We'd like to thank the editor for handling our manuscript, as well as reviewers #1 and #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 both reviewers 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.

Review of "Increasing the spatial resolution of cloud property retrievals from Meteosat SEVIRI by use of its high–resolution visible channel: Evaluation of candidate ap- proaches with MODIS observations" by Werner and Deneke.
The manuscript discusses a topic relevant for the scientific community. Various down- scaling techniques are presented and analysed in order to derive high resolution (1km) cloud properties from low resolution (3km) Meteosat SEVIRI data. Methods and motivations are layed out in detail, but the presentation suffers from the dizzying number of approaches used for downscaling, used for verification, of data spatial resolutions and of acronyms used. While I have full confidence in the authors scientific rigour, I was close to giving up reading through all the details offered. At the same time, I'm missing general interpretation of the large variety of results in some places. This leads to the unsatisfying point that the support of final and most important conclusions is not easy to find for the reader. I have the impression that this manuscript could be much improved by a major revision and tightening of the presentation, especially of the comparison results section 6.

Major points:

Especially section 6 is confusing. I would suggest to reduce the number of downscaling code versions. Especially the results section 6 even has versions not discussed anywhere. I would also suggest to reduce the number of error quantities discussed, maybe to the set shown in the tables Fig 10 and 12. Do not discuss other additional numbers in the text. Please see details below.

We agree with the reviewer and changed the revised manuscript in the following ways:

1) We focused on 3 downscaling schemes: *Statistical Downscaling Approach, Constant Reflectance Ratio Approach*, and *Adjusted LUT approach with LUT Slope Adjustment*
2) We simplified the naming scheme of each experiment. Instead of 1b, 2b, 3b, 3c, and 3d we simply refer to the three downscaling schemes as 1, 2, and 3.
3) We only briefly summarize the performance of experiments 3b and 3c (i.e., the standard *LUT Approach* and the *Adjusted LUT* approach with *Adiabatic Assumptions*). No new experiment numbers are added for those and the summarized results are only a couple of sentences long.
4) We removed all mentions of the 1st, 50th, and 99th percentiles of the retrieval differences for each of the cloud variables. We likewise changed the correlation coefficient R to explained variance R^2 in the scatterplots in Section 6, because this variable is used in the statistical comparison in the other two figures.
5) We moved the VNIR-only versus full downscaling part to its own section (Section 7). This (i) makes Section 6 shorter and easier to follow, and (ii) creates a more logical structure. After all, that new section does not compare different downscaling algorithms and instead looks at the effects of a scale-mismatch.
6) In the new Section 7 we removed the a and b experiment distinctions. Instead, we simply write out the respective experiments.

These steps should improve the readability of the comparison section. However, note that in the new Section 7 we kept the statistical measures (i.e., percentiles). We prefer the style of presentation by means of PDFs of the retrieval difference here and we believe that the statistical comparison by means of percentiles is appropriate.

Is Deneke and Roebeling 2010 the basis of this paper? I had the impression down to page 8 that many things come from this older publication. If this is the case, it would be one possibility for shortening. You have to make the connection of the two clearer in the introduction.

This study by *Deneke and Roebeling* (2010) introduced the statistical downscaling of VNIR reflectances by means of the HRV channel. It neither provided suggestions for downscaling of SWIR reflectances, nor did it analyze the effects on the subsequent cloud property retrievals.

Our study provides different approaches to downscale SWIR reflectances and evaluates the reliability of the resulting retrieval products. Moreover, it tests other downscaling techniques that ultimately yield improved results.

In this sense the *Deneke and Roebeling* (2010) study is more of a 'conversation starter'. It provides the basis of half of the *Statistical Downscaling* Approach (and parts of the LUT-based approaches).

We carefully rewrote that part of the introduction, which now says:
"The aim of this manuscript is to critically evaluate several candidate approaches for downscaling of the SEVIRI narrow–band reflectances for operational usage and to identify the most promising of these schemes, exploiting the fact that information on small-scale variability is available from its broadband high–resolution visible (HRV) channel. The study by *Deneke and Roebeling* (2010) presented a statistical downscaling approach of the SEVIRI channels in the visible to near-infrared (VNIR) spectral wavelength range. This method makes use of the fact that SEVIRI's high–resolution channel can be modelled by a linear combination of the 0.6 μm and 0.8 μm channels with good accuracy (*Cros et al.*, 2006). This study advances these efforts in three ways: (i) it explores other possible downscaling approaches, which might improve upon the statistical downscaling scheme, (ii) it introduces techniques to accurately capture information on the small–scale reflectance variability in the 1.6 μm–channel, which predominantly arises from variations in effective droplet radius, and (iii) it studies the impact of the various downscaling techniques on the subsequently retrieved cloud properties."

Technical problem is that a companion paper (Deneke et al 2019) is obviosuly not submitted at this stage. I would recommend removal of all references to it or waiting for its puplication in discussion stage.

We moved all references to the companion paper, which will be submitted at the end of January to the conclusion/outlook section. There, it basically works as a 'plans for future work' style reference. Since not a single part of our analysis depends on this (as of now) unpublished work, we think it is ok to mention it at least in the conclusions/outlook.

Specific points

p.1, l. 11 ff: Where do these numbers for tau, reff and WL and ND come from? I can not easily find these numbers in the manuscript and I hardly can find any discussion of them. Please extend discussion of these later on or remove them from the abstract.

Following other comments in this review, we removed the discussion of percentiles in section 6. This removes some of the clutter in the discussion and reduces the number of discussed statistics. However, we feel mentioning explained variance and interquartile ranges in the abstract (and to a lesser degree in the conclusions) is not really helpful. We believe the reader generally wants to know how big biases due to the lower spatial resolution are, in order to assess whether this is even an issue. We therefore prefer to keep these numbers in the abstract. We summarize these retrieval differences in the conclusions section: **"**The retrievals based on native–resolution reflectances (at a scale of ≈ 3 km) are characterized by significant deviations from the reference retrievals, especially for $\hat{\tau}$ and $\hat{W}_L$. Here, random absolute deviations as large as ≈ 14 and ≈ 89 g m^{-2} are observed, respectively (determined from the 1^{st} or 99^{th} percentiles of the absolute deviations between native and reference results for each cloud scene). For $\hat{r}_{eff}$ and $\hat{N}_D$ deviations of up to ≈ 6 μm and ≈ 177 cm^{-3} exist, respectively."

p.1, l. 20: The whole abstract reads as if it does not work very well. Maybe apart from tau. This concluding sentence reverses these statements. Please revise.

We were overly cautious when describing the performance of the higher—resolution r_{eff} retrieval in the abstract. We tried to convey the fact that the biases in the LRES retrieval are smaller than for τ and that the performance of the best downscaling technique is close to the baseline approach.

We agree with the reviewer that this is not necessary; after all there are still improvements for r_{eff} and N, just not as prominent as those for τ and WL.

We revised parts of the abstract as follows:
"… Uncertainties in retrieved r_{eff} at the native SEVIRI resolution are smaller and the improvements from downscaling the observations are less obvious than for τ. Nonetheless, the right choice of downscaling scheme yields noticeable improvements in the retrieved r_{eff}. Furthermore, the improved reliability in retrieved cloud products results in significantly reduced uncertainties in derived W_L and N_D. In particular, one downscaling approach provides clear improvements for all cloud products compared to those obtained from SEVIRI's standard-resolution and is recommended for future downscaling endeavors. This work

advances efforts to mitigate impacts of scale mismatches among channels of multi–resolution instruments on cloud retrievals. “

p. 3, l. 11: Isn’t a clear reference to Deneke and Roebeling 2010 missing here? Can you please make that clear in the introduction.

As mentioned in the reply to a previous comment in this review, we rewrote that part of the discussion and clearly pointed out the connection to the study by *Deneke and Roebeling* (2010).

p. 3, l. 20: This way you will only get relative errors. All problems retrievals at 1km resolution still suffer from are not discussed or improved . . . e.g. Zhang et al, 2009, 2011. Can you please mention that.

We agree with the reviewer. Indeed, after downscaling the reflectances (even if the results would be perfect), the same limitations exist that impact other 1 km-retrievals. In other words, the results likely would not represent the true cloud properties.

We already discussed the effects of resolved and unresolved cloud variability and cited the respective literature. In the revised manuscript we added the following text to clarify this point:
“Note, that even the retrieved cloud products from a hypothetically perfect downscaling technique would still be impacted by the effects of resolved and unresolved cloud variability. Therefore, the results of this study will not help to mitigate the uncertainties associated with the retrieval schemes of similar ≈ 1 km–sensors (e.g., clear–sky contamination, plane–parallel albedo bias, 3–dimensional radiative effects). “

p. 3, introduction in general: Can you please make clearer: What is the motivation for an improved resolution of products? What problem do you expect to improve?

The manuscript states in the introduction that:
“Use of the independent pixel approximation (IPA, see *Cahalan et al.*, 1994a, b) produces uncertainties in the retrieved cloud variables that are dependent upon the horizontal resolution of the observing sensor. “
It subsequently describes the effects of resolved and unresolved cloud variability and cites the respective literature.

Afterwards, it talks about the benefits of SEVIRI-like observations, as well as the disadvantage of the lower spatial resolution:
“However, SEVIRI pixels are characterized by a lower spatial resolution of its narrow–band channels compared to other operational remote sensing instrumentation, like the Moderate Resolution Imaging Spectroradiometer (MODIS, *Platnick et al.*, 2003) or the Visible Infrared Imaging Radiometer Suite (VIIRS, *Lee et al.*, 2006). Given the increase in retrieval uncertainty due to the IPA constraints, there is a desire to increase the resolution for geostationary cloud observations. “

We believe that this clearly motivates the study in our paper.

p. 4, l. 30: "horizontal resolution of 3 x 3". This is the nominal sub-satellite resolution. Can you please make clear whether you consider the full spatial response function for each point. Much later it sounds like, but here you widely stick to the simplified "1 km", "3 km" without further explanation.

Thanks for this comment; this point was similarly addressed by reviewer #2. 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.

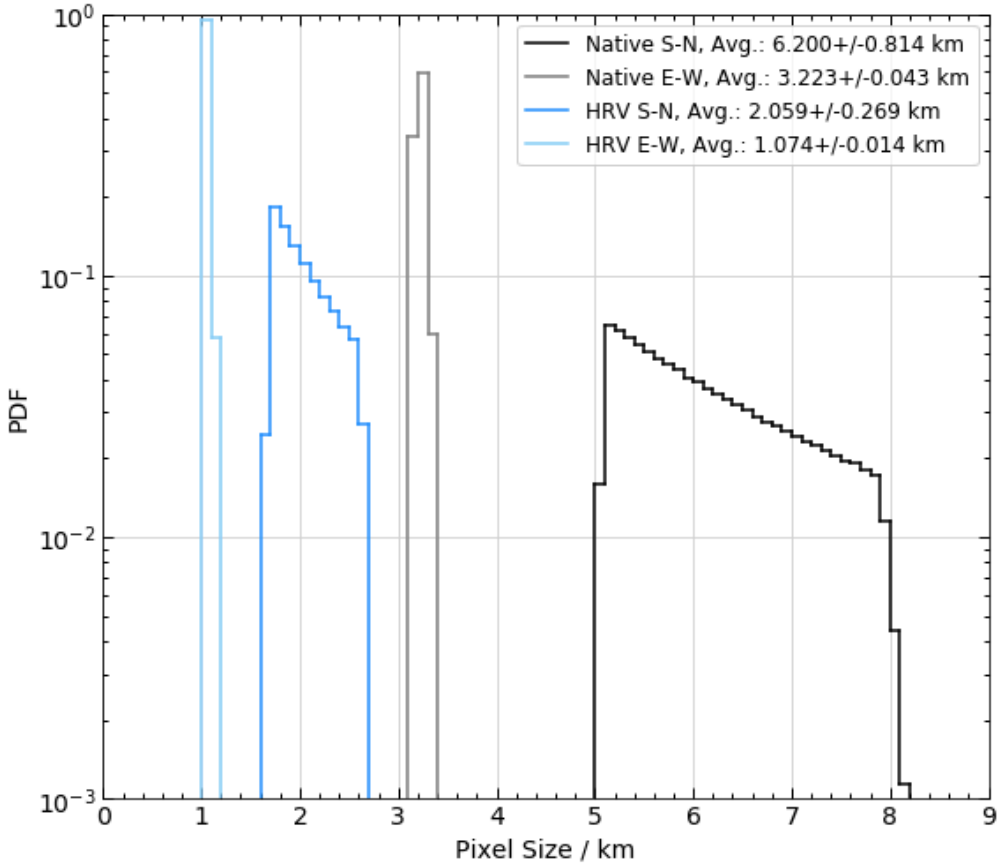

Fig 1: Statistics of SEVIRI pixel dimensions (in both latitude and longitude direction; i.e., South-North and East-West) for the native and HRV resolutions.

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 6.20×3.22 km^2 and 2.06×1.07 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 3×3km^2 and1×1km^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.

With regard to the spatial response function, we added a paragraph in the SEVIRI instrument description. Here, we mention the difference between spatial and optical resolution for SEVIRI, the definition of the spatial response function, as well as the exact treatment in our study by means of the modulation transfer function. The latter describes the convolution of the sensor signal and the point spread function (i.e., the terms in the integral of the spatial response function) in Fourier space. See also our response to a latter comment in this reply for more details.

The added section says:
"As context for the present study, the reader is reminded that the spatial resolution of geostationary satellites is significantly reduced at higher latitudes due to the oblique viewing geometry. For Germany and Central Europe as considered in this paper, the pixel size is effectively increased by a factor of two in North–South direction as a result. In addition, the distinction between sampling and optical resolution needs to be acknowledged. While the former determines the distance between recorded samples, the latter is given by the effective area of the optical system, which is larger by a factor of 1.6 than the sampling resolution for SEVIRI (*Schmetz et al.,* 2002). The spatial response of optical systems is commonly characterized by their modulation transfer function, which describes the response of the optical system in the frequency domain.

Further information about the spectral width of each SEVIRI channel, as well as the respective spatial response and modulation transfer functions, can be found in *Deneke and Roebeling* (2010). "

p. 5, l. 14/15 and 17: I do not understand the need for theses statements. These are purely technical, aren't they? First you mention a method not used in this study?! With modifications described in a study not published yet?! Then you are talking about a version control system

development branch (?) to make clear that this algorithm is not perfectly the same as in the unpublished companion paper!? I doubt that the reader needs these documentation details.

We agree with the reviewer's comment and removed the respective parts from the revised manuscript.

p. 5, l. 28: Again the reader wants to know whether you consider the different spatial response function of normal and high resolution channels? Please comment why you think you do not need to consider this or how it is considered.

The modulation transfer function, which describes the spatial response of an optical system in Fourier space, is applied during trigonometric interpolation. While the details are given in *Deneke and Roebeling* (2010), we added a reminder in the revised manuscript to ensure the reader that the spatial response of the sensor is considered:

"… implemented based on the discrete Fourier transform and multiplication with the modulation transfer function (see *Deneke and Roebeling*, 2010, for details) …"

p. 6, l. 16: "Statistical downscaling": I do not see the "statistical" element? Are not all downscaling steps fully deterministic? No random element is in there? Please clarify.

There are multiple techniques to downscale a coarse-resolution signal to a higher—resolution one. The simplest is an interpolation, which doesn't add any high-frequency variability to the new image. Another simple method would be the "Constant Reflectance Ratio Approach" discussed in this manuscript, which assumes a constant relationship between VNIR and SWIR reflectances for different spatial scales. From climate science we know of two more approaches:

(i) Dynamical downscaling, which uses a higher-resolution model to estimate the smaller-scale information, and

(ii) Statistical downscaling, which establishes statistical relationships between different data sets and uses the results in a multiple linear regression to predict the higher-resolution behavior.

We simply reminded the reader that the first technique in our manuscript is a *Statistical Downscaling Approach*, well known and used in other scientific fields. Please note that Technique #2 in our manuscript does not use any large-scale statistical relationships. Also, Technique #3 is a hybrid technique. While it uses statistical relationships between HRV and standard channels in the VNIR, it incorporates the shape of the SEVIRI LUT to determine the higher-resolution SWIR signal.

We believe that the later statement in the manuscript is sufficient to explain the naming scheme: "Note, that the use of linear models and bivariate statistics means that the downscaling algorithm described in this section is an example of statistical downscaling techniques, which are common in climate science applications (e.g., *Benestad*, 2011). "

p. 7, l. 2: IQR=0.03 of what? Daily values? Hourly?

All data points are included in this statistic, i.e., all 16-day intervals and each hourly data point. We updated the sentence in the revised manuscript as follows: "Considering all hourly data and each 16-day interval, the median …"

p. 7, l. 4, Fig 2: I can only see three colors for 14:00, 15:00 and 16:00 UTC?!

There is an overlap of data with values of about 0.51. For the 8-9 UTC and 13-16 UTC time stamps. The afternoon hours are just plotted on top of the morning hours. However, the old version of the manuscript said 15-16 UTC. We corrected this to 13-16 UTC. We also double checked all median and IQR values in this section (everything is correct here).

p. 8, l. 7: "Diurnally, the variability in the hourly derived ..." : You mean the IQR is derived over 18 or 19 hourly values over one day? Or over 16 days?

Thanks for pointing this sentence out; it indeed is confusing. We meant to say that during each 16-day interval, there is a huge diurnal variability in coefficient *c*, something not observed for coefficients *a* and *b*. We changed that sentence to: "For each 16-day interval the variability in the hourly derived *c* values…"

p. 10, eq. 9: Under which assumption does eq. line 2 follow from eq. line 1?

The long version of the equation is:

$$\frac{\hat{r}_{06,\mathrm{i}} - r_{06}}{r_{06}} = \frac{\hat{r}_{\mathrm{HV,i}} - r_{\mathrm{HV}}}{r_{\mathrm{HV}}}$$

$$\frac{\hat{r}_{06,\mathrm{i}}}{r_{06}} - 1 = \frac{\hat{r}_{\mathrm{HV,i}}}{r_{\mathrm{HV}}} - 1$$

$$\hat{r}_{06,\mathrm{i}} = \hat{r}_{\mathrm{HV,i}} \cdot \frac{r_{06}}{r_{\mathrm{HV}}}$$

The VNIR high-resolution signal is scaled by the ratio of SWIR to VNIR reflectance at the lower horizontal scale.

p. 10, l. 15, "R ∼ 1.0" : This proves that both approaches are rather equivalent, but suffer from the same core problem. the reff impact. Is this an important comparison or just a distracting sideshow?

Given that Downscaling Techniques #1 and #2 are completely different approaches (statistical downscaling versus assuming a constant ratio for each individual pixel) it is actually quite remarkable that the results agree so well.

In an older version of the manuscript we put more emphasis on the actual reflectances. However, the main point of the manuscript is the impact on retrieval products, which naturally include the changes in the reflectances. We decided to remove that part in the revised manuscript, which shortens and hopefully improves the readability of the study.

p. 10, l. 22: I can see that the first approach could produce negatives, but the second?

Thanks for noticing that mistake. Indeed, technique 2 cannot produce negative values. This was a mistake of the simple counting algorithm we used, which treated a 0-length array as 1. The result for experiment 1 was not impacted. However, following the earlier comment we decided to remove that part from the revised manuscript (including the figure).

p. 13, l. 1: Is this section including Fig 6 really needed? It confirms stuff that could be seen before and adds another side aspect.

As mentioned earlier, in a prior version of the manuscript we put more emphasis on the actual reflectances, before comparing the impact on retrieval results. In order to shorten and streamline the manuscript, we agree that this section and the figure are not needed. We removed them in the revised manuscript.

p. 14, l. 11: You did never mention failed retrievals and reasons for it. Skip this sentence?

We agree with the reviewer and skipped this sentence in the revised manuscript.

p. 14, l. 12, chapter: This section is confusing. I started reading with the understanding that you only use MODIS data in this chapter until I read the Fig. 8 caption which sounds like it shows SEVIRI data. Please make sure that this stays clear from the beginning and throughout the section. Do you use "SEVIRI data" or only "SEVIRI- resolution MODIS data" in this section?

Thanks for noticing this error. The reviewer is correct in assuming that the evaluation of downscaling techniques is done exclusively with MODIS data, resampled on the SEVIRI HRV-channel grid. In an older version of this Figure, the RGB was constructed from downscaled SEVIRI reflectances, but this turned out to be confusing (because the analysis is done without SEVIRI data). Therefore, we changed the figure and constructed the RGB with resampled ~1-km MODIS data. Unfortunately, we did not change the caption of the figure. This has been corrected in the revised manuscript and that part of the caption now says: "RGB composite image of remapped MODIS channel 6, 2, and 1 reflectances at the horizontal resolution of SEVIRI's HRV channel at a horizontal scale of 1×1 km^2 at the sub–satellite point. Data is from example scene 1 sampled on 1 June 2013 at 10:05 UTC. "

p. 14, l. 18: You mention spatial response for the first time here I guess. What about spectral differences between MODIS and SEVIRI? Please discuss.

As mentioned in the response to an earlier comment, we added information about the modulation transfer function, which described the point spread function in Fourier space. The exact spatial response for each SEVIRI channel is considered throughout the manuscript.

Regarding the spectral differences between MODIS and SEVIRI: These differences do not affect our present study. We are not comparing operational

MODIS C6 results to SEVIRI retrievals after downscaling. This technical study is purely performed with re-mapped MODIS data, as we are only interested in evaluating the different downscaling techniques. In order to evaluate the different approaches, we require a common retrieval algorithm and a ground truth, which is provided by MODIS data.

We believe that this paragraph in the manuscript is sufficient to establish that goal: “It should be noted that retrievals based upon these radiances will be different than those based upon the original MODIS C6 radiances, or from an absolutely accurate representation of the (hypothetical) truly observed, high–resolution SEVIRI samples. For one, it uses the linear model of Cros et al. (2006) and Deneke and Roebeling (2010) as a proxy for the HRV channel, thereby excluding a potentially significant source of uncertainty. Moreover, MODIS acquires these reflectances under different viewing geometries (note that the true viewing angles are used in the CPP retrieval, so within the limits of plane–parallel radiative transfer, this effect is accounted for), and the spectral characteristics of the MODIS and SEVIRI channels are not entirely comparable. However, the goal of this study is to provide a consistent reference data set for a comparison of different retrieval data sets, which are derived from a single retrieval algorithm core. The actual absolute values of the retrieved cloud products are not important here. “

The companion paper, which will be submitted at the end of January 2020, will compare downscaled SEVIRI with operational MODIS C6 retrieval results (the statistical comparison for different cloud scenes shows a significant improvement between MODIS and SEVIRI due to the downscaling efforts). The applied downscaling theme was chosen based on the results of this study. This upcoming manuscript also presents other applications of the higher-resolution SEVIRI cloud products.

p. 14, l. 33, "interpolation of SEVIRI samples": Are you talking about SEVIRI or MODIS data here? See point above.

Again, thanks for noticing these inconsistencies. Again, the analysis in section 6 is performed exclusively with remapped MODIS data. No SEVIRI reflectances are included.

We carefully read through section 6 again and removed all ambiguities, to make sure that the reader knows that only MODIS data is considered in the analysis.

p. 15, l. 4-8: Phew! Now you add sub-experiments "a, b, c, d" on top of the new nomenclature "1, 2, 3 " ... I’m struggling, to keep reading ... At least, do not use "1,2,3" acronyms alone, but write out the experiments to make them more recognizable. Please do not introduce experiments you will not even discuss (1a, 2a).

We agree that the different nomenclatures are potentially confusing. We reduced the number of experiments to three (focusing on one of the options for the *LUT Approach*). We also removed versions a and b, which were discussed in Section

6.4 (now Section 7). When we discuss the difference between full downscaling and VNIR-only results, we write out the experiment description instead.

p. 15, l. 11-12: And now ... a few new products on top. You have to mention the relevanve of these right in the introduciton.

We added the following information to the introduction, right after discussing resolved and unresolved variability in τ and reff: "These uncertainties are propagated to the liquid water content (W_L) and the droplet number concentration (N_D), which can be estimated from retrieved τ and r_{eff}. Estimates of N_D are especially susceptible to uncertainties in r_{eff}, which impacts the reliability of aerosol–cloud–interaction studies (*Grosvenor et al.*, 2018). "

Introducing these variables early should help the reader understand their importance and the need to include these parameters in the downscaling analysis.

p. 15, l. 21: This is not the first time you use the exact spatial response functions, isn't it? This is rather late to mention the reference for the first time.

As mentioned earlier, we added a paragraph about the modulation transfer function and its relationship to the point spread function and spatial response in the SEVIRI-section. In the revised manuscript we also use the modulation transfer function throughout the manuscript, which should help avoid confusion.

p. 15, l. 23, "3x3 block". This block is 333 m resolution here, right? Please make sure that this can not be confused with the other 3x3 blocks mentioned before.

That is correct. We added the following in parentheses in the revised manuscript: "(each pixel with a horizontal resolution of 333 m)"

p. 15, l. 25, "level 1b": Could be easily confused with your experiment notation. You did not use the term level 1b data before, you do not need it here.

As mentioned earlier, we removed the a, b, c, d, subcategories and focus on 3 experiments on the revised manuscript. The MODIS level 1b radiances are first mentioned at the start of Section 6.1:
"To obtain a reliable higher–resolution reference data set, MODIS level 1b swath observations (MOD021km) have been projected to the grid of the SEVIRI HRV reflectance observations …".
Since this is the correct reference to the data set, together with the easier experiment description, we believe that any potential confusion is now avoided.

p. 15, l. 28-29: What is the "modulation transfer function" good for? Why do you only mention it here, that late in the manuscript?

Thanks for this comment. Without proper context, this was indeed confusing. As mentioned earlier, we added a section about the modulation transfer function, in the SEVIRI-section.

The discussion in *Deneke and Roebeling* (2010) points out that the optical resolution of the SEVIRI channels is lower than the spatial resolution by a factor

of about 1.6. This means that the signal for each pixel is not only determined by the observations within the nominal sampling resolution (i.e., the pixel itself), but also includes contributions from neighboring pixels. This characteristic is effectively described by the spatial response function *S* of the respective SEVIRI channel:

$$S(x_0) = \int_A w(x - x_0)L(x)dx$$

$$S(x_0) = (w * L)(x_0)$$

where x_0 is the displacement from the center of the field of view, $L(x)$ is the radiance at position x, w is a weighting function commonly referred to as the point spread function, and $*$ indicates the convolution of w and L. Applying the Fourier convolution theorem means that this convolution is equivalent to a multiplication of the Fourier transforms of L and w. The modulation transfer function, which fully describes the spatial response of an imager, is the modulus of the Fourier transform of w.

As mentioned in an earlier reply, we added a new paragraph to the revised manuscript, which explains the treatment of optical resolution.

p. 16, l. 2, "spectral characteristics": This is again too late to mention such an obvious problem that late.

This might be the result of some of the confusion regarding the use of MODIS and SEVIRI data. The evaluation of downscaling techniques is based exclusively on MODIS data, which is available at ~1 km resolution. We remap this dataset to the SEVIRI geometry and apply the SEVIRI retrieval code to it. We do not expect these results to agree with operational MODIS C6 products, as the use of a different algorithm core alone will yield different results.

However, this is not the focus of this manuscript, as we are not interested in absolute values of the retrieval results. We are only interested in a comparison to the reference results. We mention this in the manuscript, when we say: "However, the goal of this study is to provide a consistent reference data set for a comparison of different retrieval data sets, which are derived from a single retrieval algorithm core. The actual absolute values of the retrieved cloud products are not important here. "

A comparison between downscaled SEVIRI retrievals, employing the most promising technique revealed by this study, and operational MODIS C6 results is performed in the follow-up paper (amongst demonstrations of other applications of the new dataset).

p. 17, l16ff: I'm missing this kind of more conclusive interpretation elements instead of adding number to number in the text.

As mentioned earlier, we removed all mentions of the 1st, 50th and 99th percentiles in Sections 6.2 and 6.3. We added some additional interpretation of the reason behind the shortcomings of the different downscaling techniques both in Section 6.2 and 6.3, as well as in the conclusions (see our replies to later comments below).

p. 17, l.31: This sounds as if 2/3 and 5/9 are magic numbers found empirically. Approx. 2/3 follow from optical properties directly and in general. 5/9 contains an empirical element. I would prefer to say "WL=2/3 ro tau reff" and adiabatic clouds have a typical additional factor of factor=5/6 due to their vertical structure.

We changed the manuscript as follows:

"$W_{\mathrm{L}} \approx \frac{2}{3} \cdot \rho_{\mathrm{L}} \cdot \tau \cdot r_{\mathrm{eff}}$.

Here, ρ_{L} is the bulk density of liquid water. Assuming adiabatic clouds, where the vertical structure of effective droplet radius follows the adiabatic growth model, introduces an extra factor of 5/6 and the coefficient 2/3 changes to $5/6 \cdot 2/3 = 5/9$."

p. 18, l. 3, eq. 18: Again, this equation seems to contain magic, but is rather simple in its core. Maybe give some additional explanation: "Droplet number could simply be derived from LWC and a droplet size. Using empirical factors accounting for typical droplet size distributions and vertical cloud structure, the following can be derived:"

We slightly disagree with the reviewer in this point. The assumptions going into the derivation of droplet number concentration are not trivial and include terms for the condensation rate, shape of the droplet number size distribution and more. Assumptions about subadiabaticity alone change statistics of droplet number concentration substantially. Likewise, going into detail about the derivation does not improve the readability of the manuscript. We believe it is enough to cite the appropriate literature here.

However, we added some clarification about the nature of the assumptions and this part of the manuscript now reads:

"Calculating N_{D} from remote sensing products requires a number of assumptions, e.g., about the vertical cloud structure and shape of the droplet number size distribution, which are summarized and discussed in *Brenguier et al.* (2000); *Schüller et al.* (2005); *Bennartz* (2007); *Grosvenor et al.* (2018). A simplified form of the resulting equation for N_{D} is:"

p. 18, l. 22, "3c overall performs worst": Why? Can you give a general explanation or guess?

As mentioned earlier, we made several changes to this section. We focus on experiments 1b, 2b, and 3d (now just 1, 2, and 3) and only briefly summarize the results for 3b and 3c.

We added a general explanation for the poor performance of 3c (and 3b in comparison) at several points in the revised manuscript:

"… We believe that this might be caused by the sensitivity of the cloud property retrieval to small reflectance perturbations, in particular for broken clouds. It is also an indication that assuming constant subpixel r_{eff} values within each LRES pixel is not sufficient. We plan to investigate this effect further in future studies. However, the second *Adjusted LUT Approach*, which determines SWIR reflectance adjustments based on adiabatic theory, performs even worse (R^2 of 0.846, 0.579, 0.741, ad 0.519 for cloud scenes 1–4, respectively). This suggests that the observed cloud fields do not follow adiabatic theory and the method is not adequate to estimate higher–resolution r^16. "

And:

"As before, we also tested the standard *LUT Approach* highlighted in Section 4.3, as well as the second *Adjusted LUT Approach*, which determines SWIR reflectance adjustments based on adiabatic theory. Due to the poor performance of the r^eff retrieval, the N^D results based on adiabatic assumptions show a similarly poor agreement to the reference results. Meanwhile, the cloud variables based on the standard *LUT Approach* never show the best or worst performance, but are almost universally worse than the *Adjusted Lookup Table Approach* with *LUT Slope Adjustment*. This again illustrates that assumptions of adiabatic clouds and constant subpixel r_{eff} values within each LRES pixel are not suitable for the cloud scenes analyzed in this study. "

The poor performance of adiabatic assumptions is not surprising. After all, the literature is filled with examples of remote sensing studies that show non-adiabatic behavior. Here is an example of MODIS data for example scene 1:

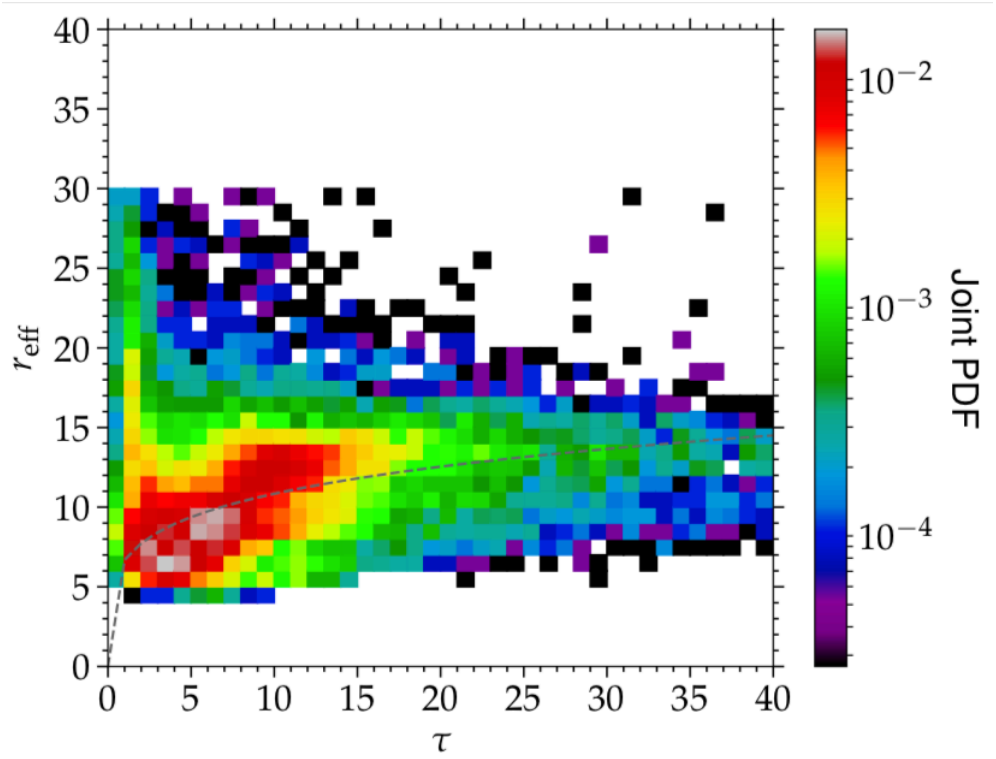

Fig 2: Example of τ-r_{eff} *relationships for example cloud scene 1.*

The grey dashed line indicates the adiabatic relationship reported by *Szczodrak et al.* (2001):

$$\ln(r_{\mathrm{eff}}) = \alpha \log(\tau) + \beta$$

There clearly are a multitude of data points not following that relationship. Here are two more examples from ASTER observations over altocumulus and from *Suzuki et al.* (2006):

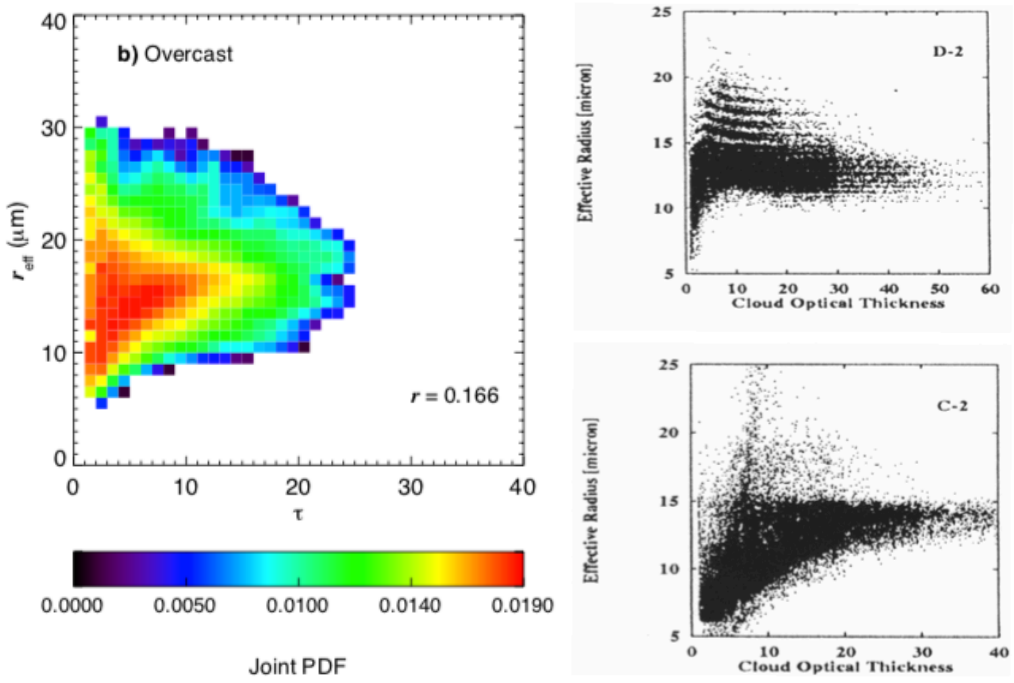

Fig 3: Examples for τ-r_{eff} relationships. (left panel) ASTER retrievals for a 50x50 km² scene. (right panel) From Suzuki et al. (2006).

It is overall not surprising that adiabatic assumptions are not ideal to describe all the different cloud types observed in the different cloud scenes shown in the manuscript.

p. 19, l. 25, "results provide strong evidence that simulateneous downscaling of the SWIR reflectances is essential": Again, why? Can you give a general explanation or guess?

Not downscaling the SWIR reflectance basically means that VNIR and SWIR reflectances exist at different spatial scales. Figure 8 in Werner et al. (2018b) compares ASTER SWIR reflectances at 240 m to artificially degraded (to 960 m) ones, as well as to values from the *Constant Reflectance Ratio Approach* (see Figure 4 below in this reviewer reply). There are significant deviations between the 240 m and 960 m SWIR reflectance, while the *Constant Reflectance Ratio Approach* provides a good estimate of the true 240 m results with a significantly reduced normalized root mean square deviation (nRMSD in that plot).

Naturally, assuming a wrong SWIR reflectance has a substantial impact on the r_{eff} retrieval, but even the cloud optical thickness will be impacted (because the isolines are generally not perpendicular; see the example SEVIRI LUT in this manuscript). It is therefore understandable that a retrieval with a scale mismatch should be avoided.
In the revised manuscript we added the following information:
"This confirms the findings in Werner et al. (2018b), who illustrated that SWIR reflectances differ significantly between the pixel-level and subpixel scale and that reliable cloud property retrievals should avoid scale mismatches between the reflectances from the VNIR and SWIR channels."

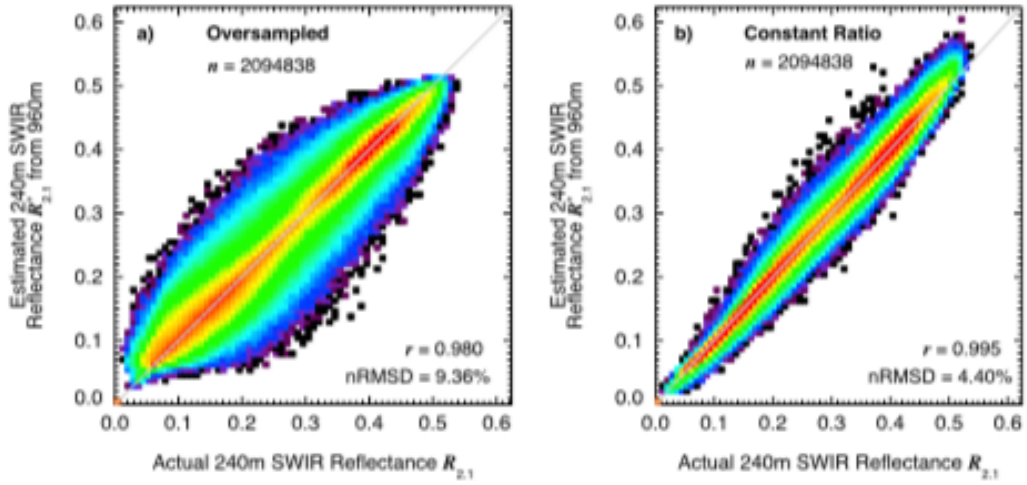

Fig 4: (a) Comparison between observed 240 m SWIR reflectances and 960 m observations, replicated to each 240 m subpixel. (b) Comparison between observed 240 m SWIR reflectances and the results of the Reflectance Ratio Approach. Adapted from Werner et al. (2018b).

p. 20, chapter: The conclusions section also needs more of this kind of general explanations and interpretations instead of repeating "x is better than y, so use x".

We extensively rewrote the conclusions section, especially the summary of the downscaling performance. We shortened the summary of the results and added some general interpretation instead.

Some of these explanations are listed below:

"This improvement can be attributed to the use of higher–resolution reflectances, which resolve the large–scale variability of the scene. It is shown that either downscaling approach, which applies estimates of the unresolved small–scale variability to the reflectance field, yields reliable retrievals of $\hat{\tau}$ at the horizontal resolution of the SEVIRI HRV channel. "

And:

"The former technique relies on large–scale statistical relationships between the reflectances, which might vary with the size of the observed region, prevalence of different cloud types and viewing geometry. The latter technique, meanwhile, was developed for optically thin clouds, where the relationship between VNIR and SWIR reflectance can be approximated by a linear function (Werner et al., 2018b). Conversely, for more homogeneous altocumulus fields the *LUT Approach* with *Adiabatic Adjustment* seems inadequate and yields the worst comparison to the reference effective radius. The study by *Miller et al.* (2016), following similar studies, illustrated that drizzle and cloud top entrainment yield vertical cloud profiles closer to homogeneous assumptions and away from the

adiabatic cloud model. Similar processes might affect the retrieval for the presented cloud scenes in this study. “

And:

“Due to the fact that these variables are derived from retrieved $\hat{\tau}$ and $\hat{r}_{eff}$, a similar behavior is observed for the derived $\hat{W}_L$ and $\hat{N}_D$. “

p. 21, l. 5/6: Many studies show that going below the 1 km scale might introduce new problems with variability which are smoothed out at this scale. Please discuss this caveat when making such a suggestion.

This suggestion was indeed not well written.

First of all, the spatial mismatch we mentioned is a direct result of the downscaling approach, which is the focus of our study (i.e., the resolution mismatch did not exist before downscaling the VNIR reflectances, yet we discussed downscaled reflectances for the purpose of this study). However, MODIS, e.g., does not have a spatial mismatch, because the VNIR data is aggregated to the horizontal resolution of the SWIR signal. We simply meant to say that if downscaling is performed, it is essential to also downscale the SWIR band reflectance, not just the VNIR band observations.

The second issue is that downscaling and retrieving at the VNIR resolution might put us close to the radiative smoothing scale; below that scale (about 200-400m, according to *Davis et al.*, 1997) the reflected field is characterized by enhanced radiative smoothing and the retrievals might be impacted by 3D radiative effects. Naturally, these facts need to be considered, before a decision about downscaling is made.

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 (≈ 200 − 400 m, see *Davis et al.*, 1997).“

Fig. 9, caption: Here "1b" is mentioned. Isn’t it "2b" in the text?

Thanks for noticing this mistake. It is indeed experiment 2b (just 2 in the revised manuscript). We corrected this mistake, which was also present in the caption of Figure 11.

Fig 10.: Please write out the experiments in words in addition to number codes.

We not only reduced the number of experiments shown (-2), but also included the experiment description in the caption of both Figure 10 and 12 (now 8 and 10 in the revised manuscript):

"… downscaling experiments 1 (statistical downscaling approach), 2 (*Constant Reflectance Ratio Approach*), and 3 (*Adjusted Lookup Table Approach with LUT Slope Adjustment*)…"

Fig. 6 and 13.: Needed?

We removed both Figure 4 and 6 in the revised manuscript. Both Figures indicated changes in the reflectance, but were not really needed for the retrieval comparison and conclusions. However, Figure 13 is needed to highlight the importance of simultaneous downscaling of the VNIR and SWIR reflectance.

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 5 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.

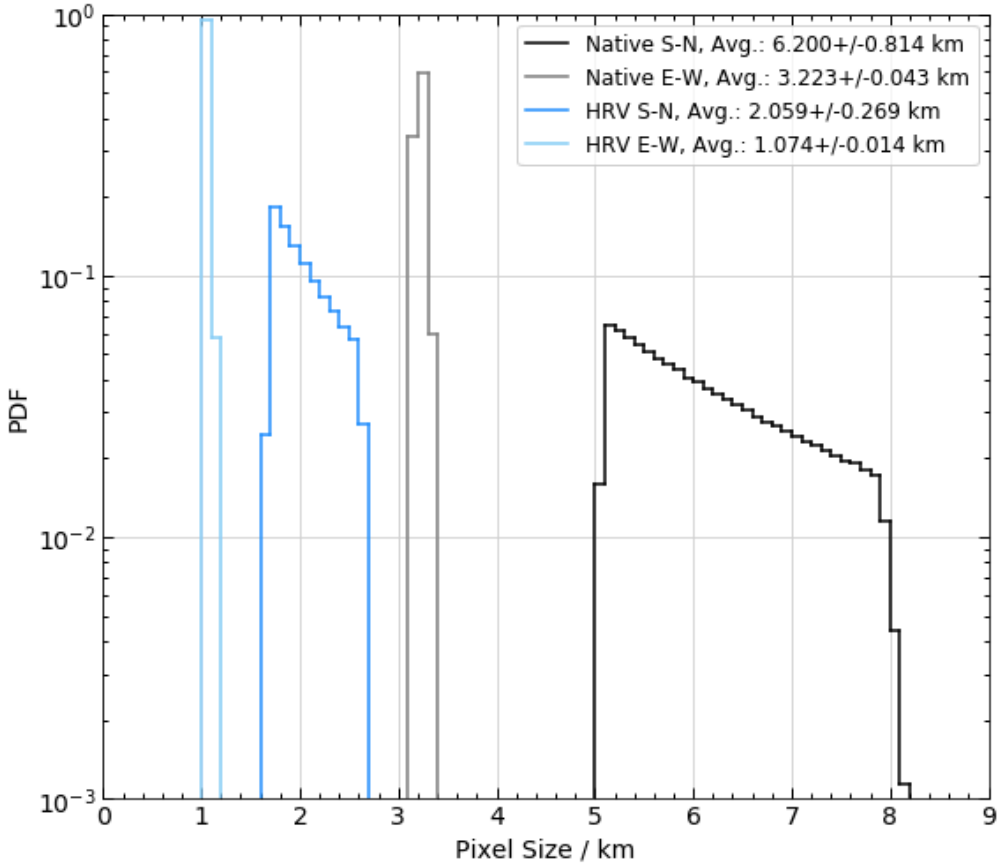

Fig 5: Statistics of SEVIRI pixel dimensions (in both latitude and longitude direction; i.e., south-north and east-west) for the native and HRV resolutions.

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 6.20×3.22 km^2 and 2.06×1.07 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 3×3km^2 and1×1km^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 $\lambda = 0.635$µm and $\lambda = 0.810$µm. SWIR reflectances (r16) are provided by channel 3 observations, which are centered around $\lambda = 1.640$ µm."

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 6 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.

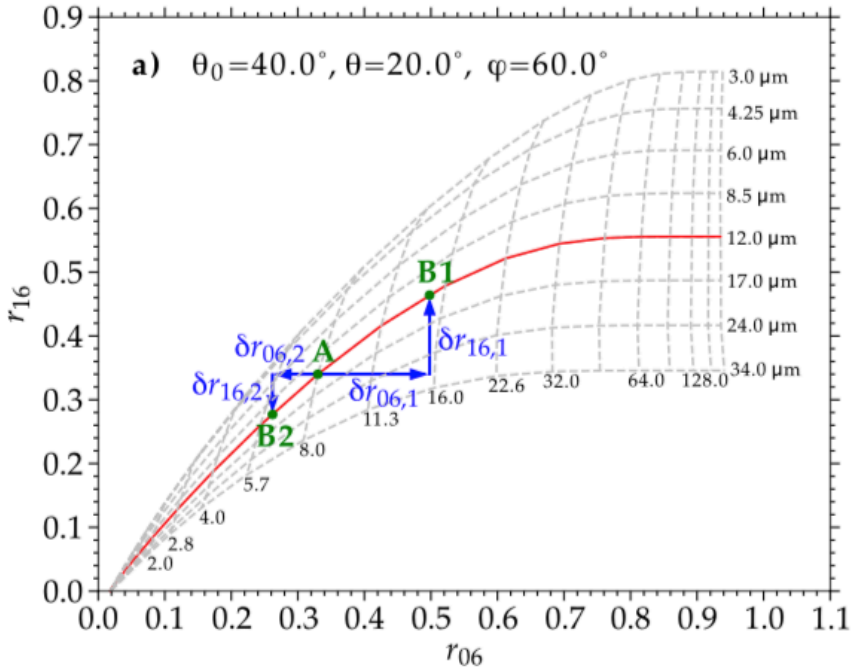

Fig 6: Example SEVIRI LUT. Isolines for constant τ and r_{eff} are shown in dashed gray lines.

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 reff 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 $\hat{r}_{\mathrm{eff}}$, as well as $\tilde{\tau}$ and $\tilde{r}_{\mathrm{eff}}$) 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 (≈ 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).

References

Suzuki, K., Nakajima, T., Nakajima, T. Y., and Khain, A.: Correlation pattern between effective radius and optical thickness of water clouds simulated by a spectral bin micro- physics cloud model, SOLA, 2, 116–119, 2006.

Szczodrak, M., Austin, P., and Krummel, P.: Variability of optical depth and effective radius in marine stratocumulus clouds, J. Atmos. Sci., 58, 2912–2926, 2001.

Werner, F., Zhang, Z., Wind, G., Miller, D. J., Platnick, S., and Di Girolamo, L.: Improving cloud optical property retrievals for partly cloudy pixels using coincident higher-resolution single band measurements: A feasibility study using ASTER observations, J. Geophys. Res. Atmos., 123, 12,253–12,276, https://doi.org/10.1029/2018JD028902, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JD028902, 2018b.

Increasing the spatial resolution of cloud property retrievals from Meteosat SEVIRI by use of its high–resolution visible channel: Evaluation of candidate approaches with MODIS observations

Frank Werner[1,2] and Hartwig Deneke[1]

[1]Leibniz Institute for Tropospheric Research, Permoserstraße 15, 04318 Leipzig, Germany
[2]Now at Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

Correspondence: Frank Werner (frank.werner@jpl.nasa.gov)

Abstract. This study presents and evaluates several candidate approaches for downscaling observations from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) in order to increase the horizontal resolution of subsequent cloud optical thickness (τ) and effective droplet radius (r_{eff}) retrievals from the native ~~$3 \times 3\,\mathrm{km}^2$~~ $\approx 3 \times 3\,\mathrm{km}^2$ spatial resolution of the narrowband channels to ~~$1 \times 1\,\mathrm{km}^2$~~$\approx 1 \times 1\,\mathrm{km}^2$. These methods make use of SEVIRI's coincident broadband high–resolution visible (HRV) channel. For four example cloud fields, the reliability of each downscaling algorithm is evaluated by means of collocated $1 \times 1\,\mathrm{km}^2$ MODIS radiances, which are re-projected to the horizontal grid of the HRV channel, and serve as reference for the evaluation. By using these radiances, smoothed with the ~~spatial response~~ modulation transfer function of the native SEVIRI channels, as retrieval input, the accuracy at the SEVIRI standard resolution can be evaluated and an objective comparison of the accuracy of the different downscaling algorithms can be made. For the example scenes considered in this study, it is shown that neglecting high-frequency variations below the SEVIRI standard resolution results in significant random absolute deviations of the retrieved τ and r_{eff} of up to ≈ 14 and $\approx 6\,\mu\mathrm{m}$, respectively, as well as biases. By error propagation, this also negatively impacts the reliability of the subsequent calculation of liquid water path (W_{L}) and cloud droplet number concentration (N_{D}), which exhibit deviations of up to $\approx 89\,\mathrm{g\,m^{-2}}$ and $\approx 177\,\mathrm{cm^{-3}}$, respectively. For τ, these deviations can be almost completely mitigated by the use of the HRV channel as a physical constraint, and by applying most of the presented downscaling schemes. ~~For the accuracy of~~ Uncertainties in retrieved r_{eff} ~~, the choice of downscaling scheme however is important: deviations are generally of similar magnitude or larger than those for retrievals at the SEVIRI standard resolution, indicative of their limited skill at predicting high–frequency spatial variability in~~ at the native SEVIRI resolution are smaller and the improvements from downscaling the observations are less obvious than for τ. Nonetheless, the right choice of downscaling scheme yields noticeable improvements in the retrieved r_{eff}. ~~A strong degradation of accuracy of r_{eff} is observed for some of the approaches, which also affects subsequent~~ Furthermore, the improved reliability in retrieved cloud products results in significantly reduced uncertainties in derived W_{L} and N_{D}~~estimates. As a result, an approach which constrains the r_{eff} to the lower–resolution results is recommended. Overall, this study demonstrates that an increase in horizontal resolution of SEVIRI cloud property retrievals can be reliably achieved by use of its HRV channel, yielding cloud properties which are preferable in terms of accuracy~~. In particular, one downscaling approach provides clear improvements for all cloud products compared to

those obtained from SEVIRI's standard-resolution and is recommended for future downscaling endeavors. This work advances efforts to mitigate impacts of scale mismatches among channels of multi–resolution instruments on cloud retrievals.

1 Introduction

In studies of the role of clouds in the climate system, the bispectral solar reflective method described by Twomey and Seton (1980); Nakajima and King (1990); Nakajima et al. (1991) is widely used to infer cloud optical and physical properties from satellite–based sensors. Based on observations of solar reflectance (r) from a channel pair at wavelengths with conservative scattering (usually around $0.6\,\mu\mathrm{m}$ or $0.8\,\mu\mathrm{m}$) and significant absorption by cloud droplets (common channels are $1.6\,\mu\mathrm{m}$, $2.2\,\mu\mathrm{m}$, and $3.7\,\mu\mathrm{m}$), respectively, this method simultaneously estimates the cloud optical depth (τ) and effective droplet radius (r_{eff}) of a sampled cloudy pixel. This method however relies on a number of assumptions which are often violated in nature: clouds are considered to be horizontally homogeneous and to have a prescribed vertical structure, which is generally assumed to be vertically homogeneous or to show a linear increase of liquid water content as predicted by adiabatic theory (see the discussions in Brenguier et al., 2000; Miller et al., 2016). Moreover, the observed cloud top reflectance field is usually described by one–dimensional (1D) plane–parallel radiative transfer, which neglects horizontal photon transport between neighboring atmospheric columns.

Use of the independent pixel approximation (IPA, see Cahalan et al., 1994a, b) produces uncertainties in the retrieved cloud variables that are dependent upon the horizontal resolution of the observing sensor. For sensors with a high spatial resolution, the observations resolve the actual cloud heterogeneity, which are unaccounted for in the IPA approach. This usually results in an overestimation of both τ and r_{eff}, as reported in Barker and Liu (1995); Chambers et al. (1997); Marshak et al. (2006). Conversely, for observations with a low spatial resolution, the actual cloud heterogeneity cannot be resolved. Moreover, the chances of clear–sky contamination within a cloudy pixel increase with increasing spatial resolution. As a result, an underestimation (overestimation) of retrieved τ (r_{eff}) is usually observed (Marshak et al., 2006; Zhang and Platnick, 2011; Zhang et al., 2012; Werner et al., 2018b). These uncertainties are propagated to the liquid water content (W_{L}) and the droplet number concentration (N_{D}), which can be estimated from retrieved τ and r_{eff}. Estimates of N_{D} are especially susceptible to uncertainties in r_{eff}, which impacts the reliability of aerosol–cloud–interaction studies (Grosvenor et al., 2018). The analysis in Varnai and Marshak (2001) suggests that a horizontal scale of around $1-2\,\mathrm{km}$ minimizes the combined uncertainty from unresolved and resolved cloud heterogeneity. While strategies to mitigate the effects of unresolved cloud variability have been recently reported in Zhang et al. (2016); Werner et al. (2018a), these techniques become less successful with lower–resolution sensors like those operated on geostationary satellites.

Remote sensing from geostationary platforms such as the Meteosat Spinning Enhanced Visible and Infrared Imager (SEVIRI) offers unique capabilities for cloud studies not available from polar orbiting satellites. These advantages include more

frequent temporal sampling of individual regions and the ability to capture the temporal evolution (Bley et al., 2016; Senf and Deneke, 2017) and diurnal cycle of cloud parameters (Stengel et al., 2014; Bley et al., 2016; Martins et al., 2016; Seethala et al., 2018). However, SEVIRI pixels are characterized by a lower spatial resolution of its narrow–band channels compared to other operational remote sensing instrumentation, like the Moderate Resolution Imaging Spectroradiometer (MODIS, Platnick et al., 2003) or the Visible Infrared Imaging Radiometer Suite (VIIRS, Lee et al., 2006). Given the increase in retrieval uncertainty due to the IPA constraints, there is a desire to increase the resolution for geostationary cloud observations.

The aim of this manuscript is to critically evaluate several candidate approaches for downscaling of the SEVIRI narrow–band reflectances for operational usage and to identify the most promising of these schemes, exploiting the fact that information on small-scale variability is available from its broadband high–resolution visible (HRV) channel. ~~Of main concern is the ability to accurately capture information on the small–scale reflectance variability in the 1.6 μm–channel, which predominantly arises from variations in effective droplet radius. Conversely, cloud optical depth is expected to be well–constrained by the HRV channel , as it~~ The study by Deneke and Roebeling (2010) presented a statistical downscaling approach of the SEVIRI channels in the visible to near-infrared (VNIR) spectral wavelength range. This method makes use of the fact that SEVIRI's high–resolution channel can be modelled by a linear combination of the $0.6\,\mu$m and $0.8\,\mu$m channels with good accuracy (Cros et al., 2006). This ~~situation is similar to that found with other satellite instruments featuring multiple resolutions for the conservative and absorbing channels, such as the MODIS instrument (with 250 m resolution versus 500 m for 1.6 μm or 1 km for 2.1 μm), VIIRS (375 m versus 750 m) , and GOES–R~~ study advances these efforts in three ways: (i) it explores other possible downscaling approaches, which might improve upon the statistical downscaling scheme, (~~500 m versus 1 km). Therefore, we believe that our findings are also relevant there. This work is a companion paper to Deneke et al. (2019), which describes the overall retrieval scheme for obtaining cloud properties and solar radiative fluxes from the Meteosat SEVIRI instrument at the spatial resolution of its HRV channel, which will be established based on the findings of this study. The companion paper also presents an important extension of this approach to the retrieval of solar surface irradiance, based on the schemes presented in Deneke et al. (2008) and Greuell et al. (2013). Satellite products with high temporal and spatial resolution are of particular interest for forecasting the production of solar power.~~ ii) it introduces techniques to accurately capture information on the small–scale reflectance variability in the 1.6 μm–channel, which predominantly arises from variations in effective droplet radius, and (iii) it studies the impact of the various downscaling techniques on the subsequently retrieved cloud properties.

A critical requirement, formulated at the start of this work, is to maintain a target accuracy for the retrieved effective radius based on the lower–resolution observations, while hoping for further improvements. This goal was set because the error in effective radius will propagate into other cloud products such as vertically integrated liquid or ice water path or the cloud droplet number concentration, thereby potentially corrupting any gains in accuracy obtained from the improved spatial resolution. However, without an independent reference data set, it is impossible to determine whether this target can be met. Thus, higher–resolution reflectance observations from Terra–MODIS are remapped to SEVIRI's HRV and standard resolution grids here as basis for a thorough evaluation of the accuracy of the retrieved cloud parameters. This allows us to objectively benchmark the accuracy of candidate approaches by comparison of results from a true ~~1 km resolution~~ $\approx 1\,$km–resolution reflectance data set, and processed with an identical retrieval scheme. Note, that even the retrieved cloud products from a hypothetically perfect

downscaling technique would still be impacted by the effects of resolved and unresolved cloud variability. Therefore, the results of this study will not help to mitigate the uncertainties associated with the retrieval schemes of similar $\approx 1\,\mathrm{km}$–sensors (e.g., clear–sky contamination, plane–parallel albedo bias, 3–dimensional radiative effects).

The results of this study are relevant for many other passive satellite sensors, which, like the SEVIRI instrument, feature multiple resolutions for the conservative and absorbing channels. Similar configurations exists for the MODIS instrument ($250\,\mathrm{m}$ horizontal resolution versus $500\,\mathrm{m}$ for the $0.6\,\mu\mathrm{m}$ and $2.1\,\mu\mathrm{m}$ channels, respectively), VIIRS ($375\,\mathrm{m}$ versus $750\,\mathrm{m}$), and GOES–R ($500\,\mathrm{m}$ versus $1\,\mathrm{km}$).

The structure of the paper is as follows: ~~section~~ Section 2 describes both the SEVIRI and MODIS instruments used as basis for this study, as well as the covered observational domain. A brief overview of the SEVIRI cloud property retrieval algorithm is given in ~~section~~ Section 3, followed by a description of the different candidate approaches for the downscaling of the narrow–band SEVIRI channel observations in ~~section~~ Section 4. An example of lower– and higher–resolution cloud property retrievals is presented in ~~section 5. Finally, a~~ Section 5. A statistical evaluation of the different downscaling approaches based on remapped MODIS observations ~~is given in section~~ follows in Section 6 for a limited number of example cloud fields. Finally, a comparison between a full downscaling scheme and a VNIR–only approach (similar to Deneke and Roebeling, 2010) is given in Section 7. The manuscript presents the main conclusions and an outlook in ~~section 7.~~ Section 8.

2 Data

This section gives an overview of both the SEVIRI and MODIS instruments in ~~section~~ Section 2.1 and 2.2. Here, the respective spectral channels of interest for this study are listed. Subsequently, the observational domain is described in ~~section~~ Section 2.3.

2.1 SEVIRI

The current version of European geostationary satellites is the Meteosat Second Generation, which has provided operational data since 2004 (Schmetz et al., 2002). The SEVIRI imager is installed aboard the Meteosat–8 to Meteosat–11 platforms, which are positioned above longitudes of $9.5°\mathrm{E}$ and $0.0°$ longitude, respectively. One SEVIRI instrument samples the full disk of the Earth from $0.0°$ longitude with a temporal resolution of fifteen minutes. However, a backup satellite positioned at $9.6°\mathrm{E}$ also scans a Northern subregion with a temporal resolution of five minutes (the so–called Rapid Scan Service). These samples – in our case from Meteosat–9 – provide the observational SEVIRI data set for the following analysis.

This study mainly considers observations from SEVIRI's solar reflectance channels 1–3, as well as from the spectrally broader HRV band. These channels cover the ~~visible to near–infrared (VNIR)~~ VNIR and shortwave-infrared (SWIR) spectral wavelength ranges. The two VNIR reflectances (r_{06} and r_{08}) are sampled in bands 1 and 2, respectively, and are centered around wavelengths $\lambda = 0.635\,\mu\mathrm{m}$ and $\lambda = 0.810\,\mu\mathrm{m}$. SWIR reflectances ($r_{16}$) are provided by channel 3 observations, which are centered around $\lambda = 1.640\,\mu\mathrm{m}$. The horizontal resolution of the channel 1–3 samples is $3 \times 3\,\mathrm{km}^2$ at the sub–satellite point and increases with higher sensor zenith angles. Conversely, the broadband reflectances r_{HV} are sampled at SEVIRI's

HRV channel at a horizontal scale of $1 \times 1\,\mathrm{km}^2$ at the sub–satellite point. These observations cover the spectral range of $0.4 - 1.1\,\mu\mathrm{m}$.

As context for the present study, the reader is reminded that the spatial resolution of geostationary satellites is significantly reduced at higher latitudes due to the oblique viewing geometry. For Germany and Central Europe as considered in this paper, the pixel size is effectively increased by a factor of two in North–South direction as a result. In addition, the distinction between sampling and optical resolution needs to be acknowledged. While the former determines the distance between recorded samples, the latter is given by the effective area of the optical system, which is larger by a factor of 1.6 than the sampling resolution for SEVIRI (Schmetz et al., 2002). The spatial response of optical systems is commonly characterized by their modulation transfer function, which describes the response of the optical system in the frequency domain.

Further information about the spectral width of each ~~channel and the respective spectral and spatial response functions~~ SEVIRI channel, as well as the respective spatial response and modulation transfer functions, can be found in Deneke and Roebeling (2010).

2.2 Terra–MODIS

The 36–band scanning spectroradiometer MODIS, which was launched aboard NASA's Earth Observing System satellites Terra and Aqua, has a viewing swath width of $2,330\,\mathrm{km}$, yielding global coverage every two days. MODIS collects data in the spectral region between $0.415 - 14.235\,\mu\mathrm{m}$, covering the VNIR to thermal–infrared spectral wavelength range. In general, the spatial resolution at nadir of a MODIS pixel is $1,000\,\mathrm{m}$ for most channels, although the pixel dimensions increase towards the edges of a MODIS granule. Only observations from the Terra satellite launched in 1999 are used here, due to broken detectors of the $1.64\,\mu\mathrm{m}$ channel of the MODIS instrument on the Aqua satellite. Information on MODIS and its cloud product algorithms is given in (Ardanuy et al., 1992; Barnes et al., 1998; Platnick et al., 2003). The current version of the level 1b radiance and level 2 cloud products used is Data Collection 6.1 (C6.1).

2.3 Domain

In this study, data from a subregion of the full SEVIRI disk has been selected. This region, which is located within the European subregion described in Deneke and Roebeling (2010), is illustrated by the red borders in Figure 1. It is centered around Germany due to its intended domain of application (thus, from here on it is referred to as Germany domain) and comprises the latitude and longitude ranges of $\approx 44.30 - 57.77°$ and $\approx -0.33 - 21.65°$, respectively. This domain includes 240×400 lower–resolution pixels (i.e., ~~samples at a horizontal resolution of $3 \times 3\,\mathrm{km}^2$~~ the native SEVIRI resolution of channels 1–3) and is far away from the edges of the full SEVIRI disk, ensuring that the observed viewing zenith angles are $< 70°$.

Due to the increased sensor zenith angles the spatial resolution of each SEVIRI pixel is degraded. The average pixel size is $6.20 \times 3.22\,\mathrm{km}^2$ and $2.06 \times 1.07\,\mathrm{km}^2$ for channels 1–3 and the HRV channel, respectively. To avoid confusion, we will use the desingations LRES (abbreviation for lower–resolution) and HRES (abbreviation for higher–resolution) scales to refer to the $3 \times 3\,\mathrm{km}^2$ and $1 \times 1\,\mathrm{km}^2$ pixel resolutions from here on.

A relatively small domain was chosen, because the number of pixels to be processed will expand by a factor of 3×3, increasing the computational costs of the subsequent cloud property retrievals by roughly one order of magnitude. Except for some regional dependencies introduced by changes in the prevalence of specific cloud types, we expect results of our study to also be valid for other domains.

3 SEVIRI cloud property retrieval algorithm

Retrieved cloud variables in this study are provided by the Cloud Physical Properties retrieval algorithm (CPP; Roebeling et al., 2006), which is developed and maintained at the Royal Dutch Meteorological Institute (KNMI). It is used as basis for the CLAAS–1 and CLAAS–2 climate data records (Stengel et al., 2014; Benas et al., 2017) distributed by the Satellite Application Facility on Climate Monitoring (Schulz et al., 2009). Using a lookup table (LUT) of reflectances simulated by the Doubling–Adding KNMI (DAK: Smith and Timofeyev, 2001) radiative transfer model, observed and simulated reflectances at $0.6\,\mu\mathrm{m}$ and $1.6\,\mu\mathrm{m}$ are iteratively matched to yield estimates of τ and r_{eff}. The CPP retrieval uses the cloud mask and cloud top height products obtained from the software package developed and distributed by the satellite application facility of Support to Nowcasting and Very Short Range Forecasting (NWCSAF), Version 2016, as input (Le Gléau, 2016). The former product identifies cloudy pixels for the retrieval, while the information on the height of the cloud is used to account for the effects of gas absorption in the SEVIRI channels. ~~An improved cloud detection scheme for the resulting higher–resolution SEVIRI retrievals based on the HRV channel based on Bley and Deneke (2013), with modifications described in Deneke et al. (2019) (i.e., the companion paper that describes the final retrieval algorithm), has been integrated into the retrieval, but has not been used for this study.~~

~~For obtaining the results presented in this study, an experimental version of the retrieval that was developed in a separate branch has been used. This algorithm deviates in some aspects from the setup described in the companion paper. Specifically, it uses the default climatology of ancillary data sets available as part of the CPP retrieval system, which have a lower horizontal resolution and do not match the specific time of the retrieval. This is expected to have only minor influence on the results presented here, because the absolute accuracy of the retrieval is not the primary focus of this study.~~

4 Candidate methods for downscaling SEVIRI reflectances

This section describes the necessary steps to convert the reflectances r_{06}, r_{08}, and r_{16}, available at ~~the native SEVIRIresolution of $3 \times 3\,\mathrm{km}^2$~~ SEVIRI's native LRES resolution, to reliable estimates of higher–resolution reflectances $\hat{r}_{06}$, $\hat{r}_{08}$, and $\hat{r}_{16}$, together with matching cloud properties, at the ~~spatial scale of $1 \times 1\,\mathrm{km}^2$ of~~ HRES scale of the HRV channel. This downscaling process utilizes the high–resolution r_{HV} observations.

As a first step, all reflectances are interpolated to the HRV grid using trigonometric interpolation, implemented based on the discrete Fourier transform and multiplication with the modulation transfer function (see Deneke and Roebeling, 2010, for details). While this step increases the spatial sampling resolution, it does not add any additional high–frequency variability. In

fact, after interpolation, the reflectance values of the central pixel of each 3×3 pixel block equal those of the corresponding standard–resolution pixel reflectances. However, the pixels apart from the central one contain information about the large–scale reflectance variabilty and can be considered as a baseline high–resolution approach. This approach already improves the agreement with true higher–resolution retrievals, as will be shown later in this study.

Three conceptually different downscaling techniques to improve upon this baseline method are described: (i) a statistical downscaling approach based on globally determined covariances between the SEVIRI reflectances in ~~section~~ Section 4.1, (ii) a local method based on assumptions about the ratio of reflectances at different scales in ~~section~~ Section 4.2, and (iii) a technique combining globally determined covariances between the VNIR reflectances and the shape of the SEVIRI LUT, while assuming a constant r_{eff} within a standard SEVIRI pixel in order to constrain the SWIR reflectance in ~~section~~ Section 4.3. As variations of this last technique, two additional approaches are considered to improve upon the constant r_{eff} constraint in ~~section~~ Section 4.4. As will be shown, each of these approaches has advantages and disadvantages, and the impact on the cloud property retrievals will be evaluated in ~~section~~ Section 6 for a number of example scenes by means of collocated MODIS observations.

~~The~~ As is discussed in Sections 4.1–4.4, the derived reflectances $\hat{r}_{06}$ and $\hat{r}_{08}$, as well as $\hat{r}_{16}$, include an estimate of the spectrally dependent, high–frequency variability of an image, and are based on the actually observed r_{HV}. These reflectances are different from those obtained by trigonometric interpolation of the respective channel observations at the native scale to the horizontal resolution of the HRV channel (i.e., the baseline approach), which are denoted by $\tilde{r}_{06}$, $\tilde{r}_{08}$, and $\tilde{r}_{16}$. While these latter variables also have a higher horizontal resolution of ~~$1 \times 1\,\mathrm{km}^2$~~ the HRV channel, they only capture the low–frequency variability resolved by ~~the SEVIRI sensor~~ SEVIRI channels 1–3.

4.1 Statistical downscaling

The statistical downscaling algorithm for the two SEVIRI VNIR reflectances was first reported in Deneke and Roebeling (2010) and assumes a least-squares linear model that links r_{06} and r_{08} to the reflectances in the HRV channel (see Cros et al., 2006) in the form:

$$\langle \tilde{r}_{\mathrm{HV}} \rangle = a \cdot r_{06} + b \cdot r_{08}. \tag{1}$$

Here, the HRV channel observations are first ~~filtered with the spatial response~~ smoothed with the modulation transfer function of the lower–resolution channels, which yields reflectances $\tilde{r}_{\mathrm{HV}}$ at the same ~~$1 \times 1\,\mathrm{km}^2$~~ HRES horizontal resolution, adjusted to the low–frequency variability at the spatial scale of the channel 1–3 observations. Subsampling the central pixel of each $3 \times 3 = 9$ pixel block subsequently yields $\langle \tilde{r}_{\mathrm{HV}} \rangle$ at the same ~~$3 \times 3\,\mathrm{km}^2$~~ LRES horizontal resolution as r_{06} and r_{08} (here, the subsampling of the field is denoted by $\langle\rangle$). The variables a and b are fit coefficients that are determined empirically by a least–squares linear fit. In order to derive a statistically significant and stable linear model, the coefficients a and b are calculated hourly between $08:00 - 16:00$ UTC within 16–day intervals. Results for the time step $08:00$ UTC are derived from 5–minute SEVIRI rapid–scan data between $08:00 - 08:25$ UTC, while the $16:00$ UTC time step is comprised of SEVIRI observations between $15:30 - 16:00$ UTC. For all time steps in between, data is from all samples after minute 25 of the prior hour up to

minute 25 of the current hour (e.g., fit coefficients for time step $09:00$ UTC are calculated from SEVIRI observations between $08:30-09:25$ UTC).

Values of hourly–derived fit coefficients for the Germany domain between 1 April and 31 July 2013 are shown in Figure 2(a) and 2(b) for a and b, respectively. Here, circles represent the respective fit coefficient for each 16–day interval, which is indicated by the first Julian day in the time period. Colors highlight the different UTC time steps. It is obvious that both coefficients a and b are very stable and show no noticeable variation from hour to hour, as well as from one 16–day interval to another. ~~The~~ Considering all hourly data and each 16–day interval, the median fit coefficients are 0.63 (for a) and 0.40 (for b), with low interquartile ranges (IQR) of 0.03. The only exceptions are the fit coefficients derived for the first time period of 1–17 April 2013, especially for the morning and afternoon hours of $08:00-09:00$ and ~~$15:00-16:00$~~$13:00-16:00$ UTC. Here, a and b deviate significantly from the other results, with values of ≈ 0.50 and ≈ 0.52, respectively, likely due to an abundance of observations with ~~a~~large solar zenith angles of $\theta_0 > 60°$ in the eastern part of the domain.

The high–frequency reflectance variations for the SEVIRI HRV channel (δr_{HV}) are calculated as the difference between the observed r_{HV} and $\tilde{r}_{\mathrm{HV}}$, which only ~~resolves~~ resolve the low–frequency variability:

$$\delta r_{\mathrm{HV}} = r_{\mathrm{HV}} - \tilde{r}_{\mathrm{HV}}. \tag{2}$$

Following the linear model in Eq.(1), the high–frequency variations of the channel 1 and 2 reflectances (δr_{06} and δr_{08}) are linked to δr_{HV} via:

$$\begin{aligned} \delta r_{06} &= S_{06} \cdot \delta r_{\mathrm{HV}} \\ \delta r_{08} &= S_{08} \cdot \delta r_{\mathrm{HV}}. \end{aligned} \tag{3}$$

The optimal slopes S_{06} and S_{08}, which minimize the least–squares deviations, can be derived from bivariate statistics:

$$\begin{aligned} k_1 &= \sqrt{\frac{b^2 \cdot \mathrm{var}(r_{08})}{a^2 \cdot \mathrm{var}(r_{06})}} \\ S_{06} &= \frac{1 + k_1 \cdot \mathrm{cor}(r_{06}, r_{08})}{a \cdot \left[1 + {k_1}^2 + 2k_1 \cdot \mathrm{cor}(r_{06}, r_{08})\right]} \\ k_2 &= \sqrt{\frac{a^2 \cdot \mathrm{var}(r_{06})}{b^2 \cdot \mathrm{var}(r_{08})}} \\ S_{08} &= \frac{1 + k_2 \cdot \mathrm{cor}(r_{08}, r_{06})}{b \cdot \left[1 + {k_2}^2 + 2k_2 \cdot \mathrm{cor}(r_{08}, r_{06})\right]}. \end{aligned} \tag{4}$$

Here, $\mathrm{cor}(r_{06}, r_{08})$ is the linear correlation coefficient between the channel 1 and 2 reflectances, while $\mathrm{var}(r_{06})$ and $\mathrm{var}(r_{08})$ are the spatial variances of the respective samples. Note, that the sampling resolution of all reflectances is ~~$3 \times 3\,\mathrm{km}^2$.~~ the LRES scale (i.e., $\approx 3 \times 3\,\mathrm{km}^2$).

As a result, the high–resolution reflectances $\hat{r}_{06}$ and $\hat{r}_{08}$, which include the high–frequency variations, can be derived from the interpolated reflectances as:

$$\begin{aligned} \hat{r}_{06} &= \tilde{r}_{06} + \delta r_{06} \\ \hat{r}_{08} &= \tilde{r}_{08} + \delta r_{08}. \end{aligned} \tag{5}$$

Note, that only $\hat{r}_{06}$ is used for the retrieval.

Similar steps can be applied for the calculation of $\hat{r}_{16}$. Again, a simple linear model is assumed to connect r_{16} to the lower–resolution $\langle \tilde{r}_{\mathrm{HV}} \rangle$ at the spatial scales of the channel 1–3 observations:

$$\langle \tilde{r}_{\mathrm{HV}} \rangle = c \cdot r_{16}. \tag{6}$$

The symbol c is used to denote the respective fit coefficient, which needs to be determined empirically. Similar to the coefficients a and b from the linear model for the VNIR reflectances, c is calculated hourly between $08:00 - 16:00$ UTC within 16–day intervals. It has to be noted, however, that in contrast to the VNIR reflectances, this fit does not have a clear physical motivation, as there is no spectral overlap with the HRV channel.

The temporal behavior of the fit coefficient c for the Germany domain for the time period between 1 April and 31 July 2013 is shown in Figure 2(c). In contrast to the coefficients a and b, there is a noticeable trend in the data, both diurnally and during the transition from 1 April to 31 July. ~~Diurnally,~~ For each 16–day interval the variability in the hourly derived c values ranges between $\mathrm{IQR} = 0.05 - 0.15$, while the median 16–day value varies between 1.04 and 1.25. Overall, the median c is 1.16, with an IQR of 0.08 (i.e., almost three times larger than the one for the coefficients a and b). The observed trends and larger IQR in the c data set shown in Figure 2(c) illustrate that the linear model in Eq.(6) is not ideal, and is expected to introduce significant uncertainties in the calculation of $\tilde{r}_{16}$. 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.

Values of $\tilde{r}_{16}$ can be derived similarly to Eqs.(3–5) for the channel 1 and 2 observations:

$$\begin{aligned} \delta r_{16} &= S_{16} \cdot \delta r_{\mathrm{HV}} \\ S_{16} &= \frac{\mathrm{cov}(r_{16}, \langle \tilde{\mathrm{r}}_{\mathrm{HV}} \rangle)}{\mathrm{var}(r_{16})} \\ \hat{r}_{16} &= \tilde{r}_{16} + \delta r_{16}. \end{aligned} \tag{7}$$

Note, that the use of linear models and bivariate statistics means that the downscaling algorithm described in this section is an example of statistical downscaling techniques, which are common in climate science applications (e.g., Benestad, 2011). While for the VNIR channels the spectral overlap with the HRV channel and the spectrally flat properties of clouds provide a sound physical justification for this technique, this is not the case for the SWIR channel.

The reliability of the linear model in Eq.(1) depends upon the correlation between channel 1 and 2 reflectances (i.e., $\mathrm{cor}(r_{06}, r_{08})$), as well as the stability of the fit coefficients a and b. The analysis in Deneke and Roebeling (2010) concludes that the explained variance in the estimates of $\hat{r}_{06}$ and $\hat{r}_{08}$ ~~are~~ is close to 1, corresponding to low residual variances, which

indicates that the linear model is robust. Moreover, the two fit coefficients are found to exhibit very low variability, as shown in Figures 2(a)–(b).

To verify the reliability of the linear model with a large SEVIRI data set, ~~joint PDFs~~ a joint PDF of the actually observed $\langle \tilde{r}_{\mathrm{HV}} \rangle$ and the results from Eq.(1) ~~are shown in Figures~~ is shown in Figure 3(a) ~~(b)~~; data is from all SEVIRI observations within the Germany domain during June 2013. In case of an ideal linear model, as well as a perfect correlation between the two reflectances, Eq.(1) would replicate the $\langle \tilde{r}_{\mathrm{HV}} \rangle$ observations. Conversely, deviations from these assumptions will yield different results from the sampled SEVIRI reflectances. It is clear that the linear model can reliably reproduce $\langle \tilde{r}_{\mathrm{HV}} \rangle$, as most of the observations lie on the 1:1 line, and Pearson's product–moment correlation coefficient (R) is $R = 0.999$. While some larger deviations exist, such occurrences are significantly less likely (i.e., the joint probability density is several orders of magnitude lower than the most–frequent occurrences along the 1:1 line). Regarding r_{16}, the assumption of a linear model is evidently flawed, because the relationship between VNIR and SWIR reflectances depends on the optical and microphysical cloud properties. As a result, a single linear slope, which describes the whole relationship between the two reflectances for all distributions of cloud properties, will introduce significant uncertainties. This is illustrated in Figure 3(~~c~~b), where the Joint PDF of $\langle \tilde{r}_{\mathrm{HV}} \rangle$ and the results from the linear model in Eq.(6) are shown. The comparison between the two data sets reveals a much larger spread around the 1:1 line and a lower correlation coefficient. Overall, the relationship resembles the shape of a LUT, displayed in form of the well–known diagram introduced by Nakajima and King (1990), where changes in r_{eff} result in a spread in the observed SWIR reflectances (see, e.g., Werner et al., 2016).

To test the impact of changes in a and b on the derived $\hat{r}_{06}$ and $\hat{r}_{08}$, two experiments are conducted: (i) the fit coefficients are derived only from cloudy pixels and are compared to the higher–resolution results from a and b, which are derived for all pixels~~.~~ ; and (ii) the Germany domain is divided into $100 \times 100\,\mathrm{km}^2$–subscenes and the fit coefficients are derived more locally within each subscene instead of globally from the full domain. Subsequently, statistics from the difference between the two data sets are calculated. Data is from 14 June 2013 at 14:05 UTC. For experiment (i), the 1^{st}, 50^{th}, and 99^{th} percentiles of the relative difference in $\hat{r}_{06}$ (defined as the difference between the reflectances from only cloudy data and the full data set, normalized by the full data set) are $-0.08, -0.02, 0.03\%$, while for $\hat{r}_{08}$ the analysis yields $-0.04, 0.02, 0.19\%$. Similarly, experiment (ii) yields relative differences of $-0.08, 0.03, 0.36\%$ and $-0.17, 0.00, 0.19\%$ for $\hat{r}_{06}$ and $\hat{r}_{08}$, respectively. These deviations are negligible compared to the measurement uncertainty and naturally, the correlation coefficients between the different data sets are $R \approx 1.00$. This confirms the robustness of the linear model described in Eq.(1). For the derivation of $\hat{r}_{16}$ from Eq.(6), a slightly increased sensitivity to the fit coefficient c is observed. Here, experiment (i) yields percentiles of the relative difference of $-0.16, 0.08, 0.86\%$, whereas experiment (ii) results in $-0.39, -0.01, 0.40\%$. While slightly higher deviations are observed compared to the linear model for the VNIR reflectances, the uncertainty in $\hat{r}_{16}$ induced by the variability in c is still significantly lower than the measurement uncertainty.

4.2 Constant Reflectance Ratio Approach

Compared to the downscaling approach in ~~section~~ Section 4.1, where fit coefficients for a linear model are derived over a large temporal and spatial domain, this second method uses local relationships (i.e., on the pixel level) between the SEVIRI

reflectances. The *Constant Reflectance Ratio Approach* was introduced by Werner et al. (2018b) and is based on the assumption that the inhomogeneity index of the HRV reflectance ($H_{\sigma,\mathrm{HV}}$, defined as the ratio of standard deviation σ_{HV} to the average, pixel–level reflectance $\langle\tilde{r}_{\mathrm{HV}}\rangle$) equals that for the channel 1 reflectance ($H_{\sigma,06}$). This implies a spectrally consistent subpixel reflectance variability. The relationship can be written as:

$$\begin{aligned} H_{\sigma,06} &= H_{\sigma,\mathrm{HV}} \\ \frac{\sigma_{06}}{r_{06}} &= \frac{\sigma_{\mathrm{HV}}}{\langle\tilde{r}_{\mathrm{HV}}\rangle} \\ \frac{\sqrt{\frac{1}{9-1}\cdot\sum_{i=1}^{i=9}(\hat{r}_{06,\mathrm{i}}-r_{06})^2}}{r_{06}} &= \frac{\sqrt{\frac{1}{9-1}\cdot\sum_{i=1}^{i=9}(r_{\mathrm{HV,i}}-\langle\tilde{r}_{\mathrm{HV}}\rangle)^2}}{\langle\tilde{r}_{\mathrm{HV}}\rangle}, \end{aligned} \tag{8}$$

where the index $i = 1, 2, \ldots, 9$ indicates any one of the nine available ~~$1 \times 1\,\mathrm{km}^2$ subpixels~~ HRES–subpixels within a lower–resolution SEVIRI pixel (i.e., at ~~a scale of $3 \times 3\,\mathrm{km}^2$~~ the LRES scale of channels 1–3). This relationship can be further simplified, assuming that this relationship is also true for individual pixels:

$$\begin{aligned} \frac{\hat{r}_{06,\mathrm{i}}-r_{06}}{r_{06}} &= \frac{r_{\mathrm{HV,i}}-\langle\tilde{r}_{\mathrm{HV}}\rangle}{\langle\tilde{r}_{\mathrm{HV}}\rangle} \\ \frac{\hat{r}_{06,\mathrm{i}}}{r_{\mathrm{HV,i}}} &= \frac{r_{06}}{\langle\tilde{r}_{\mathrm{HV}}\rangle}. \end{aligned} \tag{9}$$

The relationship in Eq.(9) suggests that the ratio of channel 1 and HRV reflectances (i.e., narrowband and broadband VNIR reflectances) remains constant for different scales. Thus, this approach is called the *Constant Reflectance Ratio Approach*.

Finally, we can mitigate some of the scale effects by substituting the lower–resolution variables with the higher–resolution reflectances that resolve the low–frequency variability (i.e., $\tilde{r}_{06}$ and $\tilde{r}_{\mathrm{HV}}$) and solve for $\hat{r}_{06}$:

$$\hat{r}_{06} = r_{\mathrm{HV}} \cdot \frac{\tilde{r}_{06}}{\tilde{r}_{\mathrm{HV}}}. \tag{10}$$

Similarly, higher–resolution SWIR reflectances $\hat{r}_{16}$ can be derived from:

$$\hat{r}_{16} = r_{\mathrm{HV}} \cdot \frac{\tilde{r}_{16}}{\tilde{r}_{\mathrm{HV}}}. \tag{11}$$

As before, the relationship implies that the ratio of VNIR and SWIR reflectances remains constant for different scales. This assumption has been shown to be reasonable, at least for ~~liquid water clouds over the ocean (Werner et al., 2018b).~~

~~A comparison of $\hat{r}_{06}$ and $\hat{r}_{16}$ from statistical downscaling and the *Constant Reflectance Ratio Approach* is presented in Figures **??**(a)–(b), respectively. For both $\hat{r}_{06}$ and $\hat{r}_{16}$ the majority of data points is positioned along the 1:1 line, and the correlation coefficient is $R \approx 1.00$. The derived reflectances from the two independent approaches are very similar, and the probability density of the few larger deviations is several orders of magnitude below the maximum probability. There are a limited number of occurrences where $\hat{r}_{06}$ and $\hat{r}_{16}$ from the statistical downscaling approach are slightly larger than the ones from the *Constant Reflectance Ratio Approach*. However, since these samples are three to seven orders of magnitude less likely than the observations around the 1:1 line, they do not change the high correlation and slope of 1.00. One minor difference between the two results concerns the number of negative $\hat{r}_{16}$, which can occur for very thin clouds~~ optically thin (i.e., ~~very low~~

~~$\tilde{r}_{\mathrm{HV}}$ and $\tilde{r}_{16}$) . For the analyzed data set, almost all such observations are the result of the statistical downscaling technique with a relative contribution of 96.98%. However, the overall fraction of data points with a negative $\hat{r}_{16}$ is very low with a value of about 0.005%.~~ $\tau \leq 10$) liquid water clouds over the ocean (Werner et al., 2018b).

4.3 Lookup Table Approach

A third method to derive high–resolution cloud property retrievals for SEVIRI utilizes an iterative approach to determine δr_{06} and δr_{16} independently, based on the shape of the LUT, while constraining the observed r_{eff} to that of the baseline approach (i.e., simple trigonometric interpolation, which yields reflectances $\tilde{r}_{06}$ and $\tilde{r}_{16}$ that only resolve the large–scale variability). While the previous approaches can be implemented as a pre–processor outside the actual retrieval, this method requires access to the LUT and has thus been implemented through modifications of the CPP retrieval algorithm.

Again, a simple linear relationship between δr_{HV}, δr_{06} and δr_{08} based on Eq.(2) is assumed:

$$\delta r_{\mathrm{HV}} = a \cdot \delta r_{06} + b \cdot \delta r_{08}, \tag{12}$$

where the fit coefficients a and b are determined from the same techniques as described in ~~section~~ Section 4.1. The variation δr_{HV} of the HRV channel is obtained from the observations following Eq.(2), while δr_{08} is calculated as the difference between r_{08} from high– and low–resolution optical thickness τ based on the functional relation $\mathcal{F}$ of the reflectances and cloud properties stored in the LUT (which motivates the name of this method). Therefore, δr_{06} can be derived from:

$$\begin{aligned} \delta r_{06} &= \frac{1}{a} \cdot \left(\delta r_{\mathrm{HV}} - b \cdot \delta r_{08}\right), \\ \hat{r}_{06} &= \tilde{r}_{06} + \delta r_{06}, \\ \delta r_{08} &= \mathcal{F}_{08}\left(\hat{\tau}, \hat{r}_{\mathrm{eff}}\right) - \mathcal{F}_{08}\left(\tilde{\tau}, \tilde{r}_{\mathrm{eff}}\right). \end{aligned} \tag{13}$$

Note that the addition of δr_{08} in the calculation of δr_{06} helps to account for the noticeable increase in surface albedo of vegetation—like surfaces at $\lambda > 700$ nm (i.e., the vegetational step). This should improve the estimation of δr_{06} for thin clouds (i.e., $\tau < 10$) and cloud–edge pixels. For the SWIR reflectance, instead of relying on the imperfect linear model in Eq.(6) or assumptions about the inhomogeneity index $H_{\sigma,16}$, the adjustment δr_{16} is determined iteratively to conserve the coarse–resolution, pixel–level (i.e., ~~$3 \times 3\,\mathrm{km}^2$~~ LRES scale of channels 1–3) value of the effective droplet radius. To reduce some of the associated uncertainties, the effective droplet radius based on the reflectances from triangular interpolation can be used instead of the LRES result. If $\tilde{\tau}$ and $\tilde{r}_{\mathrm{eff}}$ are the cloud properties based on trigonometric interpolation, and $\hat{\tau}$ and $\hat{r}_{\mathrm{eff}}$ are the higher–resolution retrievals, which are derived from an inversion of the functional relationship ($\mathcal{F}$) between the high–resolution reflectances $\hat{r}_{06}$ and $\hat{r}_{16}$ following:

$$\left(\hat{\tau}, \hat{r}_{\mathrm{eff}}\right) = \mathcal{F}^{-1}\left(\tilde{r}_{06} + \delta r_{06}, \tilde{r}_{16} + \delta r_{16}\right), \tag{14}$$

then δr_{16} can be determined as:

$$\delta r_{16} = \mathcal{F}_{16}\left(\hat{\tau}, \hat{r}_{\mathrm{eff}} = \tilde{r}_{\mathrm{eff}}\right) - \mathcal{F}_{16}\left(\tilde{\tau}, \tilde{r}_{\mathrm{eff}}\right). \tag{15}$$

This implies that a positive or negative δr_{06} is connected to a positive or negative δr_{16} using the LUT to adjust the SWIR subpixel reflectance variations in such a way to be representative of the respective standard–resolution $\tilde{r}_{\mathrm{eff}}$. As a result, we do not expect any improvement for the r_{eff} retrieval during the transition to smaller scales. Instead, we try to find a physically reasonable constraint for δr_{16} to achieve a reliable retrieval of the higher–resolution $\hat{\tau}$, while retaining the accuracy of the standard–resolution retrieval of $\tilde{r}_{\mathrm{eff}}$.

The *LUT Approach* is illustrated in Figure 4(a), where an example SEVIRI liquid–phase LUT for a specific solar zenith angle ($\theta_0 = 40°$), sensor zenith angle ($\theta = 20°$), and relative azimuth angle ($\varphi = 60°$) is shown. Vertical dashed lines and values below the grid denote fixed ~~τ~~ $\tilde{\tau}$, while the horizontal dashed lines and values right of the grid denote fixed ~~r_{eff}~~ $\tilde{r}_{\mathrm{eff}}$ in units of microns. The green dot highlighted by the capital letter "A" represents an example SEVIRI reflectance pair of approximately $\tilde{r}_{06} = 0.33$ and $\tilde{r}_{16} = 0.34$, which maps to $\tilde{\tau} = 8$ and $\tilde{r}_{\mathrm{eff}} = 12\,\mu\mathrm{m}$ (i.e., the retrieval result for the high–resolution reflectances from trigonometric interpolation). The red line highlights the $\tilde{r}_{\mathrm{eff}} = 12\,\mu\mathrm{m}$ isoline. The two horizontal, blue arrows indicate a positive ($\delta r_{06,1}$) and negative ($\delta r_{06,2}$) adjustment to $\tilde{r}_{06}$ based on Eq.(13). Without an adjustment to $\tilde{r}_{16}$, these newly derived higher–resolution $\hat{r}_{06}$ map to significantly larger and lower effective droplet radii of about $\hat{r}_{\mathrm{eff}} = 29\,\mu\mathrm{m}$ and $\hat{r}_{\mathrm{eff}} = 5\,\mu\mathrm{m}$, respectively. The adjustments $\delta r_{16,1}$ and $\delta r_{16,2}$ simply assure that the prior effective radius retrieval is preserved (i.e., $\hat{r}_{\mathrm{eff}} = \tilde{r}_{\mathrm{eff}}$). Due to the curvature of the ~~lines of fixed r_{eff}~~ isolines of fixed $\tilde{r}_{\mathrm{eff}}$ given by the LUT, small deviations of the coarse–resolution average from $\tilde{r}_{16}$ can still occur.

Note that the *LUT Approach* requires a prior cloud phase retrieval (either from the lower–resolution or interpolated reflectances) to determine the correct LUT for either liquid water or ice.

4.4 Adjusted Lookup Table Approach

In order to improve the estimation of δr_{16} in the *LUT Approach*, two modifications to the previous assumption are introduced in this section. The first one aims to provide a more realistic estimate of $\tilde{r}_{\mathrm{eff}}$ compared to the ~~$3 \times 3\,\mathrm{km}^2$~~ coarser LRES result, which subsequently is used to determine δr_{16}. The value of $\tilde{r}_{\mathrm{eff}}$ is derived from adiabatic theory, which provides a physically sound relationship between the derived high–resolution cloud variables:

$$\hat{r}_{\mathrm{eff}} = \tilde{r}_{\mathrm{eff}} \left(\frac{\hat{\tau}}{\tilde{\tau}} \right)^{a}. \tag{16}$$

Based on observations, the study by Szczodrak et al. (2001) confirmed the value of ~~$a = 0.2$~~ $a \approx 0.2$ predicted by theory for marine stratocumulus, so this is the value also adopted here. This approach is illustrated in Figure 4(b), where the $\tilde{r}_{\mathrm{eff}}$ retrieval based on the interpolated reflectances at point "A" is indicated by the red r_{eff}–isoline. During the first iteration step δr_{06} is derived from Eq. (13) and $\delta r_{16} = 0$, which maps to $\hat{\tau}^1$ in the LUT (the exponent 1 indicates the first iteration step). This value is highlighted by the vertical, blue line. Based on Eq. (16) the corresponding, adiabatic $\hat{r}^1_{\mathrm{eff}}$ is calculated (highlighted by the horizontal, blue line). This value determines the adjustment δr_{16}. Note, that the resulting reflectances at point "B" do not exactly map to $\tilde{\tau}^1$ after the first iteration. As a result, multiple iterations are necessary to derive the final cloud properties. It has however been relatively simple to merge this iteration into the iterative retrieval loop of the CPP retrieval.

A second approach to improve upon the *LUT Approach* again utilizes the shape of the LUT to derive a local slope $S = \partial r_{16}/\partial r_{06}$ from the simulated LUT reflectances. The value of S is calculated at the position denoted by $\tilde{\tau}$ and $\tilde{r}_{\mathrm{eff}}$. In the iterative CPP retrieval, this requires that both low– and high–resolution cloud properties are estimated during each iteration until convergence of both properties is achieved. This approach is illustrated in Figure 4(c). Again, the initial $\tilde{r}_{\mathrm{eff}}$ retrieval based on the interpolated reflectances at point "A1" is indicated by the red ~~r_{eff}~~ $\tilde{r}_{\mathrm{eff}}$–isoline. The slope S_{A1} at this position in the LUT is highlighted by the solid, blue line. Based on the derived slope and δr_{06} from Eq. (13) the corresponding δr_{16} can be calculated for each iteration step. Two additional examples for initial starting points ("A2" and "A3") and the respective slopes (S_{A2} and S_{A3}) are also shown. These examples indicate the change in slope for different parts of the LUT. For small $\tilde{\tau}$, the slope S_{A3} become steeper, which leads to a larger adjustment δr_{16}. Meanwhile, for large $\tilde{\tau} > 30$ (for this specific viewing geometry and LUT) the $\tilde{\tau}$ and $\tilde{r}_{\mathrm{eff}}$–isolines are nearly orthogonal and both the respective slope S_{A2} and δr_{16} are close to 0.

Both approaches introduced in this section have advantages and disadvantages, but promise to improve on the standard *LUT Approach*. While physically sound, adiabatic assumptions might not always be appropriate, especially for highly convective clouds or in the presence of drizzle. Meanwhile, large δr_{06} adjustments might map to a point in the LUT where the derived local slopes at the position of $\tilde{\tau}^{\mathrm{i}}$ and $\tilde{r}^{\mathrm{i}}_{\mathrm{eff}}$ might not be representative anymore.

4.5 ~~Comparison of interpolated and downscaled SEVIRI reflectances~~

~~In order to illustrate the difference between the various reflectances, a statistical comparison between the downscaled results for $\hat{r}_{06}$ and $\hat{r}_{16}$ and the observations at the native SEVIRI scale (i.e., r_{06} and r_{08}) is shown in Figure **??**. To allow for a pixel-to-pixel analysis, each r_{06} and r_{08} at the original horizontal resolution of $3 \times 3\,\mathrm{km}^2$ is replicated to each of the 9 available subpixels at the HRV channel resolution. To put the resulting differences into perspective, a comparison between the downscaled and interpolated high–resolution reflectances is also provided. Note that only the statistical downscaling and *Constant Reflectance Ratio Approach* are shown, because in the *LUT Approach* $\hat{r}_{06}$ and $\hat{r}_{16}$ are derived iteratively during the cloud property retrieval and are not provided as an output variable by the algorithm.~~

~~Figure **??**(a) shows a PDF of the relative difference (Δr_{06}; shown in in red), which is defined as the difference between $\hat{r}_{06}$ from the statistical downscaling approach and the resampled r_{06}, normalized by r_{06}, for an example SEVIRI scene from the Germany domain on 9 June 2013 at 10:55 UTC. Overall, $n = 696{,}879$ are included in the analysis. The distribution is centered around $\Delta r_{06} \approx 0$ and is almost symmetrical on both sides. The 1^{st}, 50^{th}, and 99^{th} percentiles of Δr_{06} are -24.17%, 0.03%, and 27.85%, respectively. This means, that statistically the two different resolution yield similar reflectance observation, but high–frequency variability, which is resolved by $\hat{r}_{06}$, introduces significant deviations from the results at the standard resolution. Overall, most of the observations, defined by the 25^{st}, 75^{th} percentiles (i.e., 50% of the data points), are in the range of -3.12% to 2.87%. These differences compare well to those observed for the downscaled $\hat{r}_{06}$ from the *Constant Reflectance Ratio Approach* (shown in blue). As expected, the relative differences between $\hat{r}_{06}$ and $\tilde{r}_{06}$ (shown in black) are visibly smaller. The 1^{st} and 99^{th} percentiles of Δr_{06} are -11.28% and 12.54%, respectively, and most observations are in the range of -1.43% to 0.99%. As before, the distribution is centered around $\Delta r_{06} \approx 0$, with a median of 0.03. The normalized root-mean-square~~

~~deviation (nRMSD; defined as the RMSD between $\hat{r}_{06}$ and $\tilde{r}_{06}$, normalized by the mean $\tilde{r}_{06}$) is nRMSD = 2.73%, which is less than half the value from the difference between $\hat{r}_{06}$ and the resampled r_{06} (nRMSD = 6.30%).~~

~~A similar analysis for the channel 3 reflectances r_{16}, $\hat{r}_{16}$ (from Eq.(7)), and $\tilde{r}_{16}$ is shown in Figure **??**(b). As before for the VNIR channel, the PDF of the relative differences (Δr_{16}) is centered around ≈ 0, and the 1^{st} and 99^{th} percentiles are -27.53% and 28.85% for the difference between $\hat{r}_{16}$ and the resampled r_{16} and -13.59% and 11.40% for the difference between $\hat{r}_{16}$ and $\tilde{r}_{16}$, respectively. The nRMSD is 3.16% (r_{16}) and 6.24% ($\tilde{r}_{16}$). Overall, 50% of the data points lie in the range of -2.67% to 2.71% (for the difference between $\hat{r}_{16}$ and r_{16}). Again, the results from the two downscaling approaches are very similar.~~

~~It has to be noted, however, that deviations of $\pm 3\%$ in the reflectances at the different spatial scales can have a significant impact on the remote sensing products of optical and microphysical cloud parameters, especially if the clouds are thin or the pixels are partially cloudy (Werner et al., 2018a, b). These impacts become even more pronounced for the samples with larger deviations between downscaled and native reflectances. Such effects are illustrated in section 5.~~

5 Example retrievals

An example of a standard SEVIRI red, green, and blue (RGB) composite and the respective cloud property retrievals, utilizing the native r_{06} and r_{16}, are shown in Figures 5(a)–(c). In comparison, the retrieval results using the downscaled $\hat{r}_{06}$ and $\hat{r}_{16}$ from the *Adjusted Lookup Table Approach*, using the *LUT Slope Adjustment*, are presented in Figures 5(d)–(f) for the same cloud field. The example is a ~~$100 \times 100\,\text{km}^2$~~$\approx 100 \times 100\,\text{km}^2$–subscene of SEVIRI observations of an altocumulus field, which was acquired on 9 June 2013 at 10:55 UTC over ocean within the Germany domain. The three illustrated parameters are an RGB composite image of SEVIRI channel 3, 2, and 1 reflectance in panels a) and c), the cloud optical thickness τ and $\hat{\tau}$ in panels b) and e), as well as the effective droplet radius r_{eff} and $\hat{r}_{\text{eff}}$ in panels c) and f). For the cloud variables only liquid–phase pixels are shown. An increase in contrast and resolved cloud structures is visible in the higher–resolution RGB composite. Regarding the retrieved cloud properties, the fields of lower–resolution τ and r_{eff} are a lot smoother and the results exhibit less dynamical range than their higher–resolution counterparts. One obvious example is the bright cloudy part along $54.6°N$, where $\tau > 45$ are observed. Moreover, the region of low r_{eff} in the north–eastern corner of the scene exhibits more nuanced values in the higher–resolution data set. ~~Note, that for this case, the number of failed retrievals is reduced for the *Adjusted Lookup Table Approach* (see south–eastern corner of the scene).~~

6 Evaluation of downscaling techniques with MODIS data

This section presents an evaluation of the different downscaling techniques, which are introduced in ~~section~~ Section 4, by means of MODIS observations. MODIS provides reflectances at a horizontal resolution of ~~$1 \times 1\,\text{km}^2$~~$\approx 1 \times 1\,\text{km}^2$. These observations are ~~re–mapped~~ remapped to the higher–resolution grid of the SEVIRI r_{HV}–band samples, ~~and provide the~~ thus simulating a hypothetical SEVIRI–like geostationary instrument, where all channels are provided at the HRES scale. This provides the means to derive reference retrievals of τ and r_{eff}. Note, that even though these reference retrievals are performed at a higher

resolution the "ˆ"–notation is omitted, because these cloud products are derived from actual observations ~~,~~and are not the estimates obtained from the various downscaling techniques. ~~Subsequently, the re–mapped, higher–resolution reflectances are smoothed using the spatial response function of the corresponding SEVIRI channels.~~

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 $\hat{r}_{\mathrm{eff}}$, as well as $\tilde{\tau}$ and $\tilde{r}_{\mathrm{eff}}$) 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 ~~these~~ the latter data are still available at a higher resolution than the native ~~$3 \times 3\,\mathrm{km}^2$~~ 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. ~~Subsampling also enables a comparison with SEVIRI's native 3 km observations.~~

~~These observations subsequently provide the means to apply the various downscaling techniques, as well as the simple triangular interpolation approach, in order to compare the retrieved cloud products (i.e., $\hat{\tau}$ and $\hat{r}_{\mathrm{eff}}$, as well as $\tilde{\tau}$ and $\tilde{r}_{\mathrm{eff}}$) to the reference results.~~ In addition, a comparison can be made to those cloud variables, which would be obtained from reflectances at SEVIRI's native spatial resolution by setting each 3×3 HRES pixel block to ~~its central~~ the LRES value.

Figure 6 shows RGB composites of the four example scenes, which comprise the data set for the evaluation of the different downscaling techniques. The scenes are increasingly more heterogeneous, starting with a rather homogeneous altocumulus field in Figure 6(a), two more heterogeneous broken altocumulus examples in Figures 6(b)–(c), and finally a broken cumulus field in Figure 6(d).

Meanwhile, ~~table~~ Table 1 summarizes the ~~ten~~ different retrieval experiments that form the comparison in this section. For the sake of completeness, the reference data (i.e., the results from the ~~re–mapped $1 \times 1\,\mathrm{km}^2$ reflectances~~MODIS reflectances, which are remapped to SEVIRI's HRES grid) are also included. ~~The~~ Retrievals based on remapped MODIS data to SEVIRI's native 3 km–scale are reproduced to each of the 3x3 subpixels to match the horizontal resolution of the reference results. Meanwhile, the cloud products derived from triangular interpolation of ~~SEVIRI~~ the remapped LRES–MODIS samples are referred to as the baseline data set, as this is the easiest approach and any reliable downscaling technique needs to add an improvement on those results. ~~These results are, however, not directly comparable with retrievals at SEVIRI's native 3 km resolution, which are added as a separate experiment and are obtained by sub–sampling the baseline results. Here, each central pixel of a 3×3 block is replicated nine times and compared to the~~ Experiment 1 ~~km reference. Experiments 1a and 1b denote~~ denotes the statistical downscaling approach from ~~section 4.1. Here, 1a is based on $\hat{r}_{06}$ and $\tilde{r}_{16}$ (i.e., only the VNIR reflectance is downscaled; the SWIR reflectance is derived from interpolation), while 1b utilizes both $\hat{r}_{06}$ and $\hat{r}_{16}$ (i.e., both reflectances are downscaled and thus include small scale reflectance variability). Similarly,~~ Section 4.1, while retrievals based on the *Constant Reflectance Ratio Approach* and the *Adjusted LUT Approach* with *LUT Slope Adjustment* are indicated as experiments ~~2a and 2b,~~ 2 and 3, respectively. Note, that we also performed analysis for the standard *LUT Approach*, as well as ~~3a and 3b, respectively. The retrievals from the two *Adjusted LUT Approaches* are denoted as experiments 3c and 3d~~the *Adjusted LUT*

Approach with *Adiabatic Adjustment*. However, we will only briefly summarize the results of these downscaling schemes where necessary.

First, the collocation and ~~re–mapping~~ remapping procedure for the native MODIS reflectances is briefly described. A comparison between the retrieved cloud products from the LRES resolution–reflectances and those from triangular interpolation, as well as the different downscaling procedures, and the reference results follows in ~~section~~ Section 6.2. These retrievals can be used to derive estimates of the liquid water content (W_{L}, $\tilde{W}_{\mathrm{L}}$, and $\hat{W}_{\mathrm{L}}$) and the droplet number concentration (N_{D}, $\tilde{N}_{\mathrm{D}}$, and $\hat{N}_{\mathrm{D}}$), which are evaluated in ~~section 6.3. While the downscaling of SEVIRI VNIR reflectances is based on their linear relationship to the observed high–resolution r_{HV}, the downscaling of SWIR reflectances is based on a number of assumptions, which might induce large uncertainties in the retrieved cloud products. Therefore, a comparison between the full downscaling techniques and the VNIR–only results is presented in section 7.~~ Section 6.3.

6.1 Reprojection of MODIS swath radiances to the SEVIRI grid

To obtain a reliable higher–resolution reference data set, MODIS level 1b swath observations (MOD021km) have been projected to the grid of ~~the SEVIRIHRV reflectance observations~~SEVIRI's r_{HV} samples, which corresponds to the Geostationary Satellite projection ~~with a pixel resolution of $1 \times 1\,\mathrm{km}^2$~~ at the HRES scale. Initially, the native HRV grid is oversampled by a factor of three in each dimension (i.e., the target grid has a ~~333 m~~ $\approx$ 333 m resolution), and nearest–neighbor interpolation is used for the projection. This oversampled field is subsequently ~~filtered with the spatial response~~ smoothed with the modulation transfer function of the HRV channel as given by (EUMETSAT, 2006), to remove high-frequency variability not resolved by the sensor and, in particular, the artifacts introduced by the nearest–neighbour interpolation technique. Finally, this field is downsampled, such that only each central pixel of a 3×3 block (each pixel with a horizontal resolution of 333 m) is retained to represent the ~~$1 \times 1\,\mathrm{km}^2$ value~~HRES–value.

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 ~~3 km channels~~ LRES–channels of Meteosat SEVIRI. This step again involves the ~~filtering~~ smoothing of the respective reflectance field with the channel–specific ~~spatial response~~ modulation transfer function of the lower–resolution SEVIRI channels (EUMETSAT, 2006). This data set represents hypothetical SEVIRI–like observations at the native LRES resolution.

In addition, a band–pass filter has been constructed from the difference between the modulation transfer functions of the HRV and the $0.6\,\mu\mathrm{m}$ and $0.8\,\mu\mathrm{m}$ channels (weighted by the coefficients of a linear model; see Deneke and Roebeling, 2010). This filter is used to extract the high–frequency signal of the HRV channel.

It should be noted that retrievals based upon these radiances will be different than those based upon the original MODIS C6 radiances, or from an absolutely accurate representation of the (hypothetical) truly observed, high–resolution SEVIRI samples. For one, it uses the linear model of Cros et al. (2006) and Deneke and Roebeling (2010) as a proxy for the HRV channel, thereby excluding a potentially significant source of uncertainty. Moreover, MODIS acquires these reflectances under different viewing geometries (note that the true viewing angles are used in the CPP retrieval, so within the limits of plane–parallel radiative transfer, this effect is accounted for), and the spectral characteristics of the MODIS and SEVIRI channels

are not entirely comparable. However, the goal of this study is to provide a consistent reference data set ~~and retrievals~~ for a comparison of different retrieval data sets, which are derived from a single retrieval algorithm core. ~~Statistical comparisons between the operational MODIS C6.1 and SEVIRI results, as well as the new high–resolution SEVIRI products , are presented in the companion paper Deneke et al. (2019). Moreover, some interesting use cases are demonstrated in that study, which can benefit from an increase in the spatial resolution of the derived SEVIRI cloud parameters.~~ The actual absolute values of the retrieved cloud products are not important here.

6.2 Results for τ and r_{eff}

Figure 7(a) shows a comparison of τ at the native ~~SEVIRI resolution~~ LRES resolution (replicated onto each subpixel), and the reference τ at the ~~1 km~~ HRES scale for the example cloud field in scene 2, which is shown as an RGB composite image in Figure 6(b). A total of over $13,000$ cloudy pixels (liquid phase) are located in this scene. While for small reference $\tau < 20$ there is a reasonable agreement between the two data sets, there is increased scatter around the 1:1 line (indicated by the gray, dashed line) for larger values of cloud optical thickness. For reference $\tau > 40$, a substantial underestimation of the ~~3 km~~ LRES–τ is observed, which yields a sizable contribution to the nRD of 15.8%. Figures 7(b)–(c) show similar scatter plots of τ and $\hat{\tau}$ from both experiment ~~2b and 3d~~2 (*Constant Reflectance Ratio Approach*) and 3 (*Adjusted LUT Approach* with *LUT Slope Adjustment*), respectively. It is obvious that the results from these two downscaling techniques improve the agreement to the reference retrievals significantly. The ~~correlation~~ explained variance (R^2, which equals the square of Pearson's product-moment correlation coefficient R) between the data sets is increased and the nRD is strongly reduced to values of 1.182% (experiment ~~2b~~2) and 1.589% (experiment ~~3d~~3).

A similar comparison between the reference r_{eff} at the HRES scale and r_{eff} at native ~~SEVIRI~~ LRES resolution, as well as $\hat{r}_{\mathrm{eff}}$ from the same downscaling experiments, is presented in Figures 7(d)–(f). Here, the native–resolution results show a much better agreement with the reference retrievals and, compared to the cloud optical thickness, the nRD$= 5.505\%$ is much lower. While experiment ~~2b~~ 2 exhibits a good agreement between reference τ and $\hat{\tau}$, the comparison of retrieved $\hat{r}_{\mathrm{eff}}$ to the reference results is less favorable. Both the reduced ~~correlation ($R = 0.943$ versus $R = 0.964$~~explained variance ($R^2 = 0.889$ versus $R^2 = 0.929$), as well as the increased scatter around the 1:1 line (nRD $= 6.630\%$) indicate that the results from experiment ~~2b~~ 2 are less reliable than the ones performed at the native ~~3 km~~ LRES resolution. Thus, the elaborate downscaling procedure actually reduces the accuracy of the ~~retrievals~~$\hat{r}_{\mathrm{eff}}$ retrieval. In contrast, the retrieved $\hat{r}_{\mathrm{eff}}$ from experiment ~~3d~~ 3 improve upon the native–resolution results, with slightly better values of ~~$R = 0.976$~~ $R^2 = 0.953$ and nRD $= 4.402\%$.

Statistics of the comparison between the reference and native ~~3 km~~LRES, baseline, and experimental retrievals are presented in Figures 8(a)–(d) for example scenes 1–4, respectively. The parameters which are used to quantify the individual comparisons are the median of the relative difference (abbreviated with p50) to indicate the average deviation from the reference results, the interquartile range (IQR; defined as the relative difference between the 75^{th} and 25^{th} percentile of the deviation to the reference retrievals) to indicate the spread between the different data sets, the nRD as a second measure of the spread of data points, and the explained variance ~~(R^2 , which equals the square of Pearson's product-moment correlation coefficient R)~~ between the different retrievals and the reference. Values with a green and red background highlight the respective experiment with the best

and worst comparison for the specific parameter. Yellow backgrounds, meanwhile, indicate all other experiments in between the two extreme results. The first noteworthy observation concerns the native and baseline retrievals of τ, which universally exhibit the largest median deviations and spread to the reference results, as well as the lowest R^2. Still, the difference between native and baseline results indicates that the trigonometric interpolation to the ~~HRV~~ HRES grid has significantly improved the comparison. ~~For scene 2, the 1st, 50th, and 99th percentiles of the absolute deviations of the native retrievals from the reference τ are −13.54, −0.08, and 6.96, respectively.~~ In contrast, each retrieval of $\hat{\tau}$ that accounts for small–scale reflectance variability ~~,~~ yields significant improvements, regardless of the approach. This is especially obvious in the parameters that characterize the spread in the deviations, i.e., IQR and nRD, which are between 2–9 and ~~3–10~~ 2–10 smaller for the various experiments and example scenes, respectively. Experiments ~~1b and 2b, as well as 3d,~~ 2 and 3 seem to achieve the best agreement to the reference retrievals. ~~For the data set from experiment 3d the 1st, 50th, and 99th percentiles of absolute deviations improve to −0.30, 0.13, and 1.36, respectively.~~

Regarding the effective droplet radius, the agreement between the native ~~3 km~~ LRES and baseline retrievals, and the reference results is significantly better. It is worth pointing out that ~~$\tilde{r}_{\text{eff}}$, obtained only by~~ , similar to the optical thickness comparison, the $\tilde{r}_{\text{eff}}$ retrieval based on interpolating reflectances to the ~~HRV grid~~ , HRES grid performs better than the native–resolution r_{eff} retrieval for all scenes. ~~As an example, the 1st, 50th, and 99th percentiles of the absolute deviations between native and reference results for example scene 2 are −1.29 μm, 0.18 μm, and 2.03 μm, respectively.~~ The most reliable downscaling approach seems to be experiment ~~3d~~3, which performs noticeably better than experiments ~~1b~~ 1 (note the increased nRD and reduced R^2 for scene 3) ~~, 3c (overall worst performance for scenes 1~~ and 2 ~~), and 2b~~ (increased spread and overall issues for the heterogeneous cloud field in scene 4). This indicates that the linear model in Eq.(6) ~~, presuming general adiabatic cloud conditions,~~ or assumptions about a constant ratio of VNIR and SWIR reflectances are not adequate to estimate higher–resolution $\hat{r}_{16}$, at least not for certain cloud conditions. In the case of experiment ~~2b,~~ 2 this is understandable, ~~since~~ because the technique was developed for partially cloudy pixels (Werner et al., 2018b). ~~For experiment 3d, the 1st, 50th, and 99th percentiles of the absolute deviations are comparable to the baseline data set, with values of −0.30 μm, 0.13 μm, and 1.36 μm, respectively.~~ These observations are characterized by a low cloud optical thickness, where the relationship between VNIR and SWIR reflectance can reliably be considered to be linear (see example LUTs in Figure 4).

~~The notably better performance of experiment 3d than 3b with respect to~~ There is a notably better performance of experiment 3, the *Adjusted LUT Approach* with *LUT Slope Adjustment*, compared to the standard *LUT Approach* highlighted in Section 4.3. Of particular note is the $\hat{r}_{\text{eff}}$ retrieval based on the standard LUT scheme, which compares significantly worse to the reference results (R^2 of 0.890, 0.648, 0.751, and 0.581 for cloud scenes 1–4, respectively). This is somewhat surprising, ~~and~~ because the specified goal ~~that experiment 3b maintains the~~ of the standard *LUT Approach* is to maintain the accuracy of the baseline $\tilde{r}_{\text{eff}}$ retrieval, which has not been fully reached. We believe that this might be caused by the sensitivity of the cloud property retrieval to small reflectance perturbations, in particular for broken clouds. It is also an indication that assuming constant subpixel r_{eff} values within each LRES pixel is not sufficient. We plan to investigate this effect further in future studies. However, the second *Adjusted LUT Approach*, which determines SWIR reflectance adjustments based on adiabatic theory, performs even worse (R^2

of 0.846, 0.579, 0.741, ad 0.519 for cloud scenes 1–4, respectively). This suggests that the observed cloud fields do not follow adiabatic theory and the method is not adequate to estimate higher–resolution $\hat{r}_{16}$.

6.3 Results for W_L and N_D

Retrievals of τ and r_eff (regardless of the resolution they are derived at) provide the means to infer other commonly used cloud variables. The W_L, which describes the amount of liquid water in a remotely sensed cloud column, can be derived as the product of retrieved cloud products (Brenguier et al., 2000; Miller et al., 2016):

$$W_\mathrm{L} = \Gamma \approx \frac{2}{3} \cdot \rho_l \cdot \tau \cdot r_\mathrm{eff}. \tag{17}$$

Here, ~~ρ_l and Γ are the~~ ρ_L is the bulk density of liquid water~~and a coefficient, which accounts for~~. Assuming adiabatic clouds, where the vertical structure of ~~the cloud profile ($\Gamma = 2/3$ for vertically homogeneous clouds, $\Gamma = 5/9$ for adiabatic clouds)~~effective droplet radius follows the adiabatic growth model, introduces an extra factor of $5/6$ and the coefficient $2/3$ changes to $5/6 \cdot 2/3 = 5/9$. Meanwhile, N_D describes the number of liquid cloud droplets in a cubic centimeter of cloudy air. Calculating N_D from remote sensing products requires a number of assumptions, e.g., about the vertical cloud structure and shape of the droplet number size distribution, which are summarized and discussed in Brenguier et al. (2000); Schüller et al. (2005); Bennartz (2007); Grosvenor et al. (2018). A simplified form of the resulting equation for N_D is:

$$N_\mathrm{D} = \approx \alpha \cdot \tau^{0.5} \cdot r_\mathrm{eff}^{-2.5}, \tag{18}$$

with $\alpha = 1.37 \cdot 10^{-5}$ (see Quaas et al., 2006). Note, that Eqs.(17)–(18) can yield both baseline and downscaled results (i.e., $\tilde{W}_\mathrm{L}$ and $\tilde{N}_\mathrm{D}$, as well as $\hat{W}_\mathrm{L}$ and $\hat{N}_\mathrm{D}$) when they are derived from the respective cloud optical thicknesses and effective droplet radii.

Similar to the comparison in ~~section~~ Section 6.2, scatterplots of the reference W_L, the native ~~3 km $\tilde{W}_\mathrm{L}$~~ LRES–W_L and the results from the downscaling experiments ~~1b and 3d~~ 2 and 3 ($\hat{W}_\mathrm{L}$) are shown in Figures 9(a)–(c), respectively. As before, data is provided by example scene 2 sampled on 9 June 2013 at 10:55 UTC. Compared to the native ~~SEVIRI~~ LRES results, a noticeable improvement in the correlation and nRD is achieved by utilizing the two downscaling experiments. Not only are retrieved $\hat{W}_\mathrm{L}$ closer to the 1:1 line, but the significant underestimation of the ~~3 km~~ LRES–W_L for larger reference results is mitigated. Especially for experiment ~~3d~~3, the spread is less than one third the value of the ~~baseline results~~ LRES–results (4.857% versus 15.234%). Regarding the comparison between reference and native N_D, as well as ~~$\hat{N}_\mathrm{D}$~~ $\hat{N}_\mathrm{D}$, downscaling experiment ~~2b~~ 2 yields less favorable results. There is a slight decrease (increase) in ~~R~~ R^2 (nRD). This is caused by the large IQR and nRD of the deviations in the retrieved $\hat{r}_\mathrm{eff}$, shown in Figure 7(e), which are amplified due to the associated power of 2.5 in Eq. (18). However, the derived values from experiment ~~3b are~~ 3 show a significantly better agreement with the reference N_D.

Values of p50, IQR, nRD, and R^2 for the W_L and N_D comparison from the four example scenes are illustrated in Figures 10(a)–(d). Due to the large deviations between the native τ and the reference retrievals, W_L for the ~~3 km~~ LRES results almost universally show the largest deviations to the reference values, and thus the largest IQR and nRD, as well as the lowest

explained variance. The exception is the heterogeneous cloud field in the fourth example scene, where the large deviations between $\hat{r}_{\mathrm{eff}}$ from experiment ~~2b~~ 2 and the reference retrievals yield the worst comparison for the respective $\hat{W}_{\mathrm{L}}$. ~~As for the statistical comparison in section 6.2, experiment 3c overall performs worst for scenes 1 and 2. However~~The estimates based on the *Adjusted Lookup Table Approach* using the LUT *LUT Slope Adjustment* (i.e., experiment 3) almost universally exhibits the best agreement to the reference results of W_{L}.

Overall, 27 of the 32 comparisons (four cloud scenes, two cloud variables, and four statistical measures) exhibit the best ~~results for experiment 3d. For the four~~ performance for experiment 3. For the example scenes considered in this analysis, it is obvious that the *Adjusted Lookup Table Approach* ~~, using the~~ with *LUT Slope Adjustment* ~~,~~ is preferable to ~~other downscaling approaches~~ the other downscaling techniques and yields more reliable high–resolution cloud variables than the ~~standard–resolution SEVIRI~~ standard LRES results.

~~For example scene 2,~~ As before, we also tested the standard *LUT Approach* highlighted in Section 4.3, as well as the second *Adjusted LUT Approach*, which determines SWIR reflectance adjustments based on adiabatic theory. Due to the poor performance of the $\hat{r}_{\mathrm{eff}}$ retrieval, the ~~the 1^{st}, 50^{th}, and 99^{th} percentiles of absolute deviations of the 3 km cloud variables from the reference W_{L} are $-88.50\,\mathrm{g\,m^{-2}}$, $0.70\,\mathrm{g\,m^{-2}}$, and $57.90\,\mathrm{g\,m^{-2}}$, respectively, which for experiment 3d changes to $-15.55\,\mathrm{g\,m^{-2}}$, $3.10\,\mathrm{g\,m^{-2}}$, and $28.95\,\mathrm{g\,m^{-2}}$. While a slight bias is introduced, the spread of deviations is significantly reduced. Meanwhilefor N_{D} these deviations are $-77.57\,\mathrm{cm^{-3}}$, $-7.44\,\mathrm{cm^{-3}}$, and $36.25\,\mathrm{cm^{-3}}$ for the 3 km results and $-55.75\,\mathrm{cm^{-3}}$~~ $\hat{N}_{\mathrm{D}}$ results based on adiabatic assumptions show a similarly poor agreement to the reference results. Meanwhile, ~~$-5.59\,\mathrm{cm^{-3}}$, and $27.04\,\mathrm{cm^{-3}}$ for experiment 3d~~the cloud variables based on the standard *LUT Approach* never show the best or worst performance, but are almost universally worse than the *Adjusted Lookup Table Approach* with *LUT Slope Adjustment*. This again illustrates that assumptions of adiabatic clouds and constant subpixel r_{eff} values within each LRES pixel are not suitable for the cloud scenes analyzed in this study.

6.4 ~~Full downscaling versus VNIR only~~

7 Full downscaling versus VNIR only

Apart from the *Constant Reflectance Ratio Approach*, the downscaling of r_{06} for each of the techniques presented in ~~section~~ Section 4 uses the well established relationship between r_{06}, r_{08}, and the averaged $\langle \tilde{r}_{\mathrm{HV}} \rangle$ (see Figure 3 and the discussion in Deneke and Roebeling, 2010). In contrast, downscaling of r_{16} is based on different assumptions about the microphysical structure and cloud heterogeneity, which induces a level of uncertainty in the subsequent cloud property retrievals. To test whether assumptions about r_{16} actually improve the retrieval of $\hat{\tau}$ and $\hat{r}_{\mathrm{eff}}$, this section presents retrievals that include the results from ~~experiment 3d~~ the *Adjusted Lookup Table Approach* with *LUT Slope Adjustment* (i.e., experiment 3) for $\hat{r}_{06}$ but do not include the respective downscaling schemes for $\hat{r}_{16}$. Instead, the SWIR reflectance for each sample is provided by the $\tilde{r}_{16}$ value derived from trigonometric interpolation.

Figure 11(a) shows PDFs of the relative difference ($\Delta\tau$) between $\tilde{\tau}$ from the baseline test (black), as well as $\hat{\tau}$ retrieved from ~~experiments 3a~~ the partial downscaling approach of only $\hat{r}_{06}$ (blue) and ~~3d~~ the full downscaling approach (red), and the

reference results (i.e., distributions of the difference between the data sets, normalized by the reference τ). Data is from example scene 2, shown in Figure 6(b), sampled on 9 June 2013 at 10:55 UTC. The largest differences to the reference retrievals are observed for the baseline results, which only account for the large–scale reflectance variability of the cloud scene. Here, relative differences cover the range of $-20.44\% < \Delta\tau < 28.22\%$ (these values indicate the 1^{st} and 99^{th} percentile of $\Delta\tau$, respectively). The distributions for ~~experiment 3d~~ the full downscaling experiment 3 is noticeably thinner and these observed ranges are reduced significantly to $-2.33\% < \Delta\tau < 3.14\%$. The differences $\Delta\tau$ for ~~experiment 3a~~ the VNIR–only approach look closer to the one from the full downscaling experiment. However, the maximum of the distribution around $\Delta\tau \approx 0$ is lower ~~than from experiment 3d,~~ and the 1^{st} percentile is actually higher than from the baseline retrievals. Clearly, the downscaling of both VNIR and SWIR reflectances is preferable for the retrieval of $\hat{\tau}$. For the effective droplet radius, the experiment comparison looks significantly different. Both relative differences Δr_{eff} based on the baseline and ~~experiment 3d~~ full downscaling experiment results exhibit a similar behavior and the full downscaling approach only yields small improvements on the retrievals from trigonometric interpolation. Conversely, Δr_{eff} from ~~experiment 3a~~ partial downscaling yields a noticeably larger spread and the retrievals become less reliable.

Regarding ΔW_{L} and ΔN_{D}, the results using the complete downscaling approach yield the narrowest distributions, with significantly smaller minimum and maximum deviations (up to a factor of 5.6) compared to the VNIR–only ~~downscaling~~ technique. Compared to the baseline results the reliability of derived liquid water path ~~from experiment 3d~~ is also improved, even though just the VNIR reflectance is downscaled.

A summary of the performance of ~~downscaling experiments 1a–3a (i.e., where only the VNIR reflectances are downscaled) compared to that of experiments 1b–3b (i.e., the full downscaling approaches) for~~ the partial and full downscaling approach for experiments 1–3 for all four example cloud scenes is given in ~~table~~ Table 2. Here, the 1^{st}, 50^{th}, and 99^{th} percentiles of the relative differences between $\hat{\tau}$ and $\hat{r}_{\text{eff}}$ and the reference retrievals are listed. An almost universal reduction in the biases is observed when both VNIR and SWIR reflectances are downscaled. These results provide strong evidence that ~~simulateneous~~ simultaneous downscaling of the SWIR reflectances is essential for providing reliable higher–resolution retrievals of $\hat{\tau}$ and $\hat{r_{\text{eff}}}$, as well as the subsequently calculated $\hat{W}_{\text{L}}$ and $\hat{N}_{\text{D}}$. This confirms the findings in Werner et al. (2018b), who illustrated that SWIR reflectances differ significantly between the pixel–level and subpixel scale and that reliable cloud property retrievals should avoid scale mismatches between the reflectances from the VNIR and SWIR channels.

This result is likely also relevant for retrieving cloud properties at highest–possible resolution from other multi–resolution sensors such as MODIS, VIIRS and GOES–R: here, VNIR reflectances are generally available at highest spatial resolution, while SWIR reflectances have a 2–4 times lower sampling resolution. Based on the previous results, smooth interpolation of the SWIR reflectances to the VNIR resolution cannot be recommended. Instead, downscaling approaches such as those presented in ~~section~~ Section 4 should be adopted to avoid a scale–mismatch in the spatial variability captured by the VNIR and SWIR channels, or equivalently, a degraded accuracy of the r_{eff}–retrieval.

8 Conclusions

In this work, several candidate approaches to downscale SEVIRI channel 1–3 reflectances ~~from their~~ are evaluated, which increases their spatial resolution from the native horizontal resolution ~~of~~ ($3 \times 3\,\mathrm{km}^2$ ~~to the horizontal $1 \times 1\,\mathrm{km}^2$ scale~~ at the sub–satellite point) to the three times higher spatial resolution of the narrowband HRV channel observations~~are evaluated~~. The goal is to identify a reliable downscaling approach to provide the means to resolve higher–resolution, subpixel reflectance and cloud property variations, which are only resolved by reflectances from SEVIRI's coincident HRV channel. The higher–resolution reflectances are subsequently used to retrieve cloud optical thickness ($\hat{\tau}$) and effective droplet radius ($\hat{r}_{\mathrm{eff}}$). These subsequently provide the means to derive estimates of the liquid water path ($\hat{W}_{\mathrm{L}}$) and droplet number concentration ($\hat{N}_{\mathrm{D}}$).

Three different methods are presented and evaluated: (i) a statistical downscaling approach using globally determined fit coefficients based on bivariate statistics, (ii) a local approach that assumes a constant heterogeneity index for different scales (i.e., the *Constant Reflectance Ratio Approach*), and (iii) an iterative approach utilizing both global statistics and the shape of the SEVIRI LUT (which consists of simulated SEVIRI reflectances for different viewing geometries and combinations of cloud properties), while assuming a constant subpixel ~~$\tilde{r}_{\mathrm{eff}}$~~ $\tilde{r}_{\mathrm{eff}}$ (i.e., the *LUT Approach*). For the latter technique, two modifications (by assuming adiabatic cloud conditions or by deriving local slopes within the LUT) are introduced, which avoid the constraint of a fixed $\tilde{r}_{\mathrm{eff}}$.

The different downscaling approaches are evaluated using MODIS observations of four example cloud fields at a horizontal resolution of ~~$1 \times 1\,\mathrm{km}^2$, which are obtained by re–mapping~~ $\approx 1 \times 1\,\mathrm{km}^2$ (i.e., comparable to SEVIRI's HRV channel), which are remapped onto the higher–resolution SEVIRI grid, followed by ~~an optional~~ smoothing with the ~~sensor spatial response function~~ modulation transfer functions of SEVIRI.This approach has the benefit of providing a reference data set to which the results ~~for~~ from the different downscaling techniques can be objectively compared. ~~Simply using~~

The retrievals based on native–resolution reflectances (at a scale of $\approx 3\,\mathrm{km}$) are characterized by significant deviations from the reference retrievals, especially for $\hat{\tau}$ and $\hat{W}_{\mathrm{L}}$. Here, random absolute deviations as large as ≈ 14 and $\approx 89\,\mathrm{g\ m}^{-2}$ are observed, respectively (determined from the 1^{st} or 99^{th} percentiles of the absolute deviations between native and reference results for each cloud scene). For $\hat{r}_{\mathrm{eff}}$ and $\hat{N}_{\mathrm{D}}$ deviations of up to $\approx 6\,\mu\mathrm{m}$ and $\approx 177\,\mathrm{cm}^{-3}$ exist, respectively.

Simply applying trigonometric interpolation of ~~radiances~~ the reflectance to the higher–resolution grid of the HRV channel (i.e., the baseline approach) provides a ~~significant improvement in~~ significantly improved agreement with the reference ~~dataset for $\hat{\tau}$ and $\hat{r}_{\mathrm{eff}}$ compared to the native 3 km resolution results.~~ data set for τ and r_{eff} (i.e., the actual higher–resolution retrievals) compared to SEVIRI's native lower–resolution results. This improvement can be attributed to the use of higher–resolution reflectances, which resolve the large–scale variability of the scene. It is shown that either downscaling approach, which applies estimates of the unresolved small–scale variability to the reflectance field, yields reliable retrievals of $\hat{\tau}$ at the horizontal resolution of the SEVIRI HRV channel. These results compare noticeably better with the reference retrievals than the ones from the baseline approach. ~~This improvement~~ The improved performance is illustrated by a lower median absolute bias and spread (factor of 2–10), as well as a higher observed correlation between the data sets. ~~Regarding~~ The reliability of

$\hat{r}_{\mathrm{eff}}$ ~~, the baseline results are found to be reliable. Depending on the cloud type, the various downscaling techniques exhibit a significantly worse agreement with the reference retrievals~~utilizing the *LUT Approach* with an adjustment based on the calculation of isoline slopes in the SEVIRI LUT is comparable to the baseline results, and improves upon the retrievals at the native LRES resolution. The performance of the other downscaling approaches depends on the observed cloud scene. For more ~~homogeneous altocumulus fields, the *LUT Approach* with adiabatic assumptions seems inadequate, while for the more~~ heterogeneous cloud fields the performance of the statistical downscaling ~~technique~~ approach and the *Constant Reflectance Ratio Approach* decreases noticeably. The ~~reliability of $\hat{r}_{\mathrm{eff}}$ utilizing the~~ former technique relies on large–scale statistical relationships between the reflectances, which might vary with the size of the observed region, prevalence of different cloud types and viewing geometry. The latter technique, meanwhile, was developed for optically thin clouds, where the relationship between VNIR and SWIR reflectance can be approximated by a linear function (Werner et al., 2018b). Conversely, for more homogeneous altocumulus fields the *LUT Approach* with ~~an adjustement based on the calculation of local slopes is comparable to the baseline results,~~ *Adiabatic Adjustment* seems inadequate and yields the worst comparison to the reference effective radius. The study by Miller et al. (2016), following similar studies, illustrated that drizzle and cloud top entrainment yield vertical cloud profiles closer to homogeneous assumptions and away from the adiabatic cloud model. Similar processes might affect the retrieval for the presented cloud scenes in this study.

Due to the fact that these variables are derived from retrieved $\hat{\tau}$ and ~~improves upon the results at the native 3 km resolution. Overall~~$\hat{r}_{\mathrm{eff}}$, a similar behavior is observed for the derived $\hat{W}_{\mathrm{L}}$ and $\hat{N}_{\mathrm{D}}$. ~~Here, the~~ Again, the *Adjusted LUT Approach* ~~,~~ in combination with the use of local slopes ~~,~~ exhibits the best agreement to the reference results for 27 out of the 32 comparisons (i.e., four example scenes, two cloud variables, and four evaluation parameters). Based on these results, this method seems to be favorable compared to the other downscaling approaches. The results are preferable to those obtained from the standard–resolution SEVIRI narrowband reflectances and pave the way for future higher–resolution cloud products by the MSG–SEVIRI imager. Especially for $\hat{\tau}$ and $\hat{W}_{\mathrm{L}}$, these improvements are significant, as even the baseline results show deviations from the reference data set of up to ≈ 11 and $\approx 70\,\mathrm{g\,m^{-2}}$ for the observed example scenes.

~~Each of the downscaling techniques utilizes~~ Most of the presented downscaling techniques utilize a well established relationship between the observed reflectance from SEVIRI channels 1 and 2, as well as the one from the broadband HRV channel. To test the validity of the different assumptions for the downscaling of the SWIR band reflectance, the reliability of VNIR–only downscaling approaches is compared to the corresponding full downscaling procedure. For the former, the ~~$1 \times 1\,\mathrm{km}^2$ SWIR~~ higher–resolution SWIR observations are provided by the baseline technique. An almost universally improved reliability of the retrieved cloud products is observed when both VNIR and SWIR reflectances are downscaled. This illustrates that, ~~for reliable~~ in order to achieve reliable higher–resolution retrievals, all channels need to capture small–scale cloud heterogeneities at the same scale. ~~This implies that, for~~ 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~~, downscaling approaches should also be adopted to~~ . 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).

Naturally, these results require more evaluation with a larger data set to validate the reliability of the approach under different observational geometries and cloud situations. If a similarly good agreement to a set of reference retrievals is found for a broad range of different test scenes, a significant step towards higher–resolution SEVIRI cloud observations is achieved. If our results are confirmed, such retrievals would ~~be a significant improvement of~~ represent a noticeable improvement upon SEVIRI's current standard–resolution retrievals. Meanwhile, more elaborate downscaling schemes could potentially improve upon the methods presented here. As an example, one possible improvement on the *Adjusted Lookup Table Approach* with ~~adiabatic assumptions~~ *Adiabatic Adjustment* would be an explicit fit of the relationship in Eq.(16) from the native, lower–resolution variables. This might also reveal valuable insights into the validity of the adiabatic assumption commonly adopted in remote sensing (Merk et al., 2016). In addition, a comprehensive evaluation of the benefits of the higher–resolution SEVIRI cloud products for the subsequent estimation of solar surface irradiance is planned. In particular, a comparison of satellite retrievals based on Greuell et al. (2013) with observations of a dense network of pyranometers following the approach of (Deneke et al., 2009) and (Madhavan et al., 2017) is planned, which will enable detailed studies of the effects of spatial and temporal resolution of satellite observations.

This work clearly demonstrated that the *Adjusted LUT Approach* with *LUT Slope Adjustment* yields reliable higher–resolution cloud products. A follow–up study by Deneke et al. (2019) will provide a comprehensive description of the overall retrieval scheme for obtaining cloud properties and solar radiative fluxes from the Meteosat SEVIRI instrument at the spatial resolution of its HRV channel, which will be established based on the findings of this study. That companion paper also includes a statistical comparisons between the operational MODIS C6.1 and SEVIRI results, as well as the new high–resolution SEVIRI products. Moreover, some interesting use cases are demonstrated in that study, which can benefit from an increase in the spatial resolution of the derived SEVIRI cloud parameters. The companion paper also presents an important extension of this approach to the retrieval of solar surface irradiance, based on the schemes presented in Deneke et al. (2008) and Greuell et al. (2013). Satellite products with high temporal and spatial resolution are of particular interest for forecasting the production of solar power.

Code and data availability. The MODIS and MSG radiance data used as input to the CPP retrieval, the Python code used for their generation, and the retrieval output are available from the authors on request, and will be made publically available through the *ZENODO* data repository for the final paper. The CPP software is copyrighted by EUMETSAT, and is not publically available.

Author contributions. Both authors have shaped the concept of this study and in particular refined the considered downscaling approaches through extensive discussions. HD implemented the processing of the high–resolution processing scheme including the different downscaling approachs. FW carried out the analysis of the output and wrote the initial draft manuscript, which was subsequently refined by both authors.

Competing interests. The authors declare no competing interests.

Acknowledgements. This study was carried out within the frame of the German collaborative project MetPVNet funded by the German Ministry of Commerce, grant number 0350009E. The use of MODIS data obtained from the Level–1 and Atmosphere Archive and Distribution System Distributed Active Archive Center, and the use of SEVIRI data distributed by EUMETSAT and obtained from the TROPOS satellite archive is gratefully acknowledged. The lead author, Frank Werner, is now employed by the Jet Propulsion Laboratory, California Institute of Technology. This work was done as a private venture and not in the author's capacity as an employee of the Jet Propulsion Laboratory, California Institute of Technology. The authors thank Anja Hünerbein, Fabian Senf, Marion Schrödter–Homscheidt, and Michael Schwartz for comments on earlier drafts of this paper, which helped to improve the submitted version.

References

Ardanuy, P. A., Han, D., and Salomonson, V. V.: The Moderate Resolution Imaging Spectrometer (MODIS), IEEE Trans. Geosci. Remote Sensing, 30, 2–27, 1992.

Barker, H. and Liu, D.: Inferring optical depth of broken clouds from Landsat data, J. Climate, 8, 2620–2630, 1995.

Barnes, W. L., Pagano, T. S., and Salomonson, V. V.: Prelaunch characteristics of the 'Moderate Resolution Imaging Spectroradiometer' (MODIS) on EOS–AM1, IEEE Trans. Geosci. Remote Sensing, 36, 1088–1100, 1998.

Benas, N., Finkensieper, S., Stengel, M., van Zadelhoff, G.-J., Hanschmann, T., Hollmann, R., and Meirink, J. F.: The MSG-SEVIRI-based cloud property data record CLAAS-2, Earth System Science Data, 9, 415–434, https://doi.org/10.5194/essd-9-415-2017, https://www.earth-syst-sci-data.net/9/415/2017/, 2017.

Benestad, R. E.: Empirical–statistical downscaling in climate modeling, Eos Trans., 85, 417–422, doi:10.1029/2004EO420 002, 2011.

Bennartz, R.: Global assessment of marine boundary layer cloud droplet number concentration from satellite, J. Geophys. Res. Atmos., 112, https://doi.org/10.1029/2006JD007547, http://dx.doi.org/10.1029/2006JD007547, 2007.

Bley, S. and Deneke, H.: A threshold-based cloud mask for the high–resolution visible channel of Meteosat Second Generation SEVIRI, Atmos. Meas. Tech., 6, 2713–2723, https://doi.org/https://doi.org/10.5194/amt-6-2713-2013, 2013.

Bley, S., Deneke, H., and Senf, F.: Meteosat-Based Characterization of the Spatiotemporal Evolution of Warm Convective Cloud Fields over Central Europe, Journal of Applied Meteorology and Climatology, 55, 2181–2195, https://doi.org/10.1175/JAMC-D-15-0335.1, https://doi.org/10.1175/JAMC-D-15-0335.1, 2016.

Brenguier, J.-L., Pawlowska, H., Schüller, L., Preusker, R., Fischer, J., and Fouquart, Y.: Radiative properties of boundary layer clouds: Droplet effective radius versus number concentration, J. Atmos. Sci., 57, 803–821, 2000.

Cahalan, R., Ridgway, W., Wiscombe, W., and Bell, T.: The albedo of fractal stratocumulus clouds, J. Atmos. Sci., 51, 2434–2455, 1994a.

Cahalan, R., Ridgway, W., Wiscombe, W., Gollmer, S., and Harshvardhan: Independent pixel and Monte Carlo estimates of stratocumulus albedo, J. Atmos. Sci., 51, 3776–3790, 1994b.

Chambers, L., Wielicki, B., and Evans, K.: Accuracy of the independent pixel approximation for satellite estimates of oceanic boundary layer cloud optical depth, J. Geophys. Res., 102, 1779–1794, 1997.

Cros, S., Albuisson, M., and Wald, L.: Simulating Meteosat–7 broadband radiances using two visible channels of Meteosat–8, Sol. Energy, 80, 361—367, 2006.

Davis, A., A. Marshak, R. Cahalan, and W. Wiscombe (1997), The Landsat scale break in stratocumulus as a three–dimensional radiative transfer effect: Implications for cloud remote sensing, *J. Atmos. Sci.*, *54*(2), 241–260.

Deneke, H. and Roebeling, R.: Downscaling of METEOSAT SEVIRI 0.6 and 0.8 μm channel radiances utilizing the high-resolution visible channel., Atmos. Chem. Phys., 10, 9761–9772, doi:10.5194/acp–10–9761–2010, 2010.

Deneke, H., Roebeling, R., and Feijt, A.: Estimating surface solar irradiance from METEOSAT SEVIRI derived cloud properties., Rem. Sens. Environ., 112(6), 3131–3141, 2008.

Deneke, H., Knap, W., and Simmer, S.: Multiresolution Analysis of the Temporal Variance and Correlation of Transmittance and Reflectance of an Atmospheric Column, J. Geophys. Res., 114, D17 206, doi:10.1029/2008JD011 680, 2009.

Deneke, H., Bley, S., Werner, F., and Senf, F.: Increasing the Spatial Resolution of Cloud Property Retrievals from Meteosat SEVIRI by Use of its High-Resolution Visible Channel: Implementation and Examples, Atmos. Meas. Tech. Discuss., in preparation, 2019.

EUMETSAT: SEVIRI Modulation Transfer Function (MTF) characterisations for MSG-1, MSG-2 and MSG-3, Tech. rep., EUMETSAT, Darmstadt, Germany, http://www.eumetsat.int/idcplg?IdcService=GETFILE&dDocName=ZIPMSGSEVIRISPECRESCHAR&RevisionSelectionMeth 2006.

Greuell, W., Meirink, J. F., and Wang, P.: Retrieval and validation of global, direct, and diffuse irradiance derived from SEVIRI satellite observations, J. Geoph. Res. Atmos., 118, 2340–2361, 2013.

Grosvenor, D. P., Sourdeval, O., Zuidema, P., Ackerman, A., Alexandrov, M. D., Bennartz, R., Boers, R., Cairns, B., Chiu, J. C., Christensen, M., Deneke, H., Diamond, M., Feingold, G., Fridlind, A., HÃŒnerbein, A., Knist, C., Kollias, P., Marshak, A., McCoy, D., Merk, D., Painemal, D., Rausch, J., Rosenfeld, D., Russchenberg, H., Seifert, P., Sinclair, K., Stier, P., vanÂ Diedenhoven, B., Wendisch, M., Werner, F., Wood, R., Zhang, Z., and Quaas, J.: Remote Sensing of Droplet Number Concentration in Warm Clouds: A Review of the Current State of Knowledge and Perspectives, Reviews of Geophysics, 56, 409–453, https://doi.org/10.1029/2017RG000593, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2017RG000593, 2018.

Le Gléau, H.: Algorithm Theoretical Basis Document for the Cloud Product Processors of the NWC/GEO, Tech. rep., Météo France, 2016.

Lee, T. E., Miller, S. D., Turk, F. J., Schueler, C., Julian, R., Deyo, S., Dills, P., and Wang, S.: The NPOESS VIIRS Day/Night Visible Sensor, Bull. Amer, 87, 191–199, https://doi.org/doi:10.1175/BAMS-87-2-191, http://dx.doi.org/10.1175/BAMS-87-2-191, 2006.

Madhavan, B. L., Deneke, H., Witthuhn, J., and Macke, A.: Multiresolution analysis of the spatiotemporal variability in global radiation observed by a dense network of 99 pyranometers, Atmos. Chem. Phys., 17, 3317–3338, 2017.

Marshak, A., Platnick, S., Varnai, T., Wen, G. Y., and Cahalan, R. F.: Impact of three-dimensional radiative effects on satellite retrievals of cloud droplet sizes, J. Geophys. Res., 111, 2006.

Martins, J. P. A., Cardoso, R. M., Soares, P. M. M., Trigo, I. F., Belo-Pereira, M., Moreira, N., and Tomé, R.: The summer diurnal cycle of coastal cloudiness over west Iberia using Meteosat/SEVIRI and a WRF regional climate model simulation, International Journal of Climatology, 36, 1755–1772, https://doi.org/10.1002/joc.4457, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.4457, 2016.

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

Miller, D. J., Zhang, Z., Ackerman, A. S., Platnick, S., and Baum, B. A.: The impact of cloud vertical profile on liquid water path retrieval based on the bispectral method: A theoretical study based on large-eddy simulations of shallow marine boundary layer clouds, J. Geophys. Res., 121, 4122–4141, https://doi.org/10.1002/2015JD024322, http://dx.doi.org/10.1002/2015JD024322, 2016.

Nakajima, T. and King, M.: Determination of the optical thickness and effective particle radius of clouds from reflected solar radiation measurements. Part I: Theory, J. Atmos. Sci., 47, 1878–1893, 1990.

Nakajima, T., King, M., Spinhirne, J., and Radke, L.: Determination of the optical thickness and effective particle radius of clouds from reflected solar radiation measurements. Part II: Marine stratocumulus observations, J. Atmos. Sci., 48, 728–750, 1991.

Platnick, S., King, M., Ackerman, S., Menzel, W., Baum, B., Riedi, J., and Frey, R.: The MODIS cloud products: Algorithms and examples from TERRA, IEEE Trans. Geosci. Remote Sens., 41, 459–473, 2003.

Quaas, J., Boucher, O., and Lohmann, U.: Constraining the total aerosol indirect effect in the LMDZ and ECHAM4 GCMs using MODIS satellite data, Atmos. Chem. Phys., 6, 947–955, https://doi.org/10.5194/acp-6-947-2006, https://www.atmos-chem-phys.net/6/947/2006/, 2006.

Roebeling, R., Feijt, A., and Stammes, P.: 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, https://doi.org/10.1029/2005JD006990, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JD006990, 2006.

Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., and Ratier, A.: An introduction to Meteosat Second Generation (MSG), Bull. Amer. Meteor. Soc., 83, 977–992, 2002.

Schüller, L., Bennartz, R., Fischer, J., and Brenguier, J.-L.: An Algorithm for the Retrieval of Droplet Number Concentration and Geometrical Thickness of Stratiform Marine Boundary Layer Clouds Applied to MODIS Radiometric Observations, Journal of Applied Meteorology, 44, 28–38, https://doi.org/10.1175/JAM-2185.1, https://doi.org/10.1175/JAM-2185.1, 2005.

Schulz, J., Albert, P., Behr, H., Caprion, D., Deneke, H., Dewitte, S., Dürr, B., Fuchs, P., Gratzki, A., Hechler, P., Hollmann, R., Jonston, S., Karlsson, K., Manninen, T., Müller, R., Reuter, M., Riihela, A., Roebeling, R., Selbach, N., Tetzlaff, A., Thomas, W., Werschek, M., Wolters, E., and Zelenka, A.: Operational climate monitoring from space: the EUMETSAT satellite application facility on climate monitoring (CM-SAF), Atmos. Chem. Phys., 9(5), 1687–1709, 2009.

Seethala, C., Meirink, J. F., Horváth, Á., Bennartz, R., and Roebeling, R.: Evaluating the diurnal cycle of South Atlantic stratocumulus clouds as observed by MSG SEVIRI, Atmos. Chem. Phys, 18, 13 283–13 304, 2018.

Senf, F. and Deneke, H.: Satellite–based characterization of convective growth and glaciation and its relationship to precipitation formation over central Europe, J. Appl. Meteorol. Climatol., 56, 1827–1845, 2017.

Smith, W. and Timofeyev, Y., A., eds.: Stammes, P.: Spectral radiance modeling in the UV–Visible range, in: IRS 2000: current problems in atmospheric radiation, IRS 2000, Deepak Publ., Hampton, Va, 2001.

Stengel, M., Kniffka, A., Meirink, J. F., Lockhoff, M., Tan, J., and Hollmann, R.: CLAAS: the CM SAF cloud property data set using SEVIRI, Atmospheric Chemistry and Physics, 14, 4297–4311, https://doi.org/10.5194/acp-14-4297-2014, https://www.atmos-chem-phys.net/14/4297/2014/, 2014.

Szczodrak, M., Austin, P., and Krummel, P.: Variability of optical depth and effective radius in marine stratocumulus clouds, J. Atmos. Sci., 58, 2912–2926, 2001.

Twomey, S. and Seton, K. J.: Inferences of gross microphysical properties of clouds from spectral reflectance measurements, J. Atmos. Sci., 37, 1065–1069, 1980.

Varnai, T. and Marshak, A.: Statistical analysis of the uncertainties in cloud optical depth retrievals caused by three-dimensional radiative effects, J. Atmos. Sci., 58, 1540–1548, 2001.

Werner, F., Wind, G., Zhang, Z., Platnick, S., Di Girolamo, L., Zhao, G., Amarasinghe, N., and Meyer, K.: Marine boundary layer cloud property retrievals from high-resolution ASTER observations: case studies and comparison with Terra MODIS, Atmos. Meas. Tech, 9, 5869–5894, https://doi.org/10.5194/amt-9-5869-2016, http://www.atmos-meas-tech.net/9/5869/2016/, 2016.

Werner, F., Zhang, Z., Wind, G., Miller, D. J., and Platnick, S.: Quantifying the impacts of subpixel reflectance variability on cloud optical thickness and effective radius retrievals based on high–resolution ASTER observations, J. Geophys. Res. Atmos., 123, 1–20, https://doi.org/10.1002/2017JD027916, https://doi.org/10.1002/2017JD027916, 2018a.

Werner, F., Zhang, Z., Wind, G., Miller, D. J., Platnick, S., and Di Girolamo, L.: Improving cloud optical property retrievals for partly cloudy pixels using coincident higher-resolution single band measurements: A feasibility study using ASTER observations, J. Geophys. Res. Atmos., 123, 12,253–12,276, https://doi.org/10.1029/2018JD028902, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JD028902, 2018b.

Zhang, Z. and Platnick, S.: An assessment of differences between cloud effective particle radius retrievals for marine water clouds from three MODIS spectral bands, J. Geophys Res., 116, 2011.

Zhang, Z., Ackerman, A. S., Feingold, G., Platnick, S., Pincus, R., and Xue, H.: Effects of cloud horizontal inhomogeneity and drizzle on remote sensing of cloud droplet effective radius: Case studies based on large-eddy simulations, J. Geophys. Res., 117, 2012.

Zhang, Z., Werner, F., Cho, H.-M., Wind, G., Platnick, S., Ackerman, A. S., Di Girolamo, L., Marshak, A., and Meyer, K.: A framework based on 2-D Taylor expansion for quantifying the impacts of subpixel reflectance variance and covariance on cloud optical thickness and effective radius retrievals based on the bispectral method, J. Geophys. Res. Atmos., 121, https://doi.org/10.1002/ 2016JD024837, 2016.

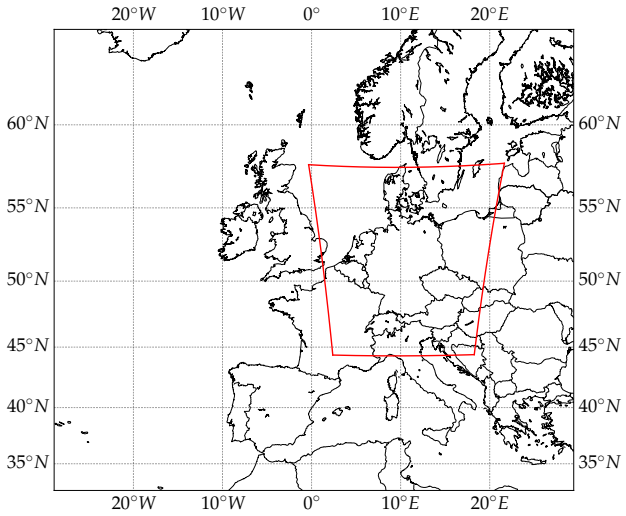

Figure 1. Map of the European SEVIRI domain, as defined in Deneke and Roebeling (2010). The red borders indicate the Germany domain, which is the focus of this study.

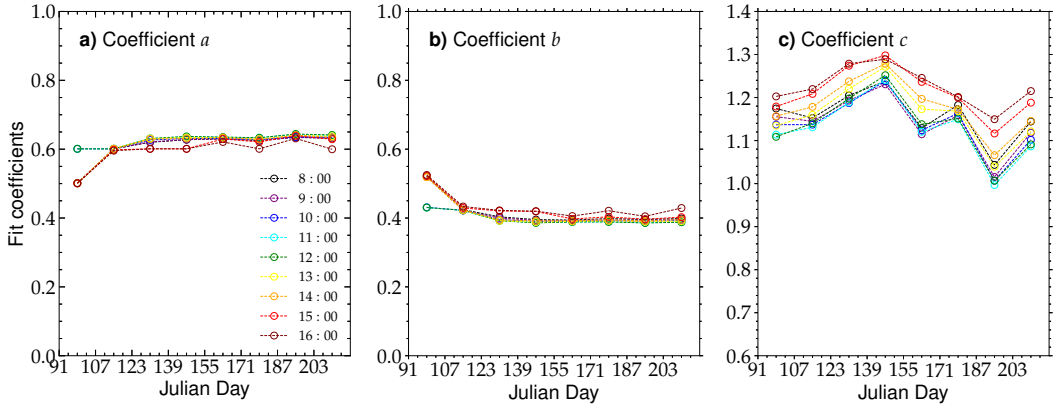

Figure 2. (a) Fit coefficients a, which are used to derive higher–resolution SEVIRI reflectances by means of statistical downscaling, as a function of Julian day. Coefficients are derived hourly and in 16–day intervals for the Germany domain between 1 April and 31 July 2013. Colors illustrate different UTC times. (b) Same as (a) but for fit coefficients b. (c) Same as (a) but for fit coefficients c.

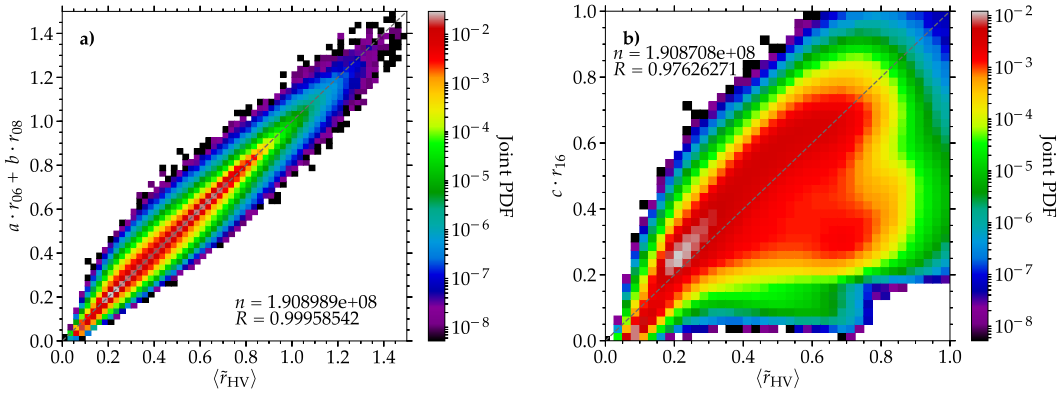

Figure 3. (a) Joint PDF of smoothed SEVIRI HRV reflectances ($\langle \tilde{r}_{\mathrm{HV}} \rangle$) and those obtained from a linear model of observed SEVIRI channel 1 (r_{06}) and channel 2 reflectances (r_{08}), specifically $a \cdot r_{06} + b \cdot r_{08}$ (see ~~section~~ Section 4.1). Data is from all 5–minute SEVIRI observations of the Germany domain during June 2013. Only cloudy pixels are considered. The number of samples (n) and correlation coefficient (R) are given. (b) Same as (a) but for a linear model for SEVIRI SWIR reflectances, specifically $c \cdot r_{16}$.

~~(a) Joint PDF of downscaled SEVIRI channel 1 reflectances ($\hat{r}_{06}$) from the *Reflectance Ratio Approach* (detailed in section 4.2) and those obtained from a linear model (described in section 4.1). Data is from all 5–minute SEVIRI observations of the Germany domain during June 2013. Only cloudy pixels are considered. The number of samples (n) and correlation coefficient (R) are given. (b) Same as (a) but for the comparison between downscaled SEVIRI channel 3 reflectances ($\hat{r}_{16}$) from the two downscaling techniques.~~

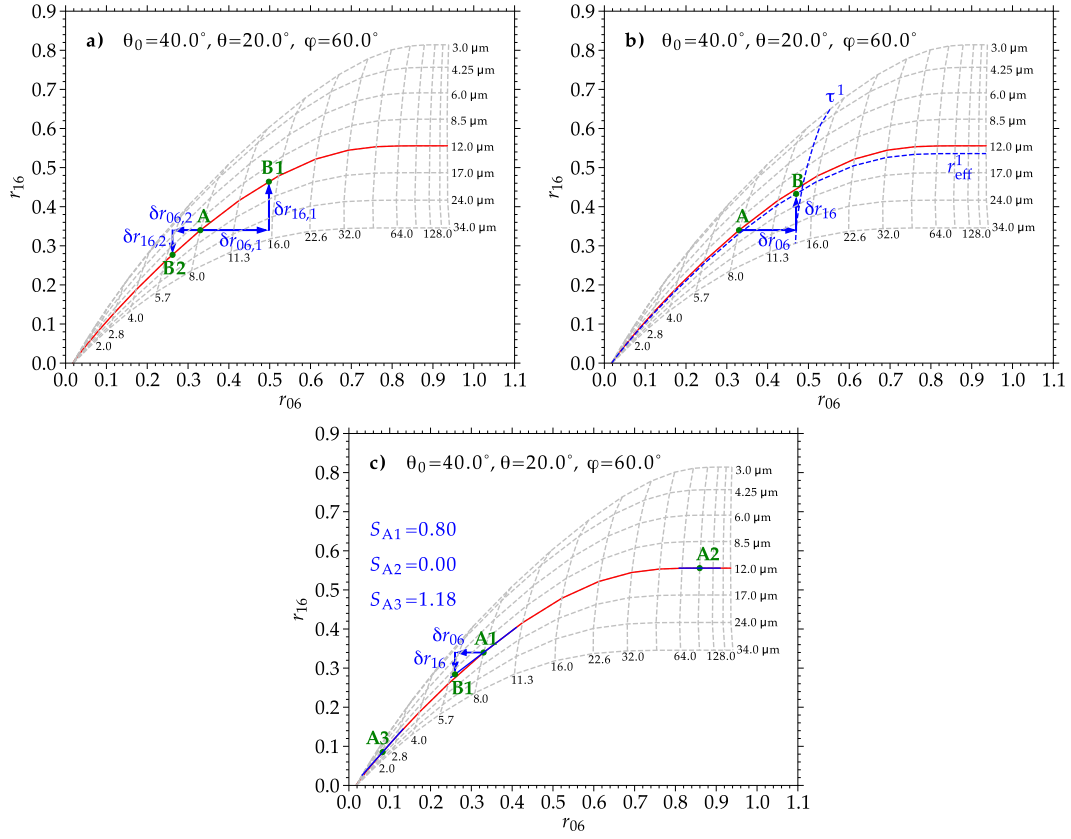

Figure 4. (a) Example SEVIRI lookup table for liquid–phase clouds, illustrating the *Lookup Table Approach* (introduced in ~~section~~ Section 4.3) for an observation highlighted by the reflectance pair indicated by point "A". For two different high–frequency variations of the channel 1 reflectance ($\delta r_{06,1}$ and $\delta r_{06,2}$) the derived high–frequency variations of the channel 3 reflectance ($\delta r_{16,1}$ and $\delta r_{16,2}$) is shown. See text for more description. (b) Same as (a) but illustrating the *Adjusted Lookup Table Approach* (introduced in ~~section~~ Section 4.4) with the *Adiabatic Adjustment* for a single δr_{06} example. (c) Same as (b) but with the *LUT Slope Adjustment*.

~~(a) PDF of the relative difference (Δr_{06}) between the downscaled SEVIRI channel 1 reflectances from the statistical downscaling approach (SD $\hat{r}_{06}$) and the higher–resolution channel 1 reflectances from trigonometric interpolation ($\tilde{r}_{06}$) is shown in black. Also shown is the relative difference between $\hat{r}_{06}$ from the statistical downscaling approach (SD; shown in red), as well as the *Constant Reflectance Ratio Approach* (RR; shown in blue), and the resampled original observations (r_{06}). Data is from SEVIRI observations of the Germany domain on 9 June 2013 at 10:55 UTC. Only cloudy pixels are considered. The 1^{st}, 50^{th}, and 99^{th} percentiles are given, as well as the normalized root-mean-square deviation (nRD; defined as the RD between the two data sets, normalized by the mean $\tilde{r}_{06}$ and r_{06}, respectively). (b) Same as (a) but for SEVIRI channel 3 reflectances (r_{16}, $\hat{r}_{16}$, and $\tilde{r}_{16}$).~~

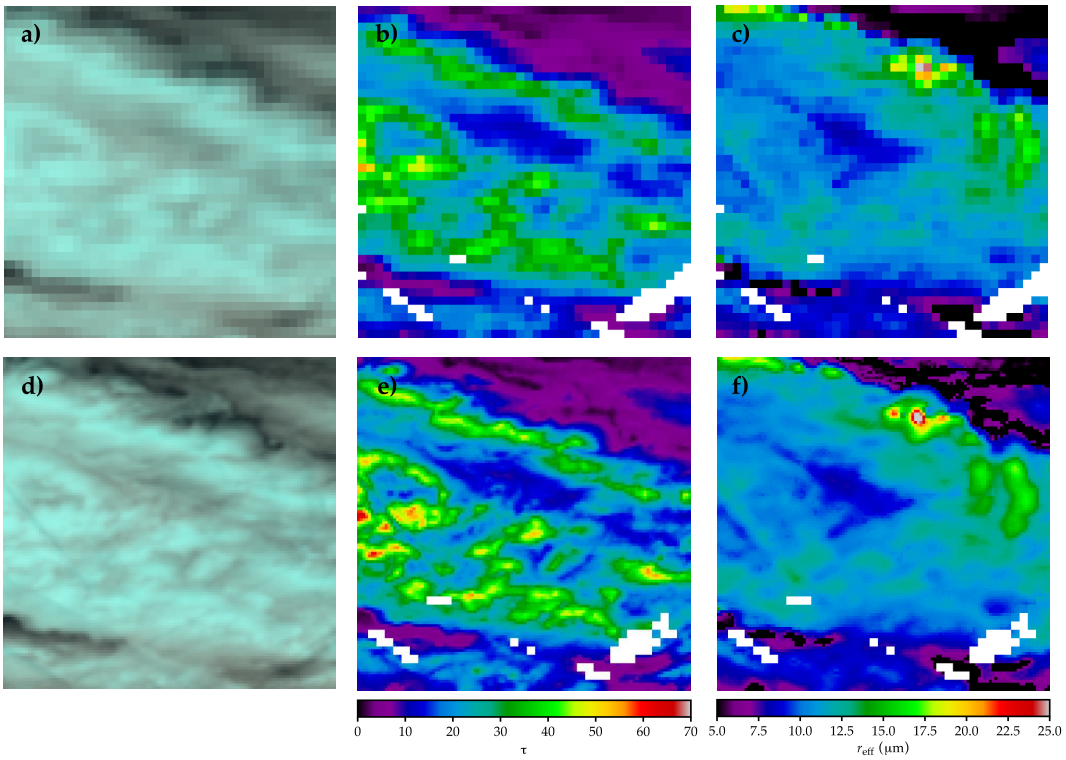

Figure 5. (a) RGB composite image of SEVIRI channel 3, 2, and 1 reflectances at the instrument's native horizontal resolution of $3 \times 3\,\mathrm{km}^2$. Data is from a $\approx 100 \times 100\,\mathrm{km}^2$ subregion within the Germany domain on 9 June 2013 at 10:55 UTC. (b) Similar to (a), but illustrating a map of the cloud optical thickness (τ). White colors indicate pixel with either a failed cloud property retrieval, a non–liquid cloud phase, or non–cloud designation by the cloud masking algorithm. (c) Same as (b) but for the effective droplet radius (r_{eff}). (d)–(f) Same as (a)-(c) but at a horizontal resolution of $1 \times 1\,\mathrm{km}^2$. The reflectances and retrievals have been derived from the *Adjusted Lookup Table Approach* as described in ~~section~~ Section 4.4, using the *LUT Slope Adjustment*.

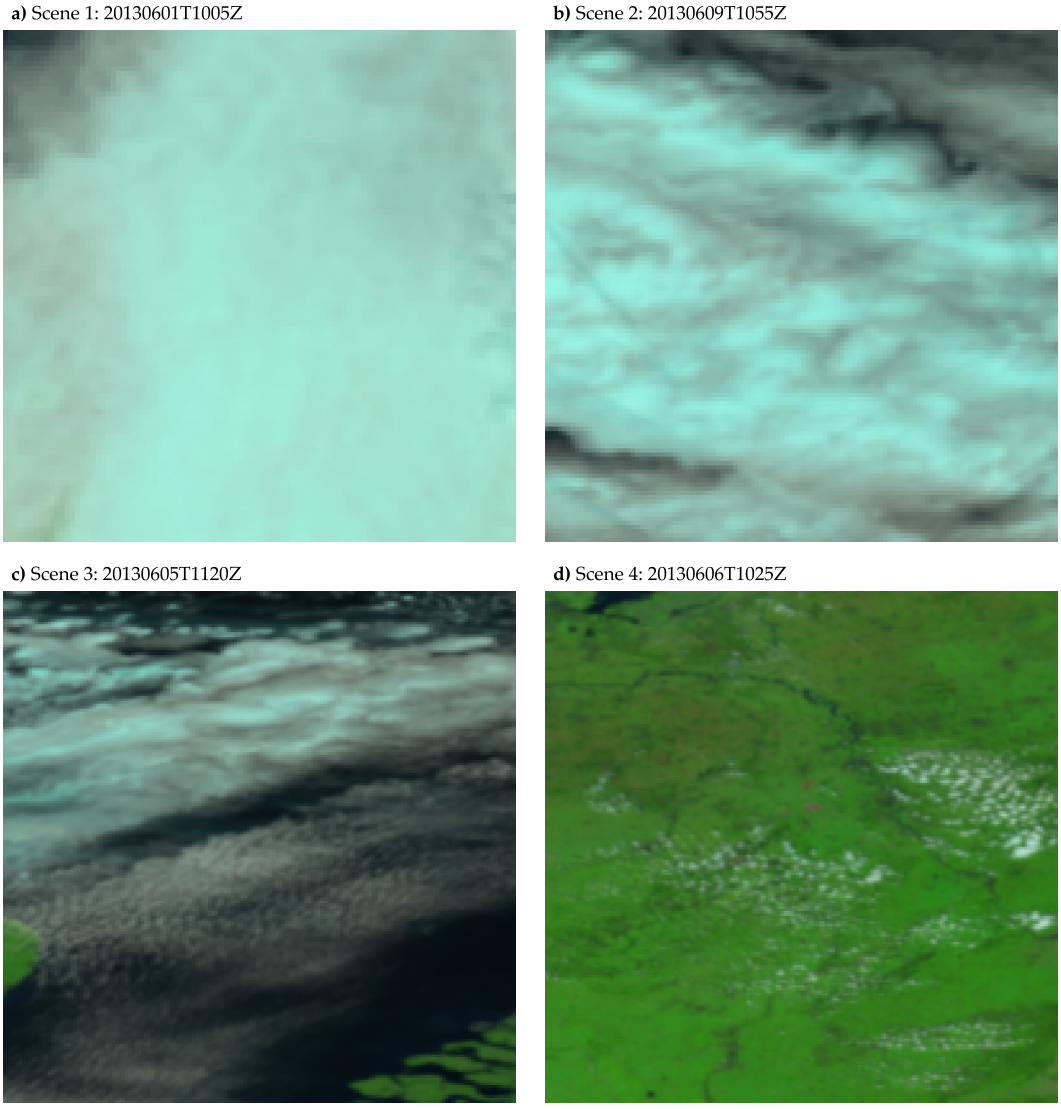

Figure 6. (a) RGB composite image of ~~SEVIRI~~ remapped MODIS channel ~~3~~6, 2, and 1 reflectances at the horizontal resolution of SEVIRI's HRV channel at a horizontal scale of $1 \times 1\,\mathrm{km}^2$ ~~for~~ at the sub–satellite point. Data is from example scene 1 sampled on 1 June 2013 ~~sampled~~ at 10:05 UTC. ~~The reflectances have been derived from the *Adjusted Lookup Table Approach* as described in section 4.4, using the *LUT Slope Adjustment*.~~ (b)–(d) Same as (a) but for example scenes 2 to 4, sampled on 9, 6, and 5 June 2013 at 10:55, 11:20, and 10:25 UTC, respectively.

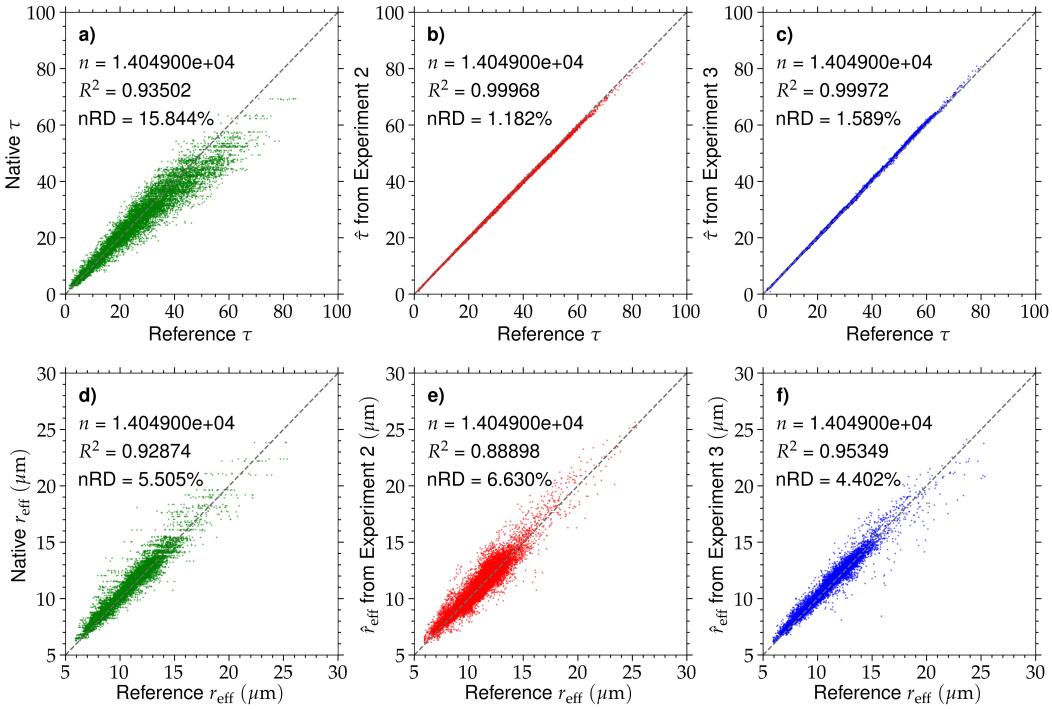

Figure 7. (a) Retrieved cloud optical thickness (τ) ~~from the~~ at SEVIRI's native ~~3 km retrieval~~ LRES resolution as a function of the reference results (τ derived from the ~~collocated~~ remapped MODIS reflectances at the ~~$1 \times 1\,\mathrm{km}^2$ scale~~HRES–scale of $\approx 1\,\mathrm{km}$). Data is from example scene 2, sampled on 9 June 2013 at 10:55 UTC. The gray, dashed line represents the 1:1 line. The number of samples (n), ~~correlation coefficient~~ explained variance (~~R~~R^2) and normalized root-mean-square deviation (nRD; defined as the RD between the two data sets, normalized by the average reference τ) are given. (b)–(c) Same as (a) but for the comparison between τ and the downscaling results ($\hat{\tau}$) from experiments ~~1b~~ 2 (*Constant Reflectance Ratio Approach*) and ~~3d~~ 3 (~~$\hat{\tau}$~~*Adjusted Lookup Table Approach* with *LUT Slope Adjustment*), respectively. (d)–(f) Same as (a)–(c) but for the effective droplet radius (r_{eff} and $\hat{r}_{\mathrm{eff}}$).

Figure 8. (a) Comparison of retrieved cloud optical thickness (τ, bottom panels) and effective droplet radius (r_{eff}, top panels) from the native ~~3~~LRES resolution (at a scale of ≈ 3 km~~resolution~~) and baseline retrievals (i.e., only accounting for low–resolution reflectance variability), as well as the ~~various~~ downscaling experiments 1 (~~1b~~statistical downscaling approach), ~~2b~~2 (*Constant Reflectance Ratio Approach*), ~~3b, 3c,~~ and ~~3d~~3 (*Adjusted Lookup Table Approach with LUT Slope Adjustment*), and the reference retrieval results. Parameters to quantify the comparisons are the median of the relative difference to the reference (p50), relative interquartile range (IQR; 75^{th}-25^{th} percentile of the relative difference to the reference), normalized root-mean-square deviation (nRD; defined as the RD between the two data sets, normalized by the average reference retrieval), and the explained variance (R^2). Green colors indicate the experiment that compares best to the reference results, i.e., highest R^2 and lowest p50, IQR, and nRD. Red colors indicate the experiment with the worst agreement to the reference retrievals, while yellow colors indicate all experiments in between. Data is from example scene 1 sampled on 1 June 2013 sampled at 10:05 UTC. (b)–(d) Same as (a) but for example scene 2 to 4, respectively.

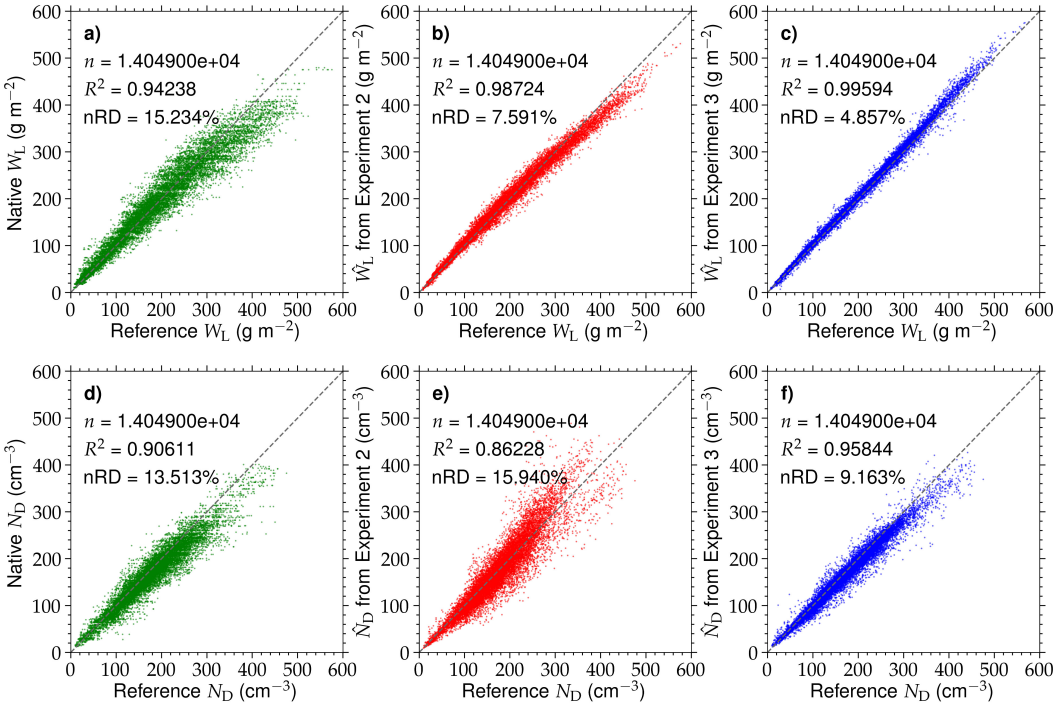

Figure 9. (a) Retrieved liquid water path (W_L) ~~from the~~ at SEVIRI's native ~~3 km~~ LRES resolution ~~retrieval~~ as a function of the reference results (W_L derived from the ~~colocated~~ remapped MODIS reflectances at the ~~1 × 1 km^2 scale~~HRES–scale of ≈ 1 km). Data is from example scene 2, sampled on 9 June 2013 at 10:55 UTC. The gray, dashed line represents the 1:1 line. The number of samples (n), ~~correlation coefficient~~ explained variance (~~R~~R^2) and normalized root-mean-square deviation (nRD; defined as the RD between the two data sets, normalized by the average reference W_L) are given. (b)–(c) Same as (a) but for the comparison between reference W_L and the downscaling results ($\hat{W}_\mathrm{L}$) from experiments ~~1b~~ 2 (*Constant Reflectance Ratio Approach*) and ~~3d~~ 3 (~~$\hat{W}_\mathrm{L}$~~ *Adjusted Lookup Table Approach* with *LUT Slope Adjustment*) , respectively. (d)–(f) Same as (a)–(c) but for the ~~effective~~ droplet ~~radius.~~ number concentration (N_D and $\hat{N}_\mathrm{D}$).

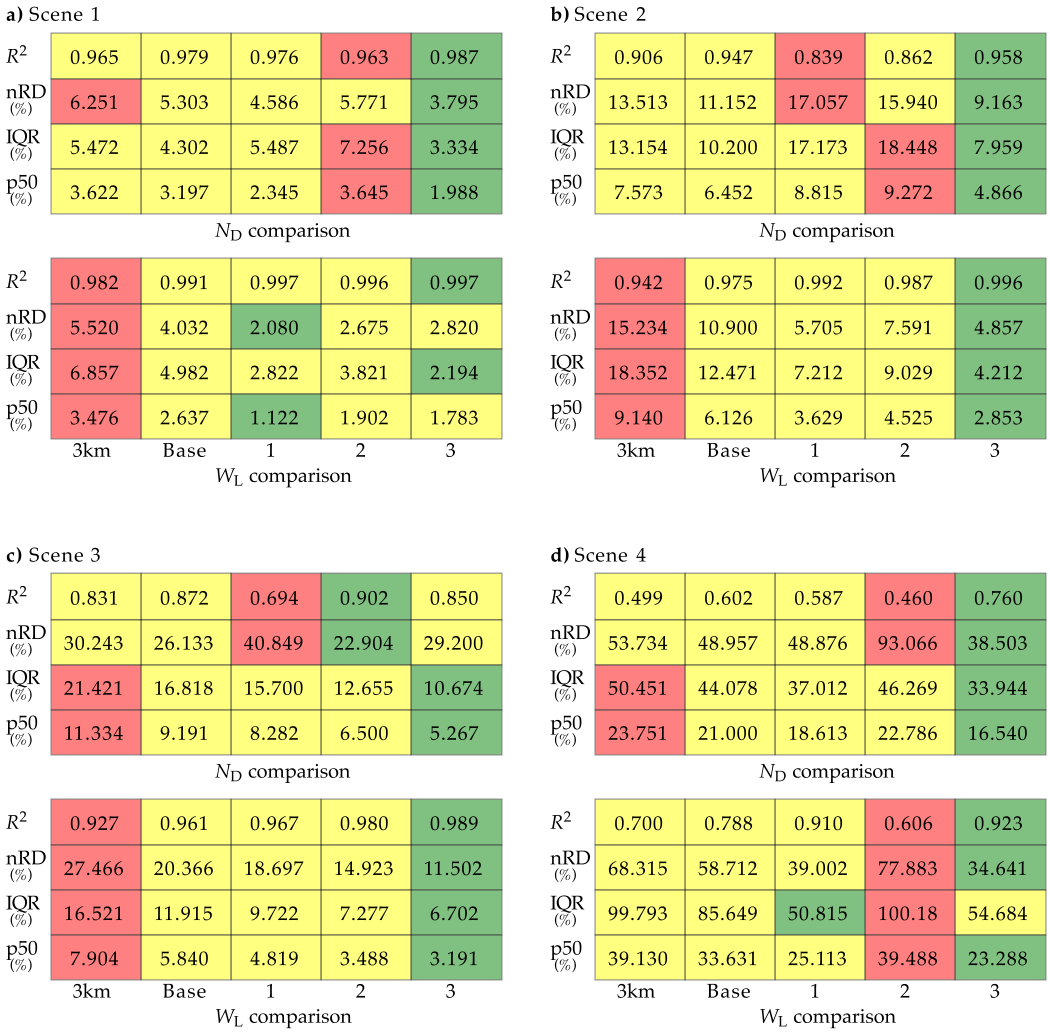

Figure 10. (a) Comparison of derived liquid water path (W_L, bottom panels) and droplet number concentration (N_D, top panels) from the native ~~3~~LRES resolution (at a scale of ≈ 3 km~~resolution~~) and baseline retrievals, as well as the ~~various~~ downscaling experiments 1 (~~1b~~statistical downscaling approach), ~~2b~~2 (*Constant Reflectance Ratio Approach*), ~~3b, 3c,~~ and ~~3d~~3 (*Adjusted Lookup Table Approach* with *LUT Slope Adjustment*), and the respective reference results. Parameters to quantify the comparisons are the median of the relative difference to the reference (p50), relative interquartile range (IQR; 75^{th}-25^{th} percentile of the relative difference to the reference), normalized root-mean-square deviation (nRD; defined as the RD between the two data sets, normalized by the average reference retrieval), and the explained variance (R^2). Green colors indicate the experiment that compares best to the reference results, i.e., highest R^2 and lowest p50, IQR, and nRD. Red colors indicate the experiment with the worst agreement to the reference retrievals, while yellow colors indicate all experiments in between. Data is from example scene 1 sampled on 1 June 2013 sampled at 10:05 UTC. (b)–(d) Same as (a) but for example scenes 2 to 4, respectively.

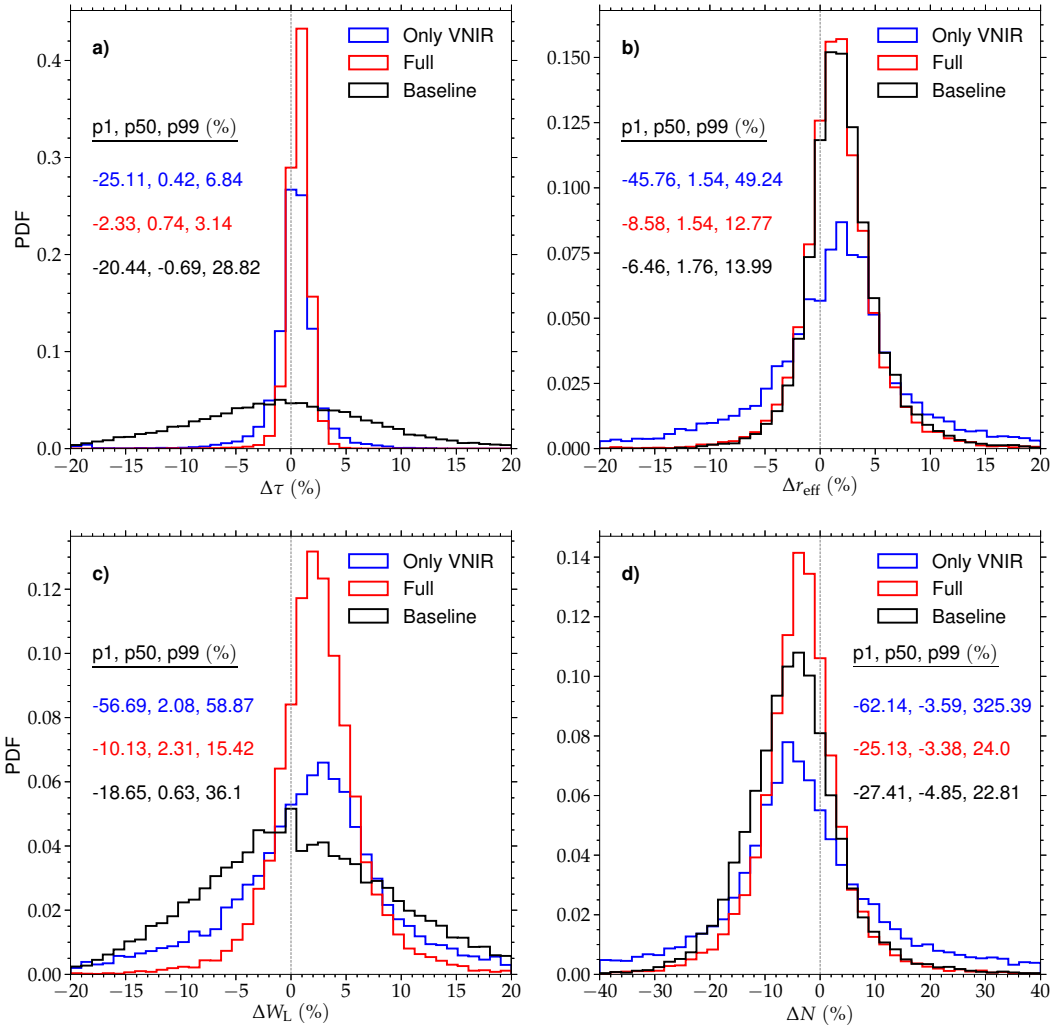

Figure 11. (a) PDFs of the relative differences ($\Delta\tau$) between the retrieved cloud optical thickness (τ) from ~~various downscaling methods (i.e.,~~ the baseline test (black), as well as ~~experiments 3a~~ a VNIR–only and ~~3d,~~ full downscaling approach for experiment 3 (shown in ~~black,~~ blue ~~,~~ and red color, respectively), and the reference results (i.e., the original 1 km–retrievals). Data is from example scene 2 sampled on 9 June 2013 at 10:55 UTC, which is shown in Figure 6(b). The 1^{st}, 50^{th}, and 99^{th} percentiles of $\Delta\tau$ for each experiment are given. (b) Same as (a) but for Δr_{eff}, which is the relative difference for the retrieved effective droplet radius (r_{eff}). (c) Same as (a) but for ΔW_{L}, which is the relative difference for the derived liquid water path (W_{L}). (d) Same as (a) but for ΔN_{D}, which is the relative difference for the derived droplet number concentration (N_{D}).

Table 1. Description for the different retrieval experiments, which are characterized by different assumptions for the downscaling of SEVIRI reflectances from the native horizontal resolution of ~~3~~≈ 3 km to the MODIS–like ~~1~~≈ 1 km scale.

Experiment	
Reference	
Native 3 km	
Baseline	
~~Native 3 km~~ 1	~~$\tilde{r}_{06}$ and $\tilde{r}_{16}$ subsampled to native SEVIRI grid, and each central value repeated 3 × 3 time~~
~~2a~~ 2	$\hat{r}_{06}$ ~~from *Reflectance Ra~~
~~3a~~ 3	$\hat{r}_{06}$ ~~from *Lookup Table Approach* as described in section 4.3; $\tilde{r}_{16}$ from trigonometric interpolation3b $\hat{r}_{06}$~~ and $\hat{r}_{16}$ from ~~*Lookup Table*

Table 2. Comparison of the cloud property retrieval results from ~~the~~ downscaling experiments ~~1a–3a~~1–3, which only account for the VNIR part, and the full downscaling experiments ~~1b–3b~~, which include adjustments to both VNIR and SWIR reflectances. The comparison shows the 1^{st}, 50^{th}, and 99^{th} percentiles of the relative differences $\Delta\tau$ (for the cloud optical thickness τ) and Δr_{eff} (for the effective droplet radius r_{eff}), which illustrate the deviation of the different retrieval approaches from the reference results, normalized by the reference retrievals. Data is from the four example scenes shown in Figure 6.

Scene	τ (%)								
	~~1a~~ 1 VNIR	~~1b~~ 1 Full	~~2a~~ 2 VNIR	~~2b~~ 2 Full	~~3a~~ 3 VNIR	~~3b~~ 3 full	~~1a~~ 1 VNIR	~~1b~~ 1 full	
#1									
1^{st}	-4.26	~~-2.59~~ -2.61	~~-3.13~~ -3.16	-1.97	~~-3.47~~ -3.48	-1.77	~~-13.20~~ -13.2	~~-5.47~~ -5.56	~~-~~
50^{th}	0.28	0.19	0.16	~~0.00~~ 0.0	0.52	0.81	0.82	0.11	
99^{th}	4.57	2.95	3.49	2.18	4.13	2.86	17.58	8.38	
#2									
1^{st}	~~-26.81~~ -26.88	~~-19.77~~ -19.82	~~-24.30~~ -24.3	~~-2.64~~ -2.63	~~-25.11~~ -25.15	~~-2.33~~ -2.36	~~-47.37~~ -47.95	~~-28.00~~ -28.29	~~-4~~
50^{th}	0.45	~~0.30~~ 0.29	~~0.21~~ 0.2	~~0.11~~ 0.12	0.42	~~0.74~~ 0.73	1.51	~~0.57~~ 0.5	
99^{th}	8.31	~~4.31~~ 4.3	6.29	~~2.66~~ 2.84	6.84	~~3.14~~ 3.13	~~53.23~~ 53.17	~~18.16~~ 18.12	~~4~~
#3									
1^{st}	~~-37.33~~ -37.34	~~-32.08~~ -31.79	~~-33.95~~ -33.96	~~-25.00~~ -24.76	-33.65	~~-20.68~~ -20.27	~~-66.37~~ -66.56	~~-45.46~~ -45.93	~~-6~~
50^{th}	~~0.00~~ 0.0	~~0.06~~ 0.0	~~0.00~~ 0.0	~~0.00~~ 0.0	~~0.25~~ 0.21	~~0.38~~ 0.35	~~0.74~~ 0.71	~~0.43~~ 0.33	
99^{th}	~~38.16~~ 38.04	~~31.44~~ 31.24	~~36.09~~ 35.97	~~23.98~~ 23.53	~~36.05~~ 36.03	~~25.56~~ 25.52	126.95	~~59.57~~ 61.12	~~11~~
#4									
1^{st}	~~-73.33~~ -78.26	~~-69.72~~ -76.37	~~-60.18~~ -66.67	~~-61.45~~ -61.98	~~-72.27~~ -76.74	~~-69.12~~ -69.13	~~-52.53~~ -53.23	~~-36.01~~ -36.14	~~-5~~
50^{th}	~~2.86~~ 2.4	~~1.47~~ 1.08	~~1.23~~ 0.65	~~7.07~~ 7.52	~~2.71~~ 2.29	~~2.61~~ 2.17	-0.13	~~0.00~~ 0.0	
99^{th}	~~309.98~~ 304.24	~~286.63~~ 284.16	~~292.25~~ 320.0	~~414.23~~ 450.08	~~308.35~~ 299.43	~~284.50~~ 280.93	~~175.70~~ 191.15	~~42.16~~ 43.01	~~12~~