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 has been updated by enlarging the validation dataset, in the attempt both to analyse more night-time scenes and increase the number of the test samples belonging to the C_2 class (both for daytime and night time). The responses to the specific comments 11 and 12 give a more in-depth explanation of how the validation dataset has been revised.

The conclusions have been extended and modified on the basis of the updated statistical scores.

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:

3.1 "**Cloud classification algorithm description**

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-RR and radar-derived RR values, if available. The collocation process both of the radar-derived RR values and the PEMW-RR values in the SEVIRI grid is described in Section 2. The SEVIRI pixel is considered for the training when:

- both the RADARinSEVIRI pixel and PEMWinSEVIRI pixel are available and the relation: (RADARinSEVIRIv $\geq$4mm$\times$h^{-1}).and.(PEMWinSEVIRIv$\geq$4mm$\times$h^{-1}) is satisfied;
- both the RADARinSEVIRI pixel and PEMWinSEVIRI pixel are available and the relation: (RADARinSEVIRIv$\geq$4mm$\times$h^{-1}).and.(PEMWinSEVIRIv<4mm$\times$h^{-1}) is satisfied and the percentage of the rainy RS samples is higher than 80%;
- only the PEMWinSEVIRI pixel is available (the AMSU-B/MHS observation is outside the area covered by the Radar Network) and the relation (PEMWinSEVIRIv$\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 1 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 20:12 UTC), on 28 April 2010 (at 12:27 UTC and 15:43 UTC), on 4 August 2010 (at 10:43 UTC and 15:12 UTC), on 2 February 2010 at 22:57 UTC, on 8 January 2010 at 13:57 UTC, on 1 October 2009 (at 05:13 UTC and 19:13 UTC). 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 has been determined for each C_MACSP class for night-time and daytime samples, separately. Moreover, a new sub-section of Section 4 has been 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:

4.1 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 300 samples taken from the SEVIRI images acquired on 12 November 2010 at 11:27 UTC, 22 November 2010 at 09:27 UTC and at 11:43 UTC, 5 May 2012 at 20:27 UTC, 19 May 2012 at 10:57 UTC, 23 July 2012 at 10:27 UTC, 5 December 2012 at 08:43 UTC, 19 September 2009 at 19:13 UTC, 6 July 2010 at 11:27 UTC and 12:27 UTC, 4 August 2010 at 14:27 UTC, 26 December 2013 at 04:57 UTC, 8 October 2013 at 18:57 UTC, 7 October 2013 at 00:57

UTC and 20 January 2014 at 11:57 UTC. 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 RR 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 7 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%)."

Table 7. Accuracy of the C_MACSP algorithm on an independent dataset

Classes	Classification accuracy (for test samples acquired during daytime)	Classification accuracy (for test dataset acquired during night-time)
Clear over land	95.0 %	95.0 %
Clear over sea	96.7 %	96.7 %
Low/middle clouds over land	91.6 %	91.0 %
Low/middle clouds over sea	92.6 %	91.3 %
High thin clouds over land	85.0 %	84.0 %
High thin clouds over sea	87.6 %	87.6 %
High thick clouds over land	98.3 %	97.3 %
High thick clouds over sea	99.0 %	99.0 %
Convective clouds over land	96.0 %	96.7 %
Convective clouds over sea	96.7 %	96.7 %

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.:
In agreement with your suggestion, the definition of the C_0 and C_1 class (line 13 on page 13680) has been 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 $\leq 4mm \times h^{-1}$) (C_1)

and, consequently, in the validation against radar-derived rain rate values the number of the non-rainy pixels as well as that of the light-to-moderate-rainy pixel has been updated. The training of the *non-rainy* and of the *light-to-moderate-rainy* class has been reconsidered on the basis of the modified RR range related to these classes.

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, 6.2 µm, 7.3 µm, 10.8 µm, and 12 µm were initially considered as components of the feature vector $\vec{x}$. Some of the above-listed spectral channels are usually utilized to infer information on cloud-top microphysical properties. In particular, the observations acquired at 10.8 µm and 12.0 µm are used to provide information on cloud top temperature and cloud optical thickness, the observations at 0.6 µm are also used to get information about cloud optical thickness, while the 3.9 µm and 1.6 µm observations are used to infer information on the cloud thermodynamic phase and cloud drop size distribution. The precipitation processes are strongly related to the cloud-top microphysical structure and, in particular, the rain rate confidence is high for cloud tops with large cloud droplets or in the presence of ice (Lensky and Rosenfeld, 1997). Consequently, in this study the use of features derived from spectral channels connected with cloud microphysical properties could allow the identification of raining 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.:

In order to elucidate the use of the 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:

"For this purpose, 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 feature vector. The definitive values of the feature vector components d and the RainCEIV k-NNM classifier k parameter have been determined as described in the following 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 has been 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.

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 PEMWinSEVIRIv and it is included in the initial training dataset. When the RADARinSEVIRIv is available and agrees with the PEMWinSEVIRI in determining the rainy/non-rainy class the SEVIRI pixel belongs to, this is included in the initial training dataset. Otherwise, when the RADARinSEVIRIv and PEMWinSEVIRIv 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 an area 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 (SZA) 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. 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
$$by_k^i = \frac{1}{r} \sum_{s=1}^r y_{k,s}^i \tag{7}$$
   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 $r = {N_{c,j}}/{5}, {N_{c,j}}/{10}, {N_{c,j}}/{2} - 8, {N_{c,j}}/{2} - 6, {Nc,j}/{2} - 4, {N_{c,j}}/{2} - 2$;
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=16, d=20 (Table 2); k=5, d=10, d=16, d=20 (Table 3); k=7; d=10, d=16, 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 40% and POD is lower than 60%, while for $k>7$ the FAR for all the classes is higher than 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=16 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 have been 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.:
All the values listed for *r* parameter were used in order to obtain an artificial test- 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 overpasses 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: Tables 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 Table 5 and Table 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 value of the ASM

is very useful in order to determine the confidence that a low/middle cloud is precipitating, but its discriminatory power is not so high as to individuate the precipitating high thick clouds. On the contrary, the minimum and maximum values of Entropy, Mean and Contrast give an useful contribution in detecting both *light-to-moderate rainy* class *and heavy-to-very-heavy-rainy* class for all the cloudy classes."

Table 5. Summary of the features considered for use in the RainCEIV k-NNM classifier during daytime. Label "A" means that the feature is used for all the C-MACSP classes; "LM" means that the feature is used for the low/middle cloud class; "HT/C" means that the feature is used for the high thick and convective cloud class.

Features	MSG-SEVIRI spectral bands (µm)							
	VIS 0.6	VIS 0.8	NIR 1.6	IR 3.9	IR 6.2	IR 7.3	IR 10.8	IR 12.0
Max Gray level							A	
Min Gray level							A	
Mean Gray level	A							
Max/Min(Gray level)								
Max(Contrast 0°, 45°, 90°, 135°)							A	
Max(Entropy 0°, 45°, 90°, 135°)			A					
Max (Mean 0°, 45°, 90°, 135°)			A			A		
Max (ASM 0°, 45°, 90°, 135°)				LM				
Min(Contrast 0°, 45°, 90°, 135°)		A						
Min(Entropy 0°, 45°, 90°, 135°)							A	
Min (Mean 0°, 45°, 90°, 135°)					A			A
Min (ASM 0°, 45°, 90°, 135°)								A
ΔTB_{15-30}					A	HT/C		
ΔTB_{15-45}					A	A		
ΔTB_{30-45}								

Table 6. Summary of the features considered for use in the RainCEIV k-NNM classifier during night-time. Label "A" means that the feature is used for all the C-MACSP classes; "LM" means that the feature is used for the low/middle cloud class; "HT/C" means that the feature is used for the high thick and convective cloud class.

Features	MSG-SEVIRI spectral bands (µm)				
	IR 3.9	IR 6.2	IR 7.3	IR 10.8	IR 12.0
Max Gray level				A	
Min Gray level	A			A	
Mean Gray level					
Max/Min(Gray level)					
Max(Contrast 0°, 45°, 90°, 135°)				A	
Max(Entropy 0°, 45°, 90°, 135°)	A				
Max (Mean 0°, 45°, 90°, 135°)			A	LM	
Max (ASM 0°, 45°, 90°, 135°)	LM				
Min(Contrast 0°, 45°, 90°, 135°)					HT/C
Min(Entropy 0°, 45°, 90°, 135°)				A	
Min (Mean 0°, 45°, 90°, 135°)		A			A
Min (ASM 0°, 45°, 90°, 135°)					A
ΔTB_{15-30}		A	HT/C		
ΔTB_{15-45}		A	A		
ΔTB_{30-45}			A		

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 the statistical scores related to the RainCEIV validation carried out against the RR radar-derived measurements have been updated for all the cases study analyzed by applying the following changes:
  - reconsidering the collocation process for the C_2 samples, that is now described at the end of section 2 as follows:
    "For simplicity, the radar samples completely included into the SEVIRI pixels will be denominated RS samples. The collocation process of the radar-derived RR measurements into the SEVIRI grid consists in associating the RS samples to each SEVIRI pixel. If the percentage of the rainy RS samples is higher than 80%, the SEVIRI pixel is considered for the validation and classified as *light-to-moderate-rainy* or *heavy-to-very-heavy-rainy* on the basis of the RS-RR value average. In some cases, the RS-RR value average is strongly influenced by the lowest RR values of the *light-to-moderate-rainy* RS samples also if the number of *heavy-to-very-heavy rainy* RS samples is higher than that of the *light-to-moderate-rainy* one. Because of this, when the percentage of the *heavy-to-very-heavy-rainy* RS samples is higher than 50% and it is higher than that of the *light-to-moderate-rainy* RS samples, the SEVIRI pixel is flagged as *heavy-to-very-heavy-rainy* regardless of the RS-RR value average. If the percentage of the non-rainy RS samples is 100%, the SEVIRI pixel is considered for the training and validation. In the other cases, the SEVIRI pixel is flagged as "uncertain" and not considered for the training and validation purposes."
  - Handling the "uncertain" RADARinSEVIRI pixels correctly. In fact, in the previous version the "uncertain" (that are the "dark-gray" pixels in the "radar-derived RR results" panels of Figures 2, 3 and 4) were not defined and were wrongly considered as non-rainy samples in the validation process. The wrong inclusion of the "uncertain" RADARinSEVIRI pixels in the validation process resulted in the high number of false alarms.
  - Updating and enlarging the training dataset on the basis of the suggestion of the referee#2.
  - Enlarging the validation dataset and determining the statistical scores for daytime and night-time samples separately.
  - Changing the RR values ranges of the *non-rainy* and *light-to-moderate-rainy* classes on the basis of your suggestion (at point 4 of this document) both for the training and the validation dataset.
  - recalculating the RainCEIV results on the basis of the updated training dataset.

In the light of the above-listed updates/changes, Table 1 (that lists the AMSU-B/MHS passes considered for the training dataset) has been updated and sub-section 4.2 now presents the following updates:
  - Table 8 (that lists the cases study used for the validation) has been updated;

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.:

The validation dataset has 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 by:

- removing the "uncertain" RADARinSEVIRI pixels from the validation samples;
- considering as *non-rainy* the RADARinSEVIRI and PEMWinSEVIRI pixels with $RR<0.5mm\times h^{-1}$ both for the training and the validation dataset;
- considering as *light-to-moderate-rainy* the RADARinSEVIRI and PEMWinSEVIRI pixels with $RR\geq0.5mm\times h^{-1}$ both for the training and the validation dataset;
- recalculating the RainCEIV results on the basis of the updated training dataset .

In particular, the number of the false alarms varies from 9 to 5 for the C_1 class and from 6 to 2 for the C_2 class, the number of the misses samples passes from 4 to 2 for the C_1 class. Consequently, the dichotomous statistical results have changed and the discussion about the case study I is modified as follows:

"The case I was chosen because it highlights the RainCEIV ability in detecting very small rainy areas. On 29th September 2009 approximately at 13:00 UTC a very rapid and heavy rainfall event affected a small area between the Basilicata and Calabria regions in Southern Italy. The accuracy score is high (99%) due to the high occurrence of the non-rainy pixels detected correctly. POD shows that RainCEIV detects 67% of the rainy samples correctly, while Bias and FAR scores reveal the RainCEIV tendency to overestimate rainy samples (the FAR score is 47% and the Bias score is 1.25). In detail, the Bias score related to the C_1 class (Bias=1.37) is higher than that related to the C_2 class (Bias=1.00), on the contrary FAR related to the C_1 class (FAR=46%) is lower than that related to the C_2 class (FAR=50%). This means that there is an overestimation of the heavy rainy area but ($C_1inC_2+C_0inC_2$) and the number of the C_2 misses is balanced with the number of the C_2 hits. This is not true for the C_1 class that shows a higher number of hits than that of the C_2 class, and this results in a higher POD (75% and 50% for the C_1 and C_2 class respectively). In remarking this statistical results, it is worth noting that they are significantly influenced by the low number both of the C_2 RADARinSEVIRI samples (4) and C1 RADARinSEVIRI samples (8). Moreover, the temporal distance between the SEVIRI and RADAR acquisitions that is about 5 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 misses 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 subsection 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 datasets 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. Here the aim of this process is to eliminate the redundant and misclassified training samples, which is similar to the CNN rule described in Hart (1968) but the main purpose of CNN is to get a training subset performing 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 given.

Let $T_o=\{(\vec{x}_i, C_j)\}$ be the original training dataset, where the pairs $(\vec{x}_i, C_j)$ indicate the training samples $\vec{x}_i$ 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 $\vec{y}$ to be the classified, the NN rule establishes that $\vec{y}$ belongs to the class C_j when the minimum distance is that from the training sample $\vec{x}_i$ that belongs to class C_j, and then $\vec{x}_i$ is the Nearest Neighbourof $\vec{y}$.

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 feature vector, the number of the components (x^i) of $\vec{x}$ has been reduced by applying the Fisher criterion (Ebert, 1987; Parikh, 1977) 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 $\vec{x}$, 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 the suggestions in Jain and Chandrasekaran (1982), 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 ith 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.

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 your 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.
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 has been 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$\times$h^{-1}) class, the *light-to-moderate rain* class (0.5 mm$\times$h$^{-1}\leq$rain rate value<4 mm$\times$h$^{-1)}$, and the *heavy-to-very-heavy rain* class (rain rate value≥4 mm$\times$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 dichotomous assessment related to daytime (night-time) validation shows that RainCEIV is able to detect rainy/non rainy areas with an accuracy of about 97% (96%), and when all the rainy classes are considered, it shows a Heidke skill score of 67% (62%), a Bias score of 1.36 (1.58), and a Probability of Detection of rainy areas of 81% (81%).”

What would be benefit of the presented rain class differentiation for further satellite based rain retrievals?
A:C.:
RainCEIV is based on a training dataset built by double-matching radar-derived rain rate values and the rain rate values obtained from the Passive MicroWave (PMW) observations from AMSU-B/MHS radiometers at a better spatial resolutions than the other PMW sensors. The PMW observations have been processed by the operative PEMW algorithm (Di Tomaso et al., 2009), whose performance has been validated by Cimini et al. (2013). Moreover, both the training phase and the RainCEIV classification algorithm are based on the C_MASCP cloud classification mask so to get more reliable results.

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 the 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 has been 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. 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 C_0, C_1, and C_2 on the basis of the corresponding PEMWinSEVIRIv and it is included in the initial training dataset. When the RADARinSEVIRIv is available and agrees with the PEMWinSEVIRI in determining the rainy/non-rainy class the SEVIRI pixel belongs to, this is included in the initial training dataset. Otherwise, when the RADARinSEVIRIv and PEMWinSEVIRIv 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 an area 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 (SZA) 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-RR and radar-derived RR values, if available. The collocation process both of the radar-derived RR values and the PEMW-RR values in the SEVIRI grid is described in Section 2. The SEVIRI pixel is considered for the training when:

- both the RADARinSEVIRI pixel and PEMWinSEVIRI pixel are available and the relation: (RADARinSEVIRIv$\geq$4mm$\times$h^{-1}).and.(PEMWinSEVIRIv$\geq$4mm$\times$h^{-1}) is satisfied;
- both the RADARinSEVIRI pixel and PEMWinSEVIRI pixel are available and the relation: (RADARinSEVIRIv$\geq$4mm$\times$h^{-1}).and.(PEMWinSEVIRIv<4mm$\times$h^{-1}) is satisfied and the percentage of the rainy RS samples is higher than 80%;
- only the PEMWinSEVIRI pixel is available (the AMSU-B/MHS observation is outside the area covered by the Radar Network) and the relation (PEMWinSEVIRIv$\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 1 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 20:12 UTC), on 28 April 2010 (at 12:27 UTC and 15:43 UTC), on 4 August 2010 (at 10:43 UTC and 15:12 UTC), on 2 February 2010 at 22:57 UTC, on 8 January 2010 at 13:57 UTC, on 1 October 2009 (at 05:13 UTC and 19:13 UTC). 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 300 samples taken from the SEVIRI images acquired on 12 November 2010 at 11:27 UTC, 22 November 2010 at 09:27 UTC and at 11:43 UTC, 5 May 2012 at 20:27 UTC, 19 May 2012 at 10:57 UTC, 23 July 2012 at 10:27 UTC, 5 December 2012 at 08:43 UTC, 19 September 2009 at 19:13 UTC, 6 July 2010 at 11:27 UTC and 12:27 UTC, 4 August 2010 at 14:27 UTC, 26 December 2013 at 04:57 UTC, 8 October 2013 at 18:57 UTC, 7 October 2013 at 00:57 UTC and 20 January 2014 at 23:57 UTC. 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 RR 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 7 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.

Table 7. Accuracy of the C_MACSP algorithm on an independent dataset

Classes	Classification accuracy (for test samples acquired during daytime)	Classification accuracy (for test dataset acquired during nighttime)
Clear over land	95.0 %	95.0 %
Clear over sea	96.7 %	96.7 %
Low/middle clouds over land	91.6 %	91.0 %
Low/middle clouds over sea	92.6 %	91.3 %
High thin clouds over land	85.0 %	84.0 %
High thin clouds over sea	87.6 %	87.6 %
High thick clouds over land	98.3 %	97.3 %
High thick clouds over sea	99.0 %	99.0 %
Convective clouds over land	96.0 %	96.7 %
Convective clouds over sea	96.7 %	96.7 %

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:

taken a PEMWinSEVIRI sample

is the RADARinSEVIRI sample available?

yes   no

(PEMWinSEVIRlv<r0).and.(RADARinSEVIRlv.<r0)

SEVIRI pixel is included in   C0   SEVIRI pixel is included in   (PEMWinSEVIRlv<r0)

((PEMWinSEVIRlv>=r0).and.(PEMWRinSEVIRlv<r1)). and.((RADARinSEVIRlv>=r0).and.(RADARinSEVIRlv<r1)). or. (PEMWinSEVIRlv and RADARinSEVIRlv do not agree and the percentage of the C1 rainy RS>80%)

SEVIRI pixel is included in   C1   SEVIRI pixel is included in   (PEMWinSEVIRlv>=r0).and. (PEMWinSEVIRlv<r1)

(PEMWinSEVIRlv>=r1).and.(RADARinSEVIRlv>=r1). or.(PEMWinSEVIRlv and RADARinSEVIRlv do not agree and the percentage of the C2 rainy RS>80%)

SEVIRI pixel is included in   C2   SEVIRI pixel is included in   (PEMWinSEVIRlv>=r1)

$r0=0.5mm/h$
$r1=4mm/h$

Initial training dataset

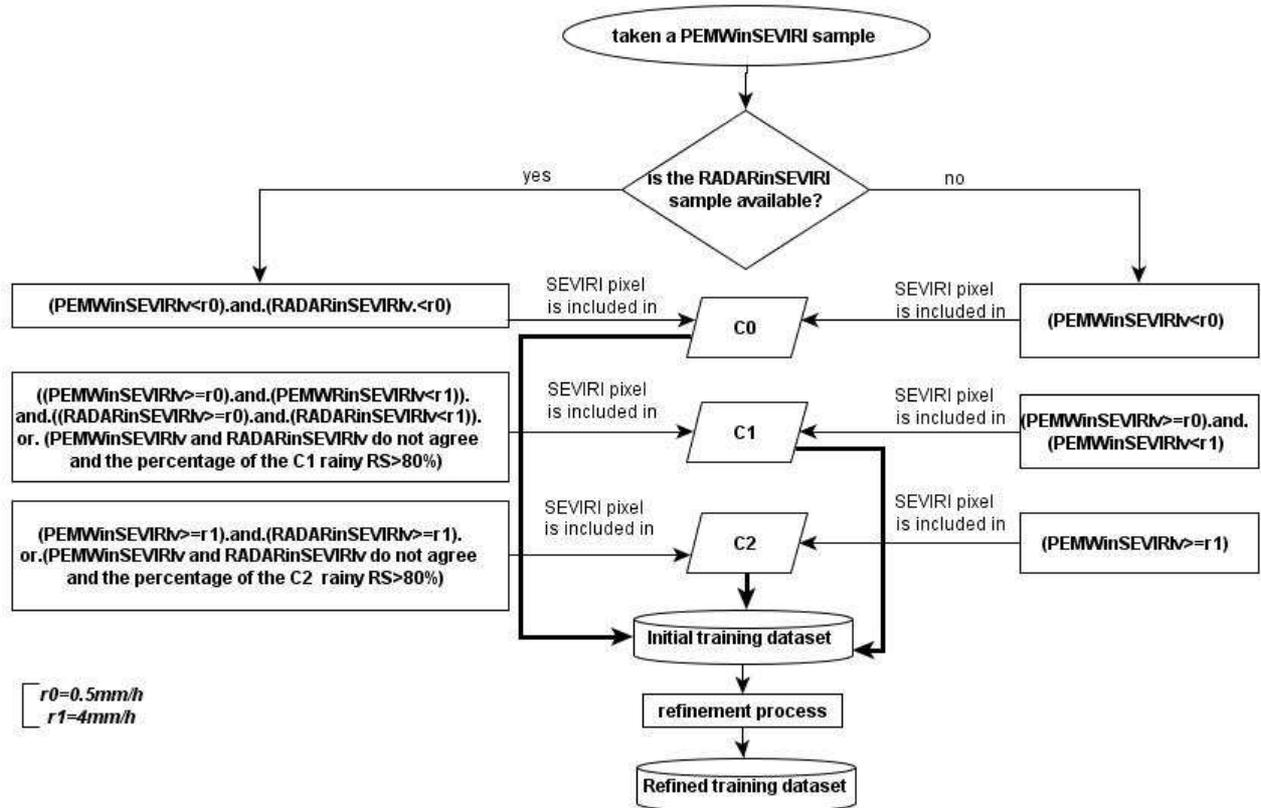refinement process

Refined training dataset

Figure 2. Flowchart of the RainCEIV training phase.

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 (a measure of the spatial randomness of the image), the Angular Second Moment (ASM, a measure of homogeneity of the image), the Contrast (a measure of local variation of the grey-level differences) and the Mean (a measure of the mean grey-level differences). The maximum and minimum values are calculated among the values calculated for the four directions (0º, 45º, 90º, 135º) in the 3×3-pixel 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.:

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 (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:

"For this purpose, 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 feature vector. The definitive values of the feature vector components d and the RainCEIV k-NNM classifier k parameter have been determined as described in the following sub-section."

Moreover, sub-section 3.2.2 has been modified to clarify how the training dataset is 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. The RainCEIV 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 has been 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 WV channel features when considered alone do not give useful information on the presence of a raining cloud, on the contrary, when considered with the other channel features, in particular those related to the 10.8 µm 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 the precipitating area. Because of this, the temporal differences $\Delta TB_{(6.2)15-30}$, $\Delta TB_{(6.2),15-45}$, $\Delta TB_{(6.2),30-45}$, $\Delta TB_{(7.3)15-30}$, $TB_{(7.3),15-45}$, $TB_{(7.3),30-45}$ between the WV brightness temperatures related to the SEVIRI acquisitions made 15, 30 and 45 minutes before the time of interest are exploited to get information on the WV temporal changes at different atmosphere levels. 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 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 C_0, C_1 and C_2 classes were trained for different ranges of Solar Zenith Angles (SZA). For this reason we analyzed more scenes during daytime than during night-time. This information has been added in sub-section 3.2.2 of the revised paper. Anyway, we accepted your suggestion to enlarge the training dataset and the updated list of the AMSU-B/MHS passes considered for the training phase is 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 of the training and the test dataset. The bootstrap procedure is applied only to the test dataset.

We accept 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 $Y = \{(\vec{y}_i, C_j)\}$ 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:

    6. the sample $(\vec{y}_k, C_j)$ was selected;

7. 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)

8. 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 $(\vec{y}_k, C_j)$.

9. Points 2 and 3 were repeated for $r = \frac{N_{c,j}}{5}, \frac{N_{c,j}}{10}, \frac{N_{c,j}}{2} - 8, \frac{N_{c,j}}{2} - 6, \frac{N_{c,j}}{2} - 4, \frac{N_{c,j}}{2} - 2;$

10. 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=16, d=20 (Table 2); k=5, d=10, d=16, d=20 (Table 3); k=7; d=10, d=16, 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 40% and POD is lower than 60%, while for $k>7$ the FAR for all the classes is higher than 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=16 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 RR 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 former Table 1 is now renamed Table 7, former Table 2 (wrongly named Table 3) is now renamed Table1.

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.:

The collocation of PEMW-derived RR values in the SEVIRI grid is now described in Section "2-Instruments and data" at line 25 on page 13677, as follows:

"The PEMW RR value is assigned to the SEVIRI pixel only when the latter is entirely enclosed in the corresponding AMSU-B/MHS FOV. PEMW rain rate values are re-sampled on the SEVIRI grid calculating the area of each AMSU-B/MHS FOV on the basis of the orbital parameters described in (Bennartz, 2000). The temporal matching is carried out considering a maximum difference of 7.5 minutes between the acquisition time of the SEVIRI pixel and that of 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:

"For this purpose, 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 feature vector. The definitive values of the feature vector components *d* and the RainCEIV k-NNM classifier *k* parameter have been determined as described in the following sub-section."

Page 13685, line 6 to 12: These lines should be included in the results section.
A.C.:
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.:
The reference dataset used is now described in sub-section 3.2.2 as follows:
"Let $Y = \{(\vec{y}_i, C_j)\}$ 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 Table 5 and Table 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 value of the ASM is very useful in order to determine the confidence that a low/middle cloud is precipitating, but its discriminatory power is not so high as to individuate the precipitating high thick clouds. On the contrary, the minimum and maximum values of Entropy, Mean and Contrast give an useful contribution in detecting both *light-to-moderate rainy* class *and heavy-to-very-heavy-rainy* class for all the cloudy classes."

Table 5. Summary of the features considered for use in the RainCEIV k-NNM classifier during daytime. Label "A" means that the feature is used for all the C-MACSP classes; "LM" means that the feature is used for the low/middle cloud class; "HT/C" means that the feature is used for the high thick and convective cloud class.

Features	MSG-SEVIRI spectral bands (µm)							
	VIS 0.6	VIS 0.8	NIR 1.6	IR 3.9	IR 6.2	IR 7.3	IR 10.8	IR 12.0
Max Gray level							A	
Min Gray level							A	
Mean Gray level	A							
Max/Min(Gray level)								
Max(Contrast 0º, 45º, 90º, 135º)							A	
Max(Entropy 0º, 45º, 90º, 135º)			A					
Max (Mean 0º, 45º, 90º, 135º)			A			A		
Max (ASM 0º, 45º, 90º, 135º)				LM				
Min(Contrast 0º, 45º, 90º, 135º)		A						
Min(Entropy 0º, 45º, 90º, 135º)							A	
Min (Mean 0º, 45º, 90º, 135º)					A			A
Min (ASM 0º, 45º, 90º, 135º)								A
ΔTB_{15-30}					A	HT/C		
ΔTB_{15-45}					A	A		
ΔTB_{30-45}								

Table 6. Summary of the features considered for use in the RainCEIV k-NNM classifier during night-time. Label "A" means that the feature is used for all the C-MACSP classes; "LM" means that the feature is used for the low/middle cloud class; **"HT/C"** means that the feature is used for the high thick and convective cloud class.

Features	MSG-SEVIRI spectral bands (µm)				
	IR 3.9	IR 6.2	IR 7.3	IR 10.8	IR 12.0
Max Gray level				A	
Min Gray level	A			A	
Mean Gray level					
Max/Min(Gray level)					
Max(Contrast 0°, 45°, 90°, 135°)				A	
Max(Entropy 0°, 45°, 90°, 135°)	A				
Max (Mean 0°, 45°, 90°, 135°)			A	LM	
Max (ASM 0°, 45°, 90°, 135°)	LM				
Min(Contrast 0°, 45°, 90°, 135°)					HT/C
Min(Entropy 0°, 45°, 90°, 135°)				A	
Min (Mean 0°, 45°, 90°, 135°)		A			A
Min (ASM 0°, 45°, 90°, 135°)					A
ΔTB_{15-30}		A	HT/C		
ΔTB_{15-45}		A	A		
ΔTB_{30-45}			A		

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 in 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.:
At first we thought of adding the accuracy as defined for the C_MACSP validation to the statistical scores used for the RainCEIV statistical assessment, but it does not provide any further information on the statistical assessment when compared with the dichotomous statistical scores already used.

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 a higher number of 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.:
In the revised version the statistical scores related to the RainCEIV validation carried out against the RR radar-derived measurements have been updated for all the cases study analyzed by applying the following changes:
- reconsidering the collocation process for the C_2 samples, that is now described at the end of section 2 as follows:
  "For simplicity, the radar samples completely included into the SEVIRI pixels will be denominated RS samples. The collocation process of the radar-derived RR measurements into the SEVIRI grid consists in associating the RS samples to each SEVIRI pixel. If the percentage of the rainy RS samples is higher than 80%, the SEVIRI pixel is considered for the validation and classified as *light-to-moderate-rainy* or *heavy-to-very-heavy-rainy* on the basis of the RS-RR value average. In some cases, the RS-RR value average is strongly influenced by the lowest RR values of the *light-to-moderate-rainy* RS samples also if the number of *heavy-to-very-heavy rainy* RS samples is higher than that of the *light-to-moderate-rainy* one. Because of this, when the percentage of the *heavy-to-very-heavy-rainy* RS samples is higher than 50% and it is higher than that of the *light-to-moderate-rainy* RS samples, the SEVIRI pixel is flagged as *heavy-to-very-heavy-rainy* regardless of the RS-RR value average. If the percentage of the non-rainy RS samples is 100%, the SEVIRI pixel is considered for the training and validation. In the other cases, the SEVIRI pixel is flagged as "uncertain" and not considered for the training and validation purposes."

- Handling the "uncertain" RADARinSEVIRI pixels correctly. In fact, in the previous version the "uncertain" (that are the "dark-gray" pixels in the "radar-derived RR results" panels of Figures 2, 3 and 4) were not defined and were wrongly considered as non-rainy samples in the validation process. The wrong inclusion of the "uncertain" RADARinSEVIRI pixels in the validation process resulted in the high number of false alarms.
- Updating and enlarging the training dataset on the basis of the suggestion of the referee#2.
- Enlarging the validation dataset and determining the statistical scores for daytime and night-time samples separately.
- Changing the RR values ranges of the *non-rainy* and *light-to-moderate-rainy* classes on the basis of your suggestion (at point 4 of this document) both for the training and the validation dataset.
- recalculating the RainCEIV results on the basis of the updated training dataset.

In the light of the above-listed updates/changes, Table 1 (that lists the AMSU-B/MHS passes considered for the training dataset) has been updated and sub-section 4.2 now presents the following updates:
- Table 8 (that lists the cases study used for the validation) has been updated;
- Table 9 (that sums up the contingency values for the RainCEIV dichotomous statistical assessment) has been updated and related only to the daytime validation dataset, while Table 10 has been added to sum up the contingency values related to the night-time validation dataset;
- Table 11 (Table 9 in the previous version that shows the statistical scores for daytime validation) has been updated and now shows the statistical scores for night-time and daytime validation separately;
- Table 12 (Table 10 in the previous version), that shows the statistical scores related to I, II and III cases study, has been updated:
- Figures 3, 4 and 5 (Figure 2, 3 and 4 in the previous version) have been updated.

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 has been be rewritten on the basis of the statistical results obtained examining more cases study.

"Conclusions

This paper proposes the RainCEIV technique as a useful tool for the continuous monitoring and characterization of the rainy areas in the Mediterranean region where there is an increased frequency of the extreme events. RainCEIV does not use any near real-time ancillary data and it exploits the temporal differences of the brightness temperatures related to the SEVIRI water vapour channels. These are indicative of the atmosphere instability and, as a consequence, could give useful information for the detection of the rainy areas when analysed with the spectral and textural features related to the other SEVIRI channels. Because of the well-known limitations of the IR/VIS observations in determining RR 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.

RainCEIV consists of two modules that use geostationary observations from SEVIRI in order to detect cloudy pixels and, successively, to associate them to a rainy/non-rainy class. RainCEIV uses both IR and VIS observations to determine if the SEVIRI pixel belongs to the *non-rainy* (C_0), *light-*

29

to-moderate-rainy (C_1) or *heavy-to-very-heavy-rainy* (C_2) class. The IR/VIS observations do not have the same potentiality as MW observations in characterizing rainy areas, but their high spatial and temporal resolution are used to get a continuous monitoring of the stratiform and convective events. RainCEIV has been trained on the AMSU-B/MHS PEMW RR values double matched with the radar-derived RR values and validated on the basis of the radar-derived RR observations. The dichotomous statistical scores indicate that a good fraction (97% for daytime validation and 96% for night-time validation) of the pixels examined are correctly identified as rainy or non-rainy by the RainCEIV. The Bias scores (1.36 for daytime validation and 1.58 for night-time validation) and the FAR scores (39% and 48%) suggest that RainCEIV tends to overestimate rainy pixels especially during the night-time, while the POD scores (81% both for daytime and night-time validation) indicate that RainCEIV detects rainy areas with a good a approximation. The rainy areas overestimation is mainly due to the misclassification of C_0 samples as C_1 samples. Moreover, the high FAR values related to the C_1 and C_2 classes are mainly due to the misclassification of the C_1 samples as C_2 samples and vice versa. The statistical scores obtained for the daytime validation are generally better than those obtained for the night-time validation. This is prevalently due to the fact that the features related to the VIS observations (unavailable during night-time) have a strong influence on the RainCEIV output because of their higher discriminatory power when compared with that of the features related to the 3.9 µm and 12.0 µm observations. In remarking upon the comparison results, it is important to bear in mind the different spatial resolutions as well as the temporal distance between radar and satellite observations that could affect the statistical scores negatively, especially for rapid convective events, even if the time distance between radar and SEVIRI acquisitions is little. As far as future developments are concerned, RainCEIV will be updated to consider in the training phase the RADARinSEVIRI samples characterized by a percentage of rainy RS samples lower than 80% so as to individuate extreme rainy events located over an area whose size is smaller than that of the SEVIRI pixel area. To this aim, information from the Visible Infrared Imaging Radiometer Suite (VIIRS) on-board the Suomi National Polar-orbiting Partnership (NPP) (characterized by higher spatial and spectral resolutions than SEVIRI) will be taken into account when available."

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 datasets 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. Here the aim of this process is to eliminate the redundant and misclassified training samples, which is similar to the CNN rule described in Hart (1968) but the main purpose of CNN is to get a training subset performing 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 given.

Let $T_o = \{(\vec{x}_i, C_j)\}$ be the original training dataset, where the pairs $(\vec{x}_i, C_j)$ indicate the training samples $\vec{x}_i$ 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 $\vec{y}$ to be the classified, the NN rule

establishes that $\vec{y}$ belongs to the class C_j when the minimum distance is that from the training sample $\vec{x_i}$ that belongs to class C_j, and then $\vec{x_i}$ is the Nearest Neighbourof $\vec{y}$.

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 feature vector, the number of the components (x^i) of $\vec{x}$ has been reduced by applying the Fisher criterion (Ebert, 1987; Parikh, 1977) 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 $\vec{x}$, 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 the suggestions in Jain and Chandrasekaran (1982), 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:

4. Considering the ith pattern $(\vec{x_i}, C_j)$ of T_o,
5. 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.
6. 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.

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 #3

The paper "A statistical approach for rain class evaluation using Meteosat Second Generation-Spinning Enhanced Visible and InfraRed Imager observations" by Ricciardelli et al., proposes a statistical technique to infer precipitation classes from SEVIRI radiances and radiance spatial and temporal features. The calibration of the technique is carried out by using AMSU derived estimates and it is validated against radar rain fields. The subject of the paper is of some interest for this journal, but is poorly written, with a number of serious weaknesses that I do not believe could be addressed through a standard major revision. I suggest to reject the paper for a number of reasons: I listed below the most relevant ones (page numbers refer to the discussion paper, from 1 to 36).

The aim of the paper seems to provide a tool to benefit short term hydrology and long term climate studies (lines 1-3 on page 3): the author should explain the usefulness of a technique that gives as output only two precipitation levels.

Author Comment (A.C.):
The abstract, the introduction as well as all other sections of the paper have been improved in order to explain the utility of RainCEIV 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 takes the temporal differences of the brightness temperatures related to the SEVIRI water vapour channels as indicative of the atmospheric instability strongly linked 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 dichotomous assessment related to daytime (night-time) validation shows that RainCEIV is able to detect rainy/non rainy areas with an accuracy of about 97% (96%), and when all the rainy classes are considered, it shows a Heidke skill score of 67% (62%), a Bias score of 1.36 (1.58), and a Probability of Detection of rainy areas of 81% (81%)."

In the introduction, there is no need to mention early works on satellite precipitation estimation in the '80s and '90s, since they used very different approaches and instruments.

On the other side, many works on SEVIRI data use for precipitation are missing (the mentioned Kidd and Levizzani reports on them).

A.C.:

The introduction has been updated following your suggestion and including other references missed in the previous version of the paper.

The correct reference for Mamoudou and Gruber is Ba and Gruber (page 4 and reference list).

A.C.: Thanks for the correction.

Section 2. The history and launch schedule of Meteosat spacecrafts are not necessary for the aim of this paper. Please, add a reference for the Italian radar network, and report on the quality of the data used. Since the radar data are used here to validate satellite product, it is mandatory a more detailed description of the radar network and its reliability.

A.C.:

Agreed. We have removed the history and launch schedule of Meteosat spacecrafts from the manuscript. References and more details concerning the Italian radar network are now provided in Section 2. The following text has been added to provide information on data quality:

"Procedures for mitigating ground clutter, anomalous propagation, beam blockage effects are applied (Vulpiani et al., 2008a). The sri product is derived applying a reflectivity-rainfall (Z-R) relationship to the Lowest Beam Map (LBM), i.e. the reflectivity values at the lowest level of the corrected radar volumes. The sri product used here represents the best estimate from the radar network available for the period under analysis, and it has been already used to validate satellite rainfall estimates (Cimini et al., 2013), including EUMETSAT H-SAF products (Puca et al., 2013). Procedures to improve the quality of the sri product, including attenuation compensation, polarimetric rainfall inversion techniques, and adaptive algorithms to retrieve mean vertical profiles of reflectivity have been recently developed at DPC (Vulpiani et al., 2012; Rinollo et al., 2013)."

Section 3. Section 3.1 roughly describes the cloud classification algorithm. Is table 1 related to this section? How is accuracy defined for cloud classes? Are clear sky pixels included in the accuracy calculation?

A.C.:

Yes, Table 1 is related to this section and lists the accuracy scores (defined as the ratio between the number of the test samples classified correctly and the total number of the test samples examined) for cloud and clear classes. In order to explain more in-depth how C_MACSP works, section 3.1 **"3.1- Cloud classification algorithm description"** has been modified in the revised version.

What are the outliers mentioned in line 13 on page 9? Are they damaged pixels, noise, or what?

A.C.:

We defined 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.

Only two images out of the nine used to validate the classification are during nighttime: are there enough pixels to verify correct classification of all cloud classes?

In the revised version, the accuracy shown in Table 7 (that was Table 1 in the previous version) has been determined for each C_MACSP class for night-time and daytime samples, separately.

I think that the validation dataset should be much larger
A.C.:
Agreed. Following your suggestion, the validation dataset has been enlarged to include more night-time scenes. In addition, we have followed the advice of referee#2 who proposes to show the C_MACSP validation results in a sub-section of Section 4. As a consequence, section 4 "Validation results" is now divided into two sub-sections: "4.1 C_MACSP validation results" and "4.2 RainCEIV validation results".

In section 3.2.1 there are a number of sentences that have to be canceled (my suggestion) or discussed with much more detail. I report here few examples, but the entire section should be rewritten or canceled. How can SEVIRI observation "individuate precipitation processes" (lines 16-17 on page 11) ? especially in convective clouds? Which processes can be individuated (coalescence, riming, breakup, melting)?
A.C.:
We apologize for the incorrect use of the English language, the term "precipitation processes" was erroneously used to mean "precipitation events". The purpose of RainCEIV is to determine a precipitation class not the precipitation process.

The radiance measured in the SEVIRI channels comes from the very top layers of the cloud. Few lines below it is said that "features related to radiances acquired at 3.9 and 1.6μm bear on the cloud drop size distribution": as a matter of fact, "cloud drop size distribution", unfortunately, cannot be derived by any feature related to SEVIRI channels.
A.C.:
The paragraph purpose was to describe the characteristics and the usefulness of the 3.9μm,1.6μm 12.0μm, 10.8μm, 0.6μm SEVIRI spectral channels to derive some cloud microphysical properties in order to make it clear that the choice of these spectral channels was made because of their connection with cloud microphysical properties so as to allow the identification of rainy clouds.
Consequently, sub-section 3.2.1 from line 11 on page 13681 to line 24 on page 13681 is rewritten 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, 6.2 µm, 7.3 µm, 10.8 µm, and 12 µm were initially considered as components of $\vec{\mathbf{x}}$. Some of the above-listed spectral channels are usually utilized to infer information on cloud-top microphysical properties. In particular, the observations acquired at 10.8 µm and 12.0 µm are used to provide information on cloud top temperature and cloud optical thickness, the observations at 0.6 µm are also used to get information about cloud optical thickness, while the 3.9 µm and 1.6 µm observations are used to infer information on the cloud thermodynamic phase and cloud drop size distribution. The precipitation processes are strongly related to the cloud-top microphysical structure and, in particular, the rain rate confidence is high for cloud tops with large cloud droplets or in the presence of ice (Lensky and Rosenfeld, 1997). Consequently, in this study the use of features derived from spectral channels connected with cloud microphysical properties could allow the identification of raining clouds."

The temperature of WV channels are related with tropospheric moisture content over clear sky areas, but in case of mid- and high- level clouds the contribution to the radiance measured by

satellite sensor has a dominant contribution from the cloud top. How can the temperature differences mentioned on lines 3-4 on page 12 "characterize convective as well as stratiform precipitation" ?

A.C.:

As before we used the verb "characterize" inappropriately. In fact, the temporal differences have been used as input for the classifier in order to associate a pixel to the class C_0, C_1, or C_2. The WV temporal differences are useful to distinguish different rainy/non-rainy classes only when used with the other components of the feature vector.

In order to clarify how the WV spectral channels have been considered for the RainCEIV purposes, sub-section 3.2.1 from line 24 on page 3681 to line 4 on page 13682 is modified 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 WV channel features when considered alone do not give useful information on the presence of a raining cloud, on the contrary, when considered with the other channel features, in particular those related to the 10.8 µm 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 the precipitating area. Because of this, the temporal differences $\Delta TB_{(6.2)15-30}$, $\Delta TB_{(6.2),15-45}$, $\Delta TB_{(6.2),30-45}$, $\Delta TB_{(7.3)15-30}$, $TB_{(7.3),15-45}$, $TB_{(7.3),30-45}$ between the WV brightness temperatures related to the SEVIRI acquisitions made 15, 30 and 45 minutes before the time of interest are exploited to get information on the WV temporal changes at different atmosphere levels. 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."

Section 3.2.2. Probably Table 3 means Table 2 (line 25 on page 13).

Correct. In the revised version, Table 1 has been renamed Table 7, thus Table 2 (wrongly named Table 3) has been renamed Table 1.

On line 26- 28 (page 13) is described the matching between SEVIRI and AMSU rain product. It seems that the rain value estimated over an area ranging between 200 km2 (at nadir) and 1000 km2 (on the edge of the swath) is assigned to a SEVIRI pixel of around 25 km2 in the considered area. This implies a number of assumptions on the rainfall spatial and temporal structure that are not usually verified in real rain.

A.C.:

The collocation of PEMW-derived RR values in the SEVIRI grid is now described in Section "2- Instruments and data" approximately at line 25 on page 13677, as follows:

"The PEMW RR value is assigned to the SEVIRI pixel only when the latter is entirely enclosed in the corresponding AMSU-B/MHS FOV. PEMW rain rate values are re-sampled on the SEVIRI grid calculating the area of each AMSU-B/MHS FOV on the basis of the orbital parameters described in (Bennartz, 2000). The temporal matching is carried out considering a maximum difference of 7.5 minutes between the acquisition time of the SEVIRI pixel and that of the AMSU/MHS FOV."

Table 6 has to be better introduced and discussed in the text, and the caption should be rewritten accordingly.

A.C.:

Table 6 is now split into two tables: Tables 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 Table 5 and Table 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 value of the ASM is very useful in order to determine the confidence that a low/middle cloud is precipitating, but its discriminatory power is not so high as to individuate the precipitating high thick clouds. On the contrary, the minimum and maximum values of Entropy, Mean and Contrast give an useful contribution in detecting both *light-to-moderate-rainy* class *and heavy-to-very-heavy-rainy* class for all the cloudy classes."

Section 4. A good validation practice requires that the datasets used for calibration and validation are independent. In the work reported in this paper, it seems this condition is not satisfied for all the considered cases. Comparing table 2 and table 6, for 4 out of 11 cases (29/09/09, 23/06/10, 04/08/10 and 10/10/10) the satellite overpasses used for validation are very close to the slot used for the calibration, and this should be avoided. I suggest to remove the mentioned cases from the validation, and to add more slots of the other cases.

A.C.:

Agreed. Although some training and validation samples have been acquired on the same day, the Solar Zenith Angle (SZA) ranges of the training and validation samples are different. Consequently, the cases study of 29 September 2009 at 13:00UTC and 23 June 2010 at 15:00UTC were not classified by using the training samples acquired on the same day.
In detail, we agree on removing the case related to 04 October 2010 at 19:30 UTC because it is very close to the training samples related to the same day, but we would rather leave the other cases for validation purposes:

- 29 September 2009 at 13:00UTC: the training samples related to 29 September 2009 at 17:00UTC have not been used as training samples to classify the SEVIRI observations acquired on 29 September 2009 at 13:00UTC because their SZA ranges do not correspond (for the samples acquired at 13:00 UTC SZA<58º, while for the ones acquired at 17:00UTC SZA>80º);
- 23 June 2010 at 15:00UTC: the SZA ranges for the training and validation samples related to 23 June 2010 are different, in fact the samples acquired at 15:00UTC for validation are characterized by a SZA>48º, while those acquired at 12:52UTC have a SZA<35º.

Moreover, the AMSU-B/MHS passes on 29 September 2009 at 15:16UTC, 4 August 2010 at 12:26 UTC and 14:46 UTC, 21 February at 13:10 UTC have been removed from Table 2 because they were used only to carry out the test dataset as described in sub-section 3.2.2 of the revised version. In fact, the AMSU-B/MHS passes used to build both training and test dataset were wrongly listed in Table 2 without distinction. This point was not explained in depth in the previous version.
Table 2 is now renamed Table 1 and has been modified on the basis of the above considerations.

In table 7 the last column title is "satellite overpass time", but the number reported in the column are probably the nominal time of delivery of the SEVIRI slot. Since the SEVIRI starts scanning the earth from the South, the Mediterranean region is scanned few minutes before the end of the scan,

at 12, 27, 42 and 57 minutes every hour. In this table should be reported the real scan time of the Mediterranean region.

The accuracy indicator is of a very limited meaning in evaluating the technique performances, since it includes the number of correct negatives, which is always very high, and can be arbitrarily increased by enlarging the considered area. See as an example table 8 and figures 2, 3 and 4.