Supporting Information for
"Increasing Resolution and Resolving Convection Improves the Simulation of Cloud-Radiative Effects over the North Atlantic"

Fabian Senf[1], Aiko Voigt[2,3], Nicolas Clerbaux[4], Anja Hünerbein[1], Hartwig Deneke[1]

[1]Leibniz Institute for Tropospheric Research, Leipzig
[2]Institute for Meteorology and Climate Research - Department Troposphere Research, Karlsruhe Institute of Technology, Karlsruhe
[3]Lamont-Doherty Earth Observatory, Columbia University, New York, USA
[4]Royal Meteorological Institute of Belgium, Brussels

1 Cloud Classification with NWCSAF v2013

1.1 Adjustments for Permanent Night Mode

For cloud classification, we apply the NWCSAF software v2013 (Derrien & Le Gléau, 2005). We keep the software itself unmodified and implement all changes via an interface that controls the input files and the execution of the software package. Within that interface, observed and synthetic infrared BTs are read from disk space and written into Meteosat SEVIRI HRIT template files (the native data format distributed by EUMETSAT). The template files themselves are valid for 0 UTC, but data embedded into the template files can have any time stamp. The NWCSAF software retrieves night-time cloud classifications independent of the actual time stamp of the embedded input data.

The Metetosat SEVIRI imager measures radiances, expressed in terms of brightness temperatures (BTs), in several channels. One of them, the 3.9 μm channel, is affected by shortwave as well as longwave radiation (see e.g. Lindsey et al., 2006). During night-time, the use of the 3.9-μm brightness temperature is beneficial for detecting clouds and their microphysical characteristics at their top (Lensky & Rosenfeld, 2003). Therefore, the 3.9 μm channel is mandatory for the NWCSAF cloud classification at night-time. Because we aim to feed the NWCSAF software with both night-time and day-time scenes, 3.9-μm radiances can be contaminated by sunlight, which might lead to erroneous cloud classifications by the NWCSAF software.

We mitigate this problem in the simplest possible way: we estimate 3.9-μm BT from BTs of the other infrared channels by means of a linear regression derived from a least-squares fit,

$$T_{3.9} = T_{10.8} + a_0 + a_1 \Delta T_{8.7-10.8} + a_2 \Delta T_{10.8-12.0} + a_3 \Delta T_{13.4-10.8} . \tag{1}$$

T_i is the infrared BT of a SEVIRI channel with central wave length i μm, and ΔT_{i-j} is the BT difference of two SEVIRI channels centered at i and j μm. The regression was applied to observed SEVIRI data over the North Atlantic analysis domain and each 0 UTC time slot within the whole NAWDEX period. The resulting average regression parameters of $(a_0, a_1, a_2, a_3) = (3\ \text{K}, 1.8, 1.5, 0.12)$ are then used to estimate the 3.9-μm BT, which is then fed into the NWCSAF software instead of observed or simulated values. The approximation of the 3.9 μm channel is done for both the SEVIRI observations and the synthetic observations derived from the ICON simulations with the SynSat forward operator.

Corresponding author: Fabian Senf, `senf@tropos.de`

The linear regression gives acceptable results, as described in more detail in the following section. Testing against observed 3.9-μm BTs at 0 UTC, explained variances are always above 99%, average biases are ~ 0.1 K and average RMSEs are below 2 K. We note that the current method is limited to ocean regions. For surfaces with a more heterogeneous surface emissivity, e.g., in the Saharan region, a more sophisticated approach would be needed.

1.2 Evaluation of NWCSAF Cloud Classification in Night Mode

As described above, we estimate 3.9-μm BTs from other channels' BTs. The extent to which this degrades the NWCSAF cloud classification is analysed below for 0 UTC at night time (see Tables S1 and S2). We use a pixel-based comparison and a binary verification concept in which a certain cloud type is considered to be present or not. Using the 2×2 contingency tables (see Wilks, 2006, p.260), five verification metrics are considered: proportion correct (PC, Wilks eq. 7.7), critical success index (CSI, Wilks eq. 7.8), BIAS (Wilks eq. 7.9), probability of detection (POD, Wilks eq. 7.12) and false alarm rate (FAR, Wilks eq. 7.13). The verification performs best if PC, CSI, BIAS and POD have values of 1m and FAR is zero. For each verifcation metric, the largest absolute deviation from these optimal values is marked in bold. In general, the performance of NWCSAF run in night-mode is very good. The degradation is strongest for fractional, very low and semi-transparent thin clouds, but even for these the performance is satisfactory. We conclude that the modified NWCSAF software will provide a robust cloud classification that can be used to assess differences between observations and simulations.

A comparison between cloud classification by our degraded NWCSAF night-time approach mode and the standard day-time NWCSAF approach is shown in Tab. S3 and S4. For the latter also solar SEVIRI channels have been used. The comparison thus shows the overall information loss when only thermal infrared BTs can be used. The verification scores are substantially worse than above. This means (i) there is a significant change in detection quality during the course of the day (which we tried to minimize with our permanent night-mode setup), and (ii) the solar channels help a lot during daytime. Again, the most affected cloud type is "fractional", followed by "very low" and "semi-transparent". The "semi. above" class is not assigned during night-time.

Table S1. Evaluation table for NWCSAF cloud masking at 0 UTC. The evaluation is based on 25 days from 20 September to 14 October, 2016.

Cloud Mask Type	$f_{degraded}$ (%)	$f_{standard}$ (%)	PC	CSI	BIAS	POD	FAR
clear	33.4	32.6	**0.967**	**0.905**	1.024	0.961	**0.030**
partially cloudy	18.7	19.5	0.989	0.945	**0.960**	**0.952**	0.002
cloudy	47.9	47.9	0.977	0.953	1.000	0.976	0.022

Table S2. Evaluation table for NWCSAF cloud masking at 0 UTC. The evaluation is based on 25 days from 20 September to 14 October, 2016.

Cloud Type	$f_{degraded}$ (%)	$f_{standard}$ (%)	PC	CSI	BIAS	POD	FAR
very low	19.5	19.4	0.975	0.881	1.002	0.937	**0.016**
low	10.3	9.8	0.980	0.817	1.047	0.921	0.014
mid-level	4.0	4.2	0.995	0.881	0.958	0.917	0.002
high opaque	6.8	6.8	1.000	1.000	1.000	1.000	0.000
very high opaque	1.2	1.2	1.000	1.000	1.000	1.000	0.000
semi. thin	5.0	5.1	0.986	0.759	0.984	0.856	0.007
semi. moderately thick	9.2	8.9	0.991	0.907	1.035	0.967	0.007
semi. thick	2.0	2.0	1.000	1.000	1.000	1.000	0.000
semi. above	0.0	0.0	1.000	-	-	-	0.000
fractional	8.6	9.9	**0.961**	**0.654**	**0.866**	**0.738**	0.014

Table S3. Evaluation table for NWCSAF cloud masking at 12 UTC. The evaluation is based on 25 days from 20 September to 14 October, 2016.

Cloud Mask Type	f_{degraded} (%)	f_{standard} (%)	PC	CSI	BIAS	POD	FAR
clear	26.4	34.9	**0.892**	**0.701**	**0.758**	**0.724**	0.018
partially cloudy	21.7	20.9	0.967	0.857	1.040	0.941	0.026
cloudy	51.5	44.2	0.919	0.845	1.165	0.992	**0.138**

Table S4. Evaluation table for NWCSAF cloud typing at 12 UTC. The evaluation is based on 25 days from 20 September to 14 October, 2016.

Cloud Type	f_{degraded} (%)	f_{standard} (%)	PC	CSI	BIAS	POD	FAR
very low	17.6	16.6	0.919	0.617	1.063	0.787	0.055
low	8.6	12.6	0.959	0.678	0.685	0.681	0.001
mid-level	3.8	4.7	0.991	0.814	0.815	0.814	0.000
high opaque	6.6	6.7	0.999	0.986	0.987	0.986	0.000
very high opaque	1.4	1.4	1.000	0.995	0.999	0.997	0.000
semi. thin	5.4	2.7	0.960	0.342	1.955	0.753	0.034
semi. moderately thick	9.1	3.0	0.939	0.329	**3.038**	1.000	**0.063**
semi. thick	2.3	2.2	0.999	0.955	1.047	1.000	0.001
semi. above	0.0	4.3	0.957	-	-	-	-
fractional	10.3	19.0	**0.825**	**0.252**	0.540	**0.310**	0.054

2 On the Bias in Simulated Clear-Sky Radiation Fluxes

Here, we provide further and more detailed information on the bias correction of simulated clear-sky radiation fluxes. It has been discussed in the main part that it is challenging to derived accurate estimates for observed clear-sky fluxes, especially due to the high cloud coverage and the rather low cloud-free fraction found in our analysis domain. For that reason we decided to use simulated clear-sky radiation fluxes as substitute for observed clear-sky fluxes. However, systematic biases in simulated fluxes need to characterized and corrected.

In the following, we consider longwave fluxes and skip the subscript "lw", but the same also applies to upwelling shortwave fluxes with the subscript "sw, up". We assume that simulated radiation fluxes have a systematic bias B and a random error ε, i.e.

$$F_{\text{ICON,clear}} = F_{\text{OBS,clear}} + B + \varepsilon \qquad (2)$$

After statistical averaging, the contribution of the random error ε gets smaller and less important. Thus, the bias B can be estimated from the average difference between simulated and observed clear-sky fluxes. The observed clear-sky flux $F_{\text{OBS,clear}}$ is however unknown. Combining cloud detection (or detection of cloud-free regions) and observational flux estimates, all-sky fluxes in cloud-free regions $F_{\text{OBS,0}}$ can derived. In addition to the clear-sky information these fluxes contain the radiative effect of undetected clouds, i.e.

$$F_{\text{OBS,clear}} = F_{\text{OBS,0}} + \Delta F_{\text{OBS}} \qquad (3)$$

The term ΔF_{OBS} characterizes our cloud detection capabilities. It is thus a characteristic property of the cloud classification algorithm. As we also derive a cloud classification based on simulations (in a very consistent way), we are able to estimate the average magnitude of the radiative effect of undetected clouds in ICON simulations as

$$F_{\text{ICON,clear}} = F_{\text{ICON,0}} + \Delta F_{\text{ICON}}. \qquad (4)$$

Both, $F_{\text{ICON,clear}}$ and $F_{\text{ICON,0}}$ are known and ΔF_{ICON} can be derived (see Fig. 5 in the main part). If we assume that the radiative effects of undetected clouds have similar magnitudes in simulations and observations, i.e. $\Delta F_{\text{OBS}} \approx \Delta F_{\text{ICON}}$, a bias correction

$$B = F_{\text{ICON,clear}} - (F_{\text{OBS,0}} + \Delta F_{\text{SIM}}) \qquad (5)$$
$$= F_{\text{ICON,0}} - F_{\text{OBS,0}} \qquad (6)$$

can be derived. This means that if a bias correction is found that adjusts differences in observed and simulated all-sky fluxes in cloud-free regions, this is equivalent to a bias correction that adjusts the radiative effects of undetected clouds in simulations and observations. For upwelling shortwave clear-sky fluxes, we applied a scaling factor and for longwave clear-sky fluxes an offset is added.

3 Additional Data Overview

We provide three additional overview plots to supplement the figures shown in the main part of the manuscript. BTs from window channel at 10.8 μm are shown in Fig. S1, BTs from the water vapor channel at 6.2 μm are shown in Fig. S2 and the dependence of cloud typing on grid spacing is visualized in Fig. S3.
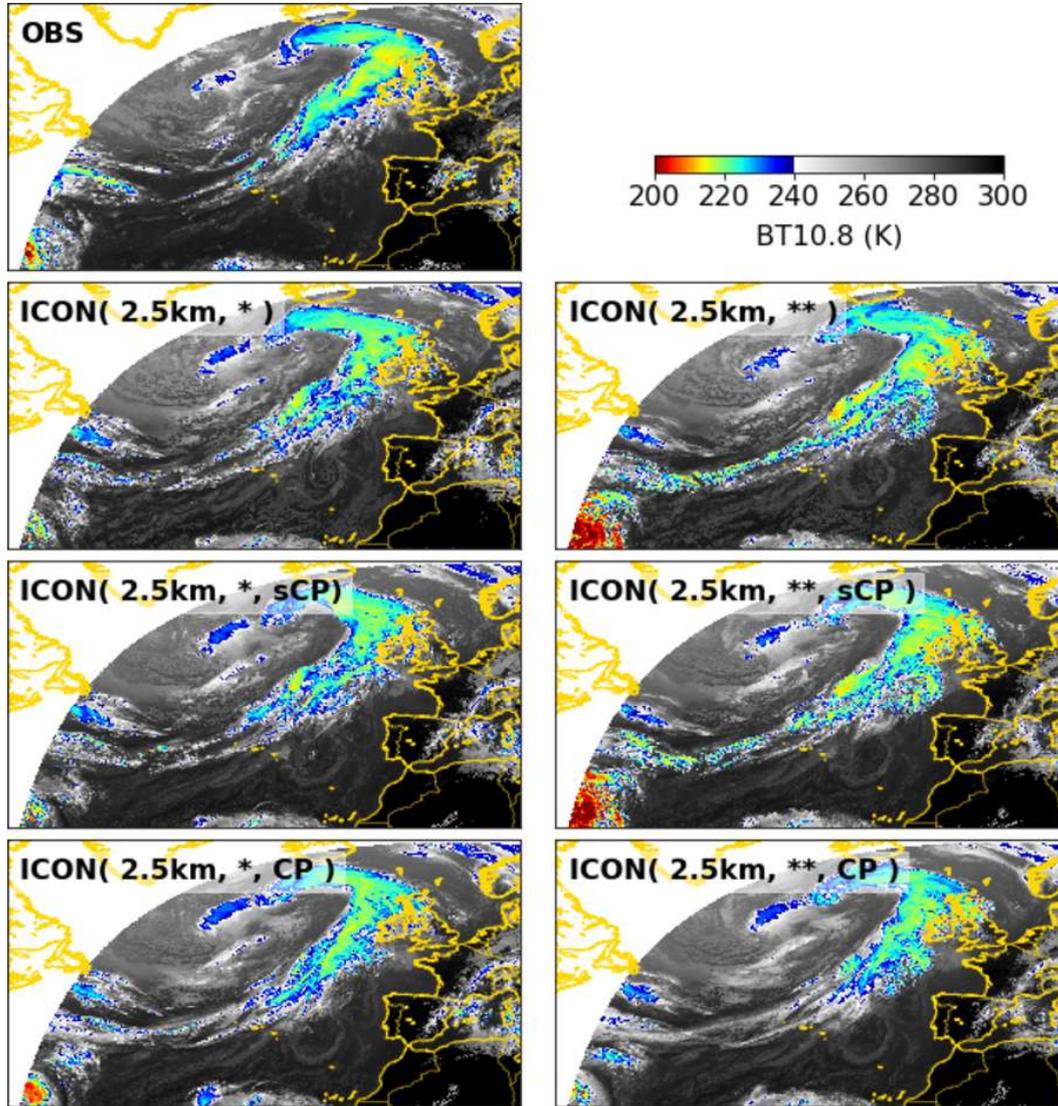

Figure S1. Overview of observed and simulated BTs from Meteosat SEVIRI's window channel at 10.8 μm for 1200 UTC 23 Sept 2016. Meteosat SEVIRI observations (top left) are compared ICON simulations with 2.5 km horizontal resolution. The left column is for simulations with one-moment cloud microphysics (*), the right column for simulations with two-moment microphysics (**). The second row is for fully explicit convection, the third row for simulations with a shallow convection scheme (sCP), and the fourth row for simulations with fully parameterized convection (CP). A special color scheme is used to highlight observed and simulated features. BTs over land are also shown to improve anticipation of the cloud scenery.
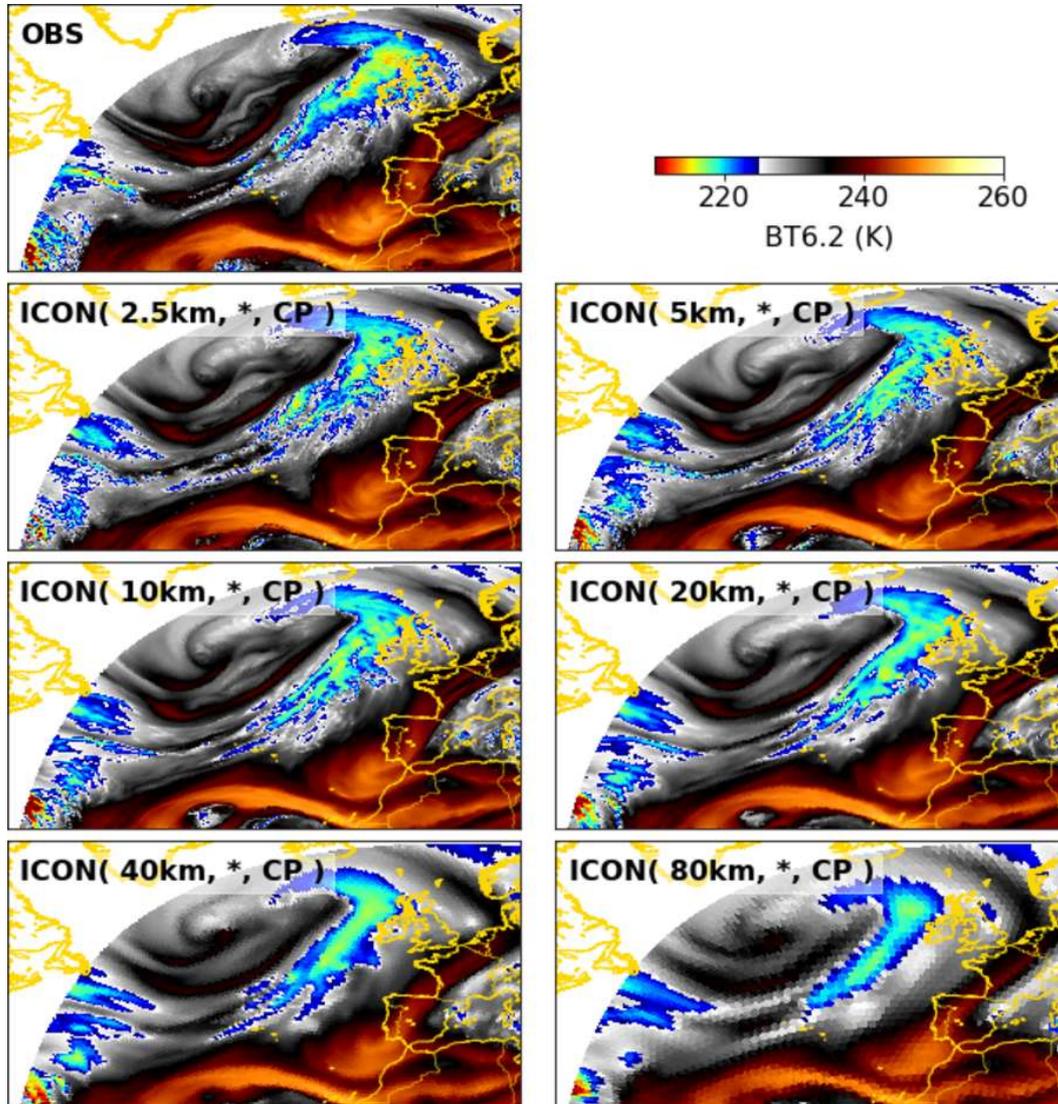
Figure S2. Overview of observed and simulated BTs from Meteosat SEVIRI's water vapor channel at 6.2 μm for 1200 UTC 23 Sept 2016. Observations are compared to ICON simulations with increasing grid spacing (left to right and downwards, from 2.5 to 80 km). Only the subset of simulations with one-moment microphysics and fully-parameterized convection is chosen for visualization. A special color scheme is used to highlight observed and simulated features. BTs over land are also shown to improve anticipation of the cloud scenery.
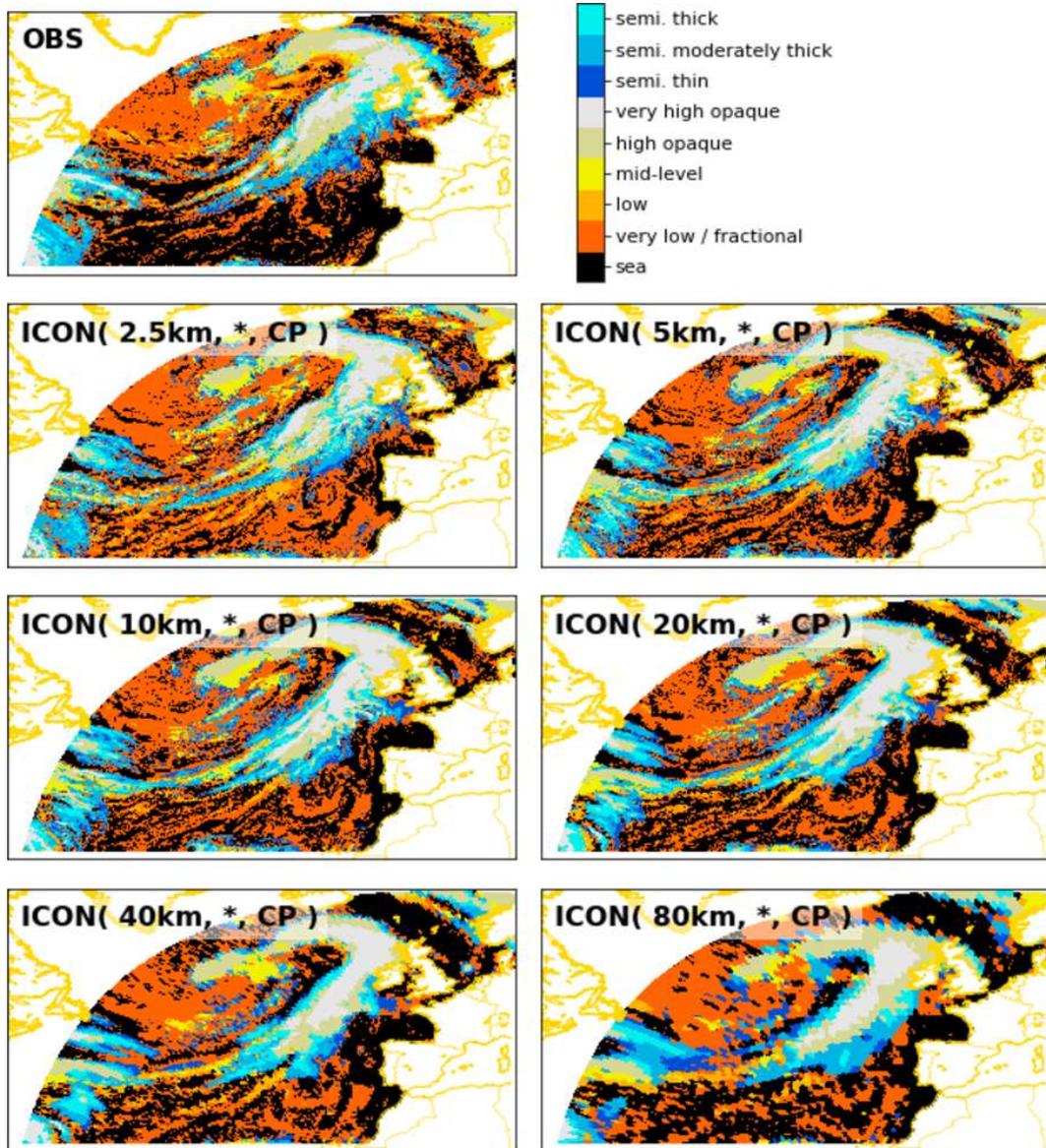
Figure S3. Overview of observed and simulated cloud types for 1200 UTC 23 Sept 2016. Cloud classification based on Meteosat SEVIRI (top row) is compared to cloud classification based on synthetic radiances derived from the ICON simulations and RTTO. The same simulations are shown as in Fig. S2.

3.1 Supplementary Data Analysis with Special Emphasis on Set 2

The following Figs. S4-S9 provides plots from additional analysis which support the arguments and conclusion made in the main part of the manuscript. Six additional numerical experiments have been performed in simulation set 2. These include ICON simulations with 2.5 km grid spacing and fully enabled convection parameterization and ICON simulations with 5 and 10 km grid spacing with only explicit convection (fully disabled convection scheme). Two main conclusion can be derived from the additional data analysis:

(i) For simulations with fully explicit convection, biases in radiation fluxes and cloud-radiative effects are reduced when the grid spacing is sequentially brought down from 10 to 2.5 km. The simulations at 2.5 km do not seem to have reached a stage where signatures of convergence can be identified. Further reduction in grid spacing is needed.

(ii) The simulation with parameterized convection and 2.5 km has similar error characteristics then its coarser counterparts. This means that difference in e.g. `ICON( 2.5km, *)` and `ICON( 5km, *, CP)` which are discussed in the main part of the manuscript are not due to difference in grid spacing.
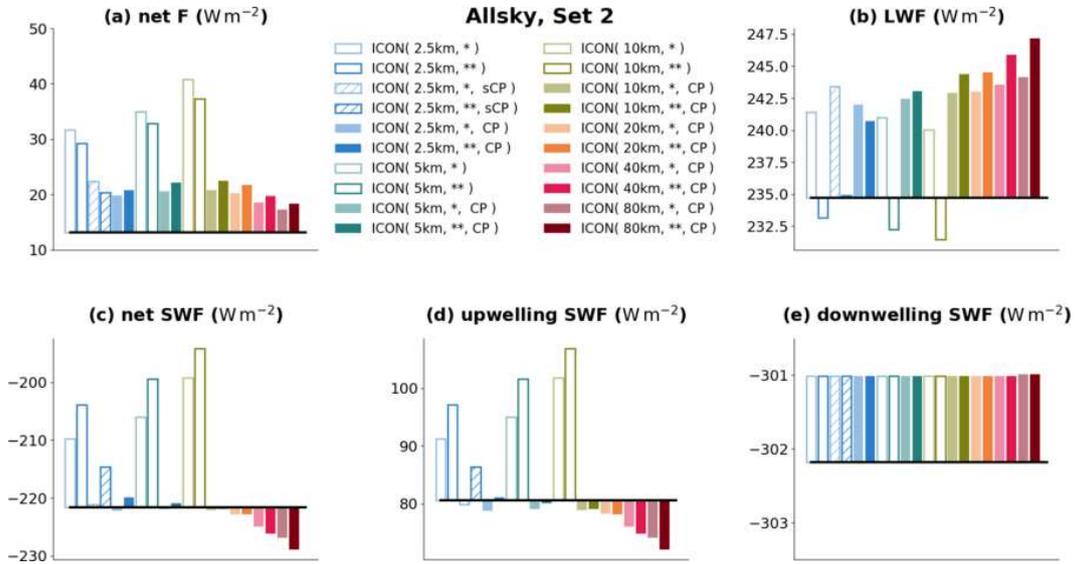

Figure S4. Analysis of domain-average allsky radiation fluxes: (a) total net flux, (b) emitted longwave flux, (c) net shortwave flux, (d) upwelling shortwave flux, and (e) downwelling shortwave flux. The Meteosat observations (black line) were chosen as reference, and deviation of simulated fluxes are shown with colored bars. The simulation experiments differ with regard to horizontal grid spacing (2.5 5, 10, 20, 40 and 80 km), and parameterization choice (one-moment vs. two-moment microphysics, with vs. without convection parameterization scheme). All values represent time averages over 3 days from simulation set 2.
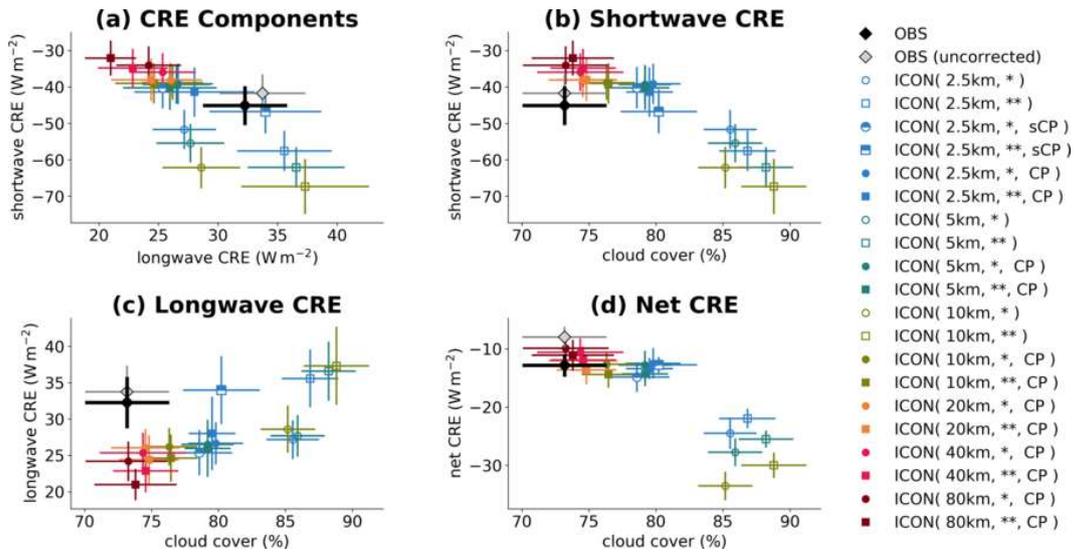
Figure S5. Comparison of domain-average allsky cloud-radiative effects and total cloud cover: (a) longwave CRE vs. shortwave CRE, and cloud cover vs. (b) shortwave CRE, (c) longwave CRE, and (d) net CRE. Symbols and error bars represent average and estimates of standard errors, respectively. With different colors and symbols styles different simulations experiments are distinguished. Please note the differences in the y-axis ranges. All values represent time averages over 3 days from simulation set 2.
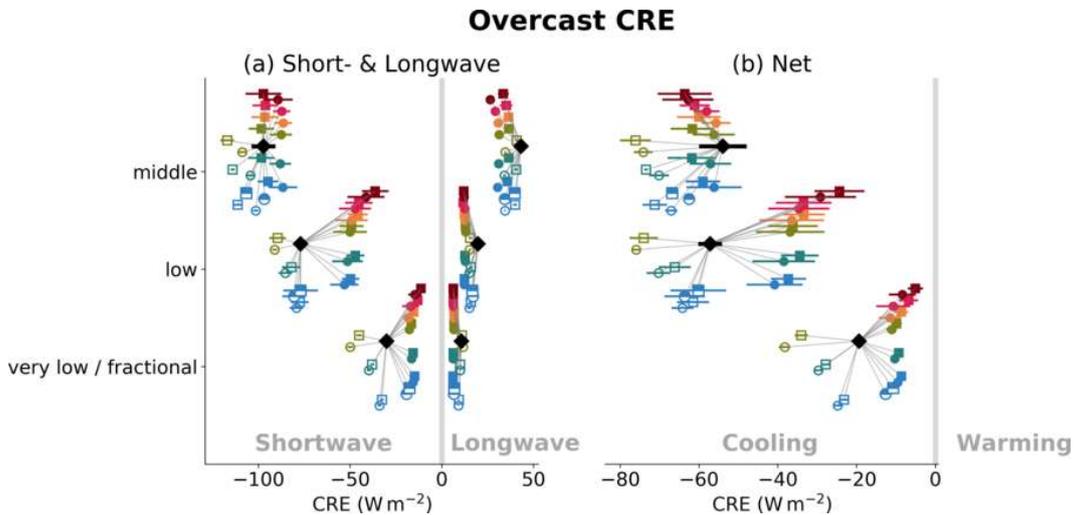
Figure S6. Overcast CRE for different shallow cloud types. A legend for color and symbols can be found in Fig. S5. All values represent time averages over 3 days from simulation set 2.

Overcast CRE Biases
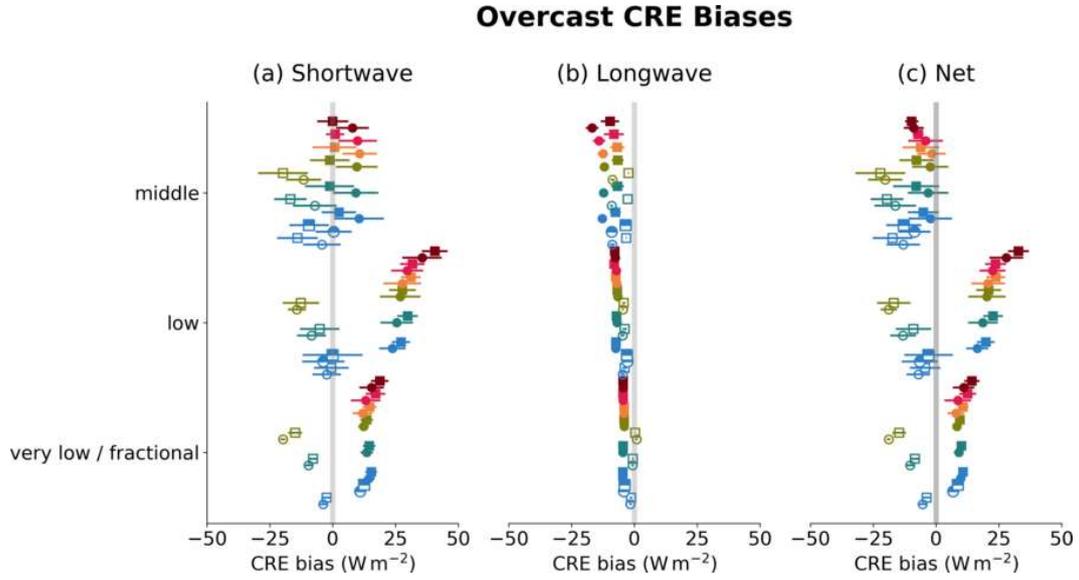
Figure S7. Biases in overcast CRE for different shallow cloud types. A legend for color and symbols can be found in Fig. S5. All values represent time averages over 3 days from simulation set 2.
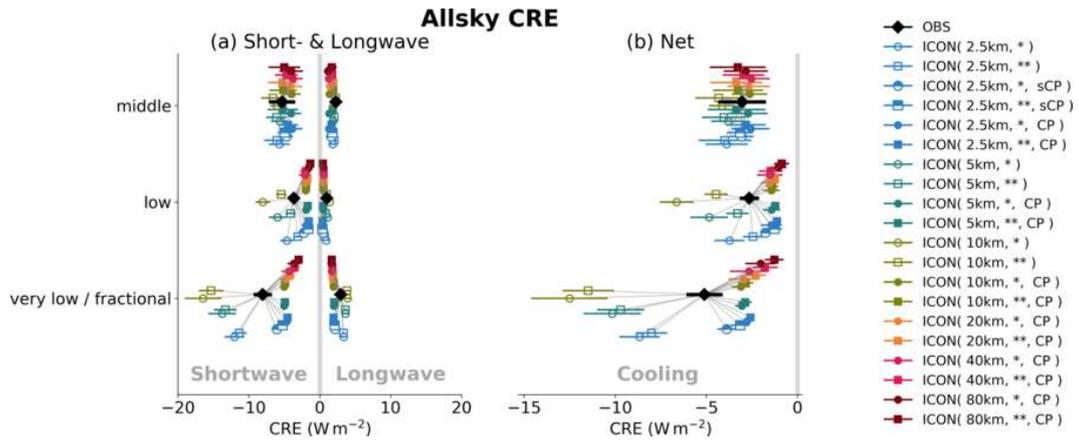

Figure S8. Allsky CRE for different shallow cloud types. A legend for color and symbols can be found in Fig. S5. All values represent time averages over 3 days from simulation set 2.
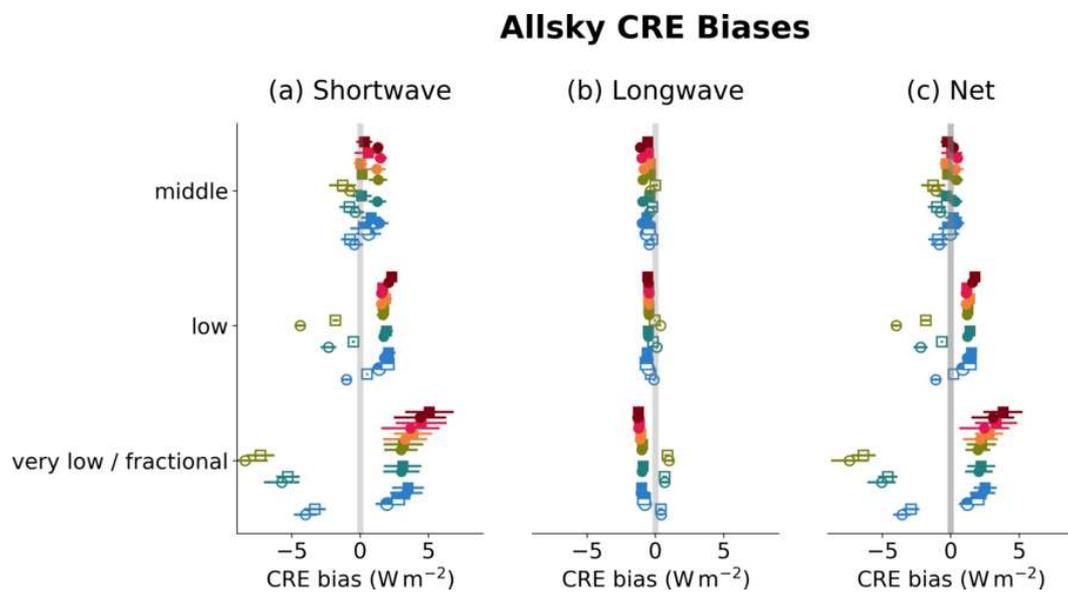
Figure S9. Biases in allsky CRE for different shallow cloud types. A legend for color and symbols can be found in Fig. S5. All values represent time averages over 3 days from simulation set 2.

References

Derrien, M., & Le Gléau, H. (2005). MSG/SEVIRI cloud mask and type from SAFNWC. *Int. J. Remote Sens.*, *26*, 4707-4732. doi: 10.1080/01431160500166128

Lensky, I. M., & Rosenfeld, D. (2003). Satellite-based insights into precipitation formation processes in continental and maritime convective clouds at nighttime. *J. Appl. Meteor.*, *42*(9), 1227 - 1233.

Lindsey, D. T., Hillger, D. W., Grasso, L., Knaff, J. A., & Dostalek, J. F. (2006). GOES Climatology and Analysis of Thunderstorms with Enhanced 3.9-μm Reflectivity. *Mon. Wea. Rev.*, *134*(9), 2342-2353.

Wilks, D. S. (2006). *Statistical methods in the atmospheric sciences* (Vol. 100). Academic Press.