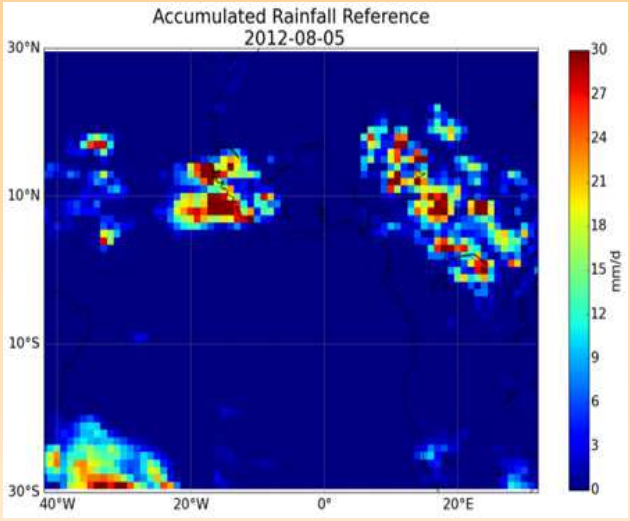
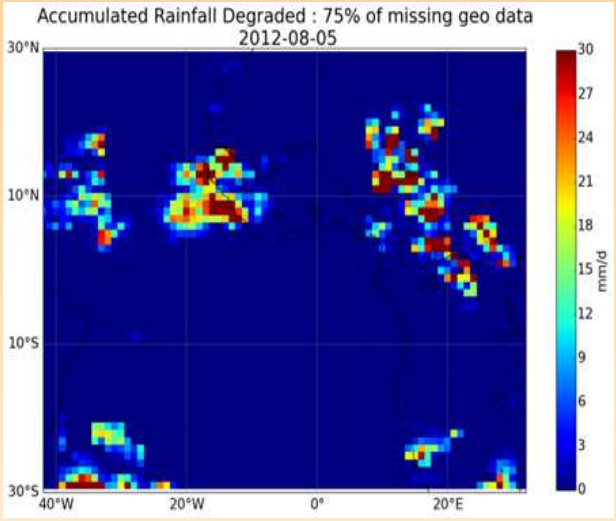
MTTM

Megha-Tropiques Technical Memorandum

Confidence Index for TAPEER-BRAIN product

Estelle Lorant, Sophie Cloché, Guillaume Urbani, Rémy Roca



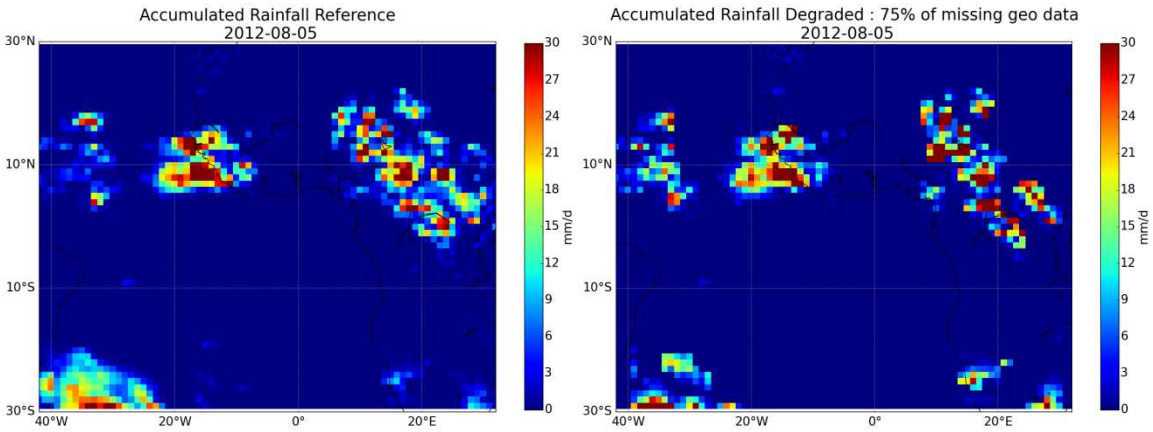


May 2017

08

Megha-Tropiques Technical Memorandum

Confidence Index for TAPEER-BRAIN Product



Estelle Lorant (CNRS/IPSL), Sophie Cloché (CNRS/IPSL), Guillaume Urbani (CNRS/IPSL),

Rémy Roca (CNRS/LEGOS)

Release 0

February 2017

Release N°	Publication date	Authors	Brief description of the change
R0	14 February 2017	E. Lorant S. Cloché R. Roca	1st version

Index

1. Introduction

The TAPEER-BRAIN algorithm combines microwave and infrared observations to provide precipitation estimates and their associated error bars at the one-degree/one-day accumulated scale, over land and ocean. Rain detection and rain rates estimates from LEO microwave imagers are merged with full space/time resolution data from the thermal window channel on board geostationary (GEO) platform (MSG, Meteosat7,GOES-E,GOES-W, MTSAT in 2012, represented in Figure 1).
The details of the algorithm are available in the ATBD document (Chambon et al.,2012a, Roca et al., 2014) while the scientific background on both the algorithm and the error modeling effort are presented in (Roca et al; 2010; Chambon et al., 2012b, Roca et al 2017).

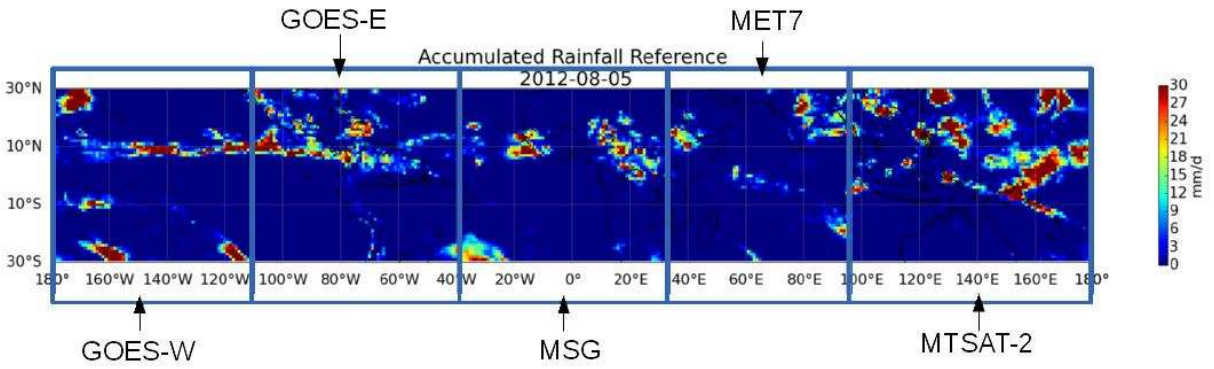


Figure 1. Example of daily accumulated rainfall map from TAPEER-BRAIN product over one day.

The retrieval of daily accumulated rainfall and its sampling error are not possible for one day if there is no GEO data for that day. Issues arise when GEO data are not fully available for the day. Daily accumulated rainfall based on few files is not representative of the meteorological situation of the day. The purpose of this study is to give a confidence index on the accumulated rainfall depending on the availability of GEO data
To derive such confidence index, a sensitivity analysis of the TAPEER-BRAIN algorithm to the missing infrared observations from GEO platforms has been performed. The missing data from GEO platform can be occurred for different reasons : interruption due to platform failure, eclipse phenomenon, maintenance ... From this analysis, the empiric thresholds of missing data percentage for the day is determined and applied in the TAPEER-BRAIN algorithm to provide confidence index for each day and each geostationary.

First a methodology of sensitivity analysis is explained then the results are shown and the determination of the thresholds is presented. Finally, the confidence index derived from these thresholds is described.

2. Sensitivity analysis of TAPEER-BRAIN algorithm to missing data and determination of degree of confidence

2.1 Methodology

To determine a confidence index for the daily accumulated rainfall, sensitivity analysis of TAPEER-BRAIN algorithm to missing data from GEO platform has been performed.
TAPEER-BRAIN products are computed from data GEO platform with full availability for the

period of study. They are called reference product for analysis. The sensitivity analysis of the algorithm is based on comparison of the TAPEER-BRAIN "reference" product with the TAPEER-BRAIN "degraded" product derived from GEO data partly available.
Several type of degradation on the input GEO data archive are applied : files missing randomly, different percentage of missing files compared to the expected number of data per day, successive missing files, simulation of eclipse phenomenon[1]...Therefore, all geostationary platform datasets are degraded for different seasons and different years.
The present study depends on how TAPEER-BRAIN algorithm works. To determine the daily rainfall accumulation, there are two parameters which are computed from different size of learning window: the rainy cloud fraction based on a brightness temperature threshold and the local mean rain rate value. The temporal size of the learning window for the rain detection is one day only whereas the temporal size of the learning window for the conditional rain rate is five day (2 days before and 2 days after). As the conditional rain rate is averaged on several days, it is more robust than the rain detection to the missing input files.
It is expected that the degradation only for the current day will impact more the daily rainfall accumulation than a degradation of the days around the day. The results from the tests of degradation of the days around one given day will not be shown in this document as they show no real impact on the daily accumulated rainfall and confirm the previous assumption. In the following, we will concentrate only on the tests and results of the degradations applied on one given day.

The geostationary platforms have a schedule of slot acquisition per day. There is a reference number of slot per day which is degraded to simulate missing data randomly.
The geostationary platforms and degradation period chosen are the following:

- GOES-E from 01/01/2012 to 10/01/2012, day of study : 05/01/2012
- MSG and GOES-E from 01/08/2012 to 10/08/2012, day of study : 05/08/2012
- METEOSAT7 from 21/07/2013 to 31/07/2013, day of study: 25/07/2013.

- Degradation of geo data for one given day :
  - simulation of eclipse phenomenon of METEOSAT 7 : remove 4 slots between 18H to 20H (noted hereafter **Eclipse 17%**)
  - remove randomly 25% of data per day (noted hereafter **25%**)
  - remove 35% of data per day in a row (noted hereafter **35% block**)
  - remove 50% of data per day, one slot in two (noted hereafter **1slot2 50%**)
  - remove 50% data per day in a row, half day from 00H to 12H (noted hereafter **50% block**)
  - remove randomly 75% of data per day (noted hereafter **75%**)
  - remove 75% of data per day in a row (noted hereafter **75% block**)

Several metrics to compare reference daily rainfall accumulation with the degraded daily rainfall accumulation are used:

For the grid points which remain rainy in the reference and the degradation and have an accumulated rainfall greater than 1 mm/day:

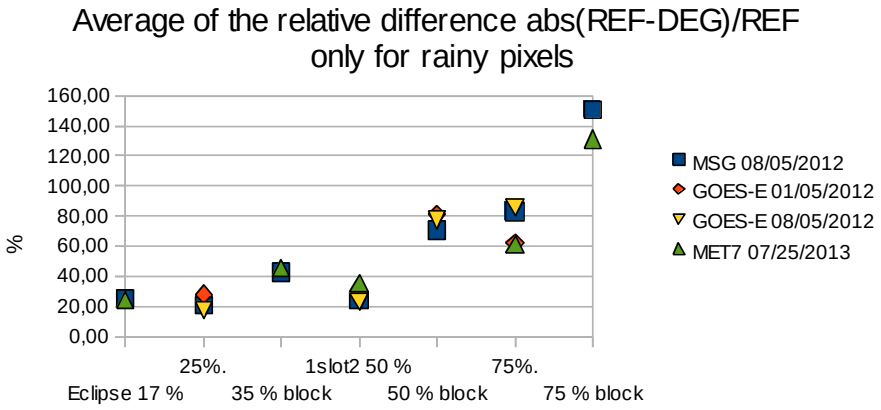
- mean and standard deviation of relative difference of daily rainfall accumulation

1 Eclipse phenomenon is a period of few hours per day where is not possible to obtain data from GEO platform. Eclipse phenomenon is common with Meteosat7-IODC.

- percentage of grid point which changed from rainy to non-rainy and vice versa.

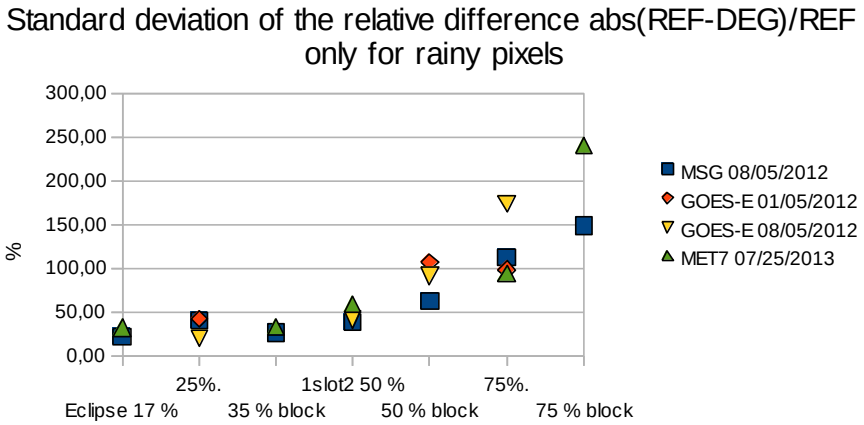
2.2 Results

First, the statistics results are presented (graphics and tables in figure 2, 3, 4). Then the maps of relative difference of accumulated rainfall and the rain detection changes are presented for several geostationary platforms, seasons and type of degradations (figures 5, 6, 7, 8).



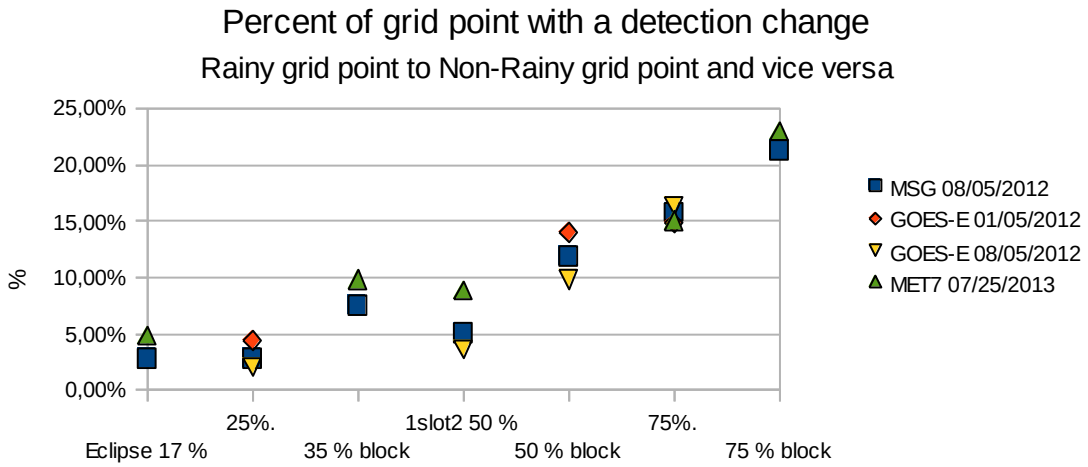
Average abs(REF-DEG)/REF (%)	MSG 08/05/12	GOES-E 01/05/12	GOES-E 08/05/12	MET7 07/25/13
Eclipse 17 %	25,11			23,79
25%.	20,56	27,93	17,29	
35 % block	42,78			45,23
1slot2 50 %	24,37		22,89	35,18
50 % block	70,81	80,74	77,25	
75%.	83,10	62,05	86,05	61,17
75 % block	151,06			130,93

Figure 2. Average of the relative difference between reference and degradation, only for rainy pixels



Standard deviation abs(REF-DEG)/REF (%)	MSG 08/05/12	GOES-E 01/05/12	GOES-E 08/05/12	MET7 07/25/13
Eclipse 17 %	21,35			31,33
25%.	39,58	41,22	19,29	
35 % block	26,17			32,58
1slot2 50 %	37,71		40,70	57,94
50 % block	62,53	106,38	90,86	
75%.	112,24	98,09	173,27	94,10

Figure 3. Standard deviation of the difference between reference and degradation only for rainy pixels



Percent of grid point With a detection Change	MSG 08/05/12	GOES-E 01/05/12	GOES-E 08/05/12	MET7 07/25/13
Eclipse 17 %	2,88%			4,89%
25%.	2,84%	4,43%	2,05%	
35 % block	7,57%			9,84%
1slot2 50 %	5,14%		3,63%	8,92%
50 % block	11,91%	14,01%	9,87%	
75%.	15,83%	14,83%	16,37%	15,09%
75 % block	21,40%			23,06%

Figure 4. Percentage of grid points with a detection change between reference and degradation : rainy grid point to non-rainy grid point and vice versa

Analysis of the statistics results.

For all the geo platform or seasons :

Whether for the detection change in Figure 4 or the relative difference average of accumulated rainfall in Figure 2 and Figure 3, a net change is observed when it exceeds a 50% of successive missing geo data (by block). The percentage of the grid points with a detection change increases from 5% to 20% (shown in Figure 4). The relative difference average increases from 50% to superior at 100% (shown in Figure 2).

Analysis of the map results.

In Figure 5, for the degradations of 25% of missing geo data and 50% of missing geo data (one slot in two), the relative difference of accumulated rainfall oscillates between -20% and +20%. For the degradations of 75% of missing geo data and 50% of successive missing geo data (by block), the relative difference of accumulated rainfall oscillates between -80% and +80%. In Figure 6, the rain detection changes for all types of degradations are not located on

specific geographical region. The main variation is the loss of rain detection (in blue on the maps, the false positive is in red: detection of rain in the degraded case whereas there is no rain detection in the reference case). The Figure 7 and Figure 8 confirm that the impact of one degradation type (here 75% of missing geo data) is the same regardless neither the geo platforms (MSG, GOES-E, METEOSAT-7) nor the seasons (in summer or winter for GOES-E).

Obviously, the more input geo data is missing, the more the rain detection is affected. But, if the input geo data is missing successively, the impact is greater on the accumulated rainfall than if it is missing at random. For the same percent of missing geo data (50%), the results of the missing data one slot in two and by block show a net difference whether the average of the relative difference (24% to 70%) or the rain detection change (11% to 15%) in Figure 2 and Figure 3.

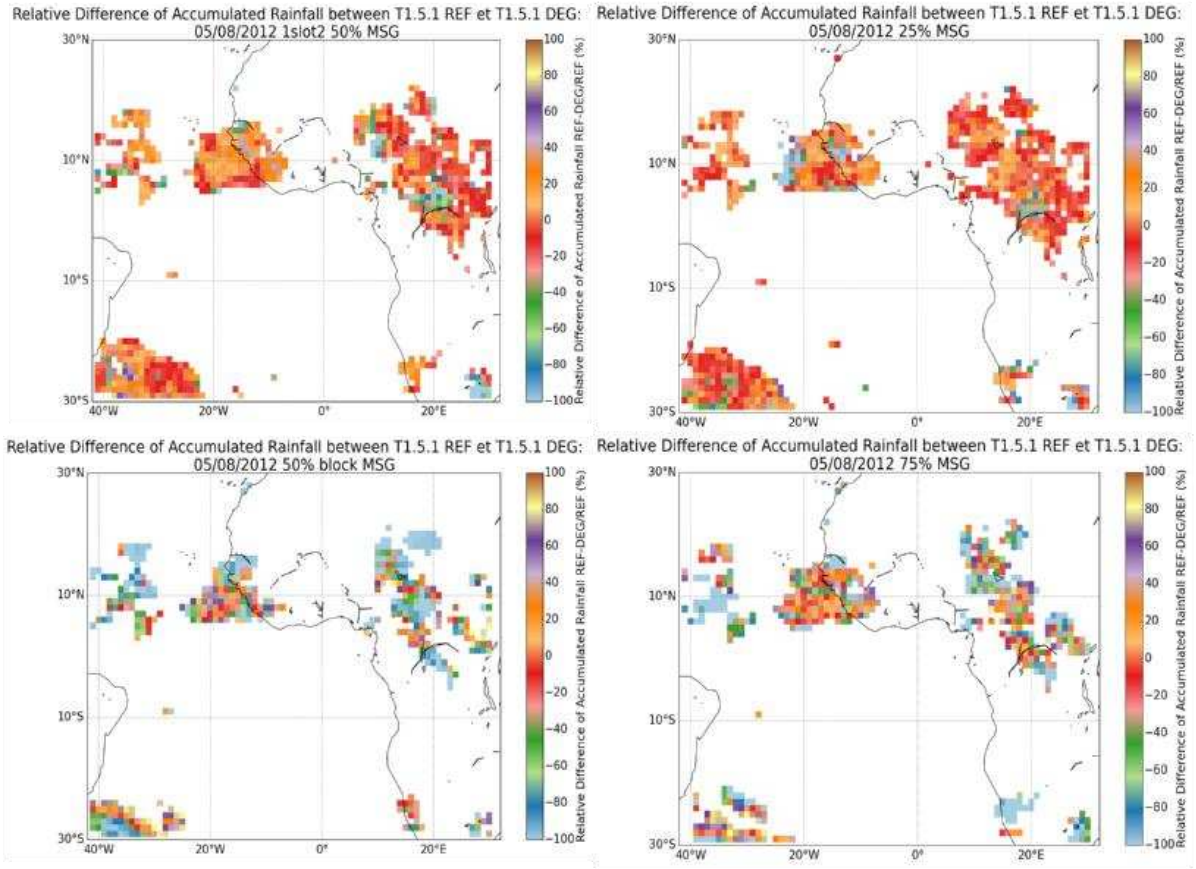


Figure 5. Map of the relative difference of accumulated rainfall between reference and several degradations on one day for MSG platform

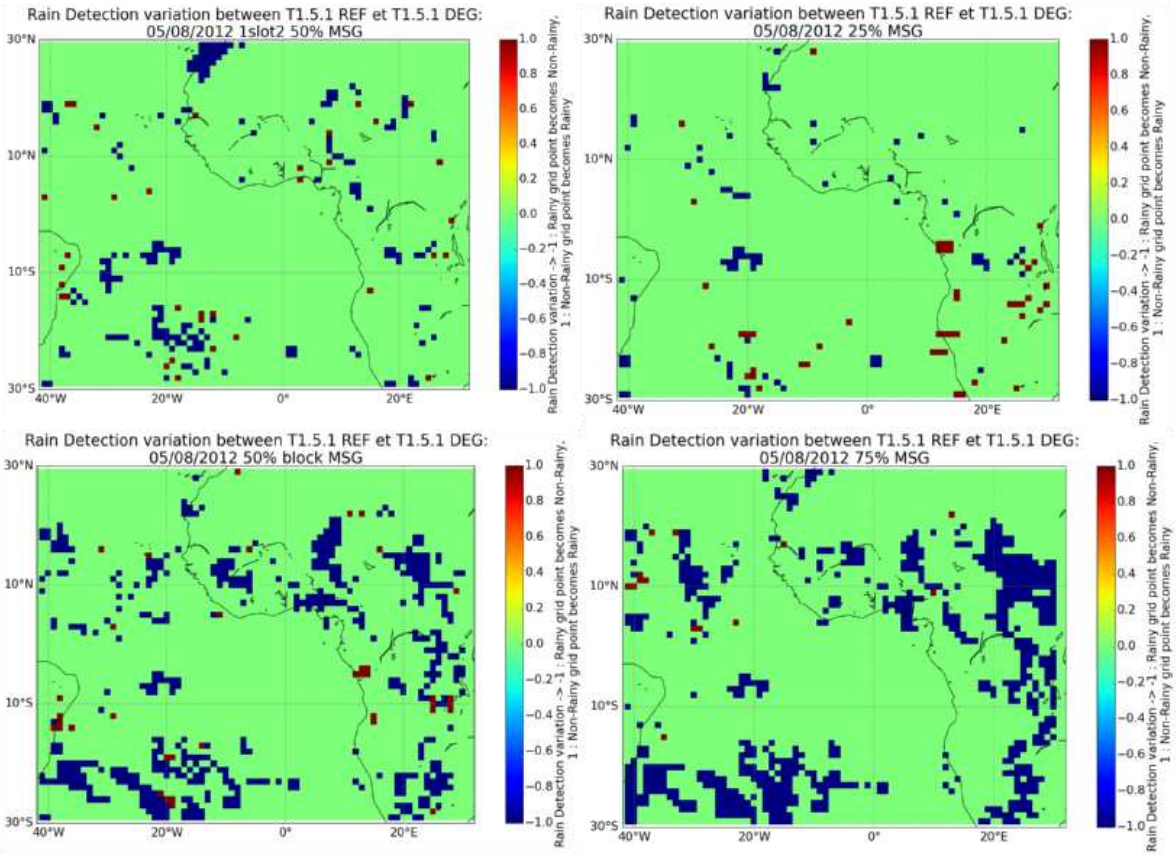


Figure 6. Map of the rain detection change between reference and several degradations on one day for MSG platform

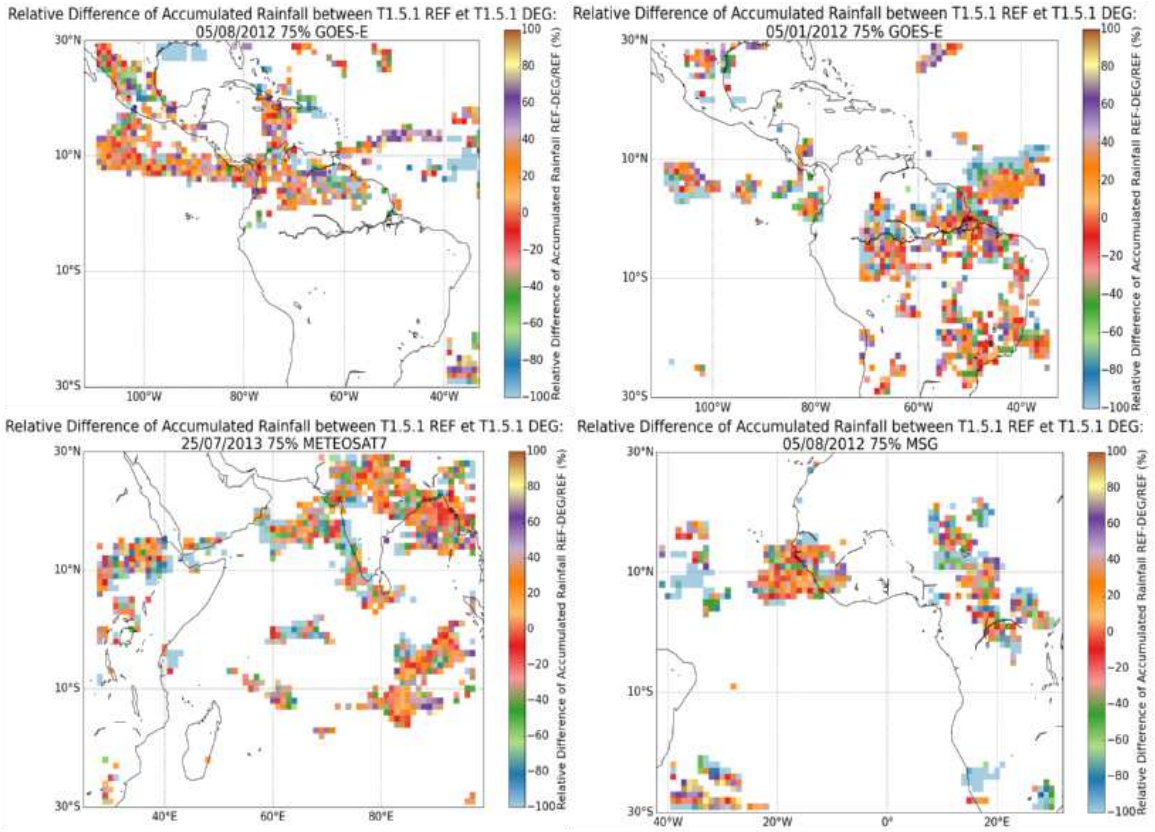


Figure 7. Map of the relative difference of accumulated rainfall between reference and one type of degradation on one day for several geo platforms and several seasons

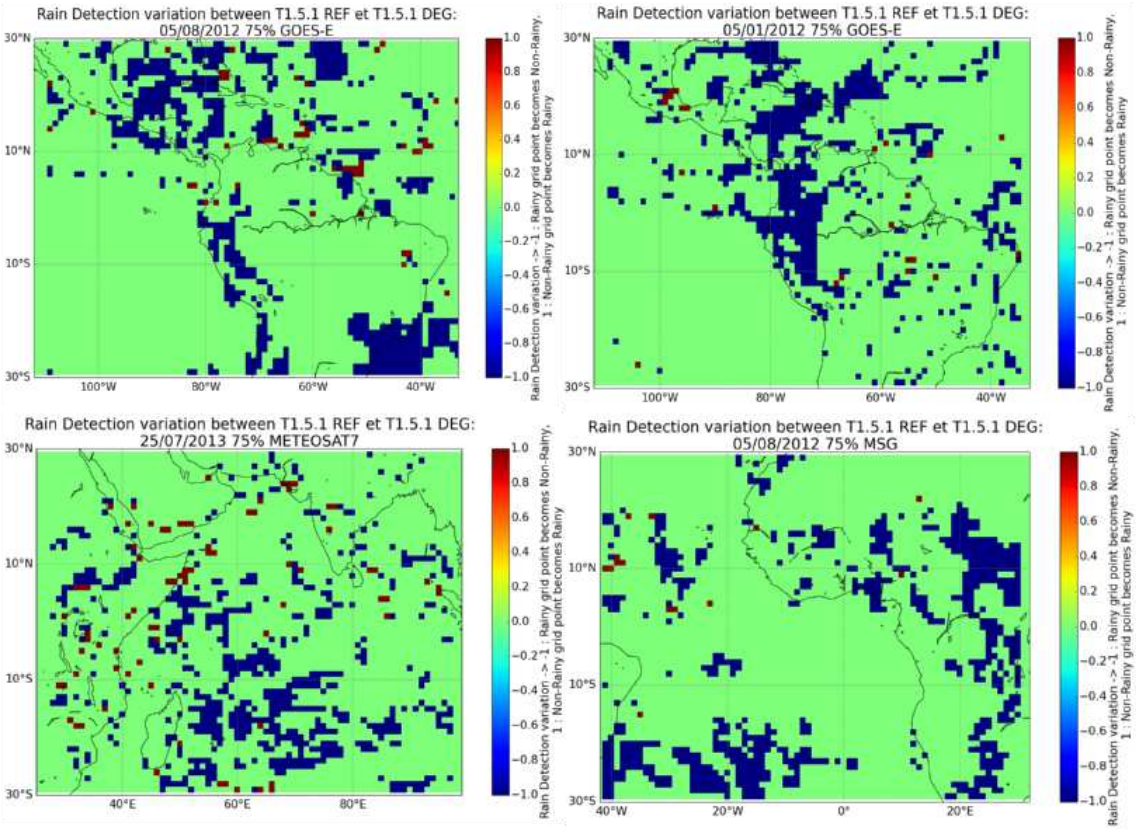


Figure 8. Map of the rain detection change between reference and one type of degradation on one day for several geo platforms and several seasons

2.3 Determination of the different thresholds of missing data percentage

All the metrics show a significant change from a degradation of 50% of successive missing data regardless season, geostationary platform. Percentage of grid points which changed rain detection increases from 5% to 20%. The relative difference average of accumulated rainfall increases from 50% to percent greater than 100%.
Missing at random and successive missing data do not impact daily rainfall accumulation equally. For the same percent of missing data (50%), the relative difference mean of daily rainfall accumulation is higher when the missing data are successive (~70-77%) than the missing data randomly (~22-24%). This is due to best chances to miss a precipitation event when the missing data are successive.

Confidence index is based on 3 levels, resulting from last comparisons:

- OK : if the percent of missing data is inferior to 25% for missing randomly or if the percent of missing data is inferior to 15% for successive missing data

- WARNING : if the percent of missing data is between 25% and 75% for missing at random or if the percent of missing data is between 15% and 50% for successive missing data

- KO : if the percent of missing data is superior to 75% for missing at random or if the percent of missing data is superior to 50% for successive missing data

The thresholds for each level of confidence index do not need to change according to the season or the geostationary platform as the comparison results show no change whatever season or geostationary platform.

3. Confidence index of the TAPEER-BRAIN product

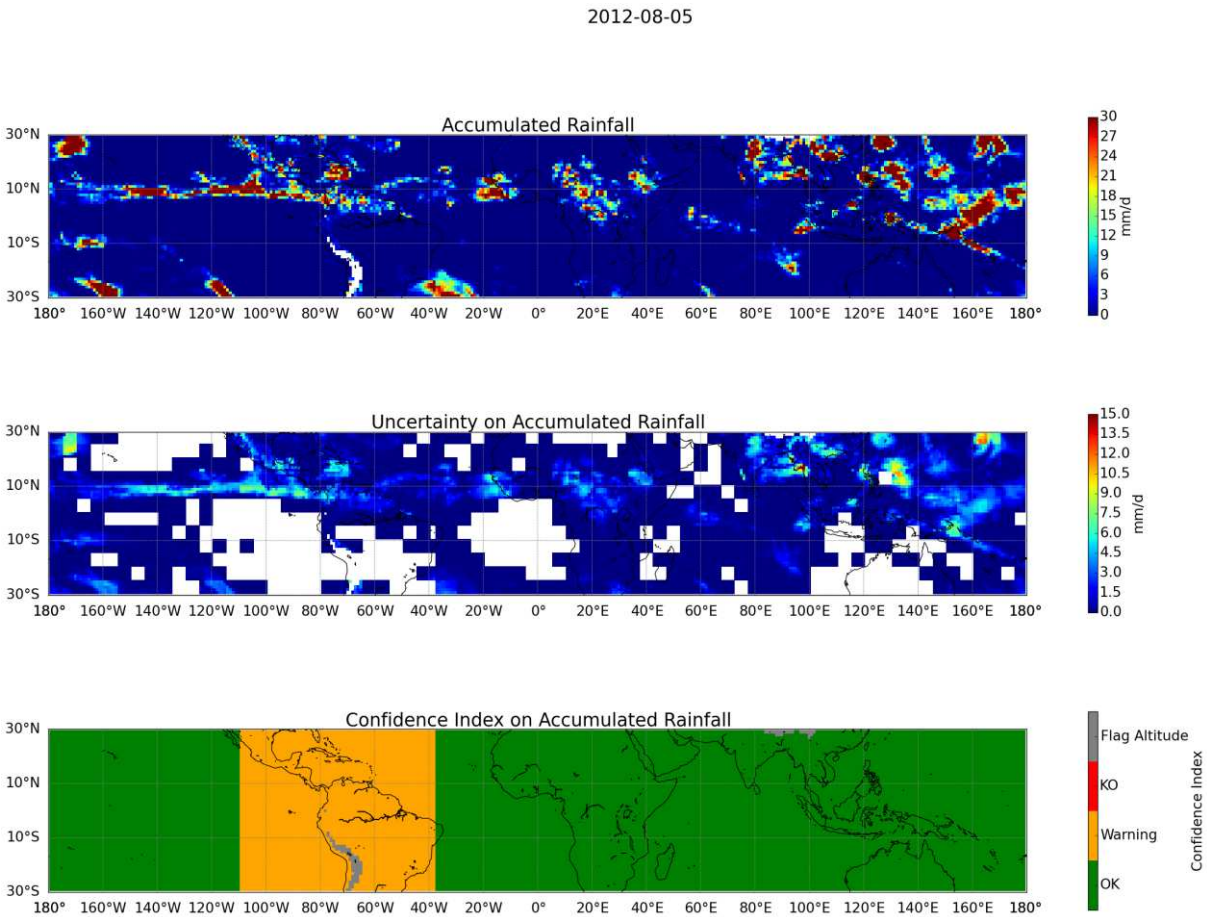


Figure 9. Example of TAPEER-BRAIN Product with confidence index on accumulated rainfall for the day

The final TAPEER-BRAIN (T-B) product for one day is presented in Figure 9. Additionally to the uncertainty, the confidence index described previously is given for each grid point of accumulated rainfall.

To give this confidence index, the control quality is based on a theoretical schedule of geo data. We compare the geo data really used in the TAPEER-BRAIN algorithm to the theoretical expected data.

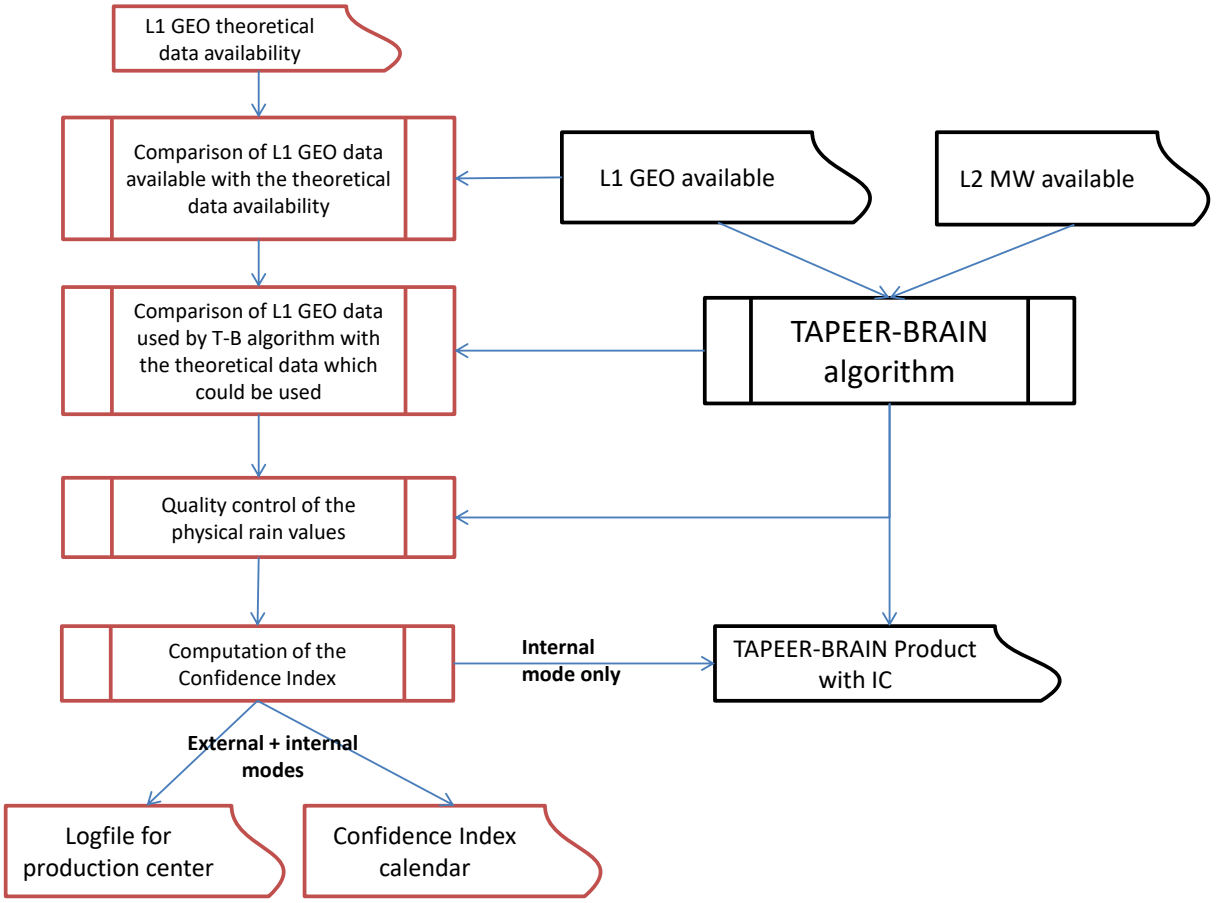


Figure 10. General flowchart of the confidence index process (red boxes) and how it is linked to TAPEER-BRAIN process (black boxes)

In Figure 10, the flowchart shows at which steps the quality control is made in the TAPEER-BRAIN algorithm. The first control of the algorithm is to check if the available input L1GEO data fit the theoretical data availability given by each agency in order to warn the production center of a potential missing data defects.
A second control of the algorithm is made in order to check if the TAPEER-BRAIN algorithm uses all the data available compared to theoretical data availability. If not, an alert is activated in the log output to warn the production center and the percent of missing data per day and per geostationary is computed from a theoretical expected number of slots per day and per geostationary and the number of geostationary slot really used by the algorithm. This percent is compared to the different thresholds (experimentally chosen after the studies presented in section 2) to determine the degree of confidence of accumulated rainfall retrieval for the day and geostationary under study.

The last control of the algorithm is made to check the physical validity of the accumulated rainfall retrieval. If not, an alert is activated in the log output to warn production center and confidence index is putted at the low rate: data is rejected.

The quality control can be included to the production of the TAPEER-BRAIN (internal mode) or can be applied after the production (external mode). Then the available outputs are different (see Figure 10). If the quality control is applied during the production, the outputs available are a confidence index inside the TAPEER-BRAIN products, a log file for the production center and a confidence index calendar for both user and production center. If the quality control is applied after the production of TAPEER-BRAIN, only the log file and the confidence index calendar are available.

Finally, for each day and each geostationary the confidence index has been merged in the Quality_Flag in the TAPEER BRAIN product. Bits number 1 to 3 describe confidence index for each grid point per day like in following table:

Bit no.	Quality Flag Name	Bit Values			Definition
1, 2 and 3	Confidence Index	Bit3	Bit2	Bit1	
		0	0	1	highest confidence data
		0	1	0	data may be degraded
		1	0	0	data has been rejected
4	Altitude Flag	0			retrieval in physical range
		1			non retrieval due to altitude > 3000 m
5	Spare bit				
6	Spare bit				
7	Spare bit				
8	Spare bit				

Table 1. Description of Quality_Flag variable in TAPEER-BRAIN product

It is highly recommended to use only data with the highest confidence index. Be careful with intermediate level. At least, it is recommended to not use data where the bit number 3 is equal to 1. Studies demonstrate data has been too much degraded to be used. The details of the confidence index format are available in the Product Definition Document of TAPEER-BRAIN.

The confidence index is also produced in calendar format (see Figure 11). It gives to the users a quick look of the availability and quality of the TAPEER-BRAIN product. In green the confidence index is at the OK level, in orange the confidence index is at the Warning level and in red the confidence index is at the KO level. Finally in grey, the TAPEER-BRAIN product is not available.

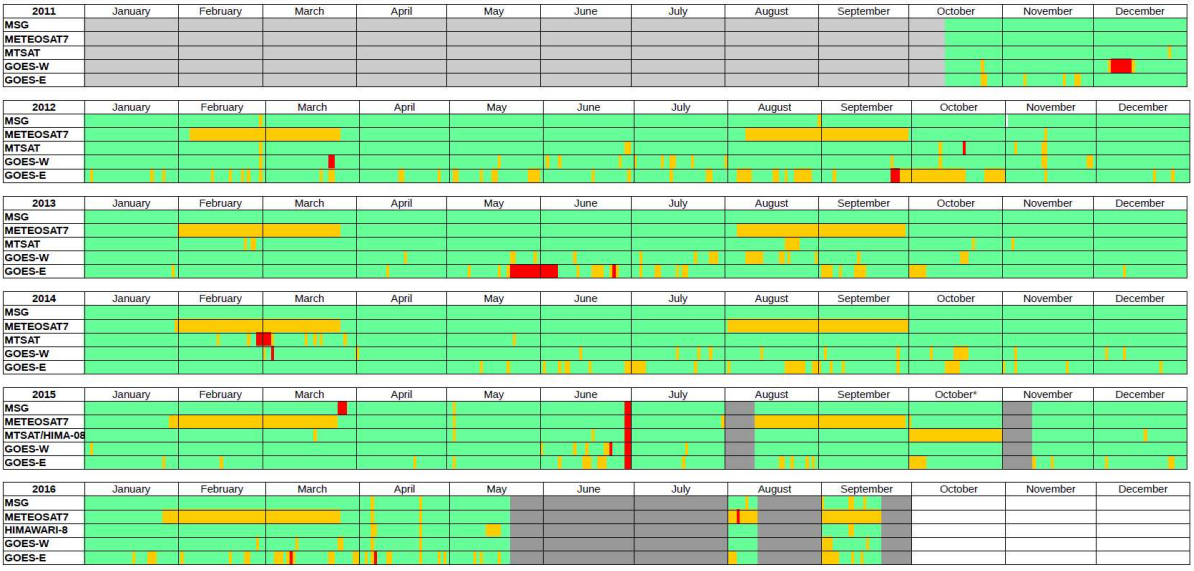


Figure 11. Example of Confidence Index Calendar for TAPEER-BRAIN products

4. Conclusion

The confidence index deduced from this study can inform the users on the TAPEER-BRAIN quality depending on the geostationary missing data.
The quality control has been tested on the recent geostationary HIMAWARI-8. The thresholds determined are also valid for the new high-resolution geostationary float (as HIMAWARI-8, GOES-16 …).
The quality control could be improved in the future by taking into account the microwave imagers missing data impact on the TAPEER-BRAIN product. The current confidence index is only available for the accumulated rainfall; this study could be also improved by studying the impact of the missing geo data on the uncertainty computation.

References

Chambon P., Roca, R., Jobard, I., Aublanc, J. (2012a). TAPEER-BRAIN product: Algorithm Theoritical Basis Document, Level-4, Megha-Tropiques Technical Memorandum #4, 13 pp. Available from http://megha-tropiques.ipsl.polytechnique.fr

Roca et al.,(2017). Quantifying the contribution of the Megha-Tropiques mission to the estimation of daily accumulated rainfall in the Tropics, in preparation for the Special Collection "Advances in Remote Sensing of Rainfall and Snowfall", Q. J. R. Meteorol. Soc.

Roca R., J. Aublanc, P.Chambon, T. Fiolleau, and N. Viltard, (2014) Robust observational quantification of the contribution of mesoscale convective systems to rainfall in the tropics, J Climate, 27, 4952–4958. doi: http://dx.doi.org/10.1175/JCLI-D-13-00628.1

Roca, R., Chambon, P., Jobard, I., Kirstetter, P.-E., Gosset, M., and Bergès, J.-C. (2010)**.** Comparing satellite and surface rainfall products over west africa at meteorologically relevant scales during the amma campaign using error estimates. J. Appl. Meteorol. and Climatol.

Chambon P, Jobard I, Roca R, Viltard N. (2012b). An investigation of the error budget of tropical rainfall accumulation derived from merged passive microwave and infrared satellite measurements. Q. J. R. Meteorol. Soc. 138: 000.000. DOI:10.1002/qj.1907

Roca R., Cloché S., Lorant E., Taburet N. (2015). Product Definition Document : Level 4 products, Surface accumulated rainfall and uncertainty, L4-TAPEER-BRAIN. Version 1, Release 2.

Figures Index

MTTM

Megha-Tropiques Technical Memorandum

Laboratoire de Météorologie Dynamique (LMD / IPSL)

Ecole Polytechnique

F-91128 Palaiseau

France

Sophie.Bouffies-Cloche@ipsl.jussieu.fr

http://meghatropiques.ipsl.polytechnique.fr/available-documents/mttm/index.html