POLITECNICO DI TORINO
Repository ISTITUZIONALE

ERDS: an Extreme Rainfall Detection System Based on Both near Real-Time and Forecast Rainfall Measurements

Original
ERDS: an Extreme Rainfall Detection System Based on Both near Real-Time and Forecast Rainfall Measurements / Mazzoglio, P., Laio, F., Balbo, S., Boccardo, P.. - In: GODISNIK NA UNIVERSITETA PO ARHITEKTURA, STROITELSTVO I GEODEZIA. - ISSN 2534-9759. - ELETTRONICO. - 52:Issue S1(2019), pp. 1423-1433.

Availability:
This version is available at: 11583/2789681 since: 2020-02-05T18:16:19Z

Publisher:
University of Architecture, Civil Engineering and Geodesy - Sofia

Published
DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

08 October 2026

ГОДИШНИК НА УНИВЕРСИТЕТА ПО АРХИТЕКТУРА, СТРОИТЕЛСТВО И ГЕОДЕЗИЯ СОФИЯ

Том Volume 52 2019 Брой Issue S1

ANNUAL OF THE UNIVERSITY OF ARCHITECTURE, CIVIL ENGINEERING AND GEODESY SOFIA

Получена: 03.06.2019 г.
Приета: 17.06.2019 г.

ERDS: AN EXTREME RAINFALL DETECTION SYSTEM BASED ON BOTH NEAR REAL-TIME AND FORECAST RAINFALL MEASUREMENTS

P. Mazzoglio[1], F. Laio[2], S. Balbo[3], P. Boccardo[4]

Keywords: *early warning system, extreme events, flood monitoring, GPM, rainfall*

ABSTRACT

Extreme rainfall may trigger some of the most catastrophic natural disasters, whose consequences may be exacerbated especially in places where an appropriate network of measurement instruments is not available. A combination of remotely sensed data and weather prediction model outputs can often help to obtain information with a global spatial coverage without the limitations that characterize other instruments. In order to achieve this goal, an Extreme Rainfall Detection System (ERDS – erds.ithacaweb.org) was developed and implemented with the aim of monitoring and forecasting exceptional rainfall events. The system was designed with the aim of providing information in an understandable way also for non-specialized users. The NOAA-GFS deterministic weather prediction model is used for the purpose of forecasting extreme rainfall events. Regarding the near real-time rainfall monitoring, the previous version of ERDS was using NASA TRMM TMPA 3-hourly data as input. Due to TRMM instrument shutdown, a different rainfall measurement must be used. NASA GPM IMERG early run half-hourly data proved to be the proper one. A comparison between GPM and rain gauge data allowed to define the minimum time aggregation intervals to be used for the detection of extreme rainfall events in order to reduce the effects of the bias due to satellite data. The same comparison was also performed using GFS data instead of GPM data. A new extreme rainfall detection methodology was also developed with the aim of

[1] Paola Mazzoglio, THACA – Information Technology for Humanitarian Assistance, Cooperation and Action, Torino, Italy (paola.mazzoglio@ithaca.polito.it)
[2] Francesco Laio, Politecnico di Torino, Torino, Italy (francesco.laio@polito.it)
[3] Simone Balbo, ITHACA – Information Technology for Humanitarian Assistance, Cooperation and Action, Torino, Italy (simone.balbo@ithaca.polito.it)
[4] Piero Boccardo,Politecnico di Torino, Torino, Italy (piero.boccardo@polito.it)

increasing system performances. The currently adopted methodology is based on the concept of event-identification threshold. A threshold represents the amount of precipitation needed to trigger a flood event induced by extreme rainfall. Specifically, if for a selected aggregation interval the accumulated precipitation exceeds the threshold, an alert is provided. Obtained results highlighted that the combination of new input data and new threshold methodology allowed one to increase system performances, both in terms of spatial and temporal resolution and in terms of identified events.

1. Introduction

Several studies highlighted an increasing trend in the number of hydrological and meteorological disasters. Early warning system based on remotely-sensed data and numerical weather prediction models can often help in the detection and monitoring of extreme rainfall at the global scale. These systems have an increasingly important role in the disaster risk management, especially for the triggering of flood preparedness actions [1].

The investigation of available dataset and approaches led several institution and research groups to develop new techniques for rainfall and/or flood monitoring. From this perspective, ITHACA developed an Extreme Rainfall Detection System (ERDS – erds.ithacaweb.org). ERDS is a demo service for the monitoring and forecasting of exceptional rainfall events, with a nearly global spatial coverage. ERDS is also an alert system designed to identify hydrometeorological events (such as hurricanes, tropical storms, convective storms, flash floods and heavy rainfall that could lead to flood events or landslides) [8]. This system provides information both on the accumulated rainfall and on heavy rainfall alerts.

The first objective of this study is the assessment of the accuracy of two different datasets used as input data for different aggregation intervals. The obtained results allowed the definition of the proper aggregation intervals usable to provide information regarding the rainfall amount and to evaluate the presence of places potentially affected by hydrometeorological disasters. The second objective, instead, is the development of an extreme rainfall detection methodology applicable at a global scale based on event-identification thresholds calibrated taking into consideration the mean annual rainfall as site-specific parameter.

One of the aspects to be taken into consideration is the non-coincidence between the place where the alert was given and the place where the flood may occur. ERDS is a tool developed for rainfall monitoring. The system provides an alert where the amount of rainfall is higher than a specific threshold. The system, however, does not take into account the morphology of the territory or information regarding basins. The flood, therefore, may occur in the alerted cell or in nearby ones. Further studies might investigate this important aspect.

2. Input Data

Regarding the near real-time rainfall monitoring, the system takes advantage of NASA GPM (Global Precipitation Measurement) IMERG (Integrated Multi-satellite Retrievals for GPM) early run half-hourly data [5]. This gridded and georeferenced product is characterized by a $0.1° \times 0.1°$ spatial resolution, a spatial coverage between 60° N and 60° S and 4-hour latency [6].

The rainfall forecasting is instead based on NOAA – NCEP GFS (Global Forecast System) deterministic weather prediction model [9], characterized by a spatial resolution of

0.25° × 0.25° and a spatial coverage between 90° N and 90° S. GFS model runs every day at 00, 06, 12 and 18 UTC.

In this study, both datasets were compared with rain gauges measurements in order to evaluate the relative accuracy for a set of different aggregation intervals. Fifty rain gauges located in different climatic zones were taken into account in order to evaluate the accuracy at a global scale (Fig. 1). Rain gauge dataset characterized by a temporal resolution of at least one hour were downloaded from the data providers websites [10, 2]. The biased spatial distribution of the rain gauges is induced by the difficulties encountered in the research of freely accessible well-recognized dataset characterized by a good temporal resolution and an almost zero percentage of missing data.

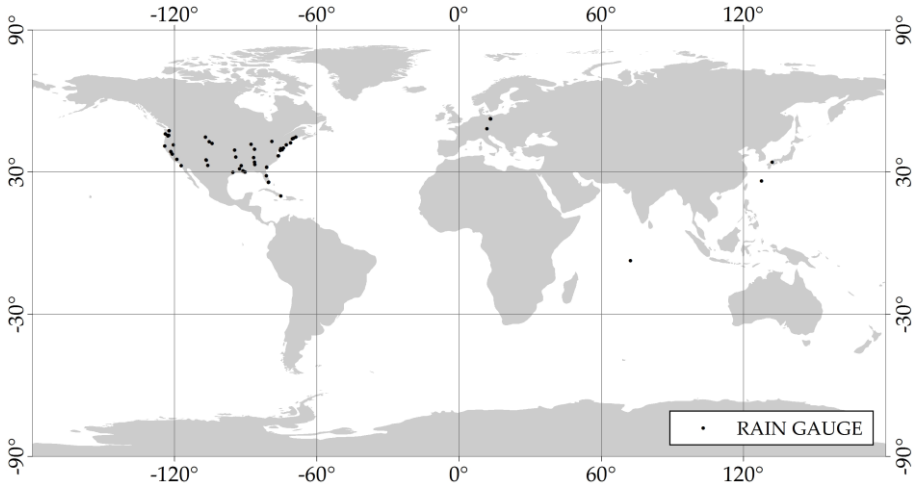

Figure 1. Spatial distribution of the rain gauges used for the evaluation of the accuracy of the input data. The reference system is WGS84

The analysis was performed using statistical performance scores and time series analyses in the period from 15th January 2015 to 30th April 2017. The following aggregation intervals were considered in order to evaluate the cumulated rainfall and the relative intensity: 1, 2, 3, 6, 12, 24, 48, 72, 96, 120 and 144 hours.

Bias and MAE (mean absolute error) were evaluated for both dataset, for each location and for the previously mentioned set of aggregation intervals, as

$$BLAS = \frac{1}{n}\sum_{t=1}^{n}\left[R_{ESTIMATOR}(t) - R_{GAUGE}(t)\right], \quad (1)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|\left[R_{ESTIMATOR}(t) - R_{GAUGE}(t)\right]\right|, \quad (2)$$

where n is the total number of time instants;

t – time;

$R_{ESTIMATOR}$ – the average rainfall intensity measured by GPM Mission or provided by GFS weather prediction model (expressed in mm/hr);

R_{GAUGE} – the average rainfall intensity measured by rain gauge in the same time interval t (expressed in mm/hr).

Both parameters were evaluated both in standard conditions (taking into account only nonzero rainfall measurements) and in heavy-rainfall conditions (considering only rainfall rate greater than 99th percentile of the intensities distribution).

The second part of the analysis consisted of the comparison of the observed and the estimated events. For every location and for every aggregation interval previously mentioned, a contingency table (Tab. 1) was created. In the contingency table, the occurrences of the following four conditions were reported:

- both rain gauge and estimator data are null (case A, correct negatives);
- nonzero rain gauge data and zero estimator data (case B, misses);
- zero rain gauge data and non-null estimator data (case C, false alarms);
- both rain gauge and estimator data are non-null (case D, hits).

Table 1. Example of a contingency table usable to evaluate the quality of the predictions

		RAIN GAUGE		
		= 0 mm/hr	**> 0 mm/hr**	
ESTIMATOR	**= 0 mm/hr**	Correct negatives (A)	Misses (B)	**Estimated non-events**
	> 0 mm/hr	False alarms (C)	Hits (D)	**Estimated events**
		Observed non-events	**Observed events**	

Three indices were derived using these contingency tables as a basis: the false alarm ratio, the probability of detection and the critical success index [11].

The false alarm ratio (FAR) represents the number of false alarms per number of estimated events and is assessable on the basis of elements contained in false alarms and hits cells. The ideal situation is characterized by approximately zero FAR.

$$FAR = \frac{FALSE\ ALARMS}{HITS + FALSE\ ALARMS}. \tag{3}$$

The probability of detection (POD) is instead evaluable by combining elements contained in the misses and hits cells. The ideal situation is characterized by a unitary POD value.

$$POD = \frac{HITS}{HITS + MISSES}. \tag{4}$$

The critical success index (CSI), unlike POD and FAR, combines the characteristics of hits, false alarms and misses and it can be expressed in terms of POD and FAR. The ideal situation is characterized by a unitary CSI.

$$CSI = \frac{HITS}{HITS + FALSEALARMS + MISSES}. \tag{5}$$

The outcomes of this analysis allowed to identify the most appropriate aggregation interval usable to provide information about the rainfall amount and the possible occurrence of disasters induced by heavy rainfall. The results are summarized in several boxplots (Fig. 2). In these figures, the first quantile, the median value of the distribution and the third quantile are identifiable. The mean values are instead summarized in a separate table (Tab. 2).

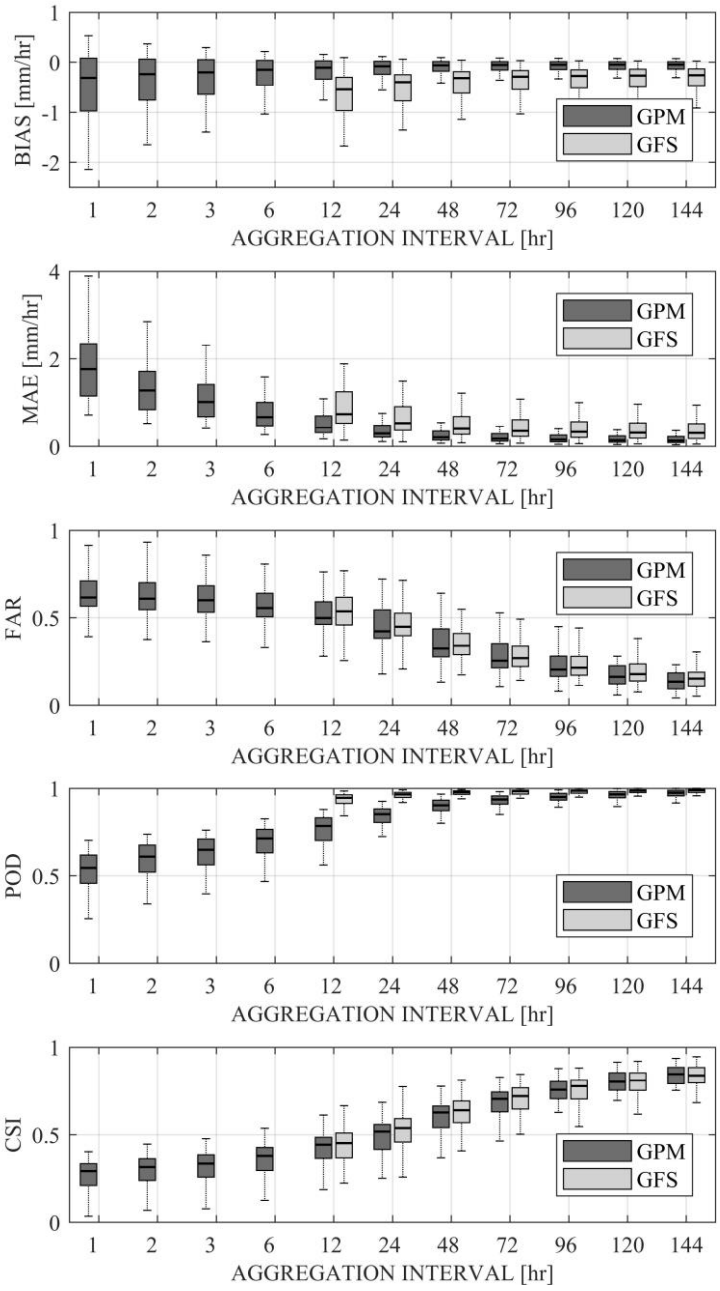

Figure 2. Bias, MAE, FAR, POD and CSI evaluated for nonzero rainfall intensities: the dark grey boxplot refers to GPM IMERG V05B data while the light grey boxplot refers to GFS data

Table 2. Bias, MAE, FAR, POD and CSI calculated taking into account only nonzero rainfall intensities both for GPM IMERG V05B early run and for GFS dataset

		1 hr	2 hr	3 hr	6 hr	12 hr	24 hr	48 hr	72 hr	96 hr	120 hr	144 hr
BIAS	**GPM**	-0.35	-0.28	-0.23	-0.17	-0.13	-0.09	-0.07	-0.06	-0.05	-0.05	-0.05
	GFS	-	-	-	-	-0.59	-0.45	-0.36	-0.32	-0.30	-0.29	-0.23
MAE	**GPM**	1.97	1.43	1.16	0.80	0.55	0.37	0.27	0.22	0.20	0.19	0.18
	GFS	-	-	-	-	0.90	0.67	0.52	0.46	0.43	0.41	0.40
FAR	**GPM**	0.63	0.62	0.60	0.57	0.52	0.45	0.36	0.30	0.25	0.21	0.18
	GFS	-	-	-	-	0.54	0.46	0.36	0.30	0.25	0.21	0.18
POD	**GPM**	0.52	0.58	0.62	0.69	0.76	0.83	0.89	0.92	0.94	0.96	0.96
	GFS	-	-	-	-	0.92	0.95	0.96	0.97	0.97	0.98	0.98
CSI	**GPM**	0.27	0.30	0.32	0.36	0.42	0.49	0.59	0.66	0.71	0.76	0.79
	GFS	-	-	-	-	0.44	0.52	0.62	0.69	0.74	0.77	0.81

From this analysis, a negative value of the bias emerged for both products. As a general rule, therefore, both the satellite-derived data and the rainfall estimation obtained through GFS weather prediction model tend to underestimate rainfall with respect to the rain gauge. For both products, as the aggregation interval increases, the bias tends to have a null value, allowing more accurate information. Aggregation intervals who matched the optimal accuracy for a near real-time application are the one greater or equal to 12 hours. Smaller time intervals have also an unsatisfactory value of false alarm ratio (greater than 0.55). As expected, FAR shows a decreasing trend for both datasets.

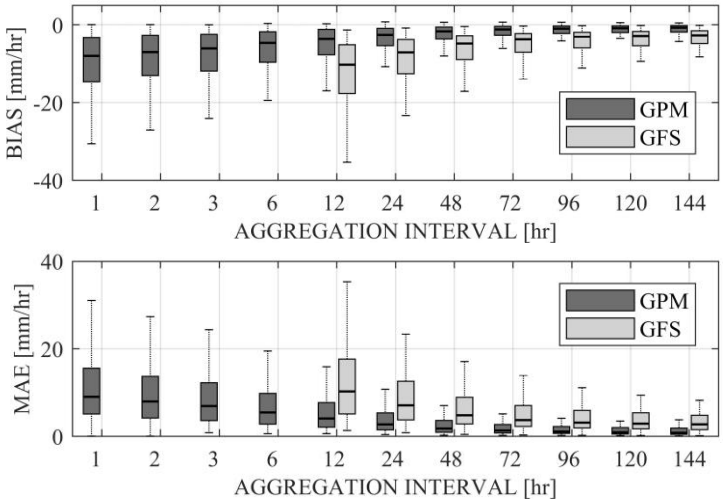

Figure 3. Bias and MAE related to heavy rainfall events. The dark grey boxplot refers to GPM IMERG V05B data while the light grey boxplot refers to GFS data

As far as GPM data concerns, the outcomes demonstrate that a 24-hours aggregation interval ensures a probability of detection greater than 80% and a critical success index equal to 50%. With an aggregation interval of 72 hours, a probability of detection greater than 90%

was reached. Shorter aggregation intervals are characterized by a not acceptable probability of detection. It is, therefore, convenient to set the minimum rainfall aggregation interval to 12 hours to be able to detect events with an acceptable accuracy.

The selected aggregation intervals were assessed taking into account several aspects, such as the requirements of the system, the ideal latency in the provision of information or the final use of the output. As a consequence, also similar aggregation intervals are suitable for these purposes. It is, however, important to highlight the limitations in terms of the accuracy of the outputs obtained using these aggregation intervals.

A further study was conducted taking into account only rainfall rate greater than the 99th percentile of the intensities distribution in order to understand the estimator performances in heavy rainfall detection. The modest underestimation of GPM and GFS data emerged from these results (Fig. 3 and Tab. 3).

Table 3. Bias and MAE related to heavy rainfall events evaluated both for GPM IMERG V05B early run and for GFS data

		1 hr	2 hr	3 hr	6 hr	12 hr	24 hr	48 hr	72 hr	96 hr	120 hr	144 hr
BIAS	**GPM**	-10.59	-9.31	-8.34	-6.54	-4.87	-3.40	-2.29	-1.84	-1.52	-1.33	-1.18
	GFS	-	-	-	-	-12.80	-9.18	-6.37	-5.15	-4.37	-3.88	-3.47
MAE	**GPM**	11.64	10.19	9.09	7.12	5.23	3.64	2.47	2.00	1.67	1.46	1.30
	GFS	-	-	-	-	12.80	9.18	6.38	5.15	4.37	3.88	3.47

3. Extreme Rainfall Detection

The main purpose of this section is to describe the extreme rainfall detection methodology developed in order to provide near real-time alerts with an almost global spatial coverage.

The whole study (described in detail in Mazzoglio et al [8]), was conducted using GPM IMERG half-hourly early run data as input. Results presented in this section refer to this data. This methodology will also be repeated using GFS data as input in order to calibrate the threshold values that will be used for alerts forecasting.

The extreme rainfall detection methodology is based on the concept of activation threshold: an event is identified when the rainfall exceeds a given threshold value. An "event-identification threshold" (EIT) represents the amount of rainfall needed to trigger a flood event induced by extreme rainfall [8]. EITs are used to define near real-time alerts about extreme rainfall. Specifically, an alert is provided if for a selected time interval the accumulated rainfall exceeds the EIT.

The calibration of these thresholds was performed using an empirical approach, analyzing rainfall events that have led to hydrometeorological disasters. These threshold values, obviously, vary over time from place to place. Extreme rainfall conditions in one place are, in fact, very different from the one that characterizes another one. Moreover, it is impossible to define a threshold if the aggregation interval is not defined. A longer time interval has, in fact, a higher EIT.

To develop this extreme rainfall detection methodology, the first step was to search databases of hydrometeorological disasters with a global spatial coverage to be used as truth data. The adopted databases were EM-DAT (The Emergency Events Database) [3], Reliefweb [12] and Floodlist [4].

The methodology consisted of the identification of the optimal EIT for different aggregation intervals (12, 24, 48, 72 and 96 hours) and for every place of the world. The study covered 85 different countries, from 12 March 2014 to 30 April 2017. For every temporal interval and for every country, a series of simulations was performed by varying the threshold values mask.

Threshold masks (different for every aggregation interval) were calibrated using a site-specific parameter: the mean annual precipitation. This total rainfall amount (Fig. 4) was calculated using 10 years of GPCC (Global Precipitation Climatology Centre) "Monitoring product" [13] with a 1° resolution. The "Monitoring product" is a monthly global data which is available about 2 months after the end of the month which it refers to. This product is recommended by GPCC to be used for applications that need high-quality gridded measures of rainfall.

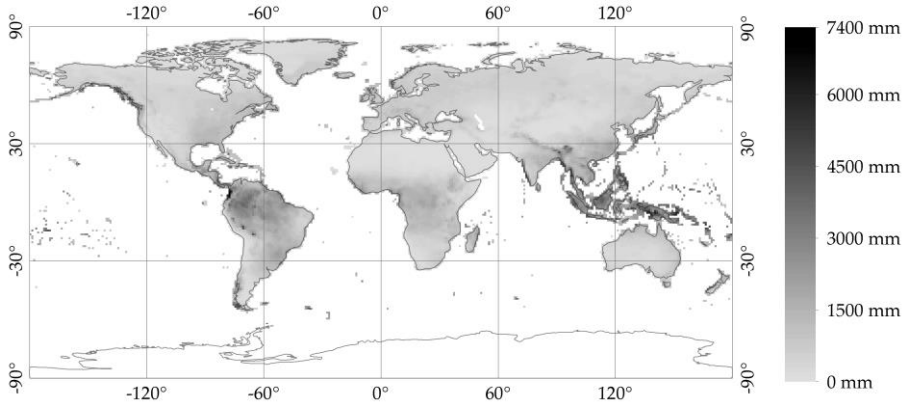

Figure 4. The mean annual rainfall calculated using 10 years of GPCC monthly "Monitoring product". The white areas are places characterized by an absence of measurement. The reference system is WGS84

For each aggregation interval, a series of simulations using this 1° × 1° total rainfall amount was performed. In every simulation, thresholds equal to a percentage of the mean annual rainfall have been adopted. Specifically, the threshold values were calculated using the equation

$$T = T.R. \cdot p_{T.R.} \tag{6}$$

where T represents the threshold;

$T.R.$ represents the total rainfall (i.e. the mean annual rainfall calculated using 10 years of GPCC data);

$p_{T.R.}$ is a parameter representing the fraction of the total rainfall leading to the extreme event identification.

The application of an upper bound and a lower bound proved to be necessary. There are places where the recorded average annual rainfall amount is very low (below 100 mm in a year), which would lead to very low threshold values, comparable with the satellite measurement accuracy, with $p_{T.R.}$ values around 0.1 – 0.2. Analogously, in places where the total annual rainfall is very high (above 2000 mm in a year), the EIT could be unrealistically high because in these places rainfall events tend to occur in the form of low intensity – high frequency events [8].

Several tests were performed. For every aggregation interval, both the criterion of minimization of missed and false alarms and the maximization of the number of identified events were taken into account in order to choose the proper threshold values.

The best results for every aggregation interval were achieved with the parameters summarized in Tab. 4.

Table 4. Threshold values currently adopted in ERDS for the near real-time alerts provision based on NASA GPM IMERG early run data

Aggregation interval (hours)	$p_{T.R.}$ (%)	Lower bound (mm)	Upper bound (mm)
12	6	100	150
24	8	120	210
48	12	140	240
72	15	170	260
96	16	190	280

Before publishing on ERDS website, the alerts produced on areas entirely occupied by sea or ocean are discarded. In order to accomplish this operation, a mask containing, in each cell, the water coverage of the area was used. This mask is freely available on NASA's website [7]. Alerts provided on cells characterized by water coverage equal to 100% were discarded. ERDS, in fact, is a tool developed for the provision of alerts on populated areas.

4. Conclusion

To sum up, the current version of ERDS automatically downloads the most recent GPM IMERG early run data and GFS data and cumulates them according to specific time intervals. More importantly, ERDS generates rainfall alerts based on GPM data where and when the rainfall amount is higher than a specific set of event-identification thresholds.

The current version of ERDS is able to provide alerts using GPM as input data every hour with 4-hour latency and a 0.1° × 0.1° resolution in the latitude range between 60° N and 60° S. The ERDS data are uploaded every hour because GPM IMERG data, despite the 30-minute resolution, is made public in pairs of two.

ERDS is an alert system designed to identify hydrometeorological events. However, there are some types of phenomena that put a strain on this alert system. Specifically, ERDS may fail in the identification of convective storms characterized by a high spatial and temporal variability and discontinuity. They can indeed be transparent to the satellite (rainfall could affect an area smaller than the size of a single cell of GPM data) or their intensity may be underestimated. ERDS may also fail in the provision of a timely alert in the case of intense flash floods affecting very small basins. ERDS, in fact, is characterized by a delay of about 4 hours (due to original GPM data delay plus the time required by the data download, processing and alert evaluation in the ERDS system). If the event is very short, very intense and can cause a flood within 4 hours, the alert will be provided too late. Conversely, the system showed a good performance regarding the identification of hydrometeorological disasters like hurricanes, cyclones, tropical storms, heavy rainfall that might lead to flood events and flash floods characterized by durations greater than the ERDS latency [8].

New improvements are still required to increase the overall accuracy of early warning systems based on remotely sensed rainfall measurements and weather prediction model outputs. The system performances are deeply influenced by the input data resolution. The system is working at the global scale with a spatial resolution of 0.1° × 0.1°. These characteristics could lead to a wrong picture of rainfall events that vary greatly on a small scale and over time. A local-scale validation is advisable. Further studies aimed at developing a temporal/spatial downscaling of the GPM data could help to provide more accurate and reliable outputs. Moreover, one of the major problems with this kind of application is the error that may be present in the rainfall measurement. The system has no control over any underestimation, overestimation or random errors. Furthermore, no remedy is in place with reference to any kind of temporary interruptions in the provision of data.

With this type of application, the model calibration and performance evaluation continue to be challenging problems in large data-scarce regions or in areas where only a few hydrometeorological events were recorded. The currently adopted thresholds may be influenced by this problem.

Acknowledgements

We express our gratitude to Paolo Pasquali, who helped us in the development of the web app.

REFERENCES

1. *Alfieri, L.; Cohen, S.; Galantowicz, J.; Schumann, G.J.P.; Trigg, M.A.; Zsoter, E.; Prudhomme, C.; Kruczkiewicz, A.; Coughlan de Perez, E.; Flamig, Z.; et al*. A Global Network for Operational Flood Risk Reduction. Environ. Sci. Policy 2018, 84, 149–158. doi.org/10.1016/j.envsci.2018.03.014.

2. Deutscher Wetterdienst Archiv der Klimadaten Deutschland – Stundenwerte. Available online: https://rcc.dwd.de/DE/leistungen/klimadatendeutschland/klarchivstunden.html%5Enn=561158 (accessed on 20 April 2019).

3. EM-DAT: The Emergency Events Database – Université Catholique de Louvain (UCL) – CRED, D. Guha-Sapir. Brussels, Belgium. Available online: www.emdat.be (accessed on 20 April 2019).

4. Floodlist. http://floodlist.com (accessed on 20 April 2019).

5. *Huffman, G.J*. GPM IMERG Early Precipitation L3 Half Hourly 0.1 Degree x 0.1 Degree V05; Goddard Earth Sciences Data and Information Services Center (GES DISC): Greenbelt, MD, USA, 2015.

6. *Huffman, G.J.; Bolvin, D.T.; Braithwaite, D.; Hsu, K.; Joyce, R.; Kidd, C.; Nelkin, E.J.; Sorooshian, S.; Tan, J.; Xie, P.* NASA Global PrecipitationMeasurement (GPM) IntegratedMulti-satellitE Retrievals for GPM (IMERG) Algorithm Theoretical Basis Document (ATBD) Version 5.2. 2018. Available online: https://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V5.2.pdf (accessed on 20 April 2019).

7. IMERG Land / Sea mask. Available online: https://pmm.nasa.gov/sites/default/files/downloads/surfrac0.1.PPS.gz (accessed on 20 April 2019).

8. *Mazzoglio, P.; Laio, F.; Balbo, S.; Boccardo, P.; Disabato, F.* Improving an Extreme Rainfall Detection System with GPM IMERG data. Remote Sens. 2019, 11, 677. doi.org/10.3390/rs11060677.

9. National Centers for Environmental Prediction/National Weather Service/NOAA/ U.S. Department of Commerce, 2015: NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, CO. Available online: https://doi.org/10.5065/D65D8PWK (accessed on 20 April 2019).

10. NOAA Local Climatological Data. Available online: www.ncdc.noaa.gov/cdo-web/datatools/lcd (accessed on 20 April 2019).

11. *Nurmi, P*. Recommendations on the Verification of Local Weather Forecasts. ECMWF Technical Memorandum n° 430. 2003. Available online: www.ecmwf.int/en/elibrary/ 11401-recommendations-verification-local-weather-forecasts (accessed on 20 April 2019).

12. Reliefweb. https://reliefweb.int (accessed on 20 April 2019).

13. *Schneider, U.; Becker, A.; Finger, P.; Meyer-Christoffer, A.; Rudolf, B.; Ziese, M.* GPCC Monitoring Product: Near Real-Time Monthly Land-Surface Precipitation from Rain-Gauges based on SYNOP and CLIMAT data. 2011, doi:10.5676/DWD_GPCC/ MP_M_V4_100.