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 #2

The authors propose a new algorithm for rainfall intensity classification with high spatial and temporal resolution based on MSG SEVIRI. The technique uses a k-nearest neighbor mean classifier that is trained with rain rate from AMSU-B data. Different spatial and spectral features extracted from MSG SEVIRI channels are considered in the classification algorithm. I think the manuscript needs some major revisions before I would recommend it for full publication.
The presentation of the different steps in section 3 should be better structured and more precise.
The authors should elaborate more on deficiencies of existing retrieval techniques and the potential benefit of the presented technique, especially of the rain intensity differentiation.
The training and validation dataset should be extended.

Author .Comment (A.C.):
We would like to thank the referee for the detailed and useful comments on our paper. We accepted all the suggestions in the revised manuscript, improving the structure of Section 3, extending the training and validation datasets, and explaining in more detail the benefits of the presented technique with respect to the existing ones.
Specific comments are addressed below.

The title ". . . rain class evaluation . . ." is misleading. I suggest changing it to ". . . rain intensity differentiation . . .".

A.C.
Agreed. The title now reads:
"A statistical approach for rain intensity differentiation using Meteosat Second Generation-Spinning Enhanced Visible and InfraRed Imager observations"

The English should be revised.
Section 1:

The authors should focus more on the deficiencies of existing satellite-based techniques.
Why is the present study necessary? What would be the advantage in contrast to other existing techniques?

A.C.:
The abstract and the introduction as well as each section of the paper is improved in order to explain the utility of the RainCEIV technique more in-depth. In particular the abstract now reads:

"This study exploits the Meteosat Second Generation (MSG)–Spinning Enhanced Visible and Infrared Imager (SEVIRI) observations to evaluate the rain class at high spatial and temporal resolutions and, to this aim, proposes the Rain Class Evaluation from Infrared and Visible observation (RainCEIV) technique. RainCEIV is composed of two modules: a cloud classification algorithm which characterizes and individuates the cloudy pixels, and a supervised classifier that delineates the rainy areas according to the three rainfall intensity classes, the *non-rainy* (rain rate value<0.5 mm×h^{-1}) class, the *light-to-moderate rain* class (0.5 mm×h^{-1}≤rain rate value<4 mm×h$^{-1)}$, and the *heavy-to-very-heavy rain* class (rain rate value≥4 mm×h^{-1}). The second module considers in input the spectral and textural features of the infrared and visible SEVIRI observations for the

cloudy pixels detected by the first module. It also uses the temporal differences of the brightness temperatures related to the SEVIRI water vapour channels indicative of the atmospheric instability strongly related to the occurrence of rainfall events.
The rainfall rates used in the training phase are obtained through the Precipitation Estimation at Microwave frequencies, PEMW (an algorithm for rain rate retrievals based on Atmospheric Microwave Sounder Unit (AMSU)-B observations). RainCEIV provides a continuous monitoring both of the cloud coverage and rainfall events without using real–time ancillary data. Its principal aim is that of supplying preliminary qualitative information on the rainy areas within the Mediterranean basin where there is no radar network coverage. The results of RainCEIV have been validated against radar-derived rainfall measurements by the Italian Operational Weather Radar Network."

The abstract will be updated by introducing the statistical scores obtained for the enlarged validation dataset, as will be explained in the discussion at the end of this document.

What would be benefit of the presented rain class differentiation for further satellite based rain retrievals?

A:C.:
RainCEIV technique is useful for the continuous monitoring of rainfall events in the Mediterranean region where there is an increased frequency of extreme events. Because of the well-known limitations of the IR/VIS observations in determining precise rain rate values, the RainCEIV main purpose is to provide a near-real time qualitative characterization of the rainy areas especially in regions not covered by the radar and rain gauge network.

Section 2:

The information on MSG is not correct. Please correct this.

A.C.:
Ok, done. 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."

It would be interesting to evaluate the performance of the proposed technique separately from uncertainties introduced by the PEMW algorithm. For comparison I suggest to train and validate the technique with independent data from the radar network.

A.C.:
The training phase has been carried out by collecting a set of SEVIRI pixels with co-located Rain Rate (RR) values inferred from AMSU-B/MHS observations processed by the PEMW algorithm, and when available with co-locate radar-derived RR values. The choice to use principally PEMW RR values instead of radar RR values for the training of RainCEIV dataset has been made because PEMW-RR values are available on a larger area than that covered by the Radar network. Nevertheless, the choice of the double matching of PEMW and radar-derived RR values, when available, in order to decide the rainy/non-rainy class of the SEVIRI pixels results very useful in the refinement of the initial training dataset. We apologize for not being clear. The paragraph that describes the training procedure 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-B/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 both for 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.”

Your suggestion is very interesting, but due to the training procedure we adopted, the comparison results obtained by training the RainCEIV with only radar-derived RR values are the same obtained by double matching PEMW and radar derived RR values during the RainCEIV training phase.

Section 3.1:

The authors should describe the extensions of the original MACSP algorithm mentioned in section 3.1 in more detail. This should include a description of the considered features as well as the approach for cloud type classification. Given the mentioned update of the MACSO algorithm the training dataset and the validation dataset should be increased.

A.C.:
We accept the suggestion; Section 3.1 has been changed as follows:

“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:

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

The validation results should be presented and discussed separately in the results section.

A.C.:
We followed this suggestion; Section 4 "Validation results" presents now two sub-sections: **4.1 C_MACSP validation results** and **4.2 RainCEIV validation results**.

4. Validation results
4.1 C_MACSP validation results

The validity of the C_MACSP algorithm has been tested by applying it to an independent dataset of which each class is made xxx samples taken from the SEVIRI images acquired on 5 May 2012 at 20:30 GMT, 19 May 2012 at 11:00 GMT, 23 July 2012 at 10:30 GMT, 5 December 2012 at 08:45 GMT, 19 September 2009 at 19:15 GMT, 6 July 2010 at 11:30 GMT and 12:30 GMT, 4 August 2010 at 14:30 GMT. The validation has been carried out separately for samples acquired during night-time and daytime by comparing the C_MACSP classification results and the samples manually collected from the independent dataset images. The manual classification has been made through a careful observation of the SEVIRI RGB composition so as to get the same number of samples for each class. The convective cloud classification results have been validated considering the rain rate maps derived both from the weather radar network and the PEMW rain rate maps. The latter have been used for the areas where radar information is missing. The accuracy (defined as the ratio between the number of the test samples classified correctly and the total number of the test samples) has been determined for each class and Table 6 shows the results obtained. On the basis of the samples examined, it is possible to assert that C_MACSP is able to classify high thick clouds as well as convective clouds, both over land and sea during daytime and night-time, with an accuracy higher than 95%. Moreover, it shows an accuracy higher than 91% in detecting low/middle clouds both during daytime and night-time over land and over sea. The accuracy in detecting high thin class over sea is 87,6% during daytime and night-time, and it is slight lower over land both during daytime (85%) and night-time (84%)."

In the revised manuscript, Table 1 is renamed Table 7 and it lists the validation results for daytime and night-time, separately. We are considering other samples to enlarge the test dataset, because of this the number of the above mentioned test samples (xxx) is not definitive and the accuracy scores discussed above will be updated.

Page 13679, line 6 to 7: Please explain in more detail how the training dataset “has been updated”.

A.C.:
Ok, done. The training dataset updating process is described in the new version of section 3.1 above reported.

Page 13679, line 5: The reference to table 2 is wrong. Please correct.

A.C.:
Ok. Table 1 (to whom we wrongly referred as Table 2) is now renamed Table 7 because the C_MACSP validation has been moved in sub-section 4.1.

Page 13679, line 12: Please specify “outliers”.

A.C.:
We define as outliers the samples that during the training phase are misclassified. (e.g. as for C_MACSP a thin cloud could be misclassified as clear, or a low/middle cloud could be misclassified as high thick cloud, as for RainCEIV heavy rain could be misclassified as moderate rainy pixel). This information is now provided in the revised version.

Page 13679, line 11 to 14: Please specify how you “refine“ the “training dataset.

A.C.:
As the procedure adopted to refine the training dataset is the same for the two modules C_MACSP and RainCEIV, this is now described in appendix A:“Procedure adopted for the training set refinement” (For convenience, Appendix A is also reported at the end of this document).
The sentence:
“In order to get a reliable training dataset, the outliers have been removed by means of the Condensed Nearest Neighbour Rule (CNN) (Hart, 1968) and the cross-validation method has been applied so to refine it.”

has been modified as follows:
“In order to refine the training dataset, by eliminating the redundant samples as well as the misclassified samples, the procedure described in appendix A has been adopted.”

Section 3.2:

Page 13681, line 5: Please provide a flowchart showing the structure and sequence of the procedure described in section 3 instead of figure 1.

A.C.:
The following flowchart, showing the training phase process, is now added to section 3:

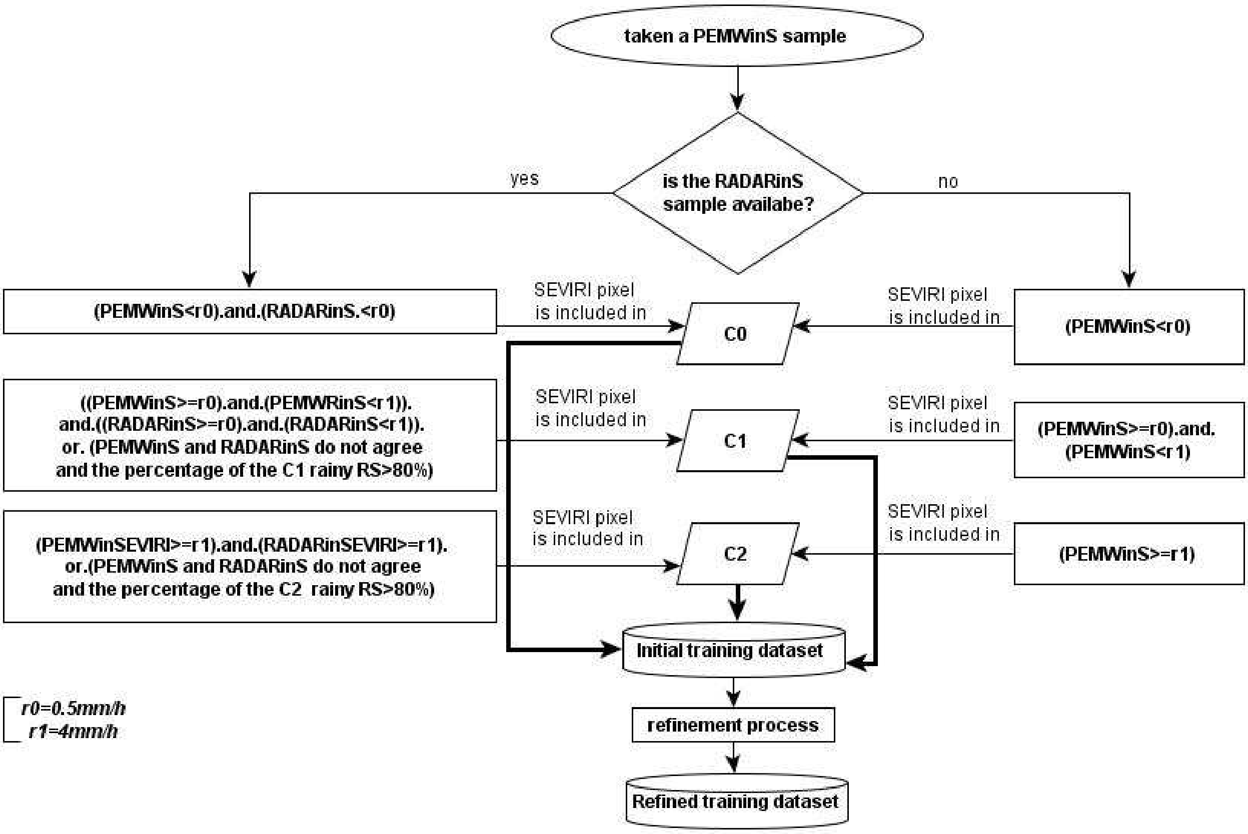

Figure 1. Flowchart of the RainCEIV training phase.

RADARinS is the SEVIRI pixel to which the radar-derived rain rate value is assigned and PEMWinS is the SEVIRI pixel to which the PEMW rain rate value is assigned, RS are the radar samples completely included in the SEVIRI pixel.

Section 3.2.1:
Please explain the considered spectral and spatial features.

A.C.:
The following text is added at the beginning of 3.2.1 sub-section:
"In detail, the spectral features used are the *maximum* and *minimum* grey levels and the ratio between them. The textural features considered are the *maximum* and the *minimum* of the Entropy, the Angular Second Moment (ASM), the Contrast (a measure of local variation of grey-level differences) and the Mean (a measure of the mean grey-level differences). The maximum and minimum values are calculated between the values determined for the four direction (0º, 45º, 90º, 135º) in the 3×3 pixels box."

Why have you chosen features for cloud detection to classify rain areas?

A.C.:
The combination of the features chosen for the classification of the rainy/non-rainy samples differs from that used in the C_MACSP statistical algorithm.
RainCEIV considers in input the maximum and minimum values among all the textural values determined for the four directions (0, 45, 90, 135). For the cloud classification purposes, the textural values are considered in the specific directions because of their usefulness in the detection of the high thin cloud. The spectral and textural features of the WV spectral channels as well as their temporal differences are considered as components of the RainCEIV feature vector, but they are not considered in the C_MACSP statistical algorithm.

An overview of the spectral and spatial features before and after the selection (Table 6) should be given. The calculated discriminatory power of the individual features should also be presented and discussed.

A.C.:
We apologize for being unclear . 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 is modified. In particular, the description of the Fisher criterion will be moved from sub-section 3.2.1 to the Appendix A (see at the end of this document). The sentence from line 14 on page 13682 to line 7 on page 13683 is now changed 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, sub-section 3.2.2 is modified 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* parameters have been chosen.

The results should be presented separately for daytime and nighttime scenes.

A.C.:

Agree. Validation results are now presented for night-time and daytime scenes separately in the revised paper.
Moreover, Table 6 is now split into two tables (Table 5 and 6) listing the features to be used during daytime and night-time, respectively.

Page 13681, line26, 27: Please explain the considered time lags of 15, 30 and 45 minutes in more detail.

A.C.:
Ok, sub-section 3.2.1, from line 24 on page 13681 to line 4 on page 13682 will be updated as follows:

"The spectral channels centred at 6.2 µm and 7.3 µm are indicative of the water vapour (WV) content in the troposphere at levels lower than 350hPa and 500hPa, respectively. The features related to WV spectral channels when considered alone do not give useful information on the presence of a rainy cloud, on the contrary when considered with the other spectral channels features, in particular with those related to the 10.8µm channel, they are useful to individuate convective events (Mosher, 2001, 2009). Moreover, the WV temporal changes are indicative of the atmospheric instability that is a useful index in the detection of precipitating area. Because of this, the temporal differences $\Delta TB_{(6.2)15-30}$, $\Delta TB_{(6.2),15-45}$, $\Delta TB_{(6.2),30-45}$, $\Delta TB_{(7.2)15-30}$, $TB_{(7.2),15-45}$, $TB_{(7.2),30-45}$, between the brightness temperature of WV channel observations acquired 15, 30 and 45 minutes before the time of interest are exploited to get information on the WV temporal changes at different levels in the atmosphere. Obviously, the temporal change of WV brightness temperature related to a pixel does not always mean that the pixel is rainy, and as for the other features it gains usefulness in discriminating rainy/non-rainy classes when used in combination with the other features opportunely chosen, as will be described in the following sub-section."

Page 13683, line 4 to 5: This sentence is not clear to me. What is meant by "training samples for each class"? I suppose the training set consists of temporally and spatially collocated MSG and AMSU-B scenes.

A.C.:
Yes, the training set consists of temporally and spatially collocated SEVIRI and AMSU-B/MHS scenes. The training samples have been chosen separately for land and sea, for night-time and daytime scenes, and they have been grouped on the basis of the 10.8µm brightness temperatures ranges and the Solar Zenith Angle (SZA) ranges.

Section 3.2.2:

The training dataset should be extended over a greater time period and include more nighttime scenes. Is the training and application done separately for land and sea areas and for daytime and nighttime scenes? If so, explain how

A.C.:
The training dataset has been built to characterize all the classes considered separately for land and sea and for daytime and night-time scenes. During daytime the C0, C1 and C2 classes were trained for different ranges of Solar Zenith Angles (SZA). For this reason we analyzed more scenes during daytime than during nighttime. This information has been added in sub-section 3.2.2 of the revised paper. Anyway, we accept your suggestion to enlarge the training dataset and the updated list of the

AMSU-B/MHS passes considered for the training phase will be shown in Table 1 of the revised version.

Please explain the bootstrap procedure in more detail using the concrete training dataset. The whole purpose is not clear to me. I think it is easier to extend the training dataset by considering more precipitation events. Could you please provide a comparison of the training dataset before and after the bootstrap procedure?

A.C.:
We apologize for the unclearness of the paragraph describing the bootstrap procedure. In the previous version, the AMSU-B/MHS passes used for defining the training and test dataset were listed in the same Table 2 and this made confusion about the function both of the training and the test dataset. The bootstrap procedure is applied only to the test dataset.

We accepted your suggestion and consider a test dataset larger than the one used in the previous version. The original test dataset and the artificial one obtained by applying the bootstrap process have been considered in order to define the best values for *k* and *d* parameters. The lines from 5 on page 13684 to 15 on page 13685 (sub-section 3.2.2) now reads as follows:

"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 $Y = \{(\vec{\mathrm{y}}_{\mathrm{i}}, C_j)\}$ 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 $(\vec{\mathrm{y}}_{\mathrm{i}}, C_j)$ indicate the test samples $\vec{\mathrm{y}}_{\mathrm{i}}$ 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 $(\vec{\mathrm{y}}_{\mathrm{k}}, C_j)$ was selected;
2. *r* was chosen equal to $N_{c,j}/4$ and the *r* nearest neighbours (NN) of the sample $(\vec{\mathrm{y}}_{\mathrm{k}}, C_j)$ (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 i^{th} 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 i^{th} 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 i^{th} 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 $r = {N_{c,j}}/{5}, {N_{c,j}}/{10}, {N_{c,j}}/{2} - 8, {N_{c,j}}/{2} - 6, {Nc,_j}/{2} - 4, {N_{c,j}}/{2} - 2$;
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 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

Page 13683, line12 to 23: These lines should be included in section 2.

A.C.:
The statistical scores shown in this paragraph have been obtained by validating PEMW-RR values against radar-derived and rain gauge-derived rain rate values. The validation was carried out by Di Tomaso et al. (2010) and Cimini et al. (2013). These statistical scores have been listed not as RainCEIV validation results but in order to give information on the PEMW accuracy, that is why this information was included in this sub-section.

Page 13683, line25: The reference to table 3 is wrong. Please correct.

A.C.:
Ok, done. Due to the fact that Table 1 is renamed Table 7, Table 2 (wrongly named Table 3) is renamed Table 1.

Page 13683, line26-27: Please explain in more detail how the MSG and AMSU-B scenes are spatially and temporally collocated for the training dataset?

A.C.:
We apologize for not making the collocation process clearer.
The collocation of PEMW-derived rain rate values in the SEVIRI grid is now described in Section "2- Instruments and data" at line 25 on page 13677, as follows:

"The PEMW rain rate value is assigned to the SEVIRI pixel only when the latter is entirely enclosed in the corresponding AMSU-B/MHS FOV. The re-sampling of the PEMW rain rate values on the SEVIRI grid was done by considering the area of each AMSU-B/FOV calculated using the orbital parameters described in (Bennartz, 2000). The temporal matching has been carried out considering a maximum delay of 7 minutes and 30 seconds between the acquisition time of the SEVIRI pixel and the AMSU/MHS FOV."

Page 13684, line 1: Please explain to what extent the k-NNM classifier is a pattern recognition classifier and how patterns are considered by the features in the training dataset.

A.C:
The k-NNM classifier in a supervised pattern recognition classifier. In this context, the term "pattern" is used to indicate the SEVIRI observation both as training sample and as sample to be classified. For each pattern (SEVIRI observation), the spectral and textural features are determined for the IR brightness temperature and/or for the VIS reflectance.

Page 13684, line 4: Please explain the application of the CNN rule in more detail.

A.C.:
As the procedure applied to refine both the C_MACSP and RainCEIV training dataset is the same, it is now described in the appendix A " Description of the procedure for the training set refinement" of the revised manuscript. For convenience, Appendix A is also reported at the end of this document.
In the light of this change, sub-section 3.2.1 from line 15 on page 13682 to line 7 on page 13683 is 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 the components of the features vector. The definitive number of the components of the feature vector, *d*, as well as the *k* parameter for the RainCEIV k-NNM classifier have been determined as will be described in the next sub-section. "

Page 13685, line 6 to 12: These lines should be included in the results section.

A.C.:
We apologize for being unclear. The statistical scores refer to the classification of the test samples (both original and artificial) and have been derived in order to determine the best combination of the *d* and *k* parameters to be used in the RainCEIV *k*-NNM classifier.

Page 13685, line 13 to 14: What reference dataset was used for the cross-validation?

A.C.:
We apologize for the confusion. The reference dataset used is now described in sub-section 3.2.2 as follows:
"Let $Y = \{(\vec{\mathrm{y}}_{\mathrm{i}}, C_j)\}$ be the independent test dataset built by examining the PEMW-RR values related to the AMSU-B/MSH overpass 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."

Page 13685, line 14 to 15: Please explain in more detail how the features in table 6 were selected. Table 6 should be revised to make it clearer. The presented feature and the expected usefulness for rain classification should be explained.

A.C.:
Sub-section 3.2.2 has been modified in order to explain more in-depth the process adopted for the selection of the features. The modified Sub-section 3.2.2 has been shown above, where the “bootstrap process” is described”.

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.”

Section 4:
Table 1 is not mentioned in the text. Please correct.

A.C.:
Thanks for spotting this typo. Table 1 is now Table 7 and it is related to the new sub-section 4.1.

Please use the same statistical scores for the validation of the cloud mask and for the validation of the rain intensity classification.

A.C.:
In the revised version, the accuracy as defined for the validation of C_MACSP (the ratio between the number of the test samples correctly detected and the total number of samples) is added to the statistical scores used for the RainCEIV results, also keeping the dichotomous statistical scores.

The validation dataset should be extended over a greater time period and include nighttime scenes.

A.C.:
The validation dataset was enlarged adding night-time scenes and choosing cases study characterized by convective events both for daytime and night-time.

The presentation of the results should include a discussion of the results in comparison to other techniques.

A.C.:
We retain that the validation of RainCEIV results against radar-derived rain rate values is sufficient for the evaluation of the RainCEIV performance. Moreover, when interpreting the statistical scores it is important to take into account that the differences in the detection of rainy areas should depend on the temporal distance and should be caused by collocation errors. The comparisons with the techniques proposed by other authors should be carried out in cooperation with the authors themselves especially regarding the choice of the cases study to be analyzed.

The interpretation of the results for the case studies is too positive. Please rephrase the respective sentences.

A.C.:
The results will be re-discussed on the basis of the statistical scores related to the enlarged dataset used for validation. 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.

Section 5:
The conclusion should be revised. At the moment it just repeats the results section.
The authors should elaborate more on further steps to improve the presented algorithm and discuss the potential benefit of the presented technique in comparison to other retrieval techniques.

A.C.:
The conclusion will be rewritten on the basis of the statistical results obtained examining more cases study.

Page 13687, line 25: "rainy/non rainy class". Please use consistent wording throughout the manuscript (e.g. "rain intensity classification").

A.C.:
Thank for the suggestion, we accept it.

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:

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

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.