Interactive comment on "A statistical approach for rain class evaluation using Meteosat Second Generation-Spinning Enhanced Visible and InfraRed Imager observations" by E. Ricciardelli et al.

Anonymous Referee #1

This paper presents a new technique (RainCEIV) to classify cloudy scenarios in terms of rain categories by exploiting the MSG-SEVIRI spectral channels. The final purpose is to provide an operational tool for continuous rainfall event monitoring (convective and stratiform), which takes advantage of the high spatial and temporal resolution of geostationary VIS and IR data in spectral and textural tests. The algorithm is composed by two modules, a cloud classification algorithm to identify clear and cloudy pixels (taking into account different cloud categories), and a second module for the delineation of the raining areas according to three rainfall intensity classes. The training processes of the two modules are presented together with the validation results for selected case studies.
General comments In my opinion the manuscript needs a deep revision to improve the description of the algorithm, which sometimes is not so sharp at the expense of the correct comprehension of the text. In particular section 3.2 and sub-sections should be improved because they represent the core of this work and I have some specific requests and/or suggestions with respect to this part. Authors should better emphasise the novelty and main strengths of their methodology with respect to similar products. Also the Conclusions section is in my opinion incomplete because it simply summarizes the results from the validation but it does not provide any perspectives about the future work. From the validation some abilities of the algorithm in discriminating raining from non-raining pixels are apparent with a tendency to the overestimation of precipitating areas, but there are problems with the precipitation class attribution, especially with C2 class. I think that the authors should include in the conclusions how you will proceed to improve the performances of your algorithm.
Moreover I suggest to the authors a general revision of English.

Author Comment (A.C.):
We would like to thank the referee for the detailed and useful comments on our paper. We accept all the suggestions as specified in our responses to the specific comments included in this document. The abstract and the introduction are modified in order to explain the utility of the RainCEIV technique better. The RainCEIV main purpose is to supply a continuous monitoring of convective and stratiform rainfall events without using any near real-time ancillary data. Its novelty is the use of the temporal differences of the brightness temperatures related to the SEVIRI water vapour channels that are indicative of the atmosphere instability and, as a consequence, give useful information for the detection of rainy areas.
The validation section is updated and we are enlarging the validation dataset, in the attempt both to analyze more night-time scenes and increase the number of the test samples belonging to the class C_2. In fact, the statistical scores obtained for the C_2 class are worse than those obtained for the C_1 class, partly due to the smaller sample size. The responses to the specific comments 11 and 12 give a more in-depth explanation of how the validation dataset is revised.
The conclusions will be extended, including a discussion on measures to improve the performances on the basis of the results obtained.

1.Specific comments 1. Page 13675 lines 5-13 The blended technique by Turk et al.(1999) was also implemented among the precipitation products of the Satellite Application Facility on Support to Operational Hydrology and Water Management (H-SAF) (Mugnai et al., NHESS, 13, 1959-1981, 2013).

A.C.:
Agreed. A sentence has been added to specify that the blended technique by Turk et al. (1999) is implemented among the precipitation products of H-SAF.

2. Page 13676 lines 23-25 Some information about MSG satellites is wrong. MSG-1 was launched in August 2002 and MSG-4 is planned for launch in 2015. I do not understand the sentence "MSG-2 was designated as the first satellite on 11 April 2007."Now the prime operational geostationary satellite is MSG-3 since January 2013, while MSG-1 data are available since January 2004.

A.C.:
Thank you very much for the correction. The sentence now reads:
"SEVIRI is the main payload on board the MSG series, composed of MSG-1 (Meteosat 8), MSG-2 (Meteosat 9), MSG-3 (Meteosat 10), and future MSG-4 (Meteosat 11), planned for launch in 2014."

3. Page 13679 lines 6-25 "The training dataset used in the previous version of MACSP has been updated in order to get a better distinction of the cloudy classes." I think that it is better at least to include a reference to Table 5 of Ricciardelli et al. (2008) to have an idea of the previous version of the training data set, and then some further details are needed about this new version of the training data set. I understand that the C_MACSP module derives from a previous work (Ricciardelli et al., 2008), but nevertheless I think that a short description of the methodology and in particular of the used spectral features are necessary.

A.C.:
Agreed. Section "**3.1- Cloud classification algorithm description**" has been modified following your suggestion and it now reads as:

"The cloud Mask Coupling of Statistical and Physical methods algorithm - MACSP (Ricciardelli et al., 2008) - is used for distinguishing *cloudy* from *non-cloudy* pixels. The version used for RainCEIV purposes is called C_MACSP, which stands for cloud Classification Mask Coupling of Statistical and Physical methods. The current version has been updated to give information about the cloud class and in particular to split the MACSP "*high cloud*" in the *high optically thin* and *high optically thick* cloud classes. Furthermore, the *convective cloud* class has been added, not just for module II but also to individuate the possible occurrence of extreme events. A pixel can be classified in 5 different classes considered both over land and sea: *clear*, *low/middle cloud*, *high optically thin cloud*, *high optically thick cloud* and *convective cloud*.

In detail, the C_MACSP physical algorithm uses the same physical threshold tests as the MACSP earlier version with the addition of a new threshold test involving the difference between the brightness temperature of the SEVIRI water vapour channel centred at 6.2µm and of the SEVIRI window channel centred at 10.8µm, $\Delta TB_{6.2\mu m-10.8\mu m}$. This difference is very small for convective cloud as asserted by Mosher (2001, 2002) in the Global Convective Diagnostic approach. The C_MACSP statistical algorithm considers in input the same spectral and textural features described and listed in section 3.2.1 and table 4, respectively, of Ricciardelli et al. (2008), but the training dataset has been updated in order to build the training samples for the *convective cloud* class. The training samples were collected in the Mediterranean basin, where RainCEIV operates. The cloud classification for the training dataset has been made through a careful visual inspection of the SEVIRI images. The clear and cloudy pixels have been selected manually after observing the spectral characteristics in SEVIRI IR/VIS images as well as in their RGB composition, a useful practice for distinguishing cloudy classes (Lensky and Rosenfeld, 2008). In order to collect the training samples for the *convective cloud* class, the cloudy SEVIRI pixels have been matched with the corresponding PEMW-rain rate (RR) and radar-derived RR values, if available. The collocation process both of the radar-derived RR values and the PEMW-RR value in the SEVIRI grid is described in Section 2. For simplicity, the pixel SEVIRI, to which the radar-derived-RR value is assigned, is denominated RADARinSEVIRI pixel, and the pixel SEVIRI, to which the PEMW-RR value is assigned, is denominated PEMWinSEVIRI pixel. Moreover, the radar samples completely included in the SEVIRI pixel are denominated RS. The SEVIRI pixel is considered for the training when:

- both the RADARinSEVIRI pixel and PEMWinSEVIRI pixel are available and the relation (RADARinSEVIRI$\geq$4mm$\times$h^{-1}).and.(PEMWinSEVIRI$\geq$4mm$\times$h^{-1}) is satisfied;
- both the RADARinSEVIRI pixel and PEMWinSEVIRI pixel are available, the relation (RADARinSEVIRI$\geq$4mm$\times$h^{-1}).and.(PEMWinSEVIRI$<$4mm$\times$h^{-1}) is satisfied and the percentage of the RS samples is higher than 80%;
- only the PEMWinSEVIRI pixel is available (the AMSU/MHS observation is outside the area covered by the Radar Network) and the relation (PEMWinSEVIRI$\geq$4mm$\times$h^{-1}) is satisfied;

When both the RADARinSEVIRI pixel and the PEMWinSEVIRI pixel are available and the relations at points 2 and 3 are not satisfied, the SEVIRI pixel is not considered for the initial training dataset. The SEVIRI images listed in table 5 of Ricciardelli et al (2008) and in particular the ones used for the training of the Mediterranean basin (enclosed in the areas B, C, and G of Figure 3 of Ricciardelli et al (2008)) have been used for the training of C_MACSP. The SEVIRI images used for the training are those acquired on 29 September 2009 at 16:57 UTC, on 01 October 2009 (at 05:12 UTC, at 08:27 UTC, and at 15:57 UTC), on 04 March 2010 (at 14:27 UTC, 15:57 UTC, and at 20:12UTC), on 28 April 2010 (at 12:27 UTC and 15:43UTC), on 04 August 2010 (at 10:43UTC and 15:12UTC). The procedure described in Appendix A has been applied in order to refine the training dataset by eliminating the redundant as well as the misclassified samples.

For RainCEIV purposes, the C_MACSP screening is useful to:

- reduce the number of the input pixels to the RainCEIV k-NNM classifier by removing the pixels classified as *clear* and *high thin cloud*;
- define the components of the feature vector in input to the RainCEIV classifier (as will be described in the following sub-section. The components chosen for each cloud class are shown in Tables 5 and 6)."

In this paragraph is presented also the validation of the C_MACSP module but without comments about the related statistical scores. These scores are shown in Table 1, which was never cited in the text.

A.C.:
In the previous version of the manuscript, Table 1 was related to Section 3.1 and listed the accuracy scores for cloud and clear classes. We considered a fixed number of test samples for each cloud class and for the clear class, making no distinction between the samples acquired during night-time and those acquired during daytime.
In the revised version, the accuracy is determined for each C_MACSP class for night-time and daytime samples, separately. Moreover, a new sub-section of Section 4 is added, dedicated to the discussion of the validation of C_MACSP. As a consequence, Section 4 "Validation results" presents now two sub-section ("4.1 C_MACSP validation results" and "4.2 RainCEIV validation results") and Table 1 is renamed Table 7.

4. Page 13680 lines 12-16 This comment concerns the rainfall intensity classes. In my opinion the non-rainy class should range from 0 to 0.1 or 0.5 mm h-1 because estimates of so light rainfall intensities (< 0.1 or 0.5 mm h-1) can be very unreliable and it could be safer to include them in the non-rainy class. Could you, please, comment on.

A.C.:
We agree. The definition of the C_0 and C_1 class (line 13 on page 13680) is modified as follows:

1. Non-rainy (rain rate $<0.5mm\times h^{-1}$) (C_0)
2. Light-to-moderate rain ($0.5\leq$rain rate$<4mm\times h^{-1}$) (C_1)

and, consequently, in the validation against radar-derived rain rate values the number of non-rainy pixels as well as that of the light-to-moderate-rain pixel is updated. We have added this information within the revised text.

5. Page 13680 line 18 “. . . determines the mean value dmin(x,Ci)” and also the eq.(1). I think that dmin should be replaced by dmean.

A.C.:
Ok, done. Thank you for spotting this typo.

6. Page 13681 line 21 “In fact, in stratiform clouds the precipitation processes are strongly related to the microphysical structure of the cloud top and, in particular, rain rate confidence is high for cloud top with large cloud droplets or in presence of ice(Lensky and Rosenfeld, 1997).” This is true not only for stratiform clouds but for all precipitating clouds. Thus considering spectral channels connected with cloud microphysical properties allows to identify raining clouds also in presence of “warm” clouds, when tests based only on IR brightness temperatures are not successful.

A.C.:
Thank you for the correction. Taking into account your correction and the suggestion of the other referees, sub-section 3.2.1, from line 12 on page 13681 to line 24 on page 13681, has been modified as follows:

“All the spectral and textural features defined for the IR/VIS SEVIRI images acquired at 0.6 µm, 0.8 µm, 1.6 µm, 3.9 µm, 6.3 µm, 7.2 µm, 10.8 µm, and 12 µm were initially considered as components of the feature vector $\vec{x}$. Among the spectral channels listed above, some are usually utilized to infer information on microphysical properties related to the cloud top. In particular, the observations acquired at 10.8µm and 12.0µm are used to provide information on cloud top temperature and cloud optical thickness, the observations at 0.6µm are used to get information on cloud optical thickness, while the 3.9µm and 1.6µm observations are used to infer information on cloud thermodynamic phase and cloud drop-size distribution. The precipitation processes are strongly related to the microphysical structure of the cloud top and, in particular, rain rate confidence is high for cloud tops with large cloud droplets or in presence of ice (Lensky and Rosenfeld, 1997). Consequently, this study considering features derived from spectral channels connected with cloud microphysical properties could allow a more accurate identification of rainy clouds.”

7. Page 13682 line 15 I do not understand when the Fisher criterion (eq. 6) is really applied in the K-NNM module to reduce the number of elements in the feature vectors, because in section 3.2.2 it seems to me that you do not use this criterion, when you describe the methodology to determine the dimension d of the feature vectors. Improve the description of this part and all sub-section 3.2.2. (especially the procedure to determine the best values of d and k).

A.C.:

We apologize for the unclear explanation of how the Fisher criterion was applied. In order to elucidate the use of Fisher criterion in determining the features to be included in the feature vector, sub-section 3.2.1 has been modified. In particular, the description of the Fisher criterion has been moved from sub-section 3.2.1 to the Appendix A (reported at the end of this document for convenience). The sentence from line 14 on page 13682 to line 7 on page 13683 has been modified as follows:

"The Fisher distance criterion (Ebert, 1987; Parikh, 1977), described in Appendix A, has been applied in order to evaluate the discriminatory power of the individual features. The Fisher distance has been determined for the following combinations: (C_0, C_1); (C_0, C_2); (C_1, C_2). The features have been ordered in a descending way on the basis of the correspondent Fisher distance value, so that the features characterized by higher Fisher distances have been chosen as components of the features vector. The definitive number of the components of the feature vector, *d* and the *k* parameter for the RainCEIV k-NNM classifier, have been determined as will be described in the next sub-section."

Moreover, in order to clarify how the training dataset has been carried out, how the process to refine the training dataset works and how the best values for d and k are chosen sub-section 3.2.2 is modified as follows:

"The training dataset has been built by coupling cloudy SEVIRI pixels with the corresponding RR value obtained by the PEMW algorithm and, where available, with the radar-derived RR values. For simplicity, the SEVIRI pixel, to which the radar-derived-RR value is assigned, is denominated RADARinSEVIRI pixel, and the SEVIRI pixel, to which the PEMW-RR value is assigned, is denominated PEMWinSEVIRI pixel. Moreover, the radar samples completely included in the SEVIRI pixel are denominated RS. When no radar-derived RR value is available (because the AMSU/MHS observation is outside the area covered by the Radar Network) the SEVIRI pixel is classified as belonging to one of the classes C0, C1, and C2 on the basis of the corresponding PEMWinSEVIRI value and it is included in the initial training dataset. When the RADARinSEVIRI value is available and agrees with PEMWinSEVIRI in determining the rainy/non-rainy class the SEVIRI pixel belongs to, this is included in the initial training dataset. Otherwise, when the RADARinSEVIRI and PEMWinSEVIRI do not agree, the SEVIRI pixel is included in the initial training dataset only if the correspondent RADARinSEVIRI pixel belongs to a rainy class C_1 or C_2 and the percentage of the rainy RS is higher than 80%. This choice is very useful for the training of the rainy events localized over areas smaller than the AMSU-B/MHS FOV area. The training samples have been considered separately for land and sea, and grouped on the basis of the solar zenith angle ranges and of the 10.8μm SEVIRI channel brightness temperature ranges. Finally, in order to refine the training dataset, the process described in Appendix A has been applied to the initial training dataset. The availability of the SEVIRI samples double matched with PEMW and radar-derived RR values is useful for both the mitigation of uncertainty due to the collocation process and the refinement of the original training dataset especially for the removal of the misclassified samples.

Successively, in order to decide the best values for *d* and *k*, a set of test samples have been classified by varying *d* and *k* combinations. Moreover, an artificial dataset, smoother and more versatile than the initial one, has been obtained by applying the bootstrap method (described by Hamamoto et al. (1997)) to the initial test samples. In order to make a more robust choice for *d* and *k*, the same *d* and *k* combinations chosen for the classification of the initial test dataset have been used to classify the artificial dataset. The best choice of *d* and *k* has been made by comparing the statistical scores obtained by classifying the two dataset separately.

Let $Y = \{(\vec{y}_i, C_j)\}$ be the independent test dataset built by examining the PEMW-RR values related to the AMSU/MSH overpasses of 12 February 2012 at 01:35UTC, 12 November 2011 at 08:50UTC, 22 November 2010 at 09:34 UTC, 4 August 2010 at 14:46 UTC, 26 April 2010 at 12:26 UTC, 01 October 2009 at 19:50UTC, 02 October 2009 at 05:00UTC. The pairs $(\vec{y}_i, C_j)$ indicate the test samples $\vec{y}_i$ belonging to the class C_j, j=1, 2, ..., N_c, N_c is the number of the classes, i=1, 2, ...,$N_{c,j}$, $N_{c,j}$ is the number of the test samples for the class C_j.

The bootstrap samples for each class have been determined as follows:

1. the sample $(\vec{y}_k, C_j)$ was selected;
2. r was chosen equal to $N_{c,j}/4$ and the r nearest neighbours (NN) of the sample $(\vec{y}_k, C_j)$ (indicated as $\{(\vec{y}_{k,s}, C_j)_{s=1,r}\}$) were found. The Nearest Neighbour decision rule is explained in Appendix A;
3. the ith component of the bootstrap sample was calculated by applying the equation

$$by_k^i = \frac{1}{r}\sum_{s=1}^{r} y_{k,s}^i \qquad (7)$$

to all the components of the $\{(\vec{y}_{k,s}, C_j)_{s=1,r}\}$ For simplicity the generic ith component of the $(\vec{y}_{k,s}, C_j)_{s=1,r}$ is indicated as $y_{k,s}^i$ without indicating the belonging class C_j, in the same way by_k^i is the ith component of the bootstrap sample $(\overrightarrow{by}_k, C_j)$ obtained by starting from the sample $(\vec{y}_k, C_j)$.

4. Points 2 and 3 were repeated for $r = {N_{c,j}}/{5}, {N_{c,j}}/{10}, {N_{c,j}}/{2} - 8, {N_{c,j}}/{2} - 6, {N_{c,j}}/{2} - 4, {N_{c,j}}/{2} - 2$;
5. the process restarted from point 1 with another sample and points 2, 3 and 4 were applied until all the test samples were considered for each class.

A careful screening has been done to eliminate the redundant *bootstrap* samples. The *bootstrap* samples and the initial test samples have been classified separately by means of the k-NNM (using the original training dataset). The statistical scores obtained for the two datasets are quite similar and they change in the same way varying *d* and *k* as can be noted in Tables 2, 3 and 4 that list the statistical scores for k=3, d=10, d=15 (Table 2); d=20; k=5, d=10, d=15, d=20 (Table 3); k=7; d=10, d=15, d=20 (Table 4). Other combinations of *d* and *k* have been investigated obtaining results worse than the ones listed in tables 2, 3 and 4. In particular, both for the original and artificial test dataset, for $k < 3, d < 10$ the FAR related to the moderate class is higher than 0.40 and POD is lower than 0.6, while for *k*>7 the FAR for all the classes is higher than 0.44 and the other statistical scores are lower than those obtained for the other *k* and *d* combinations. The statistical scores obtained by classifying the initial and artificial samples agree in suggesting k=5 and d=15 as the best choice of the parameters for the k-NNM classifier. The features chosen as components of the feature vector $\vec{x}$ related to daytime and night-time acquisition are listed in Table 5 and Table 6, respectively.**"**

In the revised manuscript Tables 3, 4 and 5 are renamed Tables 2, 3 and 4

8. Page 13685 line 1-13 "The final bootstrap training set contains the bootstrap samples obtained for r=Nj/4, Nj/5, Nj/10, Nj/2 −8, Nj/2 −6, Nj/2 −4, Nj/2 −2.". You try 7 values of the r parameter in the construction of bootstrap samples, which is the final value of r?

A.C.:
We apologize for not giving a clear explanation of the process. All the values listed for *r* parameter were used in order to obtain an artificial dataset smoother and more versatile than the initial one.

The above-reported updated version of sub-section 3.2.2 should give a more in-depth explanation of the bootstrap sample construction and of how the *r* parameter is used in the bootstrap method.

“The statistical scores obtained by classifying the bootstrap samples...” I did not understand which data were used as reference data set in the validation of the K-NNM results obtained for the bootstrap data set. Specify this point in the text.

A.C.:
We apologize for not being clear enough. To clarify this point, the test dataset in sub-section 3.2.2 is now described as follows:
“the independent test dataset built by examining the PEMW RR values related to AMSU-B/MSH overpass of 21 February 2013 at 13:10 UTC, 12 February 2012 at 01:35UTC, 12 November 2011 at 08:50UTC, 22 November 2010 at 09:34 UTC, 4 August 2010 at 12:19 UTC and 14:46 UTC, 26 April 2010 at 12:26 UTC, 01 October 2009 at 19:50UTC, 02 October 2009 at 05:00UTC, 29 September 2009 at 15:16 UTC“.
Furthermore, the AMSU-B/MSH overpasses whose samples were used to carry out the test dataset are removed from Table 2. The test dataset has been enlarged respect to the previous version, as can be noted from the above-mentioned description.

9. Page 13685 line 15 The Table 6 caption is not sufficient to explain the Table content; in particular the features are absolutely cryptic.

A.C.
Table 6 is now split into two tables: Table 5 and 6 list the features to be used during daytime and night-time, respectively. The captions of Tables 5 and 6 have been re-written so to be clearer. A description of Tables 5 and 6 is now added at the end of sub-section 3.2.2 as follows:

“The features chosen as components of the feature vector $\vec{x}$ related to daytime and night-time acquisition are listed in Tables 5 and 6, respectively. The features used over land and over sea are the same, but in some cases they vary for different cloud classes, e.g. the max and min value of the ASM are very useful in order to determine the confidence that a low/middle cloud is precipitating, but its discriminant power is not so high as to distinguish the other rainy classes. On the contrary, the minimum and maximum values of Contrast and Mean give an useful contribution in detecting both light-to-moderate rainy class and heavy rainy class for all the cloudy class.”

10. Page 13685 line 16 The title of section 4 (Validation and comparisons results) suggests that, in addition to the validation results against DPC radar rain rates, the authors present comparisons between their results and other similar products from other methodologies. But I do not see these comparisons, so I think the title should be modified by removing “comparisons”.
A.C.:
Thank you for the correction. Section 4 is now renamed “Validation results”.

11. Page 13687 lines 14-20 About the case study II you stated: “The RainCEIV is able to detect rainy samples with a POD of 85 %.” But there is still a remarkable overestimation (BIAS=1.91) of the precipitating area, and moreover the statistical scores get worse when you try the rainfall class attribution with increasing FAR and Bias values and decreasing POD and HSS. So, please, add some further comments.

A.C.:

In the revised version, we have modified the validation modality both reconsidering the collocation process for the C_2 samples and handling the “uncertain” pixels correctly as will be clarified in the following.
The collocation of the radar-derived RR values in the SEVIRI grid is now described at the end of section 2 as follows:

“For simplicity, the SEVIRI pixel to which the radar-derived-RR value is assigned is denominated RADARinSEVIRI, and the radar sample completely included in the SEVIRI pixel will be denominated RS. The collocation process of the radar-derived rain rate values into the SEVIRI grid consists in associating each SEVIRI pixel to the rain rate value obtained by averaging the RS rain rate values. If the percentage of the RS is higher than 80%, the RADARinSEVIRI pixel is considered for the validation as a rainy pixel. If the percentage of the non-rainy RS is 100%, the RADARinSEVIRI pixel is considered for the validation as a non-rainy pixel. In the other cases the SEVIRI pixel is considered “uncertain” and not included in the validation process.”

In Figures 2, 3 and 4 the “uncertain” RADARinSEVIRI pixels in the “radar-derived rain rate results” panel are the “dark gray” pixels which were not defined and were wrongly considered as non-rainy samples in the validation process. In fact, as it is possible to note in Figures 2, 3 and 4, the RainCEIV results are defined also in correspondence of the “uncertain” dark grey RADARinSEVIRI pixels. The wrong inclusion of the “uncertain” RADARinSEVIRI pixels in the validation process resulted in the high number of false alarms. In the light of this, we are reconsidering the statistical results for all the cases study analyzed. In particular for the case study II the number of false alarms for the C_1 class varies from 307 to 298 and for the C_2 class from 29 to 21. In the revised version, the correct statistical results are discussed approximately from line 14 on page 13687 as follows:

“The RainCEIV is able to detect rainy samples with a POD of 86%, that is strongly related to the correct detection of the moderate rainy samples, in fact POD is 72% for the C_1 class and 31% for the C_2 class. However, the high FAR (47% for all the rainy classes, 58% for the C_1 class and 73% for the C_2 class) attests the RainCEIV tendency to overestimate rainy samples for this case. In particular, the high FAR for the C_2 class depends on the moderate RADARinSEVIRI pixels classified as heavy by the RainCEIV. The analysis of these misclassified RADARinSEVIRI samples shows that the 67% of them contains a number of RS with $RR \geq 4 mm \times h^{-1}$ higher than the RS with $RR < 4 mm \times h^{-1}$, but the rain rate obtained by averaging all the RS rain rate values is strongly affected by the lowest values.”

“Also in this case, RainCEIV detects as rainy pixels that are no-rainy for the radar network (FAR is 0.27), but it is able to monitor the areas characterized by very heavy precipitation as well as by moderate precipitation (POD is 0.62) both on the east cost of Sicily and on Southern Calabria.” The statistical score values reported in this sentence do not agree with the values in Table 10 for the case study III (FAR=0.26 and POD=0.59 for C1,C2, FAR=0.27 and POD=0.59 for C1, and FAR=0.93 and POD=0.03 for C2). In this case the algorithm underestimate the precipitating areas, and in particular for the C2 class it seems that all precipitating pixel identified by the algorithm are actually non-precipitating (FAR=0.93), and almost all true precipitating pixels are missed (POD=0.03). Thus I think that it is not possible to state that the algorithm is able to identify regions characterized by heavy precipitation, at least for this case study.

A.C.:
Also for this case study, the correct statistical scores have been updated after the removal of the “uncertain” RADARinSEVIRI samples. Consequently, Table 10 has been modified too.

In detail, the FAR for the C_2 class is strongly affected by the number of the heavy RADARinSEVIRI pixels (63) that is lower than the number of the moderate RADARinSEVIRI pixels. 59 samples out of 63 were misclassified as moderate, thus contributing to an increase in the number of the false alarms for C_2 class only. The analysis of the 59 misclassified samples shows that approximately the 80% of them contains a number of heavy RS samples higher than that of the moderate RS samples, but their rain rate average value is strongly affected by the moderate RS sample lowest values and this is the reason why the final RADARinSEVIRI pixel is assigned to the C_1 class.
In the revised version we are reconsidering this aspect, and a SEVIRI pixel will be classified as belonging to the class C_1 or C_2 not only on the basis of the rain rate average value of the RS samples, but also on the basis of the percentage of the heavy/moderate RS samples.
We are analyzing more cases study to increase the validation dataset and in particular the number of C_2 samples.

12. Page 13688 lines 3-6 "Regarding the convective events, the RainCEIV is a useful tool for the study and characterization of the rainfall events characterized by short duration, high temporal variability, and small size area (of the order of the MSG-SEVIRI spatial resolution)." I think that it is not possible to draw this kind of conclusions on the basis of the results obtained for the case study I, statistical scores are not so good. Perhaps you could analyze other case studies of this type and consider the average behavior of the algorithm. A single case study can penalize the algorithm performances.
A.C.:
We are going to analyze further cases study in the attempt to get more convective events. The validation dataset will be enlarged by adding more daytime and night-time scenes and choosing cases study characterized by more convective events both for daytime and night-time

For the same reasons discussed in point 11, the statistical scores related to the case study I (29 September 2009 at 13:00 UTC) have been corrected after removing the "uncertain" RADARinSEVIRI pixels from the validation samples. In particular, for this case study the number of false alarms varies from 9 to 5 for the C_1 class and from 5 to 2 for the C_2 class. Consequently, the dichotomous statistical results (in particular FAR and Bias scores) have changed and the discussion about case study I is modified as follows:

"The accuracy score is high (99%) due to the high occurrence of the non-rainy pixels detected correctly. POD shows that RainCEIV detects 64% of the rainy samples correctly, while Bias and FAR scores reveal the RainCEIV tendency to overestimate rainy samples (FAR score is 0.44 and Bias score is 1.14). In detail, Bias score related to the C_1 class (Bias=1.09) is higher than that related to the C_2 class (Bias=1.33), on the contrary FAR related to the C_1 class (FAR=0.42) is lower than that related to the C_2 class (FAR=0.75). In remarking this statistical results, it is worth noting that they are strongly influenced by the low number both of the C_2 RADARinSEVIRI samples (3) and the C1 RADARinSEVIRI samples (11). Moreover, the temporal distance between the SEVIRI and RADAR acquisitions that is about 3 minutes can be determinant in the detection of the rainy events characterized by a high variability. It is argued that parts of the false alarms as well as the miss-samples are brought about by the collocation errors in the SEVIRI grid."

Technical corrections
1.Page 13674 lines 16 and 21 "Mamoudou and Gruber (2001)" The correct citation is: Ba and Gruber (2001). Please, correct also the reference in the bibliography. Ok, done.

2. Page 13676 line 4 " -20_ W and 20_ E". Replace with " 20_ W and 20_ E". Ok, done.

3. Page 13676 line 21 Pay attention to the name of algorithm modules. From the Introduction the name of the cloud classifier module is C_MACSP, not MACSP. Ok, done.

4. Page 13678 line 2 Replace DCP with DPC. Ok, done.

5. Page 13679 line 5 I think that the Table 2 cited in this sentence is not the correct one. Table 2 contains the AMSU-B overpasses used to build the training data set of the K-NNM module; I expected a table with the MSG-SEVIRI features, which actually are displayed in Table 6. Ok, done.

6. Page 13862 line 6 ". . . largest variance across the design set. . ." Is this the training data set? Replace design set with training data set. Ok, done

7. Page 13682 line 13 Replace K-NN with K-NNM. Ok, done.

8. Page 13683 line 25 AMSU-B observations used for the K-NNM training data set are displayed in Table 2, not in Table 3. Ok, it is right. Now table 2 is renamed Table 1.

9. Page 13684 line 13 The reference Efron (1979) was not included in the bibliography.
Considering that the sentence "Consequently, the *bootstrap* training set obtained is smoother than the one presented by Efron (1979)", does not add information useful for the comprehension of the bootstrap technique, we have removed this sentence from the new version of the manuscript. We apologize for the confusion.

10. Page 13684 line 21 and eq.7 I do not understand the mathematical notation used for the r nearest neighbour vectors used in the bootstrap data set construction. In my opinion yrj,y(y=1,r) should be replaced with ykj,z(z=1,...,r). bykj (line 25) should be corrected, moreover specify the range of the index i.

We apologize for the confusion. The description of the bootstrap method and the mathematical notation is now changed as described at the point 7 of this document where the updated 3.2.2 sub-section is shown**.**

13. Page 13686 line 7 "The Bias score higher for C2 ..." Replace with "The higher Bias score...".
Thank you for the correction.

14. Page 13686 lines 24-25 "The statistical scores calculated for each case are listed in Table 11 (for all classes), Table 12 (for C1 class), and Table 13 (for C2 class)." In the manuscript there is only Table 10, which summarizes the results for the three case studies, so correct the sentence accordingly. Thank you for the correction.

15. Page 13687 line 4 The Bias value (1.67) is not correct according to Table 10, which reports a Bias value of 1.64. Thank you for the correction.

16. Page 13687 line 11 Replace "...larger temporal and spatial distribution" with "...larger temporal and spatial extent". Ok, done. Thank you for the correction.

Appendix A. "Procedure adopted for the training set refinement"

The RainCEIV and C_MACSP original training dataset have been refined by applying the same procedure to the samples of each class.

The refinement process consists in using the Nearest Neighbour decision rule described by Cover and Hart (1967) in order to classify each sample of the initial training classes. The aim of this process is, in this paper, to eliminate the redundant and misclassified training samples, which is quite similar to the CNN rule described in Hart (1968) but with the difference that the main purpose of CNN is to get a training subset which performs as well as the original one. Before the description of the refinement process, a brief description of the NN decision rule and of the Fisher criterion (used to reduce the number of the components of the feature vector) will be shown.

Let $T_o=\{(\vec{x}_i, C_j)\}$ be the original training dataset, where the pairs $(\vec{\mathbf{x}}_\mathbf{i}, C_j)$ indicate the training samples $\vec{\boldsymbol{x}}_{\boldsymbol{i}}$ of the class C_j, j=1, 2, ..., N_c, N_c is the number of the classes, i=1, 2, ...,$N_{c,j}$, $N_{c,j}$ is the number of the training samples for the class C_j. Given a vector $\vec{\boldsymbol{y}}$ to be the classified, the NN rule establishes that $\vec{\boldsymbol{y}}$ belongs to the class C_j when the minimum distance is that from the training sample $\vec{\boldsymbol{x}}_{\boldsymbol{i}}$ that belongs to class C_j, and then $\vec{\boldsymbol{x}}_{\boldsymbol{i}}$ is the Nearest Neighbour of $\vec{\boldsymbol{y}}$.

Before applying the RR decision rule, it is important to define the dimension of the feature vector. In fact, since the k-NN classifier performance generally decreases with the dimension of the features vector, the number of the components (x^i) of $\vec{\mathbf{x}}$ has been reduced by applying the Fisher criterion (Ebert, 1987; Parikh, 1977) in order to evaluate the discriminatory power of the individual features and to choose the features characterized by the higher Fisher distance value. Let $\overline{x_j^i}$ and σ_j^i be the mean and standard deviation of the feature x^i for the training set from class C_j, thus the Fisher distance is defined as:

$$D_{ijk} = \frac{|\overline{x_j^i} - \overline{x_k^i}|}{(\sigma_j^i - \sigma_k^i)}. \tag{1}$$

It measures the ability of the feature x^i to differentiate class C_j from class C_k. The features x^j, within $\vec{\mathbf{x}}$, have been ordered in a decreasing way on the basis of the D_{ijk} values and the first *d* features have been chosen as the components of the feature vectors used. The dimension *d* has been fixed by following Jain and Chandrasekaran (1982)'s suggestion who point out that the ratio between the number of the training samples for each class and the feature vector dimension *d* should be at least five.

The procedure to obtain the refined training dataset, T_r, starting from the original training dataset T_o, consists in:

1. Considering the i^{th} pattern $(\vec{x}_i, C_j)$ of T_o,
2. Applying the NN decision rule and determining the following action on the basis of the three possible classification results:
   - the NN belongs to the initial belonging class C_j and the Euclidean distance is higher than zero, consequently the sample is put in T_r;
   - The NN belongs to a different class $C_i \neq C_j$, consequently the sample is reanalyzed and included in the NN class;
   - the Euclidean distance from the NN is zero, the sample is considered redundant and it is removed from T_o and not included in T_r.
3. restarting from point 2 with another sample and applying the entire process until all the training samples have been analyzed.

$T_{r,}$ determined for each class is used as the definitive training dataset.