Supplement of Hydrol. Earth Syst. Sci., 24, 2687–2710, 2020
https://doi.org/10.5194/hess-24-2687-2020-supplement

Supplement of

A daily 25 km short-latency rainfall product for data-scarce regions based on the integration of the Global Precipitation Measurement mission rainfall and multiple-satellite soil moisture products

Christian Massari et al.

Correspondence to: Christian Massari (christian.massari@irpi.cnr.it)

S1 Assessment of the validity of the TC analysis

In this section we test whether TC is representative of the results obtained using classical validation. TC provides error and correlation with an unknown truth, as a result the TC scores are generally more optimistic than those obtained with classical validation. To calculate TC correlations and errors, we built a total of 15 triplets by combining the ground-reference described in Section 2.1, IMERG-ER, IMERG-FR, ERA5, GPCC, $\mathrm{P_{R+SM}}$, SM2RAIN-AMSR2*, and SM2RAIN-ASCAT* (the last two are not calibrated datasets with global constant parameters) in different ways. We obtained a distribution of errors and correlations derived from using multiple triplets for the same product (i.e., summarized by a box plot). The underlying idea is that, if the choice of the triplet is correct, these errors and correlations should not differ too much for the same product from triplet to triplet and should provide similar conclusions to the classical validation analysis.

Moreover, we performed the TC analysis with two configurations. In the first, we included in the same triplet products that would apparently violate the TC assumptions (we only excluded those triplets where the violation is obvious, i.e., like GPCC with the ground-based reference). Specifically, we used triplets containing $\mathrm{P_{R+SM}}$ and ERA5 (i.e., the calibration dataset), as shown in Table S1. This configuration allows investigating the existence of a possible "false" beneficial increase in performance for the integrated product when it is used in conjunction with ERA5 (i.e., the calibration dataset) in the same triplet. In the second configuration, we performed the same TC analysis by removing those triplets containing ERA5 with the integrated product (Table S2). Here, we only kept triplets containing SM2RAIN-AMSR2* combined with $\mathrm{P_{R+SM}}$. Note that SM2RAIN-AMSR2* was not used in the integration and was not calibrated with ERA5.

Results show that the differences between the two configurations are negligible (compare Figure S1 vs S2). This allows assuming a negligible effect of the calibration dataset when used in the same triplet with the integrated product. However, in the second configuration, the risk of having cross-correlated error dependence is not null, as the SM2RAIN algorithm was used both in the integrated product and in SM2RAIN-AMSR2*. To check this possibility, we calculated these spurious error cross-correlations by assuming the ground-based observations for the key regions (i.e., AWAP, Stage IV, E-OBS and IMD) as a true rainfall and found them always below 0.4. This provides an additional proof about the validity of the TC analysis.

The ability of TC to provide similar conclusions of a classical validation analysis is discussed in Figure S1. The figure shows the box plots obtained by considering TC errors and correlations of Table A1 for AU, CONUS, EU and IN. The small spread around the median value suggests consistent results among the different triplets. In addition, the relative quality of the products provided by the classical validation (i.e., red dots are the median value obtained via classical validation) seems reasonably reproduced by TC both in case of correlation and error. They are generally lower (higher for the error) with respect to TC results as they contain already the uncertainty of the reference. In the case of GPCC for all the regions but India, the values are closer or higher to the one obtained via TC due to the obvious agreement with AWAP, STAGE IV and E-OBS (which share a significant number of rain gauges with GPCC). Moreover, for all the plots it can be seen that the references are characterized by a relatively high quality, which justifies their use for validation purposes.

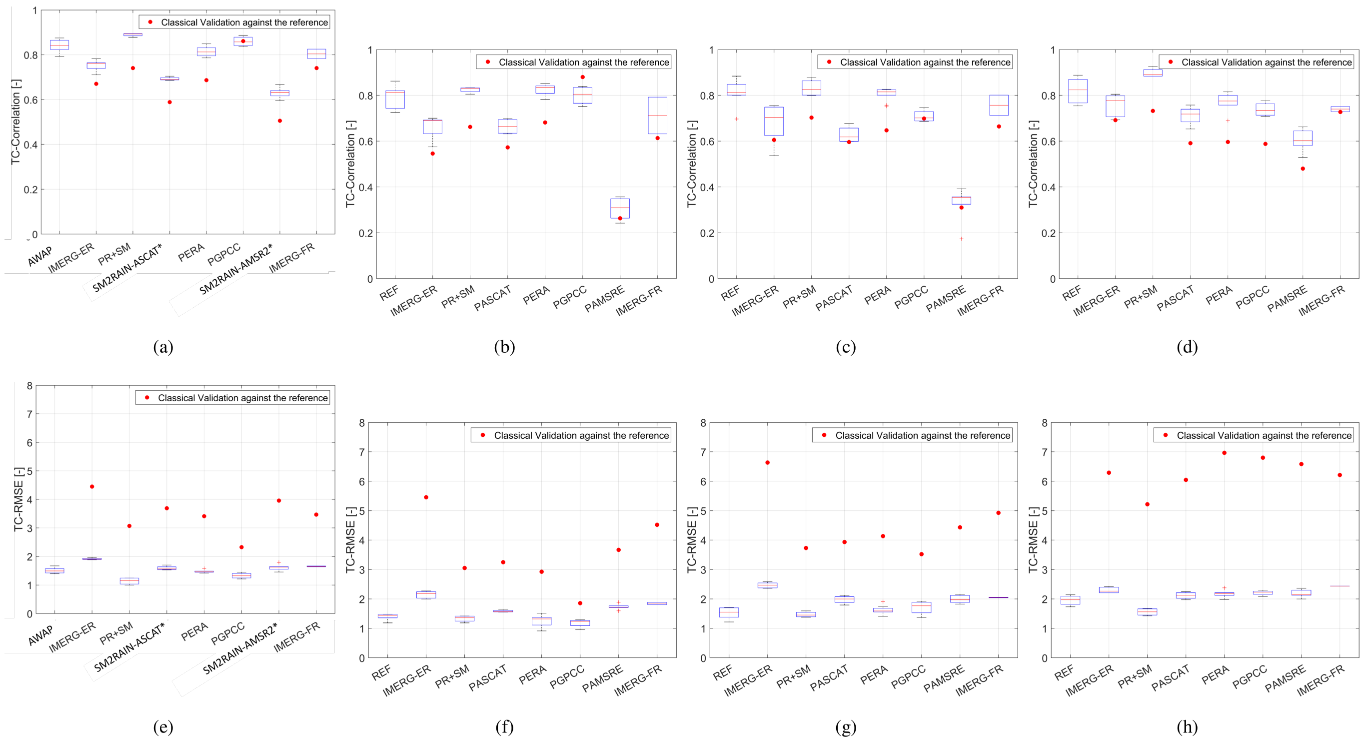


Figure S1. Box plots of TC correlation and error obtained by using the 15 triplets in Table S1, "first configuration") compared with the classical correlation and RMSE scores (red dots). (a) and (e) TC correlation and error for Australia, (b) and (f) TC correlation and error for Europe, REF=E-OBS. (c) and (g) TC correlation and error for CONUS, REF=Stage IV rainfall dataset described in Section 2. (d) and (h) TC correlation and error for India REF=Indian Meteorological Institute rainfall.The box plot refer to the $25th$ and $75th$ percentiles while the whiskers refer the minimum and maximum values. Outliers are not shown in the plot.

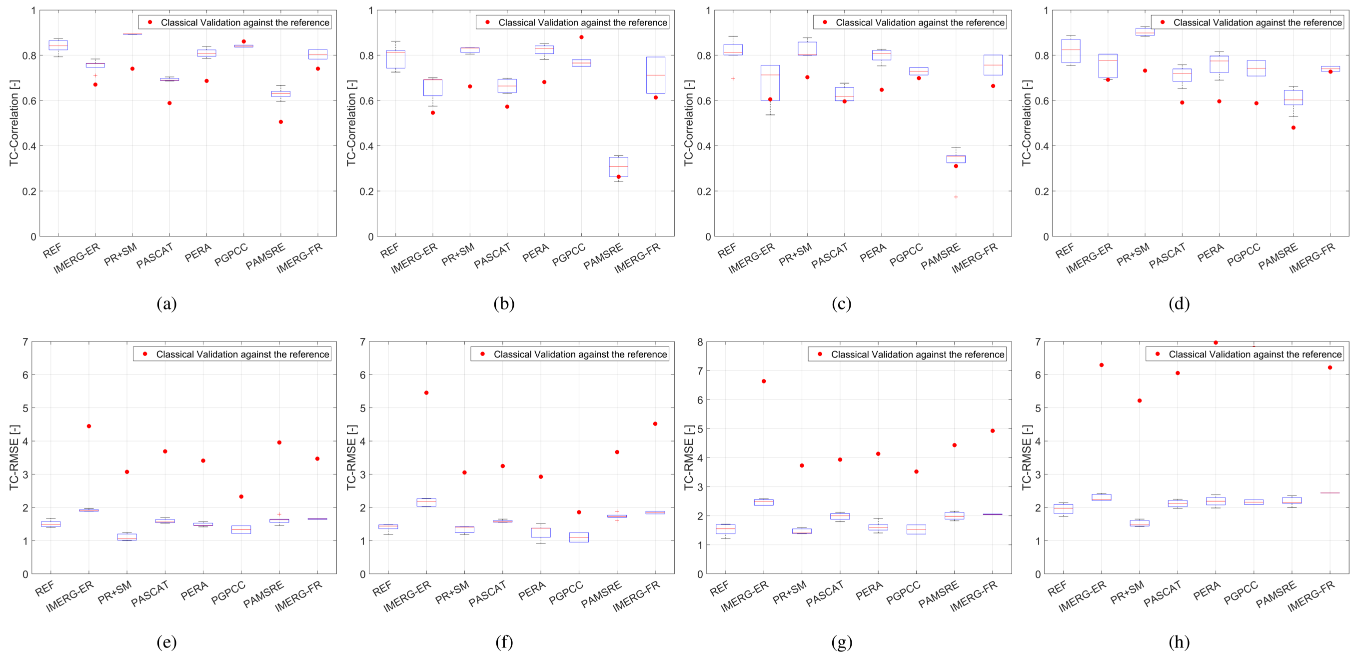


Figure S2. Box plots of TC correlation and error obtained by using the 13 triplets in Table S2, "second configuration") compared with the classical correlation and RMSE scores (red dots). a) and (e) TC correlation and error for Australia, (b) and (f) TC correlation and error for Europe, REF=E-OBS. (c) and (g) TC correlation and error for CONUS, REF=Stage IV rainfall dataset described in Section 2. (d) and (h) TC correlation and error for India REF=Indian Meteorological Institute rainfall. The box plot refer to the $25th$ and $75th$ percentiles while the whiskers refer the minimum and maximum values. Outliers are not shown in the plot.

Product A	Product B	Product C
REF	SM2RAIN-ASCAT*	ERA5
IMERG-FR	ERA5	SM2RAIN-AMSR2*
REF	ERA5	SM2RAIN-AMSR2*
IMERG-FR	ERA5	SM2RAIN-ASCAT*
GPCC	IMERG-ER	SM2RAIN-AMSR2*
GPCC	$\mathrm{P}_{\mathrm{R+SM}}$	SM2RAIN-AMSR2*
REF	IMERG-ER	SM2RAIN-ASCAT*
REF	IMERG-ER	ERA5
ERA5	IMERG-ER	SM2RAIN-ASCAT*
REF	IMERG-ER	SM2RAIN-AMSR2*
REF	$\mathrm{P}_{\mathrm{R+SM}}$	SM2RAIN-AMSR2*
REF	IMERG-ER	ERA5
REF	$\mathrm{P}_{\mathrm{R+SM}}$	ERA5
GPCC	IMERG-ER	ERA5
GPCC	$\mathrm{P}_{\mathrm{R+SM}}$	ERA5

Table S1. Triplets used in the first configuration according to section S1. Here also triplets containing ERA along with $\mathrm{P}_{\mathrm{R+SM}}$ are used. REF here refers to: 1) AWAP for Australia, 2) E-OBS for India, 3) Stage IV for CONUS and 4) IMD for India.

Product A	**Product B**	**Product C**
REF	SM2RAIN-ASCAT*	ERA5
IMERG-FR	ERA5	SM2RAIN-AMSR2*
REF	ERA5	SM2RAIN-AMSR2*
IMERG-FR	ERA5	SM2RAIN-ASCAT*
GPCC	IMERG-ER	SM2RAIN-AMSR2*
GPCC	$\mathrm{P_{R+SM}}$	SM2RAIN-AMSR2*
REF	IMERG-ER	SM2RAIN-ASCAT*
REF	IMERG-ER	ERA5
ERA5	IMERG-ER	SM2RAIN-ASCAT*
REF	IMERG-ER	SM2RAIN-AMSR2*
REF	$\mathrm{P_{R+SM}}$	SM2RAIN-AMSR2*
REF	IMERG-ER	ERA5
GPCC	IMERG-ER	ERA5

Table S2. Triplets used in the second configuration according to section S1. REF here refers to: 1) AWAP for Australia, 2) E-OBS for India, 3) Stage IV for CONUS and 4) IMD for India.

S2 Additional figures

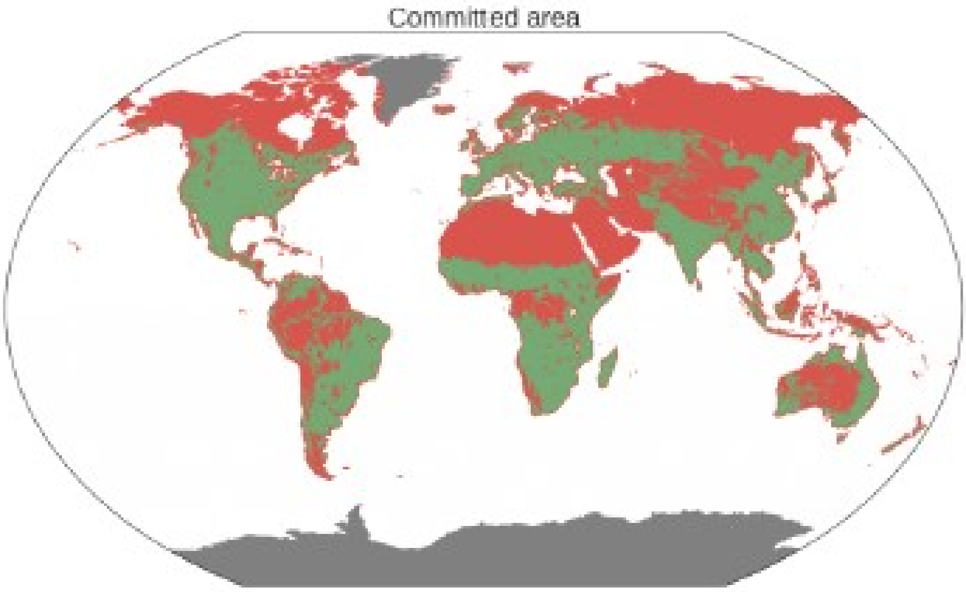
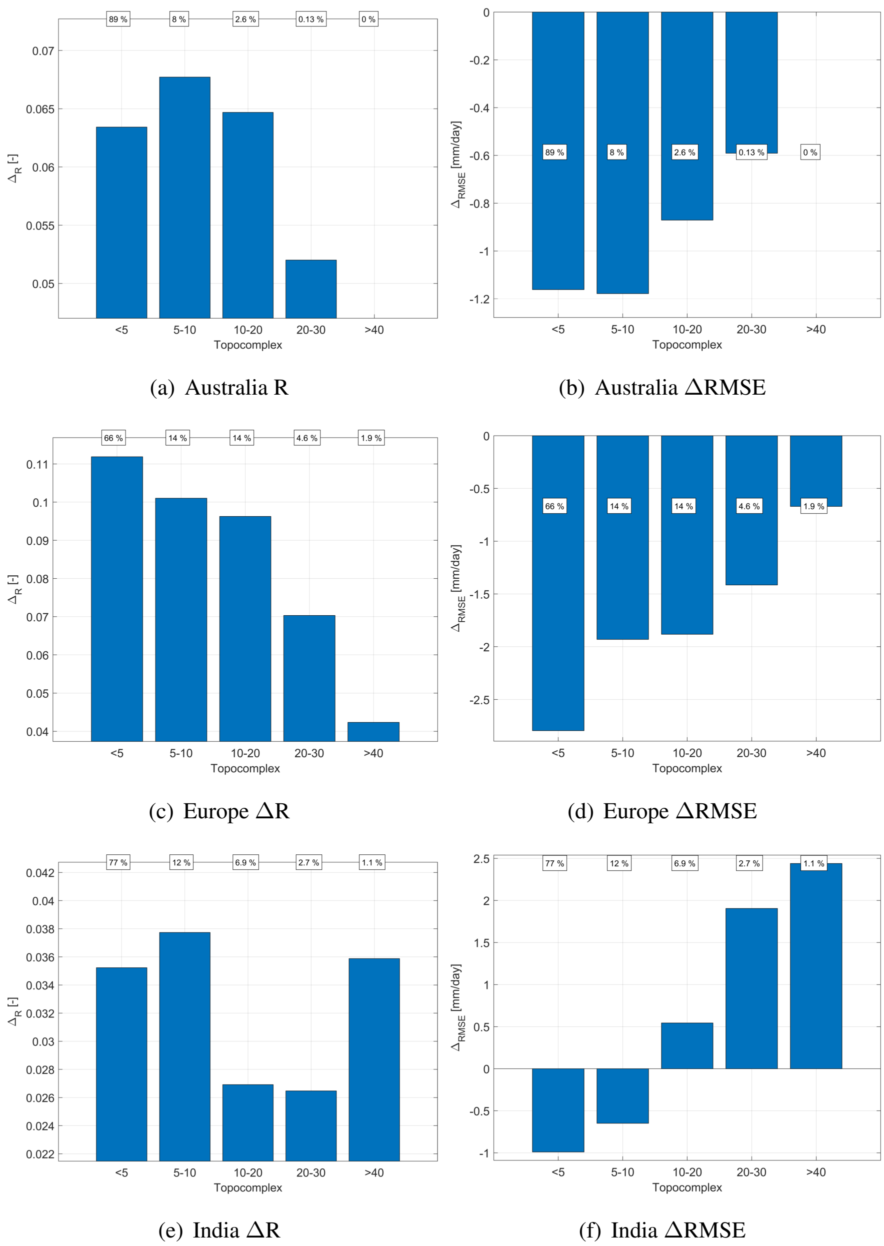


Figure S3. ASCAT committed area. In green validation points, in red areas excluded from validation but where the product is still available. The picture is taken from Han et al. (2016)



(a) Australia R

(b) Australia ΔRMSE

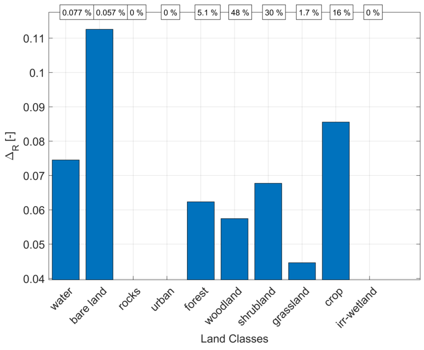
(c) Europe ΔR

(d) Europe ΔRMSE

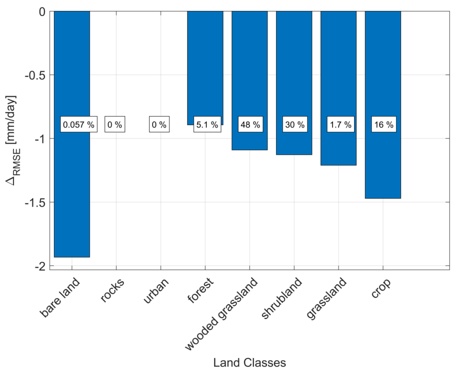
(e) India ΔR

(f) India ΔRMSE

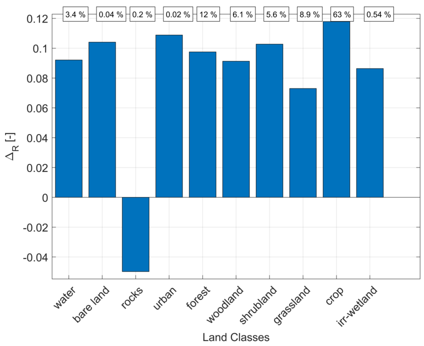
Figure S4. Difference in correlation (R, left) and error (RMSE, right) between the integrated product $\mathrm{P_{R+SM}}$ and the IMERG Early Run product (IMERG-ER) as a function of the topographic complexity. The results refer to 2015-2017 period.



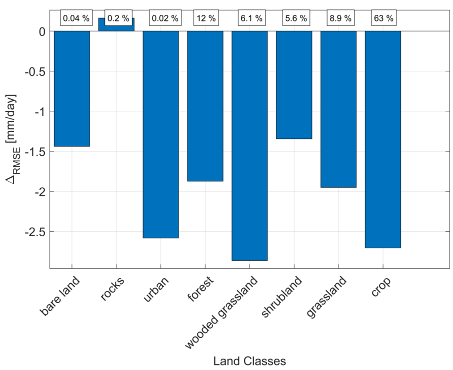
(a) Australia ΔR



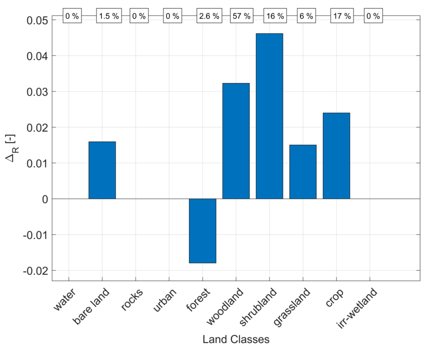
(b) Australia ΔRMSE



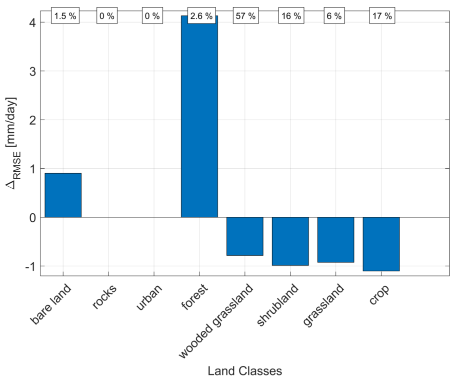
(c) Europe ΔR



(d) Europe ΔRMSE



(e) India ΔR



(f) India ΔRMSE

Figure S5. Difference in correlation (R, left) and error (RMSE, right) between the integrated product $\mathrm{P_{R+SM}}$ and the IMERG Early Run product (IMERG-ER) as a function of the land cover type. Period 2015-2017.