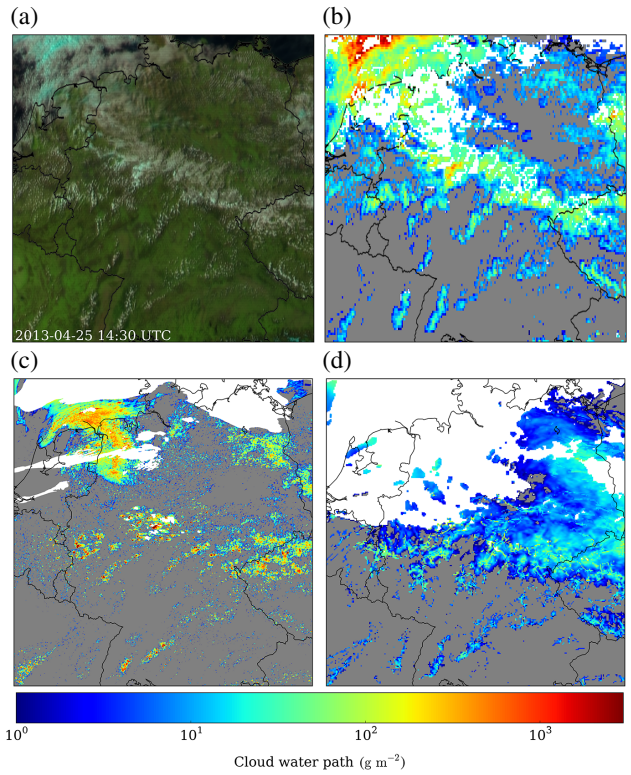

Figure 2. Case-study over the HD(CP)^2 domain on 25 April 2013 at 1430 UTC with (a) the high-resolution natural colour RGB image from Meteosat $\mathcal{O}(1\times2\,\mathrm{km}^2)$, (b) the cloud liquid water path (LWP) retrieved from Meteosat $\mathcal{O}(4\times6\,\mathrm{km}^2)$, (c) simulated by ICON-LEM $\mathcal{O}(156\,\mathrm{m})$ and (d) simulated by COSMO-DE $\mathcal{O}(2.8\,\mathrm{km})$. The cyan colours identify high cirrus clouds and deep convective cloud systems that contain ice particles, attributable to the strong absorption in the near infrared 1.6 μm channel. All frozen cloud fraction is visually filtered out by the white mask in (b)–(d). [Colour figure can be viewed at wileyonlinelibrary.com].

for Germany and the spring season. One-way nesting of ICON-LEM simulations is performed with grid refinement steps ranging from 624 to 312 m and 156 m in the innermost domain. As the grid spacings of the ICON-LEM runs are somewhat higher than those used typically in LES models for cloud studies, these runs could also be classified as 'near-LES' experiments (Mechem *et al.*, 2012).

ICON-LEM is used in a real-case configuration with prescribed lateral boundary conditions and a nesting approach (Heinze *et al.*, 2016). ICON-LEM is not run in a semi-idealized set-up like most other LES models, but uses the realistic land surface model TERRA which is also used in COSMO-DE (Heinze *et al.*, 2016). Each simulation day is initialized at 0000 UTC from the operational COSMO-DE analysis and runs for 24 h. The two-moment mixed-phase bulk microphysical parametrization of Seifert and Beheng (2005) is applied.

Apart from the great challenge to carry out such high-performance simulations at the German Supercomuting Centre (DKRZ), massive storage capacities are also needed to write the model output to disk (Heinze *et al.*, 2016). Data output of 50 terabytes was generated for one day of simulation. The three-dimensional model variables were mapped to a regular grid with 1 km grid spacing and written out every hour. Variables on the unstructured high-resolution model grid were written to files only once or twice a day during MODIS overpasses. The two-dimensional data output for the cloud properties is archived at 156, 312 and 625 m grid spacing and a 1 s time frequency, albeit sub-sampled to 1 min time steps for the present study to reduce the computing time for the analysis. Simulations from all three nests are compared. To avoid precipitation and glaciation effects that could influence the evolution of total cloud water, only the LWP is considered. Furthermore, to increase comparability to the satellite observations, we removed all columns of LWP that contain less than 1 g m^{-2} to reflect the sensitivity limit of the satellite sensors.

For comparison, COSMO-DE simulations have been carried out based on the operational set-up with 2.8 km horizontal grid resolution, but with the two-moment microphysical scheme also used in ICON-LEM (Seifert and Beheng, 2005). The initial and boundary conditions are prescribed by COSMO-EU on 7 km horizontal resolution. The COSMO-DE output has a 15 min temporal resolution, and contains the atmospheric pressure, temperature, water vapour mixing ratio and liquid water mixing ratio, which are used for calculation of the LWP. The liquid water mixing ratio already includes the subgrid-scale cloud information as used in the radiation scheme.

The performance of ICON-LEM in terms of boundary-layer variables, clouds and precipitation has already been evaluated by Heinze *et al.* (2016) using a comprehensive database including *in situ* and remote-sensing observations as well as reference model data from the COSMO-DE model. The key results of their study can be summarized as follows. COSMO-DE and ICON-LEM show a similar good performance with respect to cloud distributions and the large-scale situation, as both models are forced with nearly identical initial and boundary conditions. In ICON-LEM, shallow cumulus clouds are simulated as too large, which is likely attributable to the effective model resolution. Applying the ICON-LEM output to forward simulations relying on the look-up table-based Method for Fast Satellite Image Synthesis (MFASIS; Scheck *et al.*, 2016) shows similar cloud size distributions to the observations for cloud sizes between 1 and 100 km. Heinze *et al.* (2016) also show substantial improvements in the variability of the ICON-LEM LWP in $25\times25\,\mathrm{km}^2$ grid boxes compared to COSMO-DE.

2.2. *Observations*

Satellite data are taken from SEVIRI, which is the main payload on board the geostationary Meteosat Second Generation (MSG) satellites operated by EUMETSAT. The LWP is calculated using the Cloud Physical Properties retrieval (CPP; Roebeling *et al.*, 2006) developed in the framework of the Satellite Application Facility on Climate Monitoring (CM SAF; Schulz *et al.*, 2009). The lower part of the LWP distribution might be under-represented by MSG due to its detection limit, which mainly affects thin cirrus clouds and low small cumulus clouds. These highly variable clouds cause large uncertainties in the MSG retrieval of cloud optical thickness (τ) and droplet effective radius (r_e), and further lead to an underestimation of the LWP due to the plane-parallel albedo bias. Geostationary satellite retrievals generally underestimate the LWP especially for broken cumulus clouds (e.g. Marshak *et al.*, 2006; Wolters *et al.*, 2010). This effect also depends on the viewing geometry (Horvath *et al.*, 2014), and causes LWP uncertainties that influence the spatio-temporal characteristics of convective cloud fields. In the following, all ice-containing clouds have been identified with a cloud phase flag and filtered out. This has been done to emphasize the focus on liquid water clouds of the present study. However it should be noted that cloud phase determination from satellite is also subject to some uncertainties. The standard nadir sampling resolution of SEVIRI is $\sim 3\times3\,\mathrm{km}^2$ with a 5 min repeat cycle in the rapid scan mode. However, the horizontal resolution decreases for an increasing distance from the Equator due to the viewing geometry, having a pixel area of about $\sim3.1\times6\,\mathrm{km}^2$ in the centre of our domain. The real optical resolution of MSG is lower as characterized by the modulation transfer function (MTF) and the pixels are oversampled in the image rectification process by a factor of about 1.6 (Deneke and Roebeling, 2010). Thus the effective area of a pixel is slightly larger than the actual sampling resolution.

2.3. *Scene selection*

From the simulated ICON-LEM days, we selected two types of scene. First, domain-scale cloud scenes from one particular

day, 25 April 2013, were selected for the investigation of LWP frequency distribution (FD) and time series (TS) as well as the distribution of cloud sizes. For this day, Figure 2 shows the modelled and satellite-retrieved fields of LWP at 1430 UTC. To provide an overview of the synoptic situation and the thermodynamic phase of clouds, the satellite observations are shown as day natural colour red-green-blue (RGB) composite (Lensky and Rosenfeld, 2008). We eliminate the effect of larger-scale cloud fields that are advected into the model domain from the northwest by only considering the lower half of the domain for the calculation of FD and TS. Furthermore, cloud cover that touches the evaluation domain borders has been removed.

Second, we consider local-scale cloud scenes of a size of around $62 \times 62\,\mathrm{km}^2$ from all four simulated days. Corresponding boxes have a size of 400×400 (24×24) grid cells for ICON-LEM (COSMO-DE) simulations and are centred around developing warm convective cloud fields. This box size has been chosen to ensure preferably homogeneous advection conditions for an accurate tracking and to comprise a sufficient number of warm convective clouds at the same time to obtain robust cloud field statistics. However, the cloud cases have been subjectively selected, imposing an upper limit on cloud faction of 0.4 and 0.8 for ICON-LEM and COSMO-DE fields, respectively. A set of ten cloud cases is identified and tracked forward in time. A Fast Fourier Transform is used to calculate the spatial phase shift, determined by the maximum correlation (Anuta, 1970). This phase shift is finally transformed into a pixel shift for each of the 1 min time steps. For the local-scale cloud scenes, relations between average cloud aspect ratio and viewing angle dependence as well as the effects of resolution on spatio-temporal decorrelation scales are investigated on this data basis.

We do not expect ICON-LEM and COSMO-DE to perfectly match the MSG observations for small-scale convective clouds with respect to time and location. Rather we aim to compare the general statistics of simulated spatio-temporal decorrelation scales with the observational results obtained between April and July 2013 in our earlier study (Bley *et al.*, 2016).

2.4. Spatial resolution

The spatial resolution is a fundamental characteristic of atmospheric models and satellite instruments, which can influence not only the cloud properties but also the whole cloud field statistics, such as the distribution of cloud sizes and CF (Koren *et al.*, 2008). Particularly for the characterization of warm convective cloud fields, the spatial resolution needs to be considered carefully.

We emphasize that there can be an inherent difference between the resolution at which the data are provided, called the native resolution here, and the optical or effective resolution, for satellite observations and model simulations, respectively. For Meteosat observations, the optical resolution is lower than the native resolution, by a factor of 1.6 (section 2.2 gives details). For numerical simulations, the effective resolution is always coarser than the native grid resolution and represents a range at which the variability of physical processes can be resolved. With the term 'native resolution' we are assigning the grid spacing of the model data and the original pixel size of the satellite data in the present study. For the comparison of data at very different native resolutions, a third type of resolution, the coarsed-grained resolution comes into play which is used to convert the higher-resolved data into their lower-resolved counterpart. The ICON-LEM model has three different native grid resolutions, which are abbreviated by $\mathcal{O}(156\,\mathrm{m})$, $\mathcal{O}(312\,\mathrm{m})$ and $\mathcal{O}(625\,\mathrm{m})$ in the present study. The COSMO-DE model has a spatial resolution of $\mathcal{O}(2.8\,\mathrm{km})$ and the Meteosat native pixel resolution is $\mathcal{O}(3.6 \times 6\,\mathrm{km}^2)$ for the narrowband channels and $\mathcal{O}(1.2 \times 2\,\mathrm{km}^2)$ for the HRV channel. As stated above, the optical resolution of Meteosat is a factor 1.6 larger than the native pixel resolution. Hence, for comparison of ICON-LEM and Meteosat warm convective cloud fields, a coarse-graining of ICON-LEM cloud fields to 7 km average pixel size is performed, although ignoring the anisotropy in the observation grid.

The coarse-graining is performed sequentially. The original field is divided into small subdomains of 2×2 size for which the average is calculated and retained. After the first coarse-graining step, the resulting field hassmaller pixel size by a factor of four. σ represents the number of coarse-graining iterations. The ICON-LEM cloud fields are gradually coarse-grained to $156 \times 2^{\sigma}$ m until 7000 m spatial resolution is reached. COSMO-DE is coarse-grained to 5.6 km resolution. For the coarse-grained cloud fields, to avoid confusion over the origin of these cloud fields, the following notation is introduced: $\mathcal{C}$(native resolution$\rightarrow$coarse-grained resolution).

3. Metrics

In this section, results from the investigated metrics are presented and discussed. As metrics for the evaluation of warm convective clouds in the ICON-LEM model with Meteosat observations and COMSO-DE data, we first consider the frequency distribution and time series of LWP and CF. The effect of the cloud aspect ratio, spatial resolution and viewing geometry is subsequently addressed. In the following, the frequency distribution of cloud sizes is analyzed for the native resolution and coarse-grained to the optical resolution of Meteosat. The power laws are also compared to the observations. Convective cloud field tracking is finally applied to evaluate the decorrelation time- and length-scales. The model results are related to spatio-temporal characteristics achieved in Bley *et al.* (2016). The main focus of these investigated metrics lies on the spatial resolution sensitivity.

3.1. Frequency distribution of LWP

Figure 3 shows an example LWP distribution for a local-scale cloud field from the ICON-LEM model over southern Germany on 25 April 2013 at 1300 UTC. The image demonstrates how a warm convective cloud field is captured by the optical Meteosat resolution of 7 km (right-hand side). The fine LWP structures that are simulated by ICON-LEM $\mathcal{O}(156\,\mathrm{m})$, $\mathcal{O}(312\,\mathrm{m})$ and $\mathcal{O}(625\,\mathrm{m})$ (left-hand side) are completely smoothed down by a coarse-graining to 7 km with a substantial decreasing frequency of high LWP values. Coarse-graining of the LWP fields additionally leads to a strong increase in the CF, because larger clouds grow faster in the coarse-graining than smaller clouds disappear in the gaps (Koren *et al.*, 2008). For the case example in Figure 3, the CF of 0.23 in ICON-LEM $\mathcal{O}(156\,\mathrm{m})$ is increased to 0.69 when going to ICON-LEM $\mathcal{C}(156\,\mathrm{m} \rightarrow 7\,\mathrm{km})$. Also, the spatial decorrelation length (λ) changes drastically from $\lambda = 733\,\mathrm{m}$ to $\lambda = 4531\,\mathrm{m}$, implying that the cloud sizes are increasing and that a part of the texture information is lost. λ represents a measure of the change of spatial coherence of the horizontal cloud field structure, and indicates how far the box can be still displaced before the LWP structure between the displaced and the initial cloud field is decorrelated (Bley *et al.*, 2016). The resolution sensitivity of CF and λ is discussed in more detail in sections 3.3 and 3.5.

Most of the fine LWP structure from ICON-LEM $\mathcal{O}(156\,\mathrm{m})$ can be still obtained from ICON-LEM $\mathcal{C}(156\,\mathrm{m} \rightarrow 1250\,\mathrm{m})$. However, coarse-graining to 7 km clearly eliminates the original 156 m structure. Every pixel in the 7 km box includes a substantial amount of unresolved sub-pixel variability. Using complementary data from the HRV channel with $1.2 \times 2\,\mathrm{km}^2$ resolution can help to resolve a part of this variability (Bley *et al.*, 2016).

Figure 3 further indicates a nonlinear relation between the spatial resolution and CF as well as λ. The change in LWP structure, CF and λ is much more pronounced between $\mathcal{C}(156\,\mathrm{m} \rightarrow 625\,\mathrm{m})$ and $\mathcal{C}(156\,\mathrm{m} \rightarrow 2500\,\mathrm{m})$ than between $\mathcal{O}(156\,\mathrm{m})$ and $\mathcal{C}(156\,\mathrm{m} \rightarrow 625\,\mathrm{m})$, although the resolution is

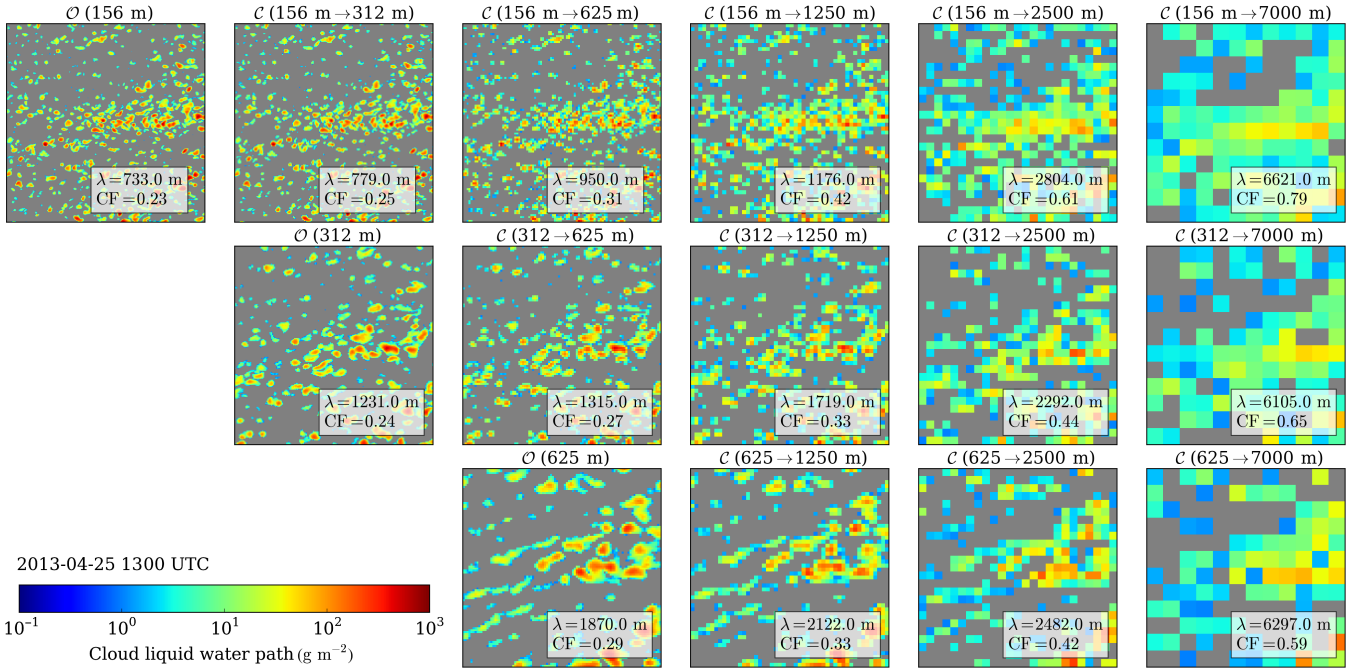

Figure 3. Example of a local-scale ICON-LEM $62 \times 62\,\mathrm{km}^2$ cloud field showing the LWP over Southern Germany on 25 April 2013. The first row shows the native resolution and coarse-grained cloud fields, originated from ICON-LEM $\mathcal{O}(156\,\mathrm{m})$, the second row from ICON-LEM $\mathcal{O}(312\,\mathrm{m})$ and the third row from $\mathcal{O}(625\,\mathrm{m})$. The grey backgroud indicates cloud-free areas; λ is the decorrelation length and CF the convective cloud fraction for each cloud field. [Colour figure can be viewed at wileyonlinelibrary.com].

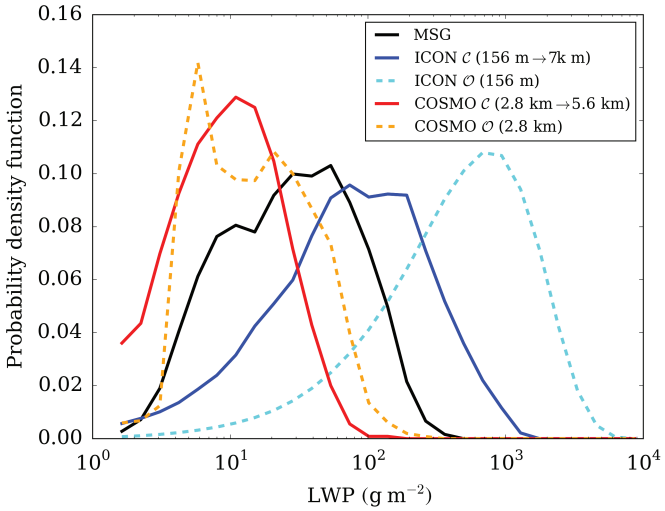

Figure 4. Probability density function (PDF) of the domain-scale LWP with warm convective clouds retrieved by MSG (black) and simulated by ICON-LEM $\mathcal{C}(156\,\mathrm{m} \rightarrow 7\,\mathrm{km})$ (blue solid) and COSMO-DE $\mathcal{C}(2.8\,\mathrm{km} \rightarrow 5.6\,\mathrm{km})$ (red solid) on 25 April 2013 between 0900 and 1530 UTC. The dashed lines indicate the ICON-LEM $\mathcal{O}(156\,\mathrm{m})$ and COSMO-DE $\mathcal{O}(2.8\,\mathrm{km})$ functions at their native resolution. [Colour figure can be viewed at wileyonlinelibrary.com].

quadrupled in both cases. This is most likely related to the effective resolution of the model and will be discussed in detail in section 3.5

The domain-scale LWP frequency distribution averaged between 0900 and 1530 UTC is presented in Figure 4. ICON-LEM and COSMO-DE are coarse-grained towards Meteosat resolution for a solid comparison. The original ICON-LEM $\mathcal{O}(156\,\mathrm{m})$ and COSMO-DE $\mathcal{O}(2.8\,\mathrm{km})$ data are also shown for a demonstration of the resolution sensitivity.

On the one hand, COSMO-DE at native resolution underestimates the typical range of LWP values compared to the observations – a situation that becomes even worse after coarse-graining to $\mathcal{C}(2.8\,\mathrm{km} \rightarrow 5.6\,\mathrm{km})$. On the other hand, the high-resolution simulations performed by ICON-LEM achieve LWP values at native resolution that are more than one magnitude larger. Coarse-graining brings the ICON-LEM LWP values closer to the observations, but still a significant overestimation remains. This might be a real model deficit and be caused by the fact that too much water is accumulated at convection scales of 1 km and more, which would have been distributed over much smaller cloud scales in reality. A comparison of observed and simulated cloud size distributions will be discussed in depth in section 3.4. In general, Figure 4 reveals the strong resolution sensitivity of LWP.

Heinze *et al.* (2016) evaluated the combined liquid and ice cloud water path (CWP) for all days over the full HD(CP)^2 domain and concluded that ICON-LEM simulates clouds better than COSMO-DE in comparison to the satellite-retrieved CWP. The MODIS CWP retrieval was found to agree better with ICON-LEM than the MSG retrieval, which is likely attributable to the higher resolution of the MODIS instrument ($1 \times 1\,\mathrm{km}^2$). While MODIS seems to be a better reference dataset to evaluate small-scale cloud structures in ICON-LEM, only MSG as a geostationary instrument with a high temporal sampling is able to resolve the temporal evolution of warm convective clouds. The temporal development of a warm convective cloud field is analyzed in the next section.

3.2. *Time series of LWP and CF*

Due to the high spatio-temporal LWP inhomogeneity of warm convective clouds, we do not expect a perfect agreement between simulations and observations with respect to time and location of individual cumulus clouds. Therefore, the evaluation of the LWP time series is performed for the domain scale.

Figure 5 shows the time series of the in-cloud averaged LWP, the CF and the area-average LWP, again for the native resolution of MSG, ICON-LEM and COSMO-DE and for COSMO-DE and ICON-LEM also coarse-grained. In contrast to the LWP frequency distribution, the ICON-LEM $\mathcal{C}(156\,\mathrm{m} \rightarrow 7\,\mathrm{km})$ time series is in a much better agreement with MSG than with COSMO-DE $\mathcal{C}(2.8\,\mathrm{km} \rightarrow 5.6\,\mathrm{km})$. However, the upper quartile range is substantially higher, in both the native and the coarse-grained ICON-LEM distribution, which is also supported by Figure 4. The MSG retrieval indicates an artifact around 1400 UTC which is most likely caused by a particular scattering angle of ∼135° around

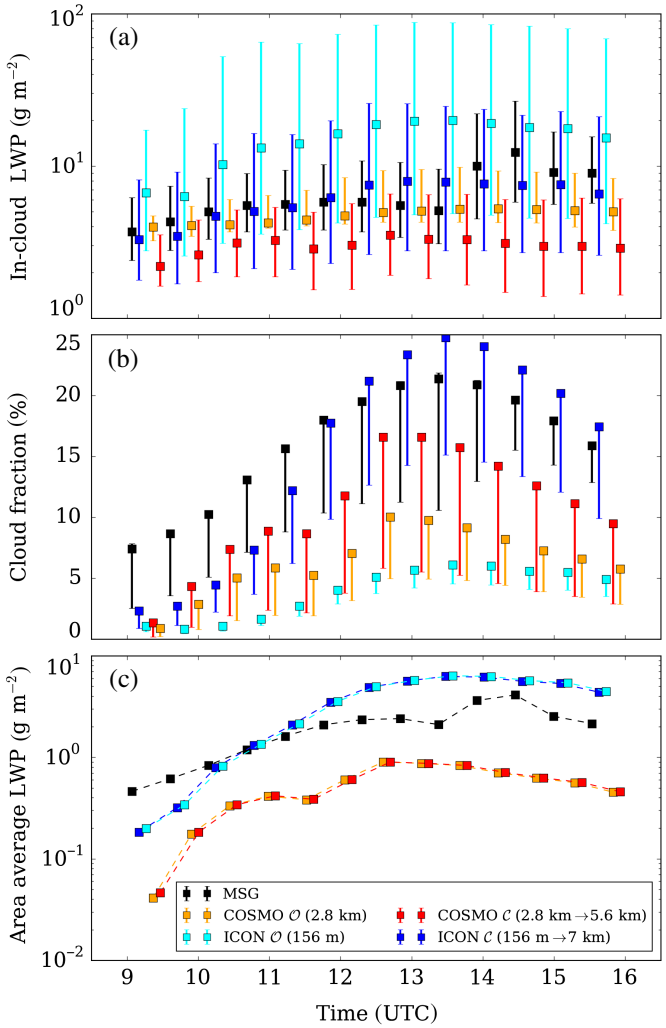

Figure 5. Time series of (a) the in-cloud LWP, (b) the CF and (c) the domain-scale average LWP retrieved by MSG (black), and simulated by ICON-LEM $\mathcal{O}$(156 m) (cyan), ICON-LEM $\mathcal{C}$(156 m → 7 km) (blue) and COSMO-DE $\mathcal{O}$(2.8 km) (orange) and COSMO-DE $\mathcal{C}$(2.8 km → 5.6 km) (red). The in-cloud LWP errorbars represent the interquartile range, CF errorbars indicate the sensitivity of the CF on the LWP threshold of LWP>1 g m^{-2} (marker) and LWP>5 g m^{-2} (lower range). [Colour figure can be viewed at wileyonlinelibrary.com].

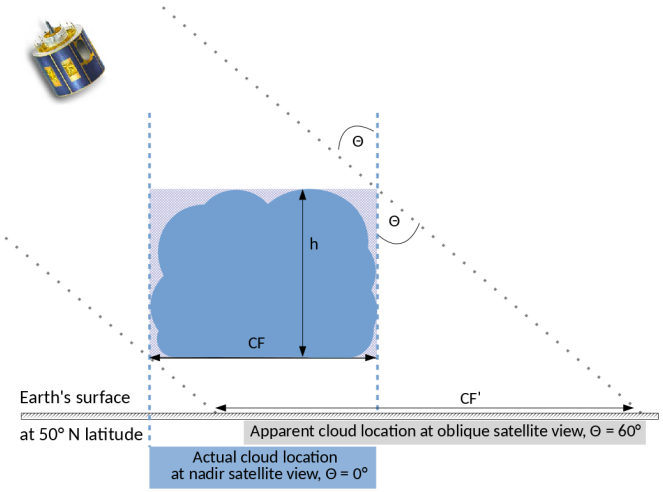

Figure 6. Sketch of the projected location of a cloud on the surface, seen from nadir perspective (blue) and from geostationary satellite perspective with a satellite zenith angle of 60° (grey). [Colour figure can be viewed at wileyonlinelibrary.com].

the cloudbow caused by liquid water droplets (Cho *et al.*, 2015). In conclusion, this shows that also the time series of quartile values of the in-cloud LWP is very sensitive to the spatial resolution.

This resolution sensitivity can be perfectly demonstrated with the time series of the CF (Figure 5(b)). Understanding the sensitivities and changes in CF is very important for climate projections, but a reliable comparison between satellite observations and model results of the CF is challenging due to different aspects. Due to the detection sensitivity, MSG basically excludes thin clouds with a LWP lower than 1 g m^{-2}. A cloud mask filter is consequently applied to the ICON-LEM, COSMO-DE and MSG cloud fields, defining LWP pixels above 1 g m^{-2} as cloudy, otherwise as cloud-free. The CF is also sensitive to the LWP threshold which is applied for the CF in our study. We therefore show the CF for LWP>1 g m^{-2} and LWP>5 g m^{-2} in the present study.

Figure 5(b) shows that observed convective CF increases from ∼ 8% in the morning hours to a peak around 20% between 1300 and 1400 UTC. The simulated CFs at native model resolution, which peak at 5% (9%) for ICON-LEM (COSMO-DE), are significantly increased by coarse-graining. COSMO-DE still remains below the observational level whereas ICON-LEM slightly overestimates the peak values in the early afternoon. Both simulations, but especially coarse-grained ICON-LEM, exhibit an overestimation of the diurnal amplitude of convective CF. However, the timing of the CF peaks seems to be in better agreement for ICON-LEM and MSG. In the late afternoon, the simulations and observation display a decreasing CF, which is likely attributable to dissolving clouds or phase transition into ice clouds due to cloud deepening. (Senf *et al.*, 2015).

Figures 5(a) and (b) clearly demonstrate a sensitivity to the spatial resolution, which affects the comparison between observations and models. Figure 5(c) presents the area-averaged LWP, which combines the in-cloud LWP and the CF and gives information about the total liquid water amount within the domain. Now the effect of the spatial resolution is eliminated, only the cloud bow artifact in the MSG retrieval is still apparent. A better agreement is found between ICON-LEM and MSG than for the COSMO-DE LWP which significantly underestimates the total amount of liquid water. From a budget point of view, the efficiency of the liquid water production of water vapour is significantly increased when going from convection-permitting scales of COSMO-DE to very high resolution of ICON-LEM (Baldauf *et al.*, 2011). This increase goes even beyond the observational values leading to an overestimation of simulated convectively generated liquid water.

In conclusion, this analysis demonstrates that higher-resolution simulations are able to reduce biases in liquid cloud properties that show up in the coarser convection-permitting simulations. An overcorrection seems to happen in which the negative bias in convective CF and total liquid water amount changes into a positive bias in the high-resolution runs.

3.3. *Cloud aspect ratio and viewing geometry*

A further important aspect that can influence the CF, but has not been discussed so far in this study, is the cloud aspect ratio, which leads to CF uncertainties due to the slanted viewing geometry of MSG. Due to the position of MSG in the geostationary orbit, observations over Central Europe are performed at high satellite viewing angles between 45° and 65°. This oblique viewing perspective leads to three-dimensional radiative interactions between cloud edges that are not considered in current one-dimensional cloud retrievals (Várnai and Marshak, 2007). That the viewing geometry also influences the CF and causes high uncertainties in data record of the average cloudiness has already been studied by Evan *et al.* (2007).

Figure 6 contrasts the projected location of a cloud over Germany, seen from nadir and from geostationary satellite view under a zenith viewing angle θ of 60°. When clouds of a certain height h are viewed from space by zenith angle θ and azimuth angle ϕ, an apparent northward shift $\Delta y \approx h \tan\theta$, the

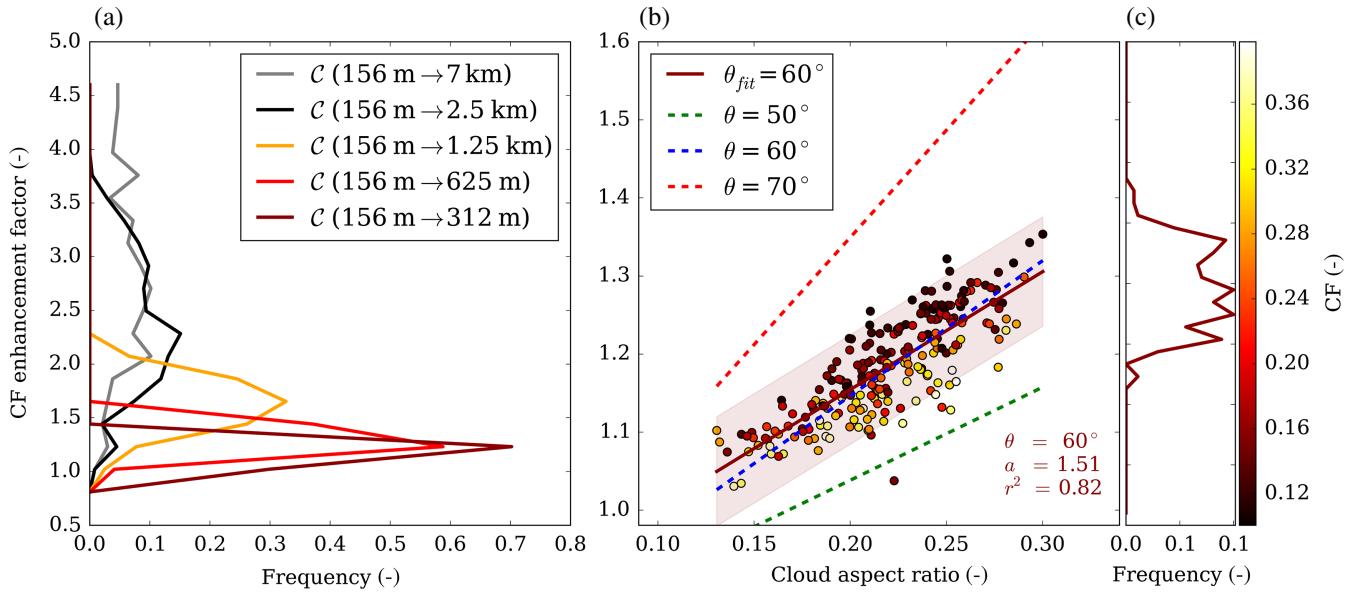

Figure 7. (a) Frequency distribution of the CF enhancement factor on a local scale for 62×62 km^2 ICON-LEM $\mathcal{O}(156\,\text{m})$ warm convective cloud fields and coarse-grained. (b) CF enhancement factor due to artificial oblique viewing geometry applied to ICON-LEM $\mathcal{O}(156\,\text{m})$ cloud fields in relation to the cloud aspect ratio for a 60° satellite zenith angle. The solid dark red line indicates the linear regression with slope a and correlation coefficient r^2. The dashed lines represent the theoretical curves for varying zenith angle, (c) is the CF enhancement factor frequency for 60°. [Colour figure can be viewed at wileyonlinelibrary.com].

so-called parallax shift, happens (e.g. Kostka *et al.*, 2014, gives further discussion). We use this relation and apply an artificial viewing angle to the ICON-LEM cloud fields to imitate a satellite instrument that observes the simulated cloud field from the geostationary orbit. Since the longitude position of the satellite (9.5°) lies in our ICON-LEM domain, we furthermore neglect shifts in the longitude direction. From Figure 6, an increased cloud diameter and corresponding cloud area is calculated. The gain in cloud area is given by CAR $\tan\theta$, where the cloud aspect ratio CAR is defined as cloud geometrical thickness divided by the cloud diameter. Furthermore, we account for the overlap of closely located cloud towers.

The enhancement factor of CF is plotted in Figure 7 for the set of local-scale cloud scenes. It represents the change in coarse-grained convective CF relative to the initial CF of the cloud field at 156 m spatial resolution. The viewing geometry effect causes CF enhancement factors of 1.2–1.4 for the ICON-LEM $\mathcal{O}(156\,\text{m})$ cloud fields. This effect appears small in comparison to the CF enhancement, supported by Figure 5(b). Figure 7(a) quantifies the increase in the CF for the same convective cloud fields as in Figure 7(b), but as the result of decreasing spatial resolution due to coarse-graining. This CF enhancement already exceeds at $\mathcal{C}(156\,\text{m} \rightarrow 1250\,\text{m})$ the viewing geometry CF enhancement, which demonstrates that coarse-graining to 7 km substantially predominates over the influence of the viewing geometry. One should note that Figure 7(b) is only shown for ICON-LEM $\mathcal{O}(156\,\text{m})$ and not for any coarser resolution, because coarse-graining is only done in the horizontal direction. This would result in considerably lower aspect ratios, which most likely lead to smaller CF enhancement factors.

One uncertainty, which we cannot account for, is the aspect ratio. The mean aspect ratio for all local-scale cloud scenes is 0.2, ranging from 0.1 to 0.27. These values only overlap at the lower end of aspect ratios calculated in (e.g.) Benner and Curry (1998) or Kassianov *et al.* (2005), who found values between 0.2 and 1. Therefore, ICON-LEM aspect ratios seem to be slightly too small, which might be caused by numerical diffusion at the scale of the effective resolution. We also suggest that the ratio between vertical and horizontal grid box size might influence the distribution of cloud aspect ratios. Uncertainties in the cloud aspect ratios can cause high uncertainties in the representation of cloud radiation interactions. Han and Ellingson (1999) found a strong influence of the estimated cloud aspect ratios on results in the long-wave radiation transfer calculations. Furthermore, adiabaticity of convective clouds and hence the liquid water distribution depends on cloud geometrical thickness (Merk *et al.*, 2016). The CF strongly increases for increasing aspect ratios and is even higher for higher zenith angles (Figure 7(b)). The highest CF enhancement appears for low CF between 0.1 and 0.2, which is consistent with results from Minnis (1989). The MSG retrieval gives reliable information neither about the cloud thickness nor the cloud-base height. Analysis of data from active satellite sensors can help to better evaluate the aspect ratios, however this goes far beyond our study focus.

In summary, coarse-graining the ICON-LEM cloud fields to MSG resolution dominates the CF enhancement factor in comparison to the viewing angle effect. Although the latter effect seems to be rather weak in the present study, it needs to be carefully considered for deeper convective clouds and long-term global trends of CF from different satellite instruments (Evan *et al.*, 2007).

3.4. *Cloud size distribution*

For evaluation of ICON-LEM cloud size distributions and the investigation of their resolution sensitivity, a binary cloud mask is generated by applying a 1 g m^{-2} threshold to the domain-scale LWP fields. From this cloud mask, connected cloud regions are labelled using a connectivity clustering method. For each cloud object, an area-equivalent diameter is calculated. The relative occurrence frequencies of the number of cloud objects within a certain size range are divided by the domain area which results in the unit km^{-3}. Note that the normalization of the resulting density functions was performed with respect to equidistant size ranges to be comparable with the studies of Heus and Seifert (2013) and Heinze *et al.* (2016), even though we present our cloud size distributions in log space.

Figure 8 shows the frequency distribution n(L) of clouds with a given cloud size L. Different studies show that a power law is the most appropriate way to fit the horizontal cloud size distribution (Benner and Curry, 1998; Zhao and Di Girolamo, 2007; Wood and Field, 2011). We calculate the power law exponent as the slope of the least squares fit to the data in logarithmic space.

The frequency distribution of cloud sizes is well represented by a power law with exponent β ranging from −2.3 to −3.2.

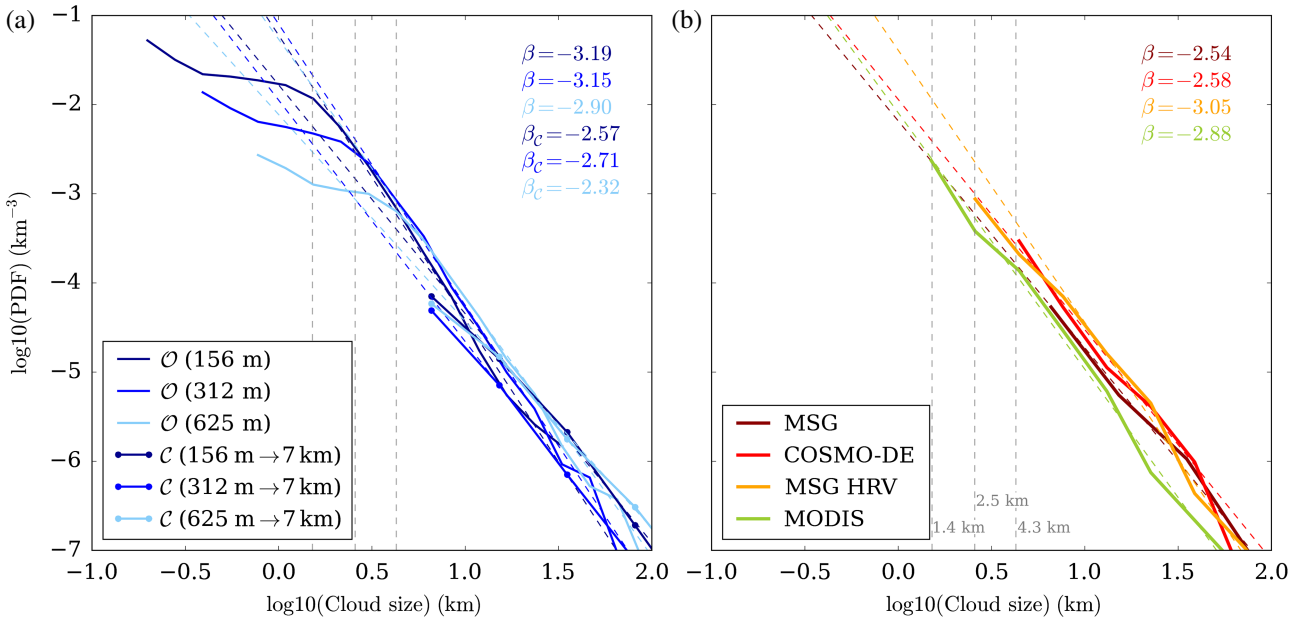

Figure 8. Domain-scale frequency distribution for cloud sizes (a) simulated by ICON-LEM and (b) observed by Meteosat (dark red), the Meteosat HRV mask (orange) and MODIS (green) and simulated by COSMO-DE (red) on 25 April 2013. The native ICON-LEM curves are shown as solid lines, and the coarse-grained distributions are additionally marked by filled circles. Linear regressions of the size distributions between different size ranges are indicated by dashed lines which can be described by the power law with exponent β. The standard error of the linear regression is ~0.1 for the original ICON-LEM distributions, ~0.15 for MSG, MODIS, MSG-HRV mask and COSMO-DE and ~0.3 for the coarse-grained ICON-LEM functions. [Colour figure can be viewed at wileyonlinelibrary.com].

Exponents between −2.3 and −2.7 are found for the coarse-grained fields and MSG, whereas a somewhat steeper decline with exponents between −2.9 and −3.2 is obtained for higher-resolution cloud fields. However, the power laws only show a moderate resolution sensitivity, which is consistent with a study by Wood and Field (2011), who demonstrated that sensor resolution is not found to strongly influence the power law. Koren *et al.* (2008) investigated the resolution sensitivity of the cloud size distribution by comparing Landsat and MODIS data with 30 m and 1 km spatial resolution, respectively. They concluded that, at any spatial resolution, a substantial number of small clouds are missed and that clouds below the sensor detection limit are more numerous than the detectable ones.

The ICON-LEM power law range between −2.9 and −3.2 is consistent with power laws obtained in a LES study by Heus and Seifert (2013) who found −2.2 to −2.9. Heinze *et al.* (2016) calculated MODIS-like ICON-LEM satellite images based on a fast radiative transfer method (Scheck *et al.*, 2016) for calculating the distribution of cloud sizes from the visible images, resulting in $\beta = -3.1$.

Using the Meteosat HRV mask for the object-based analysis yields an exponent of $\beta = -3.05$, which is also in a excellent agreement with the ICON-LEM slopes. Considering the fit standard errors of 0.1–0.3, the coarse-grained power laws remain consistent with other satellite-based studies like Zhao and Di Girolamo (2007), who estimated $\beta = -2.85$. Due to the relatively small number of data points for the COSMO-DE, MSG and coarse-grained ICON-LEM fits, these power laws should be interpreted with caution.

While the parameters of the fitted power laws show just a moderate resolution sensitivity, the ICON-LEM distributions start to strongly deviate from the power laws for sizes smaller than 1.4 km, 2.5 km and 4.3 km for ICON-LEM $\mathcal{O}$(156 m), ICON-LEM $\mathcal{O}$(312 m) and ICON-LEM $\mathcal{O}$(625 m), respectively. These cloud sizes are in the range of the effective model resolution, which is about 8–10 times the native grid resolution (Zängl *et al.*, 2015). Consequently, ICON-LEM suppresses clouds that are smaller than this effective resolution. This clearly has implications for the planning of high-resolution simulations of cloud processes: if a reliable description of cloud and precipitation processes at a 1 km scale is targeted, than the native model resolution has to be chosen to be 156 m or higher to avoid the impact of artificial numerical smoothing.

3.5. *Decorrelation length- and time-scales*

We follow the methodology of Bley *et al.* (2016) and investigate the decorrelation scales and their resolution sensitivity in ICON-LEM warm convective cloud fields. Based on MSG observations, Bley *et al.* (2016) estimated the average cloud lifetime of warm convective clouds by approaching a Lagrangian perspective. In the present study, simulated warm convective cloud fields are analyzed in a similar way. A set of ten local-scale cloud scenes are analyzed. The decorrelation length is calculated for a horizontal displacement of a box at constant time. The Lagrangian decorrelation time is calculated along the trajectory, while the Eulerian decorrelation time is obtained for a constant location. While the Lagrangian decorrelation time is expected to be generally larger than the Eulerian decorrelation time, the latter is also beneficial to estimate the impact of advective processes onto the local decorrelation behaviour.

Figure 9 shows the decorrelation length, Eulerian decorrelation time and Lagrangian decorrelation time as averages of ten ICON-LEM and COSMO-DE fields of warm convective clouds in relation to the reciprocal horizontal resolution. The black diamonds illustrate the MSG scales obtained by Bley *et al.* (2016). Two red diamonds show the COSMO-DE decorrelation lengths for its original 2.8 km and coarse-grained 5.6 km resolution, but no decorrelation time scales were calculated due to its coarse temporal sampling of 15 min, which does not allow the determination of reliable cloud field tracks (Bley *et al.*, 2016).

An excellent agreement is found between the observed and the ICON-LEM $\mathcal{C}$(7 km) scales within the error range. The spatial decorrelation scales of COSMO-DE $\mathcal{O}$(2.8 km) and COSMO-DE $\mathcal{C}$(5.6 km) also match well with the ICON-LEM scales within the respective reciprocal resolution. In contrast, ICON-LEM cloud fields at 156 m resolution exhibit substantially lower decorrelation lengths of 1 km and Lagrangian decorrelation times of 10 min. This result also demonstrates a high amount of unresolved cloud variability in the coarse-grained ICON-LEM and MSG cloud fields, which causes the high-resolution sensitivity of the decorrelation scales.

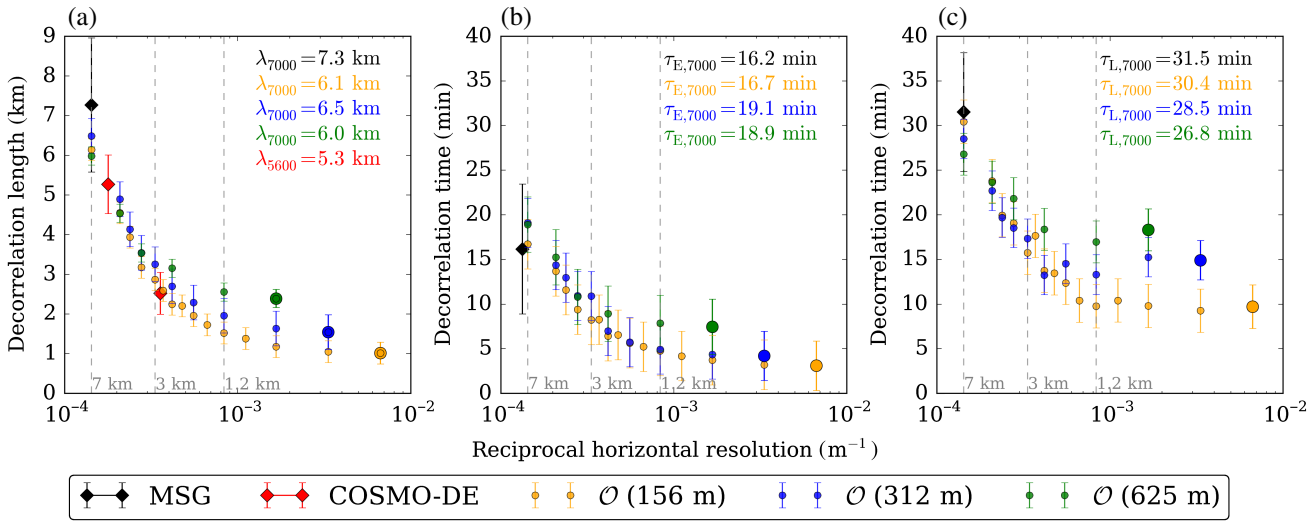

Figure 9. (a) Spatial decorrelation length and (b) Eulerian and (c) Lagrangian decorrelation times in relation to the reciprocal horizontal resolution of simulations from ICON-LEM $\mathcal{O}$(156 m) (yellow), $\mathcal{O}$(312 m) (blue) and $\mathcal{O}$(625 m) (green), COSMO-DE (red) and observations by MSG (black) of warm convective cloud fields on a local scale averaged over ten $62 \times 62\,km^2$ LWP fields. The large circles represent the ICON-LEM scales for their native resolution, and smaller circles indicate the coarse-grained ICON-LEM scales. The vertical error bars indicate the standard deviation between the different cases for ICON-LEM, COSMO-DE and MSG. [Colour figure can be viewed at wileyonlinelibrary.com].

Bley *et al.* (2016) found average spatial decorrelation scales of 7 km and Lagrangian decorrelation times of 30 min. It was further shown that the decorrelation scales start to decrease when the HRV channel with $1 \times 2\,km^2$ resolution is utilized. The metrics which are presented here for characterizing the decorrelation scales give the opportunity to understand more clearly the resolution sensitivity of decorrelation scales.

The ICON-LEM cloud fields of 156, 312 and 625 m native resolution attain similar decorrelation scales between 4 and 7 km spatial resolution. At higher spatial resolution, they start to deviate from each other. Consequently, and similar to the distribution of cloud sizes (Figure 8), the decorrelation scales below the effective model resolution are not fully resolved. The Lagrangian decorrelation times exhibit the highest deviations between the different ICON-LEM runs, which is caused by uncertainties in the tracking. Thus, an accurate tracking is necessary to study warm convective cloud fields in a Lagrangian reference frame. An extrapolation of the resolution dependency of considered decorrelation scales beyond the effective model resolution is not physically meaningful. This also holds true for all other considered metrics.

With the MSG SEVIRI rapid scan, we are able to characterize the changes in convective cloud fields with an update frequency of 5 min, which is sufficient for its spatial resolution. The biggest limit of MSG is the spatial resolution, which leads to a substantial overestimation of the CF and decorrelation scales and underestimation of the LWP. The future generation of European geostationary satellites, the Meteosat Third Generation, will give great opportunity to bring down these limits to a spatial resolution of 1 km or even 500 m for selected channels (Stuhlmann *et al.*, 2005). The analyzed scaling behaviour of Figure 9 suggests the conclusion that the planned temporal update frequency of 10 min for the operational scan schedule and 2.5 min for the rapid scan are still sufficient to thoroughly characterize the decorrelation properties, including an estimate of the cloud lifetime, of convective cloud fields at the corresponding spatial scales.

4. Summary and conclusions

In the present study, several metrics have been investigated to evaluate the representation of convective clouds in the high-resolution atmospheric model ICON-LEM with observations from the geostationary Meteosat SEVIRI instrument. ICON-LEM simulations covering a large domain at 156 m grid resolution and containing convective cloud fields have been analyzed and compared to Meteosat observations and simulations with the COSMO-DE model. A variety of metrics including the time series of convective cloud fraction and LWP, the frequency distributions of LWP and cloud size, and the spatial and temporal decorrelation length-scales, have been considered. Using the cloud fields at three ICON-LEM native grid resolutions and after coarse-graining, the sensitivity of these metrics to the spatial resolution of the model and the observations has also been quantified.

The evaluation is performed on two types of cloud scenes: the one defined on the domain scale and including a variety of local conditions and the other one defined on the local scale following a convective cloud field during its temporal evolution in a Lagrangian perspective. For the latter, a set of ten cloud cases was collected with areal coverage of $62 \times 62\,km^2$. Cloud fields containing ice have been excluded to avoid complications caused by mixed-phase and precipitating clouds. We have further imposed an upper limit on the CF for the simulated cloud fields, to focus on broken convective cloud fields with a high degree of spatial inhomogeneity.

On a domain scale, we analyzed the spatial and temporal LWP and convective CF distribution as well as the cloud size distributions of convective cloud fields. On the local scale, spatio-temporal statistics of simulated cloud fields are compared to MSG observations. Sequences of LWP fields from ICON-LEM with a 1 min repeat cycle have been used to determine cloud field tracks to evaluate along-track correlation statistics and to study the resolution sensitivity on these scales. For this issue, ICON-LEM is gradually coarse-grained to the MSG resolution.

The main results are summarized as follows:

(i) The coarse-grained ICON-LEM fields show improvements in the representation of the frequency distribution of the LWP, convective CF and cloud sizes compared to convection-permitting simulations of COSMO-DE.

(ii) A substantial resolution sensitivity is found for the convective CF, in-cloud LWP and decorrelation scales. However, the power laws of the cloud size distribution exhibit only a moderate resolution sensitivity. The different metrics show that the spatial resolution needs to be considered to avoid the interpretation of non-physical differences that are attributable to the different resolutions. One could alternatively compare time series of the LWP on a domain average which substantially reduces the resolution sensitivity.

(iii) The results suggest that Meteosat is mainly limited by its spatial resolution, which is much coarser than the decorrelation length-scales. Due to its coarse spatial resolution, a lot of the small-scale LWP variability remains unresolved, which causes large uncertainties. If the spatial resolution of a satellite instrument were highly increased, the temporal resolution needs also to be increased, to allow an accurate characterization of the spatio-temporal behaviour.

Because the present study is based on a selection of only ten cases obtained from four days of simulations, it remains unclear whether our results are statistically robust and representative, in particular for other synoptic conditions. Nevertheless, the methodology presented here can serve as an example for future studies aimed at evaluating the representation of convective clouds in high-resolution models. A larger number of simulation days will allow us to verify the robustness of our results. To address this point, it is planned to carry out more ICON-LEM runs over Germany and over the Atlantic Ocean within the second phase of the HD(CP)^2 project.

In terms of comparability, forward satellite simulators using ICON-LEM cloud properties will be important to account for inconsistencies between assumptions made in satellite retrievals and simulations of convective clouds. Such simulators will also improve the quantification of retrieval sensitivities at very small or large LWP values. As soon as the implementation of the absorbing 1.6 μ m channel in the satellite forward operator is finished, it will become possible to apply the LWP retrieval algorithm on model data and thus to investigate such inconsistencies and sensitivities.

The intercomparison of observations and high-resolution model output demonstrates that Meteosat SEVIRI observations are limited by the sensor spatial resolution in the rapid scan mode. If the spatial resolution were highly increased to 500 m or even 100 m, an enhancement of the temporal resolution would be required also, in order to consider rapid changes of convective clouds and to calculate accurate cloud field tracks. In 2020, six new satellites from the Meteosat Third Generation will be launched, which will have operational scans down to 500 m resolution (Stuhlmann *et al.*, 2005).

Recent atmospheric models are still unable to fully resolve convective cloud processes in particular for large-domain simulations. This study demonstrates that even ICON-LEM with a native resolution of $\mathcal{O}(156\,\mathrm{m})$ is not able to fully resolve the spatio-temporal variability of convective clouds in the so-called grey zone as power spectra are indicating (Dorrestijn *et al.*, 2012). The effective model resolution has implications for the planning of high-resolution simulations, where cloud processes on a 1 km scale need to be simulated at 150 m grid size or less. The insights gained here into the scaling behaviour at different model resolutions can help to improve stochastic parametrizations of cumulus convection in future modelling studies, and to better compare observational datasets and model results with different resolutions. Additionally, there is still a great step to go towards a large-domain observational dataset with a similar spatial and temporal resolution like ICON-LEM. Using data from the HRV channel with 1 km spatial and 5 min temporal resolution has already demonstrated a reduction in the tracking uncertainties (Bley *et al.*, 2016). The authors are currently developing an extension of the cloud property retrieval to the high resolution of Meteosat.

Acknowledgements

Sebastian Bley's position was funded by the German Ministry of Education and Research (BMBF) under grant 01LK1210B. The study was done within the framework of the German-wide research initiative High Definition Clouds and Precipitation for Climate Prediction (HD(CP)^2). Fabian Senf acknowledges funding by the BMBF project HD(CP)^2 phase II under grant 01LK1507C. We thank Rieke Heinze and Ksenia Gorges for providing the ICON-LEM data, and for their support regarding the management of huge datasets as well as Christian Barthlott who provided the COSMO-DE reference runs. We further acknowledge EUMETSAT for proving Meteosat SEVIRI data. Many thanks also to our colleagues for the many thoughtful comments which helped to improve this manuscript. We thank three anonymous reviewers for their supportive and constructive comments.

References

Anuta PE. 1970. Spatial registration of multispectral and multitemporal digital imagery using fast fourier transform techniques. *IEEE Trans. Geosci. Electron.* **8**: 353–368. https://doi.org/10.1109/TGE.1970.271435.

Baldauf M, Seifert A, Förstner J, Majewski D, Raschendorfer M, Reinhardt T. 2011. Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. *Mon. Weather Rev.* **139**: 3887–3905. https://doi.org/10.1175/MWR-D-10-05013.1.

Benner TC, Curry JA. 1998. Characteristics of small tropical cumulus clouds and their impact on the environment. *J. Geophys. Res. Atmos.* **103**: 28753–28767. https://doi.org/10.1029/98JD02579.

Bley S, Deneke H, Senf F. 2016. Meteosat-based characterization of the spatio-temporal evolution of warm convective cloud fields over Central Europe. *J. Appl. Meteorol. Climatol.* **55**: 2181–2195. https://doi.org/10.1175/JAMC-D-15-0335.1.

Bony S, Stevens B, Frierson DMW, Jakob C, Kageyama M, Pincus R, Shepherd TG, Sherwood SC, Siebesma AP, Sobel AH, Watanabe M, Webb MJ. 2015. Clouds, circulation and climate sensitivity. *Nat. Geosci.* **8**: 261–268.

Brown AR. 1999. The sensitivity of large-eddy simulations of shallow cumulus convection to resolution and subgrid model. *Q. J. R. Meteorol. Soc.* **125**: 469–482. https://doi.org/10.1002/qj.49712555405.

Cho HM, Zhang Z, Meyer K, Lebsock M, Platnick S, Ackerman AS, Di Girolamo L, C-Labonnote L, Cornet C, Riedi J, Holz RE. 2015. Frequency and causes of failed MODIS cloud property retrievals for liquid phase clouds over global oceans. *J. Geophys. Res. Atmos.* **120**: 4132–4154. https://doi.org/10.1002/2015JD023161.

Deneke HM, Roebeling RA. 2010. Downscaling of METEOSAT SEVIRI 0.6 and 0.8 m channel radiances utilizing the high-resolution visible channel. *Atmos. Chem. Phys.* **10**: 9761–9772.

Dipankar A, Stevens B, Heinze R, Moseley C, Zängl G, Giorgetta M, Brdar S. 2015. Large-eddy simulation using the general circulation model ICON. *J. Adv. Model. Earth Syst.* **7**: 1942–2466. https://doi.org/10.1002/2015MS000431.

Dorrestijn J, Crommelin DT, Siebesma AP, Jonker HJJ. 2012. Stochastic parameterization of shallow cumulus convection estimated from high-resolution model data. *Theor. Comput. Fluid Dyn.* **27**: 133–148. https://doi.org/10.1007/s00162-012-0281-y.

Evan AT, Heidinger AK, Vimont DJ. 2007. Arguments against a physical long-term trend in global ISCCP cloud amounts. *Geophys. Res. Lett.* **34**: L04701. https://doi.org/10.1029/2006GL028083.

Han D, Ellingson RG. 1999. Cumulus cloud formulations for longwave radiation calculations. *J. Atmos. Sci.* **56**: 837–851. https://doi.org/10.1175/1520-0469.

Heinze R, Dipankar A, Carbajal Henken C, Moseley C, Sourdeval O, Trömel S, Xie X, Adamidis P, Ament F, Baars H, Barthlott C, Behrendt A, Blahak U, Bley S, Brdar S, Brueck M, Crewell S, Deneke H, Di Girolamo P, Evaristo R, Fischer J, Frank C, Friederichs P, Göcke T, Gorges K, Hande L, Hanke M, Hansen A, Hege HC, Hoose C, Jahns T, Kalthoff N, Klocke D, Kneifel S, Knippertz P, Kuhn A, van Laar T, Macke A, Maurer V, Mayer B, Meyer CI, Muppa SK, Neggers RAJ, Orlandi Pantillon E, Pospichal F, Röber B, Scheck N, Seifert L, Seifert A, Senf P, Siligam F, Simmer P, Steinke C, Stevens S, Wapler B, Weniger K, Wulfmeyer M, Zängl V, Zhang G, Quaas D. 2016. Large-eddy simulations over Germany using ICON: A comprehensive evaluation. *Q. J. R. Meteorol. Soc.* **143**: 69–100. https://doi.org/10.1002/qj.2947.

Heus T, Seifert A. 2013. Automated tracking of shallow cumulus clouds in large domain, long duration large-eddy simulations. *Geosci. Model Dev.* **6**: 1261–1273.

Hohenegger C, Brockhaus P, Schär C. 2008. Towards climate simulations at cloud-resolving scales. *Meteorol. Z.* **17**: 383–394. https://doi.org/10.1127/0941-2948/2008/0303.

Horvath A, Seethala C, Deneke H. 2014. View angle dependence of MODIS liquid water path retrievals in warm oceanic clouds. *J. Geophys. Res.* **119**: 8304–8328. https://doi.org/10.1002/2013JD021355.

Jiang H, Xue H, Teller A, Feingold G, Levin Z. 2006. Aerosol effects on the lifetime of shallow cumulus. *Geophys. Res. Lett.* **33**: L14806. https://doi.org/10.1029/2006GL026024.

Kassianov E, Long CN, Ovtchinnikov M. 2005. Cloud sky cover versus cloud fraction: Whole-sky simulations and observations. *J. Appl. Meteorol.* **44**: 86–98. https://doi.org/10.1175/JAM-2184.1.

Koren I, Oreopoulos L, Feingold G, Remer LA, Altaratz O. 2008. How small is a small cloud? *Atmos. Chem. Phys.* **8**: 3855–3864. https://doi.org/10.5194/acp-8-3855-2008.

Kostka PM, Weissmann M, Buras R, Mayer B, Stiller O. 2014. Observation operator for visible and near-infrared satellite reflectances. *J. Atmos. Oceanic Technol.* **31**: 1216–1233. https://doi.org/10.1175/JTECH-D-13-00116.1.

Lensky IM, Rosenfeld D. 2008. Clouds–Aerosols–Precipitation Satellite Analysis Tool (CAPSAT). *Atmos. Chem. Phys.* **8**: 6739–6753.

Marshak A, Platnick S, Várnai T, Wen G, Cahalan RF. 2006. Impact of three-dimensional radiative effects on satellite retrievals of cloud droplet sizes. *J. Geophys. Res.* **111**: D09207. https://doi.org/10.1029/2005JD006686.

Mechem DB, Yuter SE, deSzoeke SP. 2012. Thermodynamic and aerosol controls in southeast Pacific stratocumulus. *J. Atmos. Sci.* **69**: 1250–1266. https://doi.org/10.1175/JAS-D-11-0165.1..

Merk D, Deneke H, Pospichal B, Seifert P. 2016. Investigation of the adiabatic assumption for estimating cloud micro- and macrophysical properties from satellite and ground observations. *Atmos. Chem. Phys.* **16**: 933–952. https://doi.org/10.5194/acp-16-933-2016.

Minnis P. 1989. Viewing zenith angle dependence of cloudiness determined from coincident GOES East and GOES West data. *J. Geophys. Res. Atmos.* **94**: 2303–2320. https://doi.org/10.1029/JD094iD02p02303.

Nagasawa R, Iwasaki T, Asano S, Saito K, Okamoto H. 2006. Resolution dependence of non-hydrostatic models in simulating the formation and evolution of low-level clouds during a 'Yamase' event. *J. Meteorol. Soc. Jpn. Ser. II* **84**: 969–987. https://doi.org/10.2151/jmsj.84.969.

Neggers RAJ, Jonker HJJ, Siebesma AP. 2003. Size statistics of cumulus cloud populations in large-eddy simulations. *J. Atmos. Sci.* **60**: 1060–1074. https://doi.org/10.1175/1520-0469(2003)60$⟨$1060:SSOCCP$⟩$2.0.CO;2.

Roebeling RA, Feijt AJ, Stammes P. 2006. Cloud property retrievals for climate monitoring: Implications of differences between Spinning Enhanced Visible and Infrared Imager (SEVIRI) on METEOSAT-8 and Advanced Very High Resolution Radiometer (AVHRR) on NOAA-17. *J. Geophys. Res.* **111**: D20210. https://doi.org/10.1029/2005JD006990.

Scheck L, Frérebeau P, Buras-Schnell R, Mayer B. 2016. A fast radiative transfer method for the simulation of visible satellite imagery. *J. Quant. Spectrosc. Radiat. Trans.* **175**: 54–67. https://doi.org/10.1016/j.jqsrt.2016.02.008.

Schulz J, Albert P, Behr H-D, Caprion D, Deneke H, Dewitte S, Dürr B, Fuchs P, Gratzki A, Hechler P, Hollmann R, Johnston S, Karlsson K-G, Manninen T, Müller R, Reuter M, Riihelä A, Roebeling R, Selbach N, Tetzlaff A, Thomas W, Werscheck M, Wolters E, Zelenka A. 2009. Operational climate monitoring from space: The EUMETSAT satellite application facility on climate monitoring (CM-SAF). *Atmos. Chem. Phys.* **9**: 1687–1709. https://doi.org/10.5194/acp-9-1687-2009.

Seifert A, Beheng DK. 2005. A two-moment cloud microphysics parameterization for mixed-phase clouds. Part 1: Model description. *Meteorol. Atmos. Phys.* **92**: 45–66. https://doi.org/10.1007/s00703-005-0112-4.

Senf F, Dietzsch F, Hünerbein A, Deneke H. 2015. Characterization of initiation and growth of selected severe convective storms over Central Europe with MSG-SEVIRI. *J. Appl. Meteorol. Climatol.* **54**: 207–224. https://doi.org/10.1175/JAMC-D-14-0144.1.

Stevens DE, Ackerman AS, Bretherton CS. 2002. Effects of domain size and numerical resolution on the simulation of shallow cumulus convection. *J. Atmos. Sci.* **59**: 3285–3301. https://doi.org/10.1175/1520-0469(2002)059$⟨$3285:EODSAN$⟩$2.0.CO;2.

Stuhlmann R, Rodriguez A, Tjemkes S, Grandell J, Arriaga A, Bézy JL, Aminou D, Bensi P. 2005. Plans for EUMETSATs Third Generation Meteosat geostationary satellite programme. *Adv. Space Res.* **36**: 975–981. https://doi.org/10.1016/j.asr.2005.03.091.

Várnai T, Marshak A. 2007. View angle dependence of cloud optical thicknesses retrieved by Moderate Resolution Imaging Spectroradiometer (MODIS). *J. Geophys. Res. Atmos.* **112**: D06203. https://doi.org/10.1029/2005JD006912.

Weisman ML, Skamarock WC, Klemp JB. 1997. The resolution dependence of explicitly modeled convective systems. *Mon. Weather Rev.* **125**: 527–548. https://doi.org/10.1175/1520-0493(1997)125$⟨$0527:TRDOEM$⟩$2.0.CO;2.

Wolters ELA, Deneke HM, van den Hurk BJJM, Meirink JF, Roebeling RA. 2010. Broken and inhomogeneous cloud impact on satellite cloud particle effective radius and cloud-phase retrievals. *J. Geophys. Res.* **115**: D10214. https://doi.org/10.1029/2009JD012205.

Wood R, Field PR. 2011. The distribution of cloud horizontal sizes. *J. Clim.* **24**: 4800–4816. https://doi.org/10.1175/2011JCLI4056.1.

Zängl G, Reinert D, Rípodas P, Baldauf M. 2015. The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Q. J. R. Meteorol. Soc.* **141**: 563–579. https://doi.org/10.1002/qj.2378.

Zhao G, Di Girolamo L. 2007. Statistics on the macrophysical properties of trade wind cumuli over the tropical western Atlantic. *J. Geophys. Res.* **112**: D10204. https://doi.org/10.1029/2006JD007371.