Interactive comment on “Global analysis of cloud field coverage and radiative properties, using morphological methods and MODIS observations” by R. Z. Bar-Or et al.

R. Z. Bar-Or et al.
Ilan.Koren@weizmann.ac.il

Received and published: 25 November 2010

We would like to thank both anonymous referees for their helpful reviews. Their comments have helped us make changes which significantly improved the manuscript. We will start our response with providing a summary of the changes we have made to the manuscript following by a respond to each of the referees comments (in italic style) one by one.

Following the comments of referee #1, we have renamed some of the presented terms, which are now much more accurate. Referee #1 also encouraged better explaining the technical procedures behind the presented algorithm, and the manuscript now includes
more details about both the algorithm itself and the input datasets or analysis used for our work. Referee #2 had many suggestions leading to a better and clearer presentation of our work. Following her/his comments we have added a new panel to Figure 1 (Fig. 1a), rephrased several paragraphs in order to better explain complicated concepts and methods that we have used, and added useful references to the Figures and to past research. In addition, referee #2 raised important questions regarding the algorithm dependence in the input cloud mask resolution and instrumentation; these questions are now discussed in detail, clarifying the advantages and the limitations of the presented methods. And finally, her/his technical comments lead to a better written manuscript.
Referee #1 - specific comments:

1. I don’t follow the definition of $A(r)$ in step 3 in page 7. Here is what I guess what it means. Suppose there is a pixel at $P(x_1, y_1)$. For $P(x_1, y_1)$ the closest cloud pixel is $Q(x_2, y_2)$. $r = \sqrt{(x_2^2 - x_1^2) + (y_2^2 - y_1^2)}$. And $A(r) = \text{total}(Pi*r^2)$. If this is what you mean, should the distribution be for the area (or $Pi*r^2$)? The integral of probability distribution should be unitless and equal to 1. The integral of $A(r)/dr * dr$ has an unit for area. Also distributions of $r^2$ and $r$ could be very different. Please clarify this.

Thank you for this important comment. For each cloud-free pixel we measure its distance from the nearest cloud and then look on the distribution of these distances. Therefore, we have rephrased all of the related terms to "distance distribution" instead of "distance probability distribution".

"The distance cumulative distribution $A(r)$ is calculated for varying distance parameter values ($r$). $A(r)$ is the total area that is closer than $r$ from any cloud in the observed domain. Then, the distance distribution is calculated as the derivative $\frac{d}{dr} (A(r))$.

2. Fig. 3 shows the sensitivity for a MODIS observed cloud field (line 17, page 8). This figure shows cloud fraction is insensitive to resolution, while the cloud field distance does. How is the cloud field distance determined? Is it at the minimum of the distribution? Authors also state “The cloud fraction is stable in the range . . . , as expected by a resolution reduction of a binary mask”. How this was done? Please clarify.

Indeed, the cloud field distance parameter ($R_0$) for each cloud field, in any resolution, is calculated using the cloud field masking algorithm described in Section 2.1, and set to be the $r$ value of the distribution’s minimum.

When a resolution reduction of a binary matrix is done, we average all sub-pixels included in the new large pixel, and set the new pixel to be cloudy if the average is higher than 0.5, and non-cloudy if the average is lower than 0.5. In order to avoid biased
results, the case of exact value of 0.5 is randomly set (with the probability of half). We have rephrased the related paragraph in section 2.2, now saying: "For these tests, the resolution of each cloud mask data was reduced by a simple averaging of pixels, and setting all pixels averaged higher than 0.5 as cloudy, and all pixels averaged lower than 0.5 to clear. Pixels whose average was exactly 0.5, were randomly set as cloudy or clear (with the probability of 1/2), keeping the reduced resolution data statistically unbiased, and verifying that the cloud fraction is constant."

3. Fig. 5 presents mean aerosol optical depth as a function of the distance from the nearest cloud. The increment of the distance is 1 km as plotted in Fig. 5. But MODIS product only provides aerosol with a resolution of 10 km x 10 km. In MODIS algorithm, if the standard deviation of reflectance of 3 x 3 pixels exceeds some critical value, the group of the nice pixels is classified as cloud (see Remer et al., 2005). Please clarify the 1 km increment in aerosol optical depth in this figure.

All input data were projected on 1km equal area grid. That means that both "1km" cloud fraction granule and "10km" AOD granule, were interpolated to fit 1 km$^2$ equal area matrix, which was separately calculated using the MODIS geo-location product. The interpolation step in the projection process is now clarified in Section 2.1: "The projection to 1 km$^2$ equal-area cloud mask and to 1 km$^2$ equal-area ocean/land masks are done using MODIS Geolocation product. The Euclidian distance of each pixel in the MODIS product to a central point is calculated, based on its exact coordinates, enabling the interpolation of the required data on a new equal-area grid of 1 km pixels. The borders of the cloud mask interpolations were rounded."

4. Fig. 5 shows systematic feature of aerosol away from cloud over land vs over ocean. It is very nice. However, it is known that clouds also act to enhance clear sky reflectance and affect aerosol retrieval in the similar temporal scale (10 km). Author should at least to need to mention this effect. (See http://modis.gsfc.nasa.gov/sci_team/meetings/201001/presentations/poster/atmos
Thank you for this reference. This effect is actually mentioned in Section 1 and included in what is known as 3D radiative effects: "… and cloud 3D radiative effects (Marshak et al., 2006; Wen et al., 2007).". The discussed scale of 10 km, in the above reference, is very interesting and significant. We have added a reference to Wen et al., (2008) in Section 1, in order to account this specific effect as well.
Referee #2 - specific comments

1. The authors address the same topic as in Bar-Or et al. (2010) in a very similar way. It should clearly stated what the differences and similarities between the two papers are and what the additional contribution provided by this manuscript is.

This paper continues, deepening and expend the ideas presented in Bar-Or et al. (2010). Here the cloud field morphology concept is pushed much further. We are clearly stating this on the first section: "For this purpose we use a morphological algorithm to analytically define cloud field boundaries, first introduced in Bar-Or et al (2010). We provide here an extended detailed description of the algorithm. We apply this algorithm to a global cloud field coverage analysis for the first time, exposing strong climatic signature of the global circulation on cloud field coverage over lands. Moreover, we add global analysis of aerosol optical depth (AOD) and aerosol fine-mode fraction (FMF) as a function of the distance from the nearest cloud, based on more than 1 million pixels data".

2. Page 19569, lines 27–29 and also later in the paper: The authors assert that a classification of cloudy and cloud-free areas should be resolution-independent. While it is clear that a real cloud has a definite extension that does not depend on resolution, it is also expected that different instruments with different spatial resolutions will give different representations of the same cloud field. Thus, a cloud field definition will be resolution-dependent yet. Please clarify what you mean by this sentence.

We completely agree with the referee. Cloud detection depends on the instrument resolution and this will be critically important for clouds whose sizes are approaching the instrument resolution. It is also true for cloud fields. However, because cloud fields naturally occupy much larger areas than a single cloud (length scale of $\sim 100 \text{ km}$), for most instruments the resolution is fine enough to make the calculated cloud field area robust.

3. Page 19571, line 10: What do you mean by detectable clouds? Do you mean clouds
that have been detected by a standard algorithm like the MODIS one (Ackerman et al., 1998)?

Right. We refer to clouds that can be detected by the algorithm in use, as oppose to clouds that falls far below the instrument resolution and/or thin clouds with very weak optical signature. In this work, a pixel is determined as a cloud if it is defined confidently cloudy according to MODIS cloud mask product (Ackerman et al., 1998).

We have emphasized this point, in Section 2.1: "The input data for this work is the MODIS cloud mask product (Ackerman et al., 1998; Platnick et al., 2003), as all pixels that are determined to be confidently cloudy according to this product were considered to be clouds."

4. Page 19571, line 21: What do you mean by “any informative input data resolution”? The referee is right to question the phrasing. Following the topic raised in the second comment, we meant to stress the information content dependence on the resolution. Simply to say that the information content decays as the scale of the studied object (clouds or cloud fields) approaching the instrument resolution. We rephrased the sentence: "the algorithm should be applicable as long as the data resolution allows for a detailed analysis of the cloud filed properties. Hence, the resolution should be on the order or finer than a characteristic cloud size."

5. Page 19571, lines 18–22: I think that a robust cloud field masking algorithm that distinguishes between cloudy and cloud-free areas in the sense described in the paper should also satisfy the requirement that it does not (strongly) depend on the input cloud mask. Although usual cloud masking algorithms can be more or less conservative depending on their target and the intended application, the resulting separation between “cloud affected” areas and really cloud-free regions should eventually be the same. Does your algorithm satisfy this requirement? Please comment on this point.

Thanks for raising this point. Our algorithm depends mostly on the organization of the clouds within the cloud field and less on the exact shape and size of each cloud. In
other words, the distribution of the distances from the nearest cloud of a cloud field is more resilient to small changes in the particular cloud size and shape. Therefore, as long as there is suitable resolution to describe the details of the distance distribution, as the referee wrote, it will be less sensitive to the cloud algorithm. This was added to the discussion in Section 2.2: "Using cloud mask as its only input data, the algorithm is sensitive to the technique and to the quality of the chosen cloud mask product. One must consider this limitation, although there is no satisfying alternative that provides better cloud spatial data."

6. Page 19572, line 28 – page 19573, line 20: These theoretical introductive paragraphs could be understood more easily by means of a practical example that is referred to in the course of the explanations. The references to Figs. 1 and 2 at the end are not enough. Furthermore, it would be nice to consider all the steps starting from an image of the cloud field which is missing in Fig. 1.

Following the referee’s suggestion, we have added to Fig. 1 the cloud mask of the discussed field in a separate panel. Section 2 was revised and it now includes more relevant references to Figure 1: "Here, the distribution of the Euclidian distance from the nearest cloud is being used for distinguishing the inner cloud field area from the surrounding cloud-free area (see Fig. 1b). Examining the whole domain (including the cloud-free area), the distance distribution shows two different regimes: (1) the intra cloud-field regime, characterized by the distance distribution of clouds inside the field (describing the cloud spatial distribution, Fig. 1c), and (2) the extra-field regime, which asymptotically approximate distance distribution of a single giant cloud (see in Fig. 1d).

While the distance distribution inside the field has a maximum point, representing the most common distance from a cloud inside the field, following by a decrease of larger distances (Fig. 1e), the distance distribution outside the fields is monotonically increasing (with a slope that asymptotically goes to \(2\pi\), away from the cloud field as the smoothed perimeter approximate a circle, see Fig. 1d)."
7. **Page 19587**: As far as Fig. 1 is concerned, why does the scale of panel b not start at zero? The distance-to-nearest-cloud should be zero when one is considering a cloud pixel. Can you please also explain why Fig. 1d is not a zoom of Fig. 1c but a different plot? What does it represent? The contour defined by R0 could be added as well to both the cloud field representation (not plotted yet) and the cloud field distance map (Fig. 1a).

I. In order to increase the dynamic range, Figure 1b-1c present distance map in a logarithmic scale, we added this information in the caption of Fig 1, and in the figure titles. Their scale actually starts at 0, but since the given value for r=0 on these maps is log(0), we chose to label only values greater than 1 on our scale.

II. Figures 1d and 1e present counting histograms of the distance maps 1b and 1c. The difference between the domains of 1b and 1c leads to different total number of pixel count in each bin, and to different range of distances (the zoomed domain does not consist large distances). 1e represents an approximated scaled zoom of 1d, neglecting all distances related to cloud-free areas. We have added that information to the caption of Figure 1: "Note that while Fig. 1c represents a zoomed image of Fig. 1b, the corresponding Fig. 1e represents only an approximated scaled zoom of the interior distances in Fig. 1d (up to \(\sim 22\) pixels), neglecting all distances that describe cloud-free areas."

III. As recommended, in order to give better feeling about the input and output of the synthetic case, we have added panel 1a, demonstrating the input cloud mask of the field, and the extracted cloud field boundary for Fig. 1b.

8. **Page 19588**: Please add the terms A(r) and dA/dr to the caption of Fig. 2 in the appropriate way. The red curve seems very smooth while its derivative, the blue curve, is very noisy. Why is it like this?

We have modified the caption of Figure 2, and it now includes A(r) and dA(r)/dr: "Figure 2. Analysis of an observed cloud field, including the distance distribu-
tion $dA(r)/dr$ (blue line, normalized), the filtered distance distribution function (green line, normalized), the distance cumulative distribution $A(r)$ (red line), and the transition point, identified by the minimum of the filtered distance distribution function, and defining the field distance parameter (marked with an orange arrow)."

The derivative $dA(r)/dr$ is defined by the differences and therefore even small changes in the growth/decrease rate of $A(r)$ are emphasized in its graph. In addition, one should remember that both $A(r)$ and $dA(r)/dr$ suffer from the geometrical noise cause by the operation of a circular symmetry calculation (the Euclidian distance) on a discreet cubic symmetry input (the input matrices that use square pixels).

9. Page 19573, lines 13–17: The meaning of the distance parameter $R_0$ should be better clarified. It represents a characteristic length of the cloud field but it is not clear why it marks the cloud field boundary. Furthermore, it is a purely geometric quantity that in the first instance is not associated to any physical cloud or atmospheric property. Of course, clouds only form where the necessary meteorological conditions are satisfied... Please explain.

The field distance parameter $R_0$ is indeed a pure geometrical measure and this is part of its strength. It shows the transition distance from the inner organization of the cloud field to the outer region (cloud-free area). $R_0$ dependence on the inner organization of the clouds is not strong. The main assumption of this method is that cloud fields are organized in a way that limits the typical distances between neighboring clouds. Because this assumption is applicable to all cloud types and it has the flexibility to avoid assuming any preferred shape of a cloud field (circular of elliptical), it can map any cloud field (with some limitation on cases of isolated clouds discussed in Section 2.2). We added this to the paper in Section 2: "The distance value corresponding to the local minimum is defined here as the field distance parameter ($R_0$), and it represents the largest distance-from-the-nearest-cloud that is still considered to be part of the cloud field, based on the assumption that each cloud field has
a limited distance between neighboring clouds, and therefore limited intra-field distance-from-the-nearest-cloud. The contour defined by R0 marks the cloud field boundaries, distinguishing the cloud field from the surrounding cloud-free area, and demonstrating the ability of R0 to define cloud field boundaries without making any prior assumption about their geometrical shape."

10. **Page 19573, line 25**: Please add some detail about data projection. Do you use the 1 km MODIS cloud mask or the 250 m cloud mask? Which method do you use? Nearest neighbor? Do you consider sub pixel cloudiness in the resulting cloud mask? Do you correct for parallax effects?

In Section 2.1 the algorithm is described in details, without any dependency in the input data source or data resolution. Later, in our work, we use the 1 km MODIS cloud mask. The interpolation method is "triangle-based linear" and sub-pixel cloudiness is rounded to 0 or 1, using a threshold of 0.5. We did not consider for parallax effects, since these are usually included, if needed, in the cloud masking algorithms. As the referee mention earlier, part of the advantages of this algorithm is resiliency to small details of the cloud mask.

We have added further technical details about the procedure of data projection to equal-area matrices in Section 2.1: "The projection to 1 km equal-area cloud mask and to 1 km equal-area ocean/land masks are done using MODIS Geolocation product. The Euclidian distance of each pixel in the MODIS product to a central point is calculated, based on its exact coordinates, enabling the interpolation of the required data on a new equal-area grid of 1 km pixels. In case of binary interpolated data, all values between 0 and 1 are rounded."

11. **Page 19574, line 5**: Do you compute the distance in km or pixels? Is the distance a floating point? Do you consider distances from the center of a pixel to the center of the closest cloud pixel? Is there a fast way to do these computations? Can you please give at some point in the paper an indication about time consumption of the algorithm?
Does the physical extension of the data set (i.e. the size of the MODIS granule) play a role for the determination of the distance map?

I. We are calculating distances in km. We first project the data on a 1 km equal-area matrix and then calculate distances. The distance is calculated from the center of the pixel to the center of the closest cloudy pixel, and it is a floating point (Section 2: "The best metric that meets this requirement was found to be the distribution of the distance-from-nearest-cloud (Koren et al., 2007), where each element represents the Euclidian distance of the center of the pixel to the nearest cloud.").

II. The most computationally expensive step in our algorithm is the projection of the data on equal-area matrices. Luckily, this step is about to be unnecessary as soon as MODIS datasets will include "distance from the nearest cloud" product, as planned.

But nevertheless generally speaking, this step takes average of 6 min per one MODIS granule, using a single 2.66 GHz quad-core Xeon CPU.

III. The physical extension of the datasets does not play any significant role for the determination of the distance map, as we carefully cut any pixel that its distance from the nearest cloud couldn’t be determined because it was close to the granule edge. We clarify this in Section 2.1, step 2: "Additional domain size correction is done by cutting out from the data all pixels whose calculated distance from the nearest cloud is larger than their distance to the granule edge."

12. Page 19574, line 8: Can you please better explain the meaning of \( A(r) \) and its derivative and their relationship to each other? Do you first compute the distance probability distribution, as suggested by the fact that in step 2 you determine the distance map, or the distance cumulative distribution, as indicated in step 3?

The cumulative distance distribution \( A(r) \) is the total area in the examined domain that its distance from the nearest cloud is equal or smaller than \( r \). It’s derivative \( \frac{dA(r)}{dr} \) describes the distance-from-the-nearest-cloud distribution of the examined domain,
and in a discrete form - the histogram of distances from the nearest cloud.

Once the distance from the nearest cloud map is calculated and each pixel is assigned with its minimal distance to a cloud, both $A(r)$ and $dA(r)/dr$ can be directly produced. We have added a short explanation in step 3: "This procedure is equivalent to computing the distance distribution as the histogram of the distance map, but is technically faster and more sensitive to the minimum location."

13. Page 19574, line 13: How large is the Gaussian filter used? Can you please discuss here whether a minimum value $R_0$ can always be found (see also page 19575, after line 23)? How do you proceed in case no minimum could be found?

The Gaussian filter is used once in the algorithm, before determining the location of the distribution’s minimum. It’s easy to prove that such minimum exists for any $N \geq 2$ separated cloudy pixels distributed in an infinite domain. $R_0$ will not be found in the singular case of a cloud field that contains one perfectly round cloud. In that case, which is easily detectable, the minimum is at $R_0=0$. This case is also presented in Section 2.2.

14. Page 19575, line 13: How do you determine the cloud mask at lower resolutions from the average of the high resolution mask? Which threshold do you use to say that a low resolution pixel is cloudy? How do you verify that cloud fraction is constant?

For the theoretical exercise of resolution reduction we did the following: to lower a resolution of a binary mask by a factor of $k$, we average every $k \times k$ pixel box into one pixel value in the new matrix. Then we set a threshold of $1/2$, and in the new matrix, all pixels that have a value above this threshold are set to 1, and all pixels that have a value lower than this threshold are set to 0. In case that the new pixel value is exactly $1/2$, it is randomly set to 0 or to 1, with a probability of 50%, to avoid biased results. We later checked that the portion of cloudy pixels over the total number of pixels remained unchanged.
It is now clarified in Section 2.2: "For these tests, the resolution of each cloud mask data was reduced by a simple averaging of pixels, and setting all pixels averaged higher than 0.5 as cloudy, and all pixels averaged lower than 0.5 to clear. Pixels whose average was exactly 0.5, were randomly set as cloudy or clear (with the probability of 50%), keeping the reduced resolution data statistically unbiased, and verifying that the cloud fraction is constant."

15. Sections 3.2 and 3.3: In the aerosol discussion all possible effects for the aerosol-cloud interaction in the twilight zone are mentioned (page 19578, lines 25–29), including also 3D radiative transfer enhancement effects. However, I have the impression that in the evaluation of the results the 3D effects are neglected, although Wen et al. (2007, 2008) have shown that they can play a very important role. In particular, these 3D effects can produce inaccuracies in the retrieved aerosol properties that are not related to any physical characteristics of the aerosol in the twilight zone. Hence, I do not think that the assertion on page 19579, lines 22–23 (“because of the significant difference in aerosol properties and its measured optical characteristics near detectable clouds”) is correct. In fact, these observed differences could be artifacts caused by the 1D nature of aerosol retrievals (i.e. by the neglect of 3D radiative transfer effects) or by undetected clouds that are misinterpreted as aerosol particles (as mentioned at the end of Sect. 3.3 but not in Sect. 3.2). Please comment on this and, in case, modify your conclusions.

The 3D radiative effect is definitely considered as one of the twilight zone’s affecting components. By saying "measured optical characteristics near detectable clouds" we claim that the measured optical properties (unlike the actual optical properties) are affected. We have clarified this statement saying: "... because of the significant difference in both actual aerosol properties and the measured aerosol optical characteristics near detectable clouds." Furthermore, Section 3.2 begins with a paragraph including undetected cloud effects: "The aerosol properties’ retrievals in the vicinity of clouds are affected by aerosol humidification processes (Fein-
gold and Morley, 2003; Twohy et al., 2009), by signal contribution of undetected clouds (Koren et al., 2007; Koren et al., 2008; Koren et al., 2009), and by cloud 3D radiative effects (Marshak et al., 2006; Wen et al., 2007)."

We have also added the 3D radiative effect to the discussion in Section 3.3 as an additional contributor to the FMF increase vs. distance from the nearest cloud: "**The results of the first regime in the graph can be explained by the theoretical superposition of three effects.** The first is the aerosol swelling process that produces sharp exponential decay in the aerosol size, as the distance to the nearest cloud grows. The second is the effect of undetectable clouds that “increase” the aerosol apparent size, and the third is the cloud 3D radiative effect (Wen et al., 2007), which may enhance the apparent aerosol optical depth near clouds."

Referee #2 - minor revisions:

1. **Page 19570, line 12:** Most algorithms do not only work with solar channels but also with thermal information. Moreover, these algorithms can work in different ways and produce different cloud masks depending on the application they are thought to serve. Please modify the text accordingly and comment whether this fact affects your algorithm.

In the text, we are not limiting the wavelength range of the cloud masking algorithms that can be used by our cloud masking algorithm. However, we added a sentence to this paragraph, in order to include the effect of the final target of each algorithm on the threshold and method selection: "**The majority of these algorithms use threshold techniques which separate cloud from cloud-free atmosphere by their reflectance values in different wavelengths or their variances; the cloud masking algorithms may also use different thresholds or techniques according to the final purpose they serve (Ackerman et al., 1998; Platnick et al., 2003; Dybbroe et al., 2005; Luo et al., 2008).**"

2. **Page 19572, line 8:** Which algorithms do you mean by “Other methods”? Please
We are now citing Koren et al., (2009) as a basis of an algorithm to determine cloud field boundaries by marking a circle around its "center of gravity".

3. **Page 19572, lines 11–14**: Please explain this concept a little more clearly.

We have rephrased this sentence, now saying: *Here, that requirement is defined as "locality", i.e. the algorithm should be sensitive to scales which are higher than the scale of the whole cloud field in order to mask fields with relatively complex shape.*

4. **Page 19572, line 20**: “Euclidean distance transform” is a term that appears here for the first time without explanation, although the meaning has been illustrated above. Please add some clarification.

This is a good point. Since this term is not used in any other place, we dropped the word "transform" from this sentence. We have also changed all "distance transform" terms in the figures to "distance map".

5. **Page 19577, line 3**: Is it $29 \pm 1 \text{ km}$ or $29 \pm 9 \text{ km}$?

$29 \pm 9 \text{ km}$ is the right range. The text is corrected now.

6. **Page 19577, lines 20–24**: How do you infer the CFF values for the various $R_0$ of 20 and 30 km?

The CFF values for constant $R_0$ are derived by calculating the area of the detectable clouds, surrounded by an $R_0$ km thick belt. The ratio between that area and total domain's area is the CFF.

7. **Page 19580, line 4**: Please give a reference for the assertion that the FMF product over land is affected by larger inaccuracies.

We are now referring to Remer et al., (2005) in the mentioned sentence.
Referee #2 - technical corrections

1. *Please check for the consistent usage of “cloud-free” or “cloud free” throughout the paper.*

   The text includes only "cloud-free" now.

2. **Page 19568, line 21:** I think that “oceans” should not be written in capital letters.
   Corrected.

3. **Page 19571, line 13:** as one approaches detectable clouds...
   Corrected.

4. **Page 19571, line 18:** should comply with the...
   Corrected.

5. **Page 19572, line 2:** one proposed a method using...
   Corrected.

6. **Page 19572, line 7:** in the cloud field area.
   Corrected.

References


Feingold, G., and Morley, B.: Aerosol hygroscopic properties as measured by li-


