We'd like to thank the editor for handling our manuscript, as well as reviewer #2 for reading our manuscript and providing numerous, helpful comments. We have carefully read through all the comments and questions and revised the manuscript accordingly. Please find our point-to-point response to reviewer #2 below. Here, the reviewer's general remarks are formatted to be left-aligned text in italic font, the specific questions/comments are shown in left-aligned text in bold and italic font, while our responses are indented and formatted in regular font.

Here is a summary of the major changes in the revised manuscript:

1) We rewrote the abstract to better summarize the results for r_{eff}, W_L and N_D.
2) We rewrote the introduction so the connection to the study by *Deneke and Roebeling* (2010) becomes clearer.
3) We added a paragraph about the difference between spatial and optical resolution and made clear that we account for this difference by means of the modulation transfer function.
4) We removed Figures 4 and 6 and the respective text describing it.
5) In Section 6 we focus on 3 downscaling schemes (instead of 5) and simplified the experiment designations from "1a, 1b, 2a, 2b, 3a, 3b, 3c, and 3d" to the simpler naming scheme of "1, 2, and 3".
6) In section 6 we removed some of the statistical measures (i.e., the percentiles of retrieval differences) to simplify the analysis and focus on 4 statistical measures only.
7) We found a small bug in Figures 10 and 12 (now 8 and 10), where the nRD was normalized twice (and the factor 100 for the calculation of percentages) was applied twice. Naturally, this only affects the values, but not the interpretation.
8) We moved the VNIR-only versus full downscaling approach to its own Section.
9) The values for Table 2 also slightly changed (as for point 7, this did not affect the interpretation of results), as there was an additional filter applied that was not needed.
10) We rewrote parts of the conclusions and added more interpretation instead of just summarizing the findings of Section 6.

My opinion is that this manuscript presents significant work well worth publishing. The key achievement lies the development and testing of methods for using geostationary satellite data to obtain cloud properties at a three times higher spatial resolution than the current standard. The methodology is sound, and the presentation is generally clear. I recommend a number of minor refinements (mainly to improve clarity), but there is one issue I'd like to single out in particular.

The text says throughout the manuscript (starting with Lines 3-4 of the abstract) that the proposed methods can increase the spatial resolution of SEVIRI cloud products from 3 km to 1 km (from the resolution of most SEVIRI bands to the resolution of the SEVIRI HRV band). My understanding, however, is that the resolution of SEVIRI observations is 3 km and 1 km only at the sub-satellite point, and that this resolution degrades with the cosine of the viewing zenith angle. (See, for example, http://www.esa.int/esapub/bulletin/bullet111/chapter4_bul111.pdf or http://www.icare.univ-lille1.fr/projects/seviri-aerosols.) For the test area around Germany, this can increase the meridional extent of SEVIRI pixels by 40% or more. For the most part, considering this effect would require only a clarification in the text; the only part where this becomes a substantial issue is the comparison with MODIS data. Considering that the meridional resolution of MODIS images should remain around 1 km even if the SEVIRI resolution became 40% coarser, it could be more appropriate to use a larger (e.g., 4 X 3) array of 1 km-size MODIS pixels to cover a coarse-resolution SEVIRI pixel. My own guess is that a such modification would not bring substantial changes to the overall outcomes (e.g., it would not change which method is deemed best), and I am not certain that considering the exact pixel sizes and using 4X3 arrays of MODIS pixels would yield more appropriate comparisons to 3X3 arrays of SEVIRI HRV pixels. Even so, it seems important to clarify in the manuscript the actual SEVIRI resolution around Germany, and to discuss any limitations or problems the different pixel sizes may introduce into the comparison of small-scale variability in SEVIRI and MODIS data.

The reviewer is absolutely correct. The actual spatial resolution is dependent on the viewing geometry and thus on geolocation. By sticking with the simplified description of 3x3 km^2 and 1x1 km^2 we tried to make the manuscript less confusing, but apparently achieved the opposite.

Statistics of pixel size for the Germany domain are shown in Figure 1 of this reply. The 3x3 km^2 pixels are closer to 6.2x3.2 km^2, while the higher—resolution pixels cover an average area of 2.1x1.1 km^2. However, the factor 3 between the spatial resolutions of channels 1-3 and the HRV channel remain. Similar stretching is observed for the MODIS pixels of the four example scenes. For scene 1 pixels are 1.5x2.4 km^2 large (comparable to the SEVIRI HRV resolution), while the other scenes are characterized by 1.1x1.2 km^2 pixels.

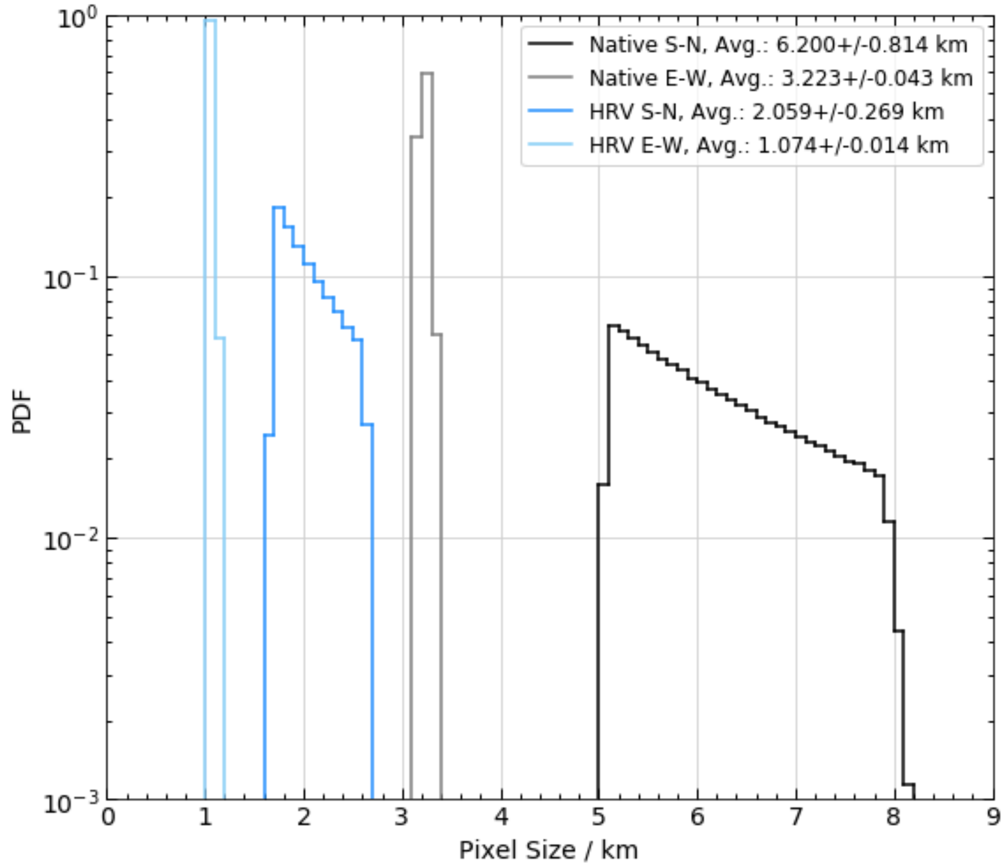

Fig 1: Statistics of SEVIRI pixel dimensions (in both latitude and longitude direction; i.e., south-north and east-west) for the native and HRV resolutions.

With regard to the evaluation of the downscaling techniques, these differences have no effect. There are two reasons for that: (i) We do not aggregate/colocate the MODIS data on the SEVIRI geometry. Instead, we first interpolate the MODIS reflectances on a higher-resolution grid and subsequently re-map these higher-resolution samples with the help of the sensor characteristics and open-source gdal libraries. (ii) We do not compare SEVIRI to MODIS; in fact, the actual values of the re-mapped MODIS reflectances are not important. They simply serve as a ground-truth for SEVIRI r_{06}, r_{08} and r_{16} reflectances at the HRV geometry, which is subsequently degraded (using the SEVIRI spatial response characteristics) by means of the same Fourier transforms (i.e., trigonometric interpolation) we describe throughout the manuscript. In other words, we degrade a ground-truth according to the SEVIRI characteristics and subsequently try to replicate the ground-truth again by means of the different downscaling techniques.

An actual comparison between downscaled SEVIRI and operational MODIS results is presented in the companion paper, which will be submitted by the end of January 2019 (this paper will also present other applications for this high-res SEVIRI data set). Here, we are just interested in finding a suitable technique.

We decided on a number of changes for the revised manuscript.

- We added the "≈" Symbol to the pixel scales in the abstract.
- We added "at the sub-satellite point and increases with higher sensor zenith angles" at the SEVIRI instrument description.
- We added a paragraph to the domain description in Section 2.3: "Due to the increased sensor zenith angles the spatial resolution of each SEVIRI pixel is degraded. The average pixel size is $6.20 \times 3.22\ km^2$ and $2.06 \times 1.07\ km^2$ for channels 1–3 and the HRV channel, respectively. To avoid confusion, we will use the designations LRES (abbreviation for lower–resolution) and HRES (abbreviation for higher–resolution) scales to refer to the $3{\times}3km^2$ and$1{\times}1km^2$ pixel resolutions from here on. "
- We replaced all other mentions of $1x1\ km^2$ and $3x3\ km^2$ with LRES and HRES abbreviations, or descriptive explanations. This should help avoid possible confusions by the reader.

Additional suggestions for minor revisions are listed below:

Page 3, Line 4: The resolution of 2.1 μm MODIS data is 500 m (and not 1 km).

Thanks for noticing this mistake. We corrected that error and it now says: "250 m horizontal resolution versus 500 m for the 0.6 μm and 2.1 μm channels, respectively"

Page 5, Line 23: It could help to clarify that the subscripts 06, 08, and 16 indicate 0.6 μm, 0.8μm, and 1.6 μm.

We actually mention that in the SEVIRI description in section 2.1, where it says: "The two VNIR reflectances (r06 and r08) are sampled in bands 1 and 2, respectively, and are centered around wavelengths $\lambda = 0.635\mu m$ and $\lambda = 0.810\mu m$. SWIR reflectances (r16) are provided by channel 3 observations, which are centered around $\lambda = 1.640\ \mu m$."

Page 6, Lines 11-12: I suggest starting the paragraph with something like "As is it discussed in Section 4,", just so readers know they will be able to learn about the exact estimation methods later on.

We added the following before that paragraph: "As is discussed in sections 4.1–4.4, the derived reflectances…"

Page 6, Line 14: For added clarity, I suggest inserting "latter" in front of "variables".

We added the word "latter", as suggested.

Page 8, Lines 5-10: It would be interesting to add a few words about what may cause the variations in c. For example, could it be variations in typical cloud droplet size?

We agree with the reviewer that it is worthwhile to discuss the behavior of parameter c a bit more.

The answer can be found in the shape of the SEVIRI LUT (see Figure 2 of this response, which is adapted from the manuscript). For a constant effective radius and increasing VNIR reflectance (r_{06}), which indicates an increase in cloud optical thickness, the SWIR reflectances (r_{16}) at first increase almost linearly ($r_{06} < 0.3$). However, for $r_{06} > 0.3$ there is a curvature in the isolines and the linear relationship between r_{16} and r_{06} becomes non-linear. For even larger optical thicknesses ($r_{06} > 0.7$) the isolines become orthogonal and r_{16} remains constant with increasing r_{06}. For the latter case positive or negative changes in subpixel VNIR reflectances would be translated into positive and negative SWIR reflectance deviations, even though for large optical thicknesses r_{16} becomes independent of r_{06}. This means that assuming a linear relationship in the form $\langle r_{\mathrm{HV}} \rangle = c \cdot r_{16}$ is a flawed assumption outside of optically thin clouds.

Thus, scenes with convective clouds, where the optical thickness can be larger than 20 (even larger than 100) are not well described by this relationship. As a result, the fit coefficient c is not well constrained and can vary widely from hour to hour. However, stratus and altocumulus cloud fields are usually characterized by $\tau \approx 10$ and for these types of clouds this relationship should work rather well. As a result, varying cloud types will determine the reliability of this relationship. Over central Europe we often observe altocumulus and stratus clouds and thus for a large number of pixels the linear relationship works quite well (see the dark red and silver area around the 1:1 line in Figure 3b of the manuscript). For small cumulus clouds and convective thunderstorms, however, we will get large deviations from the linear relationship.

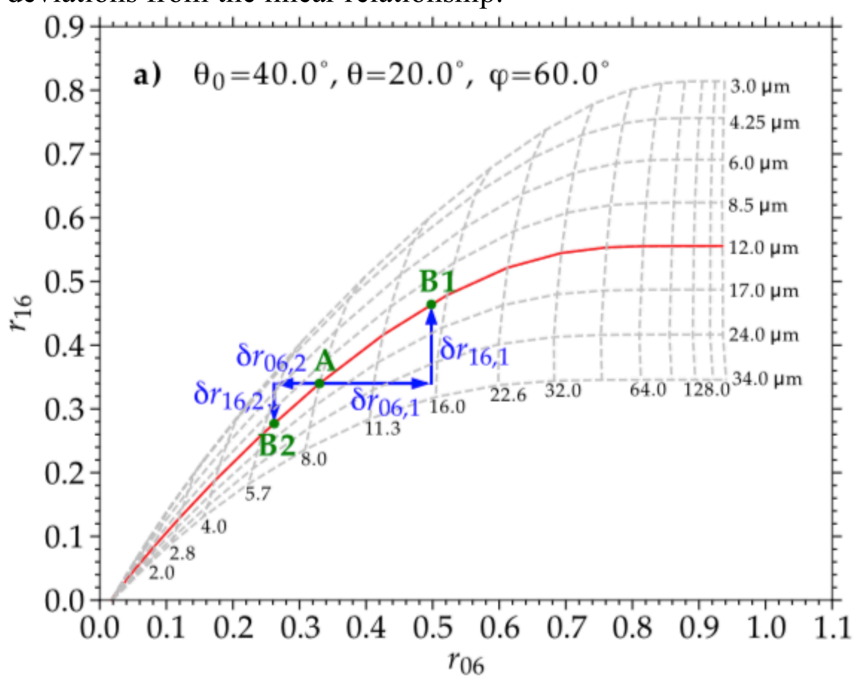

Fig 2: Example SEVIRI LUT. Isolines for constant τ and r_{eff} are shown in dashed gray lines.

In the revised manuscript we added the following explanation:

"This behavior is expected, as the relationship between VNIR and SWIR reflectance can usually not be described by a linear function (see discussions in *Werner et al.,* 2018a, b, as well as the LUT examples in Figure 4 later in this study). For a constant r_{eff} there is a linear increase in r_{16} with increasing r_{06}, as the cloud optical thickness increases. However, the slope of this linear relationship increases with decreasing r_{eff}. For $\tau > 10$ the relationship between r_{16} and r_{06} is characterized by a prominent curvature, while for $\tau >> 10$ the r_{16} become independent of r_{06}. Therefore, the fit coefficients c depend on the distribution of cloud optical and microphysical parameters, which varies widely with cloud type, meteorological conditions and different dynamic processes.

Page 14, Line 22: Wouldn't spatial averaging of MODIS data provide a better comparison than subsampling?

Thanks for this comment. This part of the manuscript was actually a bit confusing in the original manuscript.

It turns out that trigonometric interpolation (i.e., Fourier transform of the image and the inverse on a higher-resolution grid), combined with the application of the modulation transfer function (i.e., the spatial response function in Fourier space) yields an interpolated image, where the reflectance of the central pixel of each 3x3 pixel block corresponds to the lower—resolution reflectance value. In other words, by subsampling we combine the effects of spatial and optical resolution of the SEVIRI imager and get the exact reflectances that the lower-resolution SEVIRI channels would see. By carefully applying the two different modulation transfer functions (from the HRV channel and channels 1-3) and subsampling of the central pixel of each 3x3 pixel block we could simulate the reflectances at the lower spatial resolution (i.e., the native resolution of SEVIRI channels 1-3).

However, this is not the pathway we chose for this study. As mentioned in Section 6.1, we generated a second data set, where the MODIS level 1b observations where remapped to the lower-resolution (~3 km) grid (in the same way the reference data set was created at the HRV grid). The baseline results where then calculated by trigonometric interpolation and smoothing with the modulation transfer function.

Note, that both pathways are valid and yield the same baseline reflectances.

Somehow, the old manuscript version described both pathways and the result was rather confusing. We rewrote parts of both the general introduction to Section 6, as well as Section 6.1, where the remapping is described:

"Remapping MODIS reflectances to SEVIRI's LRES grid (i.e., the native resolution of channels 1–3) subsequently provides the means to apply the various downscaling schemes, as well as the simple triangular interpolation approach, in order to compare the retrieved cloud products (i.e., $\hat{\tau}$ and $\hat{r}_{eff}$, as well as $\tilde{\tau}$ and $\tilde{r}_{eff}$) to the reference results. Naturally, the ideal downscaling approach would yield results that closely resemble the MODIS–provided HRES observations. Furthermore, the ideal downscaling approach would also represent an improvement upon the simple interpolation technique. The reader is reminded, that the latter data are still available at a higher resolution than the native LRES grid of the SEVIRI r_{06}, r_{08}, and r_{16} channels, but no longer contain any information about the high–frequency reflectance variability. As the simplest approach to derive higher–resolution cloud products, these results are called the baseline results. "

And:

"To perform the subsequent downscaling experiments, a second set of level 1b radiances are generated, where the spatial variability is reduced to match that of the LRES–channels of Meteosat SEVIRI. This step again involves the smoothing of the respective reflectance field with the channel–specific modulation transfer function of the lower–resolution SEVIRI channels (EUMETSAT, 2006). This data set represents hypothetical SEVIRI–like observations at the native LRES resolution. "

Page 14, Line 29: The part "(a)" seems to be missing from "Figure 8(a)".
Thanks for pointing out this mistake, we added "(a)" to the text.

Page 14, Line 31: The "t" in "table 1" should be capitalized.
We capitalized the "t". We also capitalized it for other table references in the manuscript.

Page 15, Line 4: The "s" in "section" should be capitalized.
We capitalized the "s". We also capitalized it for other section references in the manuscript.

Page 20, Line 9: It would help to clarify what is meant by SEVIRI LUT (what specific look-up table is referred to).
Given the next comment (that a lot of readers jump from the Abstract to conclusions), we agree to add a clarification here. We added the following information in parentheses after "SEVIRI LUT":
"(which consists of simulated SEVIRI reflectances for different viewing geometries and combinations of cloud properties)".

Page 20, Line 17: Some readers jump from the Abstract straight to the conclusions and read the rest only afterwards. For the sake of these readers, it is important to clarify in the conclusions section what is meant by the caret accent over tau and reff.
Again, we agree. We added the following information in parentheses:
"(i.e., the actual higher-resolution cloud properties)"

Page 20, Line 25: It would help to clarify that "local slopes" refer not to the slopes of the cloud top surface, but to the steepness of curves in the used LUT.

We changed the sentence as follows:
"with an adjustment based on the calculation of isoline slopes in the SEVIRI LUT".

Page 21, Line 6: The spatial averaging used by MODIS is a reasonable alternative to downscaling. Although at visible wavelengths MODIS reflectances are available at a higher resolution, the MODIS cloud algorithm degrades the resolution of all input reflectances to a common 1 km resolution. Therefore, while downscaling could certainly help, the resolution mismatch can also be avoided by averaging, without the downscaling approach. Accordingly, at least the word "should" should be replaced.

We agree with the reviewer, even though the spatial mismatch is a direct result of the downscaling approach, which is the focus of this study (i.e., the resolution mismatch did not exist before downscaling the VNIR reflectances, yet we want downscaled reflectances for the purpose of this study). The sentence is indeed misleading. We meant to say that it is essential to also downscale the SWIR band reflectance, not just the VNIR band ones.

We rewrote the respective paragraph and it now says:
"This illustrates that, in order to achieve reliable higher–resolution retrievals, all channels need to capture small–scale cloud heterogeneities at the same scale. These results confirm the findings of Werner et al. (2018b), who compared SWIR reflectances at different spatial scales and demonstrated the need for effective downscaling approaches to match the spatial scale of the VNIR reflectance. This also has implications for other multi–resolution sensors, such as MODIS, VIIRS, and GOES–R ABI. To avoid a scale–mismatch of resolved variability in the VNIR and SWIR channels, the higher–resolution observations can either be degraded to match the lower–resolution samples (which yields overall lower–resolution cloud property retrievals), or downscaling techniques are applied to one or both channel reflectances, which yields matching scales and higher–resolution estimates of cloud properties. It is important to note that downscaling might result in increased retrieval uncertainties, if the spatial resolution is below the radiative smoothing scale ($\approx 200 - 400$ m, see *Davis et al.*, 1997)."

Page 32, Lines 4-5 of Figure 2 caption: It would help to clarify that the blue lines show the relative difference between the Constant Reflectance Ratio Approach and the resampled original data. To this end, the words "relative difference" should be included, and the mention of color should be moved to the end of the sentence.

Thanks for this comment. In the revised manuscript we decided to remove that figure and the respective section discussing it (following advice from reviewer #1).