Dear Referee #2,

thanks for your helpful comments. We address them in the following. Your comments are printed in italics.

Best Regards,
Daniel Merk

Major Comment (1)
For the problem of high false alarm rates (FARs), it is my suggestion (albeit this is likely a lot of work) to determine a better way of verifying a "successful hit" of CI. The way it is now, as I understand things, is that only one pixel is allowed to link to one radar echo of CI, leaving the other satellite pixels in the near vicinity to just get labeled as "false alarms." In a similar exercise I was part of two years ago, using NWP model data to help verify satellite-based CI nowcasts, the same thing occurred. Hence, finding an appropriate way to validate this Cb-TRAM methodology would be preferred, which would significantly lower your FARs in a physically more meaningful way. One possibility would be to group adjoining pixels along the edges of (or within) convective clouds (re: Figs. 1 and 2) so that a future radar echo would mean that all pixels are successful hits. This implies that an object tracking framework be employed as part of the validation. This discussion related to the definition on page 1789, lines 18-19.

It seems that the discussion on page 1789, lines 18-19 is a bit misleading. Although detection is made on a pixel-scale, tracking and verification is actually done with cell objects consisting of neighboring pixels as suggested. Therefore the absolute scores are already improved over a per-pixel verification. In our method we do not verify against radar data. See also page 1782, lines 24-26.

We tried to clarify our approach. We changed line 18-22 to the following:

```
Our verification method allows only an allocation of one CI object, that
consists of neighboring pixels, to one thunderstorm object. From the whole area
that shows signs of early development, usually only one object in the near
surrounding is selected by very localized characteristics. The preferred CI
object will soon dominate all other CI objects in the surrounding, as low level
convergence and upper level divergence suppresses other updrafts.
```

We also changed p.1799, l.14-18 to the following:

```
Some CI cell objects within an area of rapidly growing cumulus clouds may be
classified as false alarms in a conservative verification setup. Here we track
each objects development history individually. It could also be as legitimate to
classify a whole area of multiple CI objects as hits if only one single strong
thunderstorm develops out of the area.
```

We further change p.1777, l.12 to the following:

```
In their study Mecikalski and Bedka (2006) investigate the precursor signals for
CI and therefore the applied criteria can be directly compared to Cb-TRAM's
first stage detection.
```

It is true that our object verification method is a simple but at the same time very demanding approach. Directly connected pixels are clustered into single objects. We require a specific cell object to show further development. Other objects nearby would be ignored. Nonetheless, this way the verification is easy to understand and applying to intuition. To consider a larger area around the

original cell object would require the quantification of an optimal area size.
Certainly a comparison of absolute numbers to other publications has to consider the characteristics of this method.

Major Comment (2)

The flow of the document would be substantially improved if commas were used. This is a difficult thing to correct, in that literally 50+ commas may need to be added as a means of making the sentences read easier, and in less confusing ways at times. A few examples are given in the minor comments below.

We went trough the paper again. Commas have been inserted or longer sentences have been splitted. See minor comments, too.

p.1772, l.4:

```
It uses the channels of the Spinning Enhanced Visible and Infrared Imager
(SEVIRI) onboard Meteosat Second Generation (MSG).
```

p.1772, l.10:

```
The different criteria include timetrends of the 10.8 IR channel, and IR channel
differences, as well as their timetrends.
```

p.1772, l.12:

```
To provide the trend fields an optical flow based method is used: the Pyramidal
Matching algorithm, which is part of Cb-TRAM. The new detection scheme is
implemented in Cb-TRAM, and is verified for seven days which comprise different
weather situations in Central Europe.
```

p.1773, l.6:

```
NWP models have to deal with nonlinear dynamic processes that act on short time
scales and limited spatial resolution. This makes it necessary to parameterize
convective processes.
```

p.1774, l.4:

```
12 different channels are available for the Meteosat Spinning Enhanced Visible
and InfraRed Imager (SEVIRI). Image refresh rates are 15min for normal scan mode
(NS), and 5min for rapid scan mode (RS).
```

p.1774, l.6:

```
While some concentrate on tracking of mature thunderstorms, such as RDT and
MASCOTTE, others use radar data only, e.g. CONRAD or RadTRAM.
```

p.1774, l.15:

```
They found rather high false alarm ratios. These can be explained by the
physical characteristics of convection.
```

p.1774, l.21:

```
Through the work described in this manuscript, Cb-TRAM is provided with a day
and night-time detection of early convection stages. An estimate of the CI
detection skill is obtained with a verification setup utilizing the detection of
later stages within Cb-TRAM for NS and RS Meteosat data.
```

p.1774, l.24:

```
The development of the new detection and verification schemes, as well as a
detailed description, are presented in Sect. 3. The verification including a
comparison of the existing Cb-TRAM CI detection and the new algorithm, as well
as the comparison of the NS and RS data are presented in Sect. 4. Afterwards a
```

summary of the method and results, and a discussion of the remaining sources of uncertainty is given in Sect. 5.

p.1775,.l4:
The Cumulonimbus Tracking and Monitoring (Cb-TRAM) algorithm is introduced in Zinner et al. (2008). Changes to this detection schemes are presented in Zinner et al. (2013).

p.1775, l.9:
Cb-TRAM is an algorithm for the detection, tracking, and nowcasting of intense convective cells, using the data from Meteosat SEVIRI.

p.1775, l.13:
(2) the detection of convective cells at different stages of their life cycle, and (3) the tracking and nowcasting up to 60 min using the motion vector field.

p.1775, l.15:
It is used in the EU projects RiskAware (2004–2006), FLYSAFE (2006–2009, Tafferner et al., 2008), and ongoing DLR project "Wetter und Fliegen" (Forster and Tafferner, 2009, 2012).

p.1775, l.18:
From these a disparity vector field $\boldsymbol{V}$ is derived by warping one image on the other so that either the differences of the image intensities are minimised, or the local correlation is maximised.

p.1776, l.18:
An additional criterion used therein, that improves the limitation of the detection to active convective cores, is the HRV channels texture. Texture is quantified by a normalised local standard deviation field (WV6.2 instead of HRV channel at nighttime).

p.1776, l.24:
Existing cell objects at time $t-1$ are extrapolated using the disparity vector. Afterwards, the overlap with cells of time t is analysed.

p.1778, l.20:
For the different channel values, channel differences, and derived time trends, fixed tresholds are set.

p.1780, l.22:
To achieve this aim, an analysis of strengths and weaknesses is necessary at first.

p.1781, l.7:
The motion vector field in Cb-TRAM is derived on a pixel basis. It is independent of trackable features.

p.1781, l.9:
The field does not only include the pure advection, but also local development. This has to be considered correctly.

p.1781, l.19:
In addition the importance of individual interest fields for CI detection in SATCAST was already investigated by Mecikalski et al. (2008). This provides a starting point for the further implementation of selected interest fields into a new method.

p.1782, l.26:

If the Cb-TRAM's CI detections are perceived as forecasts of further development into thunderstorms, and stages 2 and 3 for the related Cb-TRAM object as proof of convective development, one can define the following categorical variables regarding the development of each individual cell object

p.1784, l.11:

Values of the BIAS above 1 constitute over-forecasting, and below 1 under-forecasting.

p.1784, l.13:

A Cb-TRAM cell object can represent several consecutive cell life-cycles of a multi-cell thunderstorm. This is because the tracking algorithm will allocate a new development, close enough to the expected track, to an already existing cell in a decaying stage.

p.1785, l.9:

Cloud top temperatures above 278 K refer to very low clouds, such as cumulus or stratus.

p.1785, l.18:

The convective process does not produce a smooth cloud top structure which is more likely for, e.g., cirrus and stratus.

p.1785, l.21:

While the IR10.8 channel gives the possibility to detect all clouds in the lower troposphere (if no overlaying clouds in the upper layer exists), the WV7.3 is used to guarantee that clouds have reached a significant altitude at a lower mid-troposphere level (approx. 3000 m).

p.1787, l.4:

We decided to omit these criteria for our purposes. This is due to the slightly different emphasis on early development and, consequently, the fact that such development is already covered by the stage 2 detection within Cb-TRAM, and the result of the combinations.

p.1789, l.3:

The results for daytime conditions in NS mode are investigated as total values over all seven test days, as sub-totals for the three synoptic conditions, and for each test day individually. This is done to get an overall impression of the new detection algorithm compared to the existing one, and to explore the behaviour under different synoptic weather conditions.

p.1791, l.9:

For the new algorithm, the FAR is lowest at high pressure cases, followed closely by cold front cases.

p.1791, l.11:

For the new algorithm, the best CSI values are found for high pressure, followed by cold front cases.

p.1791, l.13:

The highest reduction of BIAS compared to the original algorithm is found for cold air cases (by nearly one third), where the influence of the 12 June 2009 is most striking.

p.1792, l.3:

The main difference to the original detection scheme seems to be the improved detection of orographically induced CI. This results in an increased number of hits, but also a reduction of false alarms.

p.1793, l.11:
Considering the longer time-frame for night (14 h), and that activity on the night onset may be still higher than later at night, the tendency of decreasing activity at nighttime can still be seen by the smaller number of hits and misses (representing the total active cells).

p.1794, l.9:
Roberts and Rutledge (2003) found that after 15 min the first precipitation is observed under sustained cooling conditions. We think that in this way setting IF 2 to 15 min, we get a physical meaningful combination of the two criteria for detecting strong convective cells in an early stage.

p.1798, l.4:
FAR for NS mode under daytime conditions decreases from 91 to 81 %, and CSI increases from 7.4 to 12.7 %, while BIAS is reduced from 320 to 146 %, using the new detection algorithm instead of the original Cb-TRAM method.

p.1798, l.12:
The CI detection newly implemented in Cb-TRAM shows a reduction of false alarms for all test days, both for NS (15 min data) and RS (5 min) when compared to the original detection.

p.1798, l.23:
Generally, the high values of FAR for most of the days lead to the question on the limitations of the detection of Convective Initiation using geostationary data only, and their use as an early warning for stronger convective storm development.

p.1799, l.5:
Unfavourable conditions for further development, such as the interruption of a sufficient supply of warm moist air, a stable layer at some height above the convective cloud top, or the advection of the whole cell into an area with unfavourable conditions for convection, could stop the development at any time of the cell life-cycle.

p.1799, l.9:
Together with the choice of the verification method, this all sets narrow boundaries for the quality values possible.

p.1800, l.24:
Using strict thresholds does also not account for developments just below these given limits. Thus some hits can be missed.

p.1800, l.27:
Considering the difficulties arising when detecting and verifying CI, this leads to the question on how these could be further improved.

Major Comment (3)
The text becomes less technical and more conversational between pages 1794 and 1796. Suggest re-writing these pages with an emphasis on technical content only.

We have rewritten the text on the pages as suggested. The focus should now be on technical content only.
The new text reads as follows:
p1794 l15ff:

The total number of cell objects increases due to the larger number of timesteps investigated. More short-living convective storms can be observed. Under daytime conditions the sum for all six RS test days shows an increase of hits by 465, a decrease of false alarms by 22908 (relative decrease of more than 60%) and an increase of misses comparing original and new algorithm. The latter is a technical side effect which results from an increased number of stage 2 and 3 cell objects. Although the detection of stage 2 and 3 itself is not modified, changes of stage 1 detection influence the stage 2 and 3 objects. Each object includes all neighboring pixels with all earlier detection stages. Missing stage 1 pixels formerly connecting stage 2 and 3 pixels lead to a break up these of stage 2 and 3 objects into more individual objects. This results in a higher amount of misses. This issue is more pronounced in RS mode, but also is observed in NS mode. This effect leads to the slight decrease of POD on 4 of 6 days although the number of hits increases on 5 days. FAR decreases from 96 to 86%, CSI increases from 4 to 11.5%, BIAS decreases from 999 to 295%. The number of hits increases for 4 out of the 6 days, while it shows a decrease on 2 days. False alarm numbers are improved for all 6 days. On 12 June 2009 only about 10% of the original false alarms remain.
Best values of POD (around 50% for RS) were found for previous and new algorithms under high pressure conditions. With almost a doubling of hits the biggest increase in POD is yield for cold air cases here. Applying the new detection method, a decrease of POD for the two other classes is observed. This is due to the discussed technical issue resulting in more misses. Best FAR values for the previous algorithm are found for cold front cases. Applying the new algorithm, best FAR values are found for the cold air cases. The improvement of FAR for the cold air cases is the best of all three classes, comparing the old and new algorithm. The same is true for CSI values. This supports the normal scan results, where the FAR improved mainly for the upper cold air mass conditions as well. In Fig. 4 it is obvious that the overall behaviour of FAR, POD, CSI and BIAS does not show any major differences between day and night.In the following, we adress the question whether the use of RS data improves the detection skill over the use of NS data. As we apply some of the criteria from Mecikalski and Bedka (2006) to 5min data with adjusted thresholds (table 3), there is the chance to observe rapidly developing cells. The higher time resolution further allows to evaluate the development over three timesteps within the last 15 minutes. Short-living cells that may be missed by using only 15 min timesteps are detected this way. Even with the overall increase of detectable cells in RS mode, we achieved a lower BIAS compared to the previous algorithm in NS mode. A forecaster using the new algorithm will get a clearer picture of the situation updated every 5 min. For the new algorithm compared to the old one, BIAS is a factor of 2 higher in RS mode. This is due to the higher number of hits and the higher chance to provide false alarms.
We observe higher POD values for all three synoptic classes in RS compared to NS mode. FAR and CSI are only better for the cold air class. With the drastically decreasing number of false alarms and the the related reduction of tracking processes for these cells, we also obtain an improvement in processing time. This is an important point for a nowcasting tool providing real-time warnings.

Major Comment (4)

In numerous places acronyms which were previously defined are not used. Convective Initiation (CI) should be defined once and then used as "CI" everywhere, and certainly, there is no need to capitalize "Convective Initiation" as it is not a proper noun.

We went through the document again and substituted full terms with the previously defined acronyms.
These acronyms are:
CI for Convective initiation
RS for rapid scan
NS for normal scan

NWP for numerical weather prediction
IF for interest field
IR for infra-red

Major Comment (5)

Related to the synoptic descriptions of the days of interest. This discussion needs to include references, and be cleaned up. In particular, I have no idea what is meant by "backward upper cold air masses". Please include a figure(s) or a reference, or just describe in more synoptically appropriate terminology. All three types of convective days should be described similarly, with more detail and references.

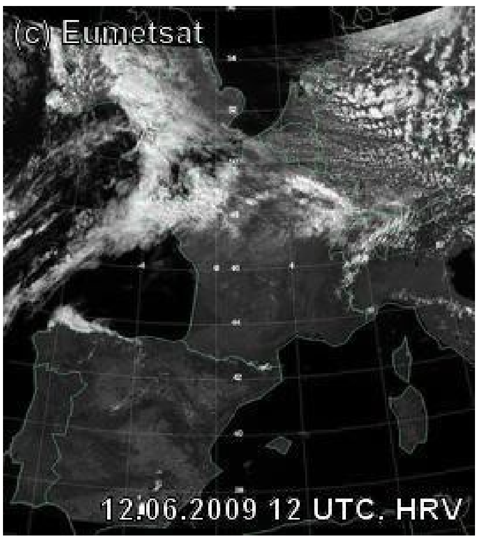
The discussion have been rewritten and the synoptic conditions have been described in more detail using typical terminology. The new classification description on p.1788, l.5ff reads as follows:

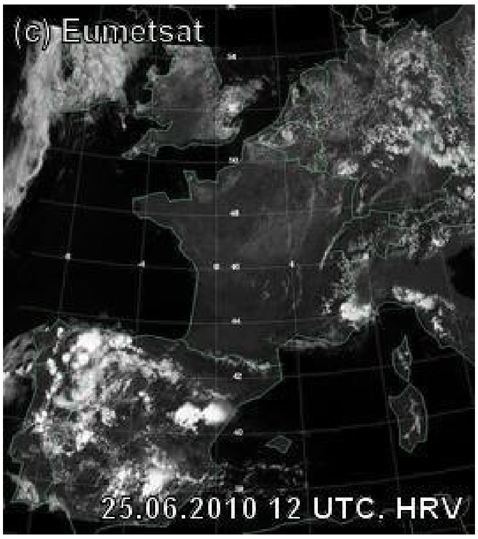
- Class "cold front": Convection connected with a cold front passage. The typical pattern is an upper-level trough over the west-coast of the Atlantic, passing over the area of interest during the observation period. Convection is triggered directly at the frontal region and in the prefrontal moist and destabilized airmass. Prefrontal high CAPE values can be observed and also deep layer shear is present. (day 1, 2, 3)
- Class "cold air": Advection of cold air masses together with an active upper-level trough and presence of surface heating leads to instability and triggering of many convective cells. Both thunderstorm and non-electric rain-showers occur in a typical honey-combed structure. (day 4,5)
- Class "high pressure": Convection connected with weak-forcing conditions. A typical pattern is an upper-level ridge stretching over central Europe,connected to a low-level high pressure area. Generally suppressing cloud formation trough to subsidence. Convection is mainly triggered orographically or by sufficient surface heating to overcome CIN.

p.1788, l.14 is changed to:

For example, the 14 July 2010 shows convection along the cold front and convection triggered by advection of cold-air masses behind the cold front.

We further added three exemplary HRV satellite images for the synoptic classes.
Fig.: High Resolution Visible (HRV) Meteosat Images for 26.05.2009, 12.06.2009, 25.06.2010, 12 UTC respectively.





Minor Comment (a)
Abstract, line 4: Initiation is spelled wrong, and no need to capitalize.
Corrected.

Minor Comment (b)
Abstract, line 9: change to "...criteria aims to identify..."
Changed.

Minor Comment (c)
page 1773, line 5: Define "NWP" on line 2 and use "NWP" only on line 5, and everywhere else in the document.
Changed. NWP has been used throughout the whole paper.

Minor Comment (d)
page 1773, line 18: Change to "As a result, nowcasting, i.e., ..." Add comma to make sentence flow better.
Changed.

Minor Comment (e)
page 1773, line 29: change to "data are better"
Corrected.

Minor Comment (f)
page 1774, line 2: Change to "An advantage..."
Corrected.

Minor Comment (g)
page 1774, lines 8-10: Define all acronyms, e.g., RDT, MASCOTTE, etc

Acronyms have been defined:
Rapid Developing Thunderstorms (RDT)
MAximum Spatial COrellation Tracking TEchnique (MASCOTTE)
KONvektionsentwicklung in RADarprodukten, convection evolution in radar products (KONRAD)
Radar Tracking and Monitoring (Rad-TRAM)
SATellite Convection AnalySis and Tracking (SATCAST)
Cumulonimbus Tracking and Monitoring (Cb-TRAM)

Minor Comment (h)
page 1774, line 18: change to "In the following study, a combination..." Add comma.
Comma added.

Minor Comment (i)
page 1774, line 20: add common after "manuscript"
Comma added.

Minor Comment (j)
page 1774, lines 27-29: There are perhaps five places where commas can be added in these two sentences.
Commas added.

Minor Comment (k)
page 1775, line 14-15: Define all acronyms.

Acronyms have been defined:
European Union (EU)
Project “RiskAware” - no acronym, project name. Added quotes.
Project “FLYSAFE” - no acronym, project name. Added quotes.
Deutsches Luft- und Raumfahrtzentrum (DLR, German Aerospace Center)
Wetter und Fliegen (Weather and Flying)

Minor Comment (l)
page 1775, lines 20-22: Correct sentence -- presently it does not read well.

New version of sentence:
```
Typically, cloud motion on small scales is dominated by the large scale flow
pattern (spatial autocorrelation). This is considered by the optical flow method
used in Cb-TRAM.
```

Minor Comment (m)
page 1775, line 26: change to "...allows for calculation of..."
Changed.

Minor Comment (n)
page 1777, line 8: Define "AMV" in this location (not on line 26), and add comma in location after "vectors". See also page 1777, line 12.
Changed.

Minor Comment (o)
page 1777, line 15: suggest referencing "Berendes et al. (2008)" after the word "mask" [Berendes, T. A., J. R. Mecikalski, W. M. Mackenzie, K. M. Bedka, and U. S. Nair, 2008: Convective cloud detection in satellite imagery using standard deviation limited adaptive clustering. J. Geophys. Res., 113, 20207, doi:10.1029/2008JD010287.]
Reference added.

Minor Comment (p)
page 1777, lines 27-28: Do not reference "UW-CIMSS", rather use Velden et al. (1997, 1998) for AMV algorithm. These are AMS publications and are ease to location.
Changed reference as suggested.

Minor Comment (q)
page 1778, lines 7 and 11: Do not reference "UW-CIMSS" AMV approach, just use "AMV".
Changed.

Minor Comment (r)
page 1778, line 12: Use the reference "Bedka et al. (2005)" after defining "MAMV". This is an AMS references, and should be easy to location.
Reference added.

Minor Comment (s)
page 1778-1780: Suggest removing text from lines 22 (page 1778) to line 18 (page 1780). This is straight from several references, including Mecikalski and Bedka (2006), Mecikalski et al. (2008, 2010), and Roberts and Rutledge (2003). This would shorten the paper. Maybe only highlight the main points from this section in one shorter paragraph.

The paragraph hasbeen shortened according to your suggestions and states only the main points. Now the paragraph reads as follows:

```
With IF (1) and IF (4) the special importance of the freezing level is
considered. IF (2) refers to the vertical cloud growth which results in cloud
top cooling. IF (3) tests for persistence of cloud growth to assure that the
observed cooling is not only a random pattern. IF (5) gives information about
the cloud top height relative to the troposphere, or a very dry layer in the mid
to upper troposphere. IF (7) investigates the time trend of the channel
difference in IF (5). IF (6) and IF (8) are used to highlight cloud pixels that
are likely to develop into a precipitating cloud (see table 1, Mecikalski and
Bedka 2006, Mecikalski et al 2008, 2010, Roberts and Rutledge 2003, Mueller et
al 2003). As in Mecikalski and Bedka (2006), in order to issue a CI warning 7
out of 8 criteria per pixel have to be met.
```

Line 12ff have been kept as the reasons for not including additional criteria given there.

Minor Comment (t)
page 1781, line 3: add comma after "algorithm"
Comma added.

Minor Comment (u)
page 1781, line 27: change to "An additional objective..."
Changed.

Minor Comment (v)
page 1782: add comma on line 5 (after Typically), line 6 (after CI), line 9 (before CI), line 12 (after Thus), line 19 (after 2013)
Commas added.

Minor Comment (w)
page 1782: no need to capitalize names of clouds, like "cumulonimbus"
We checked the whole paper for capitalized names of clouds and changed accordingly.

Minor Comment (x)
page 1783, equations (1)-(4): These could almost just be references, since they are very common, and hence there is no need to redefine them here.
Equations have been substituted by a reference to Roebber, Paul J. "Visualizing multiple measures of forecast quality." *Weather and Forecasting* 24.2 (2009): 601-608.
Substituting p 1783, line 17-22 with:
`Different verification statistics can be calculated (Roebber, 2009) with these categorical variables. In this paper the following are used: POD (Probability of Detection), FAR (False Alarm Ratio), CSI (Critical Success Index), BIAS.`

Minor Comment (y)
page 1786, lines 14 & 16 (and elsewhere in the document): What does "IF 2", "IF 3", etc. mean? Please define. See page 1794 as well.
Interest field (IF) has been defined at the first occurrence of the word on page 1778.

Minor Comment (z)
page 1787: Add comma on line 2 (after 2008), and line 12 (after conditions)
Added Commas.

Minor Comment (aa)
page 1788: Capitalize "Europe" everywhere in the document
Capitalized.

Minor Comment (bb)
page 1788, line 15: remove "for this feature"
Removed.

Minor Comment (cc)
page 1788, line 15-16: remove "in the further investigation"
Removed.

Minor Comment (dd)
page 1789, line 28: "halfed" should be "halved"
Corrected.

Minor Comment (ee)
page 1791, line 25: remove the word "the" before "3 July 2010"
Removed.

Minor Comment (ff)
page 1794, line 10: change "founding" to "finding"
Corrected.

Minor Comment (gg)
page 1794, line 17: change "developments" to "convective storms"
Changed.

Minor Comment (hh)
page 1795, lines 15-16: re-write sentence beginning with "Best FAR values..." It does not make sense as written.
Corrected. See Major Comment 3.

Minor Comment (ii)
page 1795, lines 27-28: Combine paragraphs -- no need for a 1 sentence paragraph.
See Major Comment 3.

Minor Comment (jj)
page 1796, lines 13-22: Correct sentences -- at least two sentences are fragments, and not complete. Also, add commas to improve sentence flow.

See Major Comment 3.

l20ff:
`With the drastically increasing number of false alarms, and the omission of tracking all these cells, we also obtain an improvement in calculation time.`

Minor Comment (kk)
page 1797, line 1: was "infra-red" every defined as "IR"? if so, use "IR" .
Used IR instead of infra-red.

Minor Comment (ll)
page 1797, line 28, and onto page 1798: no need to redefine POD, FAR, CSI
Redefinition removed.

Minor Comment (mm)
page 1798, line 5: How can the BIAS be reduced from 320 to 146%? I thought BIAS was a number around 1? Please make more clear and/or correct.

Optimal values are now expressed as percentage and used as percentage throughout the whole paper.
p1784, l1ff reads now:
`The ideal value is 100% for the Probability of Detection (POD), Critical Success Index (CSI), and BIAS and is 0% for the False Alarm Ratio (FAR).`

Minor Comment (nn)
page 1799, line 5: change "Unfavourably" to "Unfavourable"
Changed.