

Article

A Performance Evaluation and Statistical Analysis of IMERG Precipitation Products During Medicane Daniel (September 2023) in the Thessaly Plain, Greece

Evangelos Leivadiotis  and Aris Psilovikos * 

Laboratory of Ecohydraulics & Inland Water Management, Department of Ichthyology & Aquatic Environment, University of Thessaly, Fytokou St., 38446 Nea Ionia, Magnisias, Greece; eleivadiotis@uth.gr

* Correspondence: psiloviko@uth.gr

Abstract

The precise estimation of precipitation is key to understanding and mitigating the effects of extreme weather conditions, especially in areas susceptible to Mediterranean cyclones. This work assesses the performance of the integrated multi-satellite retrievals for GPM (IMERG) precipitation products during the extreme Mediterranean cyclone “Medicane Daniel” that affected the Thessaly Plain in Central Greece in early September 2023. Three IMERG versions (final run (FR), early run (ER), and late run (LR)) were inter-compared with gauge-based interpolated rainfall estimates using inverse distance weighting (IDW) and ordinary kriging techniques. Pixel-wise and categorical verification metrics, such as the probability of detection (POD), false alarm ratio (FAR), critical success index (CSI), and Peirce skill score (PSS), were calculated for rainfall thresholds between 50 mm and 400 mm. It was found that the IMERG final run agreed most with the ground observations, with a correlation coefficient (R) of 0.87, RMSE of 138.8 mm, and CSI up to 0.995 at the 100 mm threshold when the IDW interpolation was used. Kriging produced slightly better spatial accuracy overall, as indicated by a lower RMSE (14.5 mm) and higher correlation ($R = 0.99$). The results indicate the benefit of combining satellite precipitation data with ground-based observations through spatial interpolation for the enhanced monitoring of extreme weather events over complex terrain. Kriging is suggested when greater spatial reliability is needed, while IMERG-FR is found to be a reliable satellite product for quick response analysis during heavy precipitation events. The study emphasizes the importance of blending satellite precipitation estimates and ground observations via spatial interpolation methods, i.e., kriging and IDW, allowing for a more localized and precise validation of intense weather events.

Keywords: IMERG; Daniel; extreme; precipitation; satellite; performance; metrics



Academic Editors: Paul Kucera, Haiyang Ma and Yaosheng Wang

Received: 18 June 2025

Revised: 4 August 2025

Accepted: 12 August 2025

Published: 14 August 2025

Citation: Leivadiotis, E.; Psilovikos, A. A Performance Evaluation and Statistical Analysis of IMERG Precipitation Products During Medicane Daniel (September 2023) in the Thessaly Plain, Greece. *Water* **2025**, *17*, 2401. <https://doi.org/10.3390/w17162401>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Mediterranean basin is a climate change hotspot due to its intricate geographical, climatic, and topographical features and its position as the convergence point of tropical and mid-latitude systems. This has resulted in the increasing prevalence of severe weather phenomena, often referred to as “Mediterranean Cyclones” or “Medicanes”, which share similarities with tropical cyclones but form in the Mediterranean Sea [1–3]. These Medicanes, which are characterized by intense rainfall, high winds, flash floods, erosion, and sediment deposition phenomena, have significant impacts [1] on the populations, ecosystems, and economies of regions along the Mediterranean coastline, such as Greece [4].

Greece's recent history in hydrometeorological hazards is highly connected to these events [5]. Especially, between 2016 and 2023, five powerful medicanes hit Greece, with Medicane Daniel (September 2023) being the most destructive. Between 3 and 7 September, extreme rainfall ranging from 305 mm to 1096 mm led to widespread destruction of infrastructure, agricultural areas, tourism facilities, and buildings [2,6–8]. Understanding the influence of climate change on extreme phenomena like “Daniel” requires crucial and thorough analysis of precipitation patterns, which is very important and will lead to a path of preparedness and resilience for future extreme weather events. To accurately measure and analyze precipitation, especially in regions susceptible to extreme weather events like the Mediterranean, it is critical to employ advanced remote sensing techniques, which offer a high resolution and near real-time capabilities [9–12].

Accurate precipitation measurement is essential for weather forecasting, hydrological modeling, and flood risk assessment [5,12]. Ground-based rain gauges provide precise, point-specific data but lack comprehensive spatial coverage, particularly in complex terrains [13]. Weather radars offer broader observational capabilities, capturing storm structures and precipitation patterns in real time [14,15]. However, their effectiveness is often limited by network coverage, operational constraints, and terrain-induced measurement distortions. In many regions, including Greece, outdated radar infrastructure and sparse gauge networks hinder the accurate monitoring of extreme events, such as Medicane Daniel [14,16].

Satellite-based precipitation estimates help to address these limitations by offering continuous, large-scale observations [12,17]. However, their accuracy varies, especially in mountainous and coastal areas where localized rainfall is difficult to detect. Advancements in satellite technology, including higher spatial and temporal resolutions, along with sophisticated algorithms such as artificial neural networks (ANNs), have significantly improved rainfall estimations. Nonetheless, challenges persist due to differences in data resolution and uncertainties in precipitation modeling [18–20].

Numerous satellite products have been utilized for regional, national, and global studies about hydrological simulation, drought management, and extreme precipitation [21,22]. Among the evaluation and comparison of various products, IMERG (integrated multi-satellite retrievals for GPM), the GPM (global precipitation measurements) mission's widely used products, presents excellent precipitation detection performance [23]. It is a gridded rainfall product with a spatiotemporal resolution of $0.1^\circ \times 0.1^\circ$ every 30 min between 60° N and 60° S. Data acquisition can almost be achieved in real time, with only a 4 h latency between capture and acquisition and 3.5 months latency for the availability of post-real-time research data. Due to this versatility, IMERG products have received great interest from researchers in recent years [24–26].

Prior studies have highlighted both the strengths and limitations of satellite rainfall products in capturing heavy precipitation [8,12,27–30]. For instance, research on Medicane Daniel by Kolios [27] demonstrated that while GSMaP offers realistic cumulative precipitation, it tends to overestimate in certain areas. Similarly, Katsanos et al. [28] found that satellite-derived products often underestimate precipitation, particularly in extreme events, due to challenges in detecting cloud convection and peak rainfall intensities. These findings are in line with more general criticisms of IMERG performance in the detection of extreme rainfall, as presented by Yu [29] and Sakib [30], who identified discrepancies between real-time products (IMERG-ER and IMERG-LR) and enhanced but not flawless accuracy in the final-run (IMERG-FR).

The intense Medicane Daniel in 2023 was marked by its strong landfall intensity and widespread impact. The objective of this study was to evaluate how effectively the V06 of IMERG-ER, IMERG-LR, and IMERG-FR capture the heavy precipitation caused by Daniel.

Utilizing rain gauge data as a reference for observed precipitation, IMERG products from various storm phases affected regions of the storm. The evaluation focused on precipitation retrieval characteristics, verification through assessment indices, rainfall process variations, and extreme precipitation statistics. The study investigated Daniel's precipitation dynamics and assessed the effectiveness of IMERG products in monitoring storm-related rainfall.

2. Materials and Methods

2.1. Study Area

The case study area is the Pinios River basin (GR16), which is the largest drainage basin in the Thessaly Water District (GR08), with an area of 10,700 km² (Figure 1). The average annual precipitation is 779 mm, and the mean annual runoff is 3499×10^6 m³/y (327 mm) or 110,95 m³/s. Geographically, it lies within latitudes 39°11'–39°58' N and longitudes 21°52'–22°45' E (Figure 1), bordered by extensive plains and mountainous regions, including Mount Olympus (2918 m), Greece's highest point. The Thessaly Plain in the center is one of Greece's most fertile agricultural regions, consisting mainly of flat land covered by alluvial deposits [31]. Agriculture occupies 49.5% of the land cover in the Thessaly Water District, and the plain accounts for 14.2% of Greece's principal agricultural output (Ministry of Environment and Energy, Special Secretariat for Water, www.ypeka.gr). The Pinios River rises in the Pindus Mountains, Greece's largest mountain range, and outflows into the Aegean Sea. The basin's geomorphological structure, defined by a round catchment area and low-lying topography, renders it highly susceptible to flooding (Figure 1). Historically significant floods have been documented, such as those of 4 June 1907, 27 October 1980, 23 March 1987, and 22 October 1994, as per the EU Floods Directive (2007/60/EC), the preliminary flood risk assessment by Greece's Ministry of Environment and Energy. Additionally, 31.7% of the Thessaly Water District is classified as an area with high flood risk.

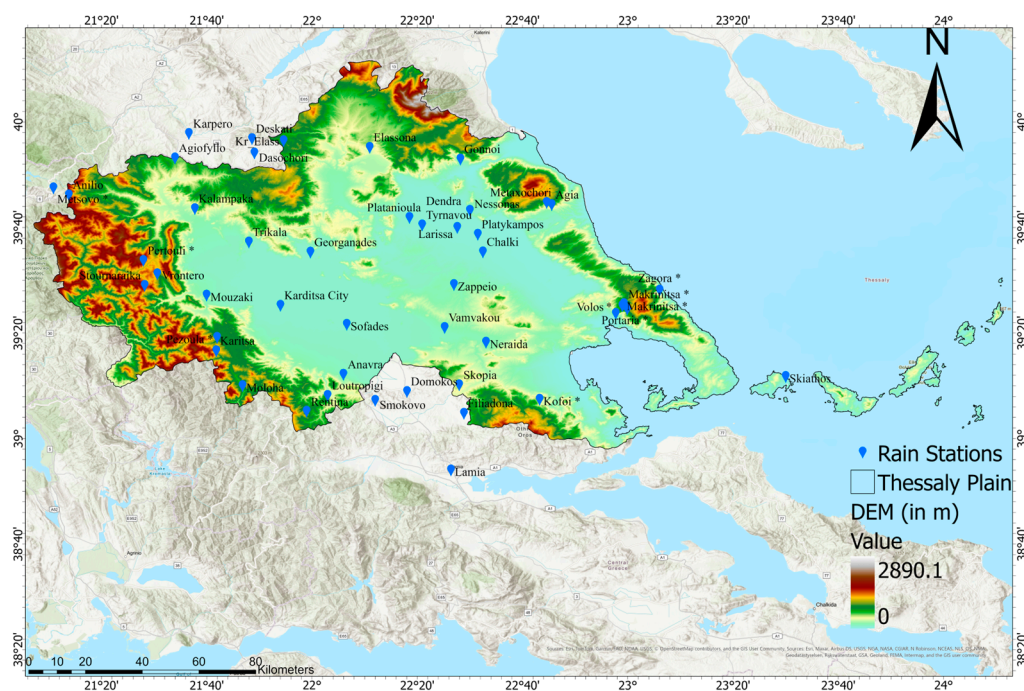


Figure 1. The study area of the Thessaly plain and the distribution of rain stations.

2.2. IMERG Precipitation Data

The IMERG precipitation product is derived from the GPM constellation, integrating data from passive microwave sensors on low-Earth orbit satellites and infrared sensors

on geostationary satellites. It is further refined using high-resolution spatial and temporal precipitation observations. The latest IMERG version, V06, was released in 2019, introducing improvements over V05 in data processing, algorithm optimization, and verification. The inclusion of the GPM microwave imager (GMI) and dual-frequency precipitation radar (DPR) has enhanced its ability to detect both liquid and solid precipitation, significantly improving accuracy [32]. IMERG provides three distinct products, accessible at (<https://gpm.nasa.gov/data-access/downloads/gpm>) access on 1 March 2025. IMERG-ER and IMERG-LR are quasi-real-time datasets with release latencies of 4 h and 12 h, respectively, while IMERG-FR is a research-grade product available with a 3.5-month delay. All IMERG products have a spatial resolution of 0.1° and offer multiple temporal resolutions. IMERG-ER is available in 30-min, 3-h, daily, and weekly resolutions, while IMERG-LR includes an additional monthly option. IMERG-FR provides 30-min, daily, and monthly resolutions. These products cover global latitudes from 60° N to 60° S. IMERG-ER employs forward propagation, whereas IMERG-LR incorporates both forward and backwards propagation to refine accuracy [33]. IMERG-FR, the most comprehensive version, integrates a broader dataset and is calibrated using the Global Precipitation Climate Center (GPCC) monthly precipitation analysis. This progressive enhancement in data processing from IMERG-ER to IMERG-FR results in improved accuracy, though with increasing data latency.

2.3. Rain Gauge Data

The daily rain gauges were selected as the observation data. These data were derived from the thorough analysis of Dimitriou et al. [34]. Overall, 47 meteorological stations were considered, and their characteristics and rainfall depths during the four-day period (from 4 to 7 September 2023) are presented in Table 1.

Table 1. Characteristics of the rain stations.

Station	Z (m)	Owner	Longitude	Latitude	4 Sep	5 Sep	6 Sep	7 Sep	Total
Agia	200	NOA	22.76	39.72	11.2	218.6	15.4	90.6	335.8
Agiofylo	584.1	MinEnv	21.57	39.87	60	40	54	-	154.5
Anavra	196.3	MinEnv	22.10	39.18	98	131	218	0	447
Anilio	1660	NOA	21.23	39.75	20.6	148.2	374.6	149.2	692.6
Chalki	75	NOA	22.54	39.57	25.2	180.8	68.8	30.4	305.2
Dasochori	737	NOA	21.82	39.88	13.8	98.6	46.4	23	181.8
Dendra Tyrnovou	75	NOA	22.35	39.65	19.2	86.4	64.8	22.8	193.2
Deskati	769	NOA	21.81	39.93	10.4	96	50.6	25.6	182.6
Domokos	570	NOA	22.30	39.13	40.4	110	225.2	6.4	382
Elassona	282	NOA	22.18	39.90	26.4	72.2	19.6	40.8	159
Filiadona	487	WU	22.48	39.06	23.1	72.5	125.5	9.1	230.2
Georganades	92	HCMR	22.00	39.57	24.8	116.3	195.8	28.9	365.8
Gonnoi	111	NOA	22.47	39.86	14	147.4	31.4	71.2	264
Kalampaka	238	NOA	21.63	39.71	10.8	94.2	165.8	85.2	356
Karditsa City	121	NOA	21.90	39.40	42.4	185.2	404.4	26.8	658.8
Karitsa	1074.3	MinEnv	21.70	39.26	90	110	110	300	610
Kofoi	500	NOA	22.72	39.10	23.4	152.6	342.2	32.2	550.4
Loutropigi	722.1	MinEnv	22.05	39.12	0	98.5	86.5	99.6	284.6
Makrinita	850	NOA	22.99	39.41	125.2	757.4	273.6	79.2	1235.4
Makrinita	685.4	MinEnv	22.98	39.40	75	82	-	38.4	-
Metaxochori	340	WU	22.74	39.72	14	170.7	14.7	60.7	260.1
Metsovo	1240	NOA	21.18	39.77	20.2	91.8	204	75.8	391.8
Mouzaki	175	NOA	21.67	39.43	23.8	163.8	321.8	89	598.4
Neraida	243	NOA	22.55	39.28	19.6	226.6	91.2	23.8	361.2
Nessonas	92	NOA	22.50	39.70	11.6	78	3.6	71.6	164.8

Table 1. *Cont.*

Station	Z (m)	Owner	Longitude	Latitude	4 Sep	5 Sep	6 Sep	7 Sep	Total
Pertouli	1170	WU	21.47	39.54	0.6	58.6	415.2	165.8	640.2
Pezoula	891	NOA	21.70	39.30	43.2	250	378.4	90.8	762.4
Platanioula	83	NOA	22.31	39.68	18.6	75.4	91.2	25.4	210.6
Platykampos	72	NOA	22.52	39.62	20.8	107.6	0.4	76.4	205.2
Portaria	600	NOA	23.00	39.39	105.4	764.7	14.4	0	884.5
Rentina	884.9	MinEnv	21.98	39.07	0	35	120	135	290
Skopia	444.7	MinEnv	22.47	39.15	107	46.8	60	27.5	241.3
Smokovo	444	NOA	22.20	39.10	40.2	97	89.4	14	240.6
Trikala	163	NOA	21.80	39.60	17.6	116.6	256.8	86.4	477.4
Vamvakou	148	NOA	22.42	39.33	19.4	191.4	129.8	29	369.6
Volos	52	NOA	22.96	39.37	35.2	450.8	121	10.4	617.4
Zagora	505	NOA	23.10	39.45	134.6	759.6	3.8	197.6	1095.6
Zappeio	172.3	MinEnv	22.45	39.47	89.9	274	139	12.8	515.7
Anchialos			23.50	39.18	6	146.4	155.1	43.4	350.9
Larissa			22.46	39.65	15	101.3	24.6	37.4	172.5
Lamia			22.44	38.88	32.1	12.9	92.4	29	165.4
Kr_Elass			21.91	39.92	11.7	69.7	27.6	30.8	136.3
Skiathos			22.11	39.34	9.8	198.9	34.4	27.2	270.3
Karpero			21.61	39.94	0	38.8	51.5	33.4	123.7
Vrontero			21.51	39.50	0	89	205	113	575
Moloha			21.78	39.15	0	110	110	197	534.7
Stournaraika			21.47	39.46	0	89	205	113	575

2.4. Performance Indices

In this study, the IMERG products were evaluated using two types of statistical comparative metrics to assess their performance (Table 2), using the rain gauge stations as a reference. The simple statistical metrics comprised Pearson’s correlation coefficient (CC), BIAS (BIAS), and root mean squared error (RMSE). The statistical metrics were calculated to quantify the agreement of the data with the precipitation records obtained from rain gauges. To analyze the probability and accuracy of precipitation detection by IMERG products with respect to reference radar data, various probabilistic statistical indices were calculated. They include the probability of detection (POD), false alarm ratio (FAR), critical success index (CSI), and Peirce skill score (PSS). These metrics are utilized extensively to evaluate the precision of satellite-based precipitation estimates [29,30,35]. POD gauges the capability of IMERG satellites to detect rainfall events when precipitation is observed by radar. Conversely, FAR measures instances where IMERG detects rainfall when there is no radar-observed precipitation. The CSI measure accumulates instances of rainfall observed by both satellite and radar platforms and thus offers a comprehensive evaluation of the concordance of the two data sources. Such a measure offers valuable insights into the reliability of precipitation reporting under varying spatiotemporal scales [15]. The PSS measure, on the other hand, defines the overall accuracy of the system in detecting rainfall events as the difference between the probability of correct detection and the probability of false detection. Furthermore, the Kling–Gupta efficiency (KGE) coefficient, a powerful statistical metric, was employed to evaluate the model’s performance. Originally proposed by Gupta et al. [36] as a better alternative to conventional metrics such as mean squared error (MSE) and Nash–Sutcliffe efficiency (NSE) for hydrologic modeling, KGE takes into account three fundamental aspects: correlation, bias, and variability.

Table 2. Formulas of statistical and performance indices used for the evaluation of satellite products.

Index	Formula	Range
CC	$\frac{\sum (P_i - \bar{P})(G_i - \bar{G})}{\sqrt{\sum (P_i - \bar{P})^2} \sqrt{\sum (G_i - \bar{G})^2}}$	(−1)–(+1)
RMSE	$\sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - G_i)^2}$	[0–+∞)
BIAS	$\frac{\sum (P_i - G_i)}{\sum G_i} - 1$	−∞/+∞
KGE	$\frac{1 - \sqrt{[s_r \times (r - 1)]^2 + [s_\beta \times (\beta - 1)]^2 + [s_\gamma \times (\gamma - 1)]^2}}{1}$	(−∞)–(+1)
POD	$\frac{H}{H+M}$	0–1
FAR	$\frac{F}{H+F}$	0–1
CSI	$\frac{H}{H+M+F}$	0–1
PSS	POD–FAR	(−1)–(+1)

2.5. Interpolation Methods

A handful of studies have used interpolation techniques and statistical parameters and indexes to evaluate the performance of hydrometeorological parameters [37–40]. To compare the IMERG products with our rain-gauge data, we followed the process below: firstly, we interpolated the ground truth data using the IDW and ordinary kriging methods for the total study period and also for the daily level. Additionally, we resampled our IMERG products according to the new IDW and kriging raster’s resolution in order to have as accurate a comparison as possible. Finally, we evaluated the interpolation methods and IMERG product with the aforementioned methodology.

In Table 2, P represents the IMERG, and G represents the rain gauge. H is the frequency of accurate precipitation forecasts, F is the frequency of false precipitation reports, and M is the frequency of missed precipitation reports.

3. Results

The findings are presented in Tables 3–7 with the outcomes of statistical metrics among the three IMERG products from the total to the daily level. The performance assessment of inverse distance weighting (IDW) and kriging interpolation techniques across various IMERG datasets (final run, early run, and late run) demonstrates variations in their respective efficacies (Figures 2–6). Table 3 shows the analysis of IMERG products through IDW and kriging methodologies, indicating that IMERG-ER demonstrates better performance relative to the other products, whereas IMERG-LR exhibits the least favorable performance. IMERG-ER presents the highest Kling-Gupta efficiency (KGE) values (0.72 for IDW and 0.70 for kriging), denoting the best agreement with reference data. Moreover, it displays a moderate root mean square error (RMSE) (151.46 for IDW and 152.96 for kriging) and a comparatively minor bias (−31.02 for IDW and −30.36 for kriging), implying a marginal underestimation. Meanwhile, IMERG-FR ranks as the second most effective performer product, showing lower RMSE values (138.87 for IDW and 133.11 for kriging) alongside the highest correlation coefficients (0.87 for IDW and 0.89 for kriging), thereby indicating a robust association with the observed data. However, it is noteworthy that its bias is considerably positive (54.04 for IDW and 54.25 for kriging), indicating a slight overestimation of precipitation. In contrast, IMERG-LR demonstrates the least optimal performance, characterized by the highest RMSE values (215.34 for IDW and 217.06 for kriging) and the most noticeable negative bias (−91.2 for IDW and −90.3 for kriging),

thereby illustrating a substantial underestimation of rainfall. Furthermore, its KGE values (0.48 for IDW and 0.47 for kriging) are the lowest among the evaluated products, rendering it the least reliable option. In conclusion, IMERG-ER is identified as the most effective product due to its elevated KGE and moderate error rates. IMERG-FR closely follows with a strong correlation while exhibiting increased bias, and IMERG-LR is deemed the least accurate owing to considerable underestimation and elevated RMSE values.

Table 3. The areal average of statistical indices for the total rainfall and different interpolation methods.

Method	IMERG-FR		IMERG-ER		IMERG-LR	
	IDW	Kriging	IDW	Kriging	IDW	Kriging
Index						
CC	0.87	0.89	0.85	0.85	0.86	0.86
RMSE	138.87	133.11	151.46	152.96	215.34	217.06
BIAS	54.04	54.25	−31.02	−30.36	−91.2	−90.3
KGE	0.18	0.19	0.72	0.70	0.48	0.47

Table 4. The areal average statistics for total rainfall on 4 September for different interpolation methods.

4 September						
Metric	IMERG-FR	IDW	IMERG-LR	Kriging	IMERG-ER	IMERG-LR
		IMERG-ER		IMERG-FR		
RMSE	15.47	19.78	39.73	15.61	18.88	40.78
Bias	−4.03	−8.58	−21.12	−4.7	−8.76	−21.5
CC	0.80	0.83	0.84	0.85	0.90	0.89
KGE	0.61	0.42	0.11	0.53	0.38	0.08

Table 5. The areal average statistics for total rainfall on 5 September for between different interpolation methods.

5 September						
Metric	IMERG-FR	IDW	IMERG-LR	Kriging	IMERG-ER	IMERG-LR
		IMERG-ER		IMERG-FR		
RMSE	54.06	72.17	88.25	58.50	79.61	93.83
Bias	14.02	−17.6	−41.43	−12.92	−17.09	−40.46
CC	0.90	0.86	0.92	0.88	0.83	0.88
KGE	0.47	0.67	0.47	0.51	0.67	0.48

Table 6. The areal average statistics for total rainfall on 6 September for different interpolation methods.

6 September						
Metric	IMERG-FR	IDW	IMERG-LR	Kriging	IMERG-ER	IMERG-LR
		IMERG-ER		IMERG-FR		
RMSE	61.28	57.83	72.05	55.38	47.96	71.08
Bias	25.52	−2.31	−23.19	25.03	−3.93	−23.04
CC	0.87	0.78	0.80	0.89	0.84	0.83
KGE	−0.55	0.77	0.56	−0.42	0.81	0.53

Table 7. The areal average statistics for total rainfall on 7 September for different interpolation methods.

7 September						
Metric	IMERG-FR	IDW IMERG-ER	IMERG-LR	Kriging IMERG-FR	IMERG-ER	IMERG-LR
RMSE	38.57	46.62	43.25	38.74	46.18	43.20
Bias	13.05	1.06	0.87	13.14	1.23	0.95
CC	0.40	0.38	0.43	0.39	0.39	0.44
KGE	−0.59	0.37	0.43	−0.54	0.38	0.43

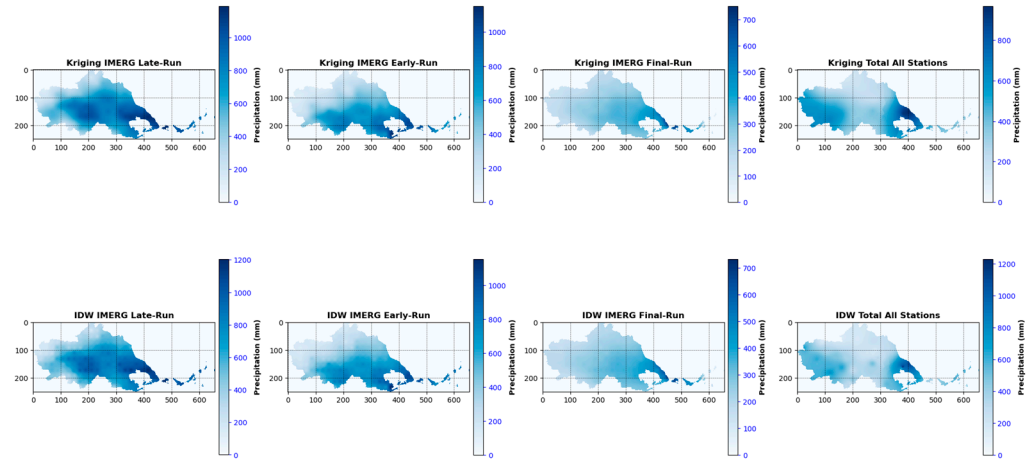


Figure 2. Spatial distribution of total rainfall (in mm), as estimated by IMERG products (final, early, and late runs) and rain gauge interpolation (IDW and kriging). The x- and y-axes represent pixel grid coordinates of the resampled raster (not geographic coordinates). Rainfall intensity is color-coded.

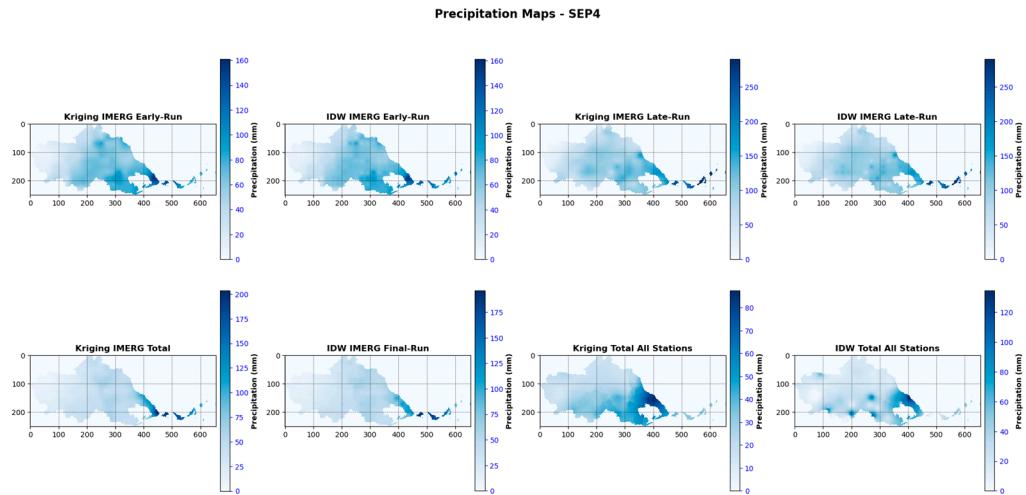


Figure 3. Spatial distribution of total rainfall (in mm) on 4 September 2023, as estimated by IMERG products (final, early, and late runs) and rain gauge interpolation (IDW and kriging). The x- and y-axes represent pixel grid coordinates of the resampled raster (not geographic coordinates). Rainfall intensity is color-coded.

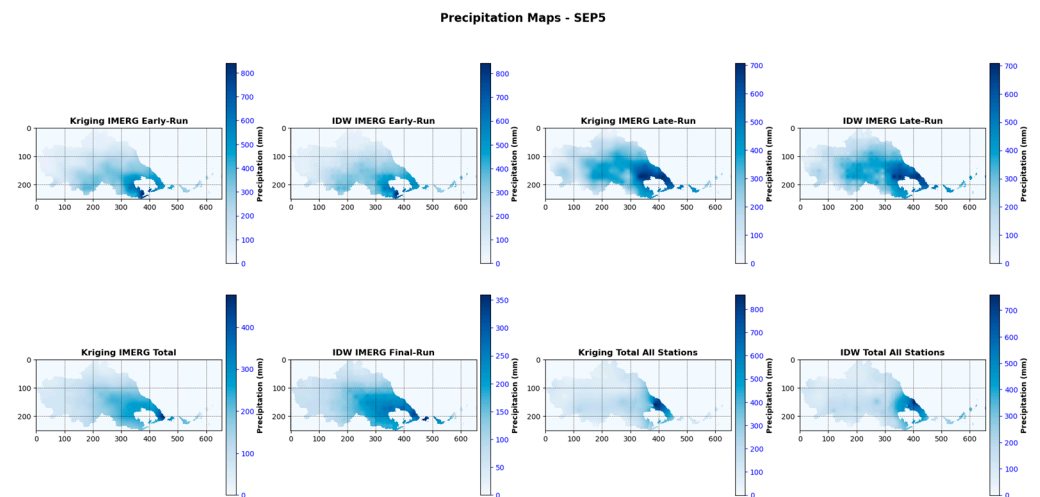


Figure 4. Spatial distribution of total rainfall (in mm) on 5 September 2023, as estimated by IMERG products (final, early, and late runs) and rain gauge interpolation (IDW and kriging). The x- and y-axes represent pixel grid coordinates of the resampled raster (not geographic coordinates). Rainfall intensity is color-coded.

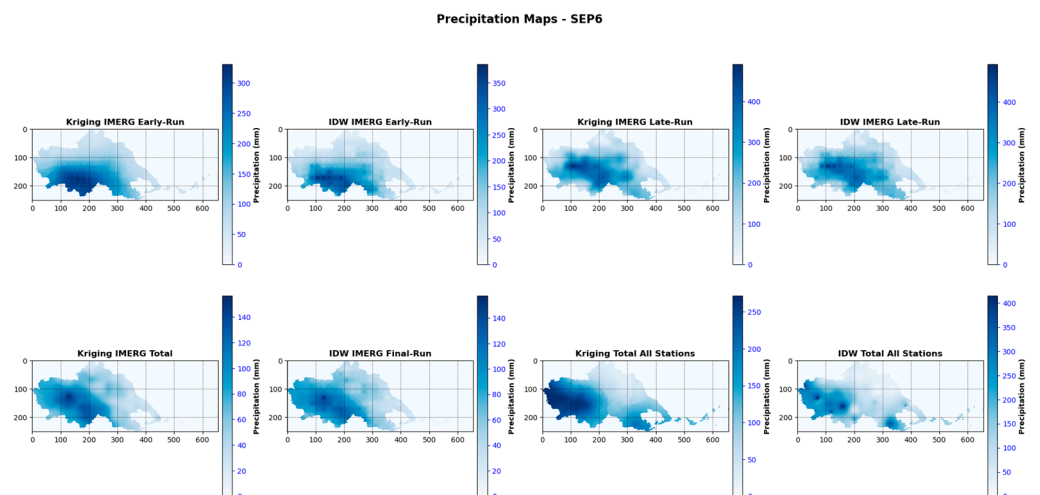


Figure 5. Spatial distribution of total rainfall (in mm) on 6 September 2023, as estimated by IMERG products (final, early, and late runs) and rain gauge interpolation (IDW and kriging). The x- and y-axes represent pixel grid coordinates of the resampled raster (not geographic coordinates). Rainfall intensity is color-coded.

Daily performance (the same for all) on 4 September is presented in Table 4. IMERG-FR, among the three IMERG products examined, has the least root mean square error (RMSE) of 15.47 for inverse distance weighting (IDW) and 15.61 for kriging. Additionally, low values of bias (−4.03 for IDW and −4.7 for kriging) show no overestimation or underestimation of precipitation to a significant extent. Moderate values are shown for the correlation coefficient (CC) (0.80 for IDW and 0.85 for kriging) and Kling–Gupta efficiency (KGE) (0.61 for IDW and 0.53 for kriging), reflecting the best overall agreement with reference data. IMERG-ER’s overall performance comes second, with a slightly higher RMSE (19.78 for IDW and 18.88 for kriging) and a stronger bias (−8.58 for IDW and −8.76 for kriging), which points to underestimation. However, its correlation coefficient is the highest of the three products (0.83–0.90), which reflects that it is closely aligned with reference data. Nonetheless, its KGE values are lower (0.42 for IDW and 0.38 for kriging), which reflects that its overall performance is less reliable compared to IMERG-FR. IMERG-LR presents the least reliable performance, with the highest RMSE values (39.73 for IDW and 40.78 for

kriging) and the largest bias values (-21.12 for IDW and -21.5 for kriging), among the IMERG products, reflecting a large underestimation of precipitation. Despite its relatively high correlation (0.84 for IDW and 0.89 for kriging), its KGE is the lowest (0.11 for IDW and 0.08 for kriging), indicating the least favorable product.

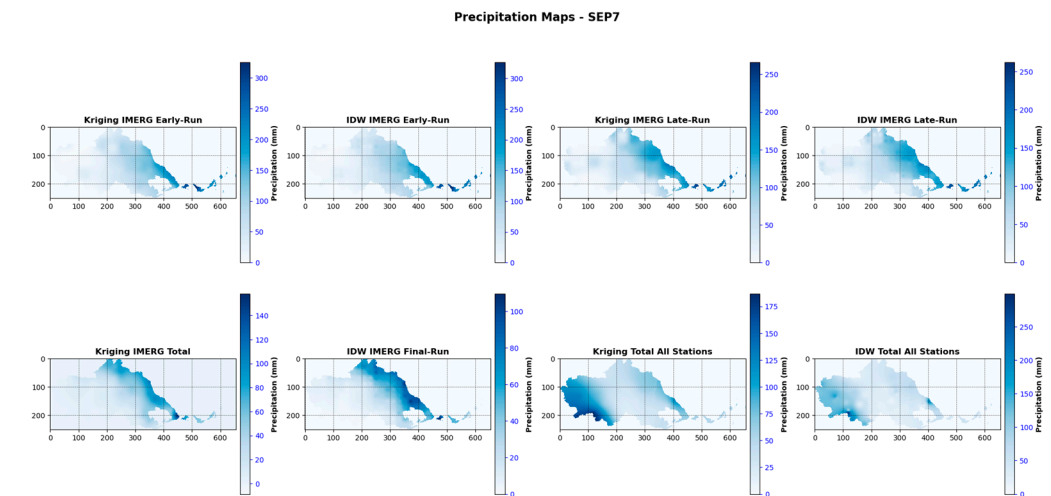


Figure 6. Spatial distribution of total rainfall (in mm) on 7 September 2023, as estimated by IMERG products (final, early, and late runs) and rain gauge interpolation (IDW and kriging). The x- and y-axes represent pixel grid coordinates of the resampled raster (not geographic coordinates). Rainfall intensity is color-coded.

For 5 September, from Table 5 it can be noticed that IMERG-FR outperforms (the same) the other products in most of the metrics regardless of the interpolation method applied. The highest correlation coefficient (CC) values of 0.90 and 0.88 for IDW and kriging, respectively, indicate a very good agreement with the rain station data. Additionally, IMERG-FR has lower RMSE than IMERG-ER and IMERG-LR, providing smaller rain estimation errors. Yet, IMERG-LR achieves the highest CC (0.92) with IDW but undermines the use of kriging (0.88), suggesting that method choice affects accuracy in lower-resolution data. In terms of bias, both interpolation methods suggest that IMERG-FR has the lowest absolute bias, whereas IMERG-LR exhibits severe underestimation (-41.43 for IDW and -40.46 for kriging). This underestimation is confirmed by its high RMSE values, assuring that IMERG-LR cannot strongly represent rainfall intensities. The KGE values also highlight the strengths of IMERG-ER, with the highest (0.67) for both IDW and kriging, which signifies that it provides a fair trade-off between correlation, bias, and variability.

The results for 6 September have the same distinct differences between the IMERG products and interpolation methods. IMERG-ER consistently shows superior performance compared to the other products with the highest KGE values of 0.77 for inverse distance weighting (IDW) and 0.81 for kriging, revealing a closer approximation with the reference data. It also has the least amount of bias, with values of -2.31 for IDW and -3.93 for kriging, showing little systematic overestimation or underestimation. IMERG-FR has a high CC (0.87) for IDW and 0.89 for kriging, but is marked by high bias (25.52 for IDW and 25.03 for kriging), indicating a tendency to overestimate precipitation consistently. Its negative (KGE) values (-0.55 for IDW and -0.42 for kriging) further reinforce this issue, showing that despite the high correlation, the overall reliability of the model is compromised by high bias and variability. IMERG-LR presents moderate correlation (CC ranging from 0.80 to 0.83) but with the largest RMSE (72.05 for IDW and 71.08 for kriging) and a notable negative bias (-23.19 for IDW and -23.04 for kriging). This implies that it has a tendency to consistently underestimate rainfall and is the least dependable product.

Results for 7 September indicate below-average (the same) performance by all IMERG products with low CC values and moderate levels of error. Among the three IMERG products, IMERG-LR performs best overall, with the highest CC (0.43) for IDW and (0.44) for kriging, as well as optimum KGE 0.43 for both interpolation methods. This indicates that IMERG-LR has a generally consistent relationship with the reference data, even when its performance is low. IMERG-ER closely tracks in terms of correlation (CC = 0.38–0.39), as well as values in terms of KGE (0.37 for IDW and 0.38 for kriging). Its highest bias, however, is 1.06 for IDW and 1.23 for kriging, meaning that it neither consistently overestimates nor underestimates rain. IMERG-FR, however, does the poorest, with the KGE values (−0.59 for IDW and −0.54 for kriging), reflecting poor correlation with the reference data. While it is improved in terms of RMSE than IMERG-ER (38.57 for IDW and 38.74 for kriging), it is plagued by high bias (13.05 for IDW and 13.14 for kriging), reflecting systematic overestimation of rainfall.

3.1. Categorical Indices

3.1.1. Overall Performance Evaluation of IMERG Products for Total Rainfall

Results derived from the three experiment runs (early, late, and final) offer valuable information on the capacity of IMERG in depicting rainfall intensities across a set of thresholds (50–400 mm) based on inverse distance weighting (IDW) and kriging interpolation techniques. Since the study area experienced different amounts of rainfall, a variety of thresholds (50–400 mm) have been selected (Figures 2–6). Various performance measures, such as the probability of detection (POD), false alarm ratio (FAR), critical success index (CSI), and Peirce skill score (PSS), were evaluated to determine the validity of IMERG for the detection of extreme precipitation events. Overall, at lower thresholds (50–100 mm), both IDW and kriging have nearly perfect detection ($POD \approx 1$) with few false alarms, reflecting that IMERG is successful in capturing moderate rain events (Figure 7). Beyond a threshold increase above 200 mm, however, performance declines with a notable decline in POD and an increase in FAR, especially for IDW. At the 400 mm threshold, POD drops to as low as 0.15–0.18, whereas FAR remains relatively stable (~ 0.05 – 0.07 in Final Run, but increases up to 0.53 in the previous runs). This indicates that IMERG struggles to accurately portray extreme precipitation events with increasing uncertainty at very high thresholds. As for interpolation technique comparison, kriging has a consistently slightly better POD and lower FAR at extreme thresholds (300–400 mm) than IDW, which implies a slight advantage in capturing heavy rainfall events. Both techniques, however, display a decrease in CSI and PSS at higher intensities, highlighting the intrinsic limitation of IMERG in resolving extreme precipitation with great accuracy. The observed temporal variations within runs further suggest that additional calibration or bias adjustment is required when using IMERG products for extreme rain analysis. Lastly, between the different interpolation methods, kriging shows a slightly better ability to capture extreme events; nonetheless, both procedures indicate that IMERG underestimates high-density rainfall.

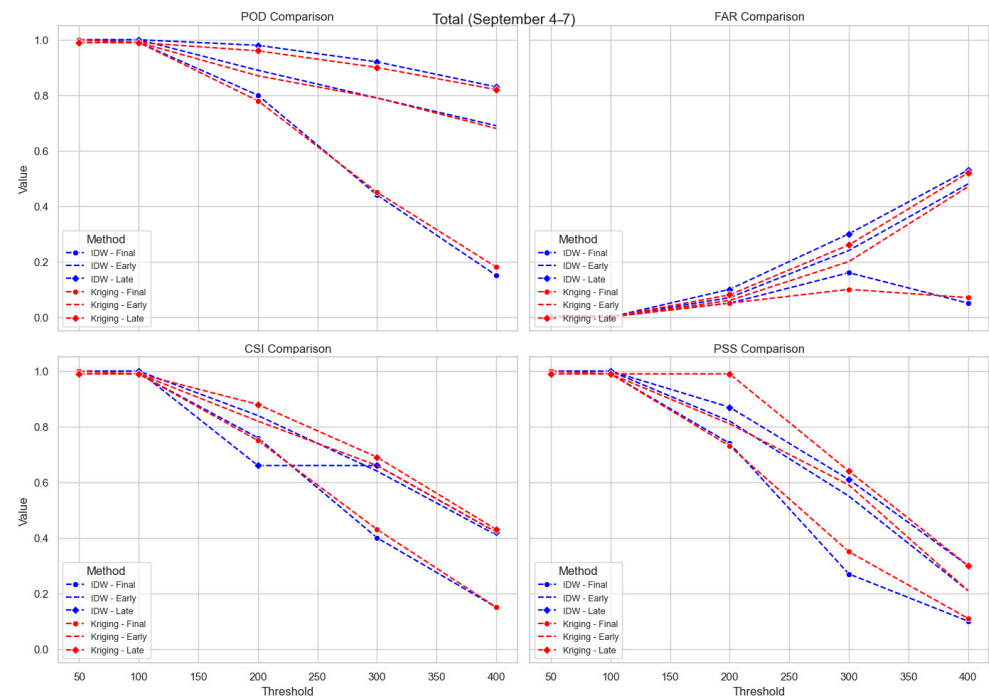


Figure 7. Performance indices of total rainfall IMERG products for different interpolation methods.

3.1.2. Spatial Distribution of Statistical Performance Indices of Total Rainfall

The three composite images of Figure 8 depict the spatial performance of IDW and kriging interpolation techniques over a spectrum of rainfall intensity thresholds (50 mm to 400 mm) through essential verification metrics: POD, FAR, CSI, and PSS. Throughout all runs, IDW outperforms kriging, especially in the central and western regions of Thessaly, where station density is greater and terrain more homogeneous. At lower thresholds (50 mm and 100 mm), all three runs exhibit generally good skill for both methods, with high POD and CSI values and low FAR, suggesting consistent detection of rainfall. However, as thresholds exceed 200 mm, the differences between runs and methods intensify. In the final run, IDW retains greater spatial coherence and higher skill compared to kriging, especially at 300 mm and 400 mm, where kriging is characterized by spatial fragmentation and numerous undefined (NaN) areas in the northeast and east. The early run presents similar spatial features to the final run, albeit with somewhat greater degradation of kriging performance at elevated thresholds. The late run reveals the largest disparity between the two methods, with kriging failing to identify high-intensity rainfall over extensive areas of the study region, especially in the eastern coastal and northern mountainous regions. IDW, however, maintains comparatively high POD and CSI in the southwest and western plains, with more stable PSS values for all thresholds. Overall, comparative analysis indicates that IDW yields a more reliable and spatially stable portrayal of extreme rainfall events, whereas kriging is more sensitive to rainfall threshold, spatial heterogeneity, and likely sparse observational input, particularly in the late run scenario.

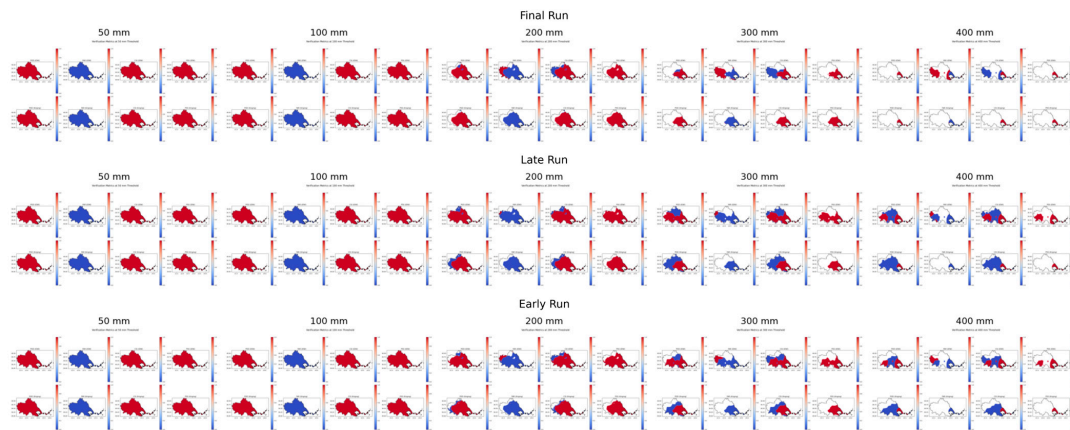


Figure 8. Spatial distribution of the performance indices of total rainfall of IMERG products for different interpolation methods.

3.2. Daily Comparison

3.2.1. Overall Performance Evaluation of IMERG Products on 4 September

The comparative assessment of interpolation techniques, IDW and kriging, for different IMERG products (final, early, and late) of 4 September is presented in Figure 9. POD remains at high levels across all the thresholds, while kriging consistently performs better than IDW, especially at lower rainfall thresholds where it nearly reaches a value equal to 1 ($POD \approx 1$). Nevertheless, with increasing threshold, there is a slight reduction in POD for both methods, with IDW showing a more drastic reduction, particularly above the 30 mm threshold. FAR increases as thresholds increase, indicating a greater chance for misclassification at higher rainfall rates. Here, kriging again has a slight edge, with a lower FAR for the majority of thresholds, in agreement with its superior spatial prediction performance. It can be noticed in Figure 9, where the kriging-based estimates (red lines) are consistently lower than the IDW estimates (blue lines) in the FAR panel. The CSI and PSS are additional results of the greater capability of kriging, whose CSI values are higher than the other methods at all thresholds, most prominently in the 10–30 mm range, where the two methods differ most. Beyond this range, both methods experience a sharp decrease in CSI and PSS, reflecting greater difficulty in the accurate detection of extreme precipitation events.

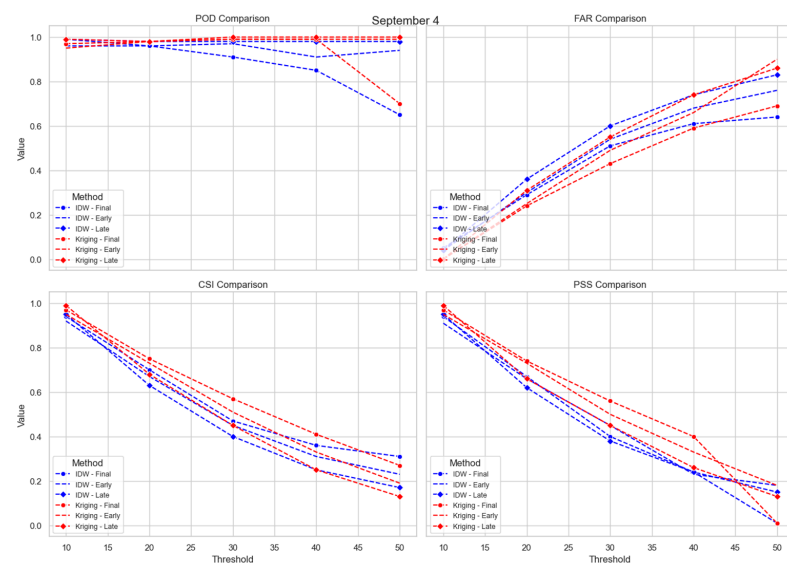


Figure 9. Performance indices of the total rainfall on 4 September of IMERG products for different interpolation methods.

3.2.2. Spatial Distribution of Categorical Indices on 4 September

The validation measures of the 4 September event (projected over the early run, final run, and late run composite plots (Figure 10)) emphasize pronounced spatial variation among interpolation techniques (IDW and kriging) and the stability of rainfall detection at rising thresholds from 10 mm to 50 mm. Throughout the three runs, IDW consistently yields wider spatial coverage with more POD and CSI values, especially over the western and central plains of Thessaly, where there is more favorable rainfall station density. For low thresholds (10–20 mm), both techniques capture widespread precipitation skillfully; however, kriging already starts revealing localized underperformance, with FAR values increasing in the northeastern and eastern sectors, and scattered areas of reduced skill in POD and PSS. By 30 mm and above, the performance gap increases spatially. The final run exhibits the strongest behavior of IDW throughout all thresholds, maintaining high skill (deep red POD/CSI areas) in central and western Thessaly, whereas kriging illustrates more patchy and fragmented skill, especially in the north and east, where excessive smoothing and poor station influence lead to missed events (low POD) and high false alarms (high FAR). The early run shows similar tendencies, with IDW faring well in most respects, although kriging starts to fail more noticeably in eastern and mountainous regions as the threshold rises. The late run, however, has the poorest overall kriging performance, with extensive areas, particularly in the southeast and northeast, having low POD and high FAR even at 20–30 mm, and rapidly diminishing skill at 40–50 mm. In short, for the 4 September, IDW sustains spatial consistency and skill across all thresholds and runs**, whereas kriging's performance is more threshold-dependent and regionally heterogeneous, with specific difficulties in low-observation or heterogeneous terrain. The final run-IDW couple emerges as the most spatially consistent and skillful setup for this extreme rainfall event.

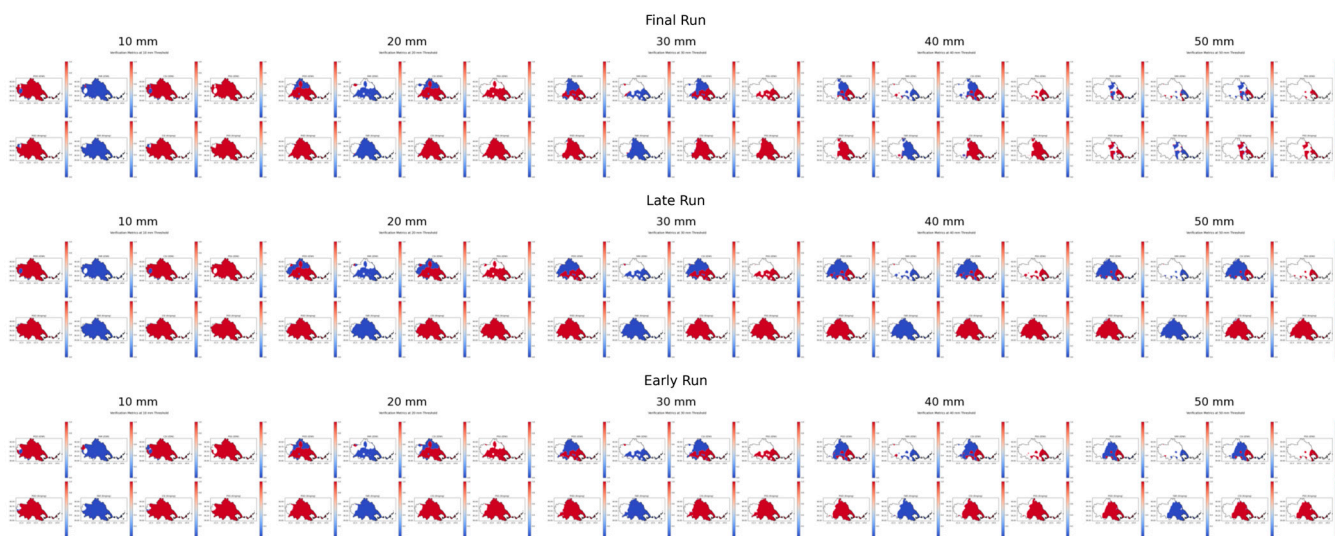


Figure 10. Spatial distribution of the performance indices of total rainfall on 4 September of IMERG products for different interpolation methods.

3.2.3. Overall Performance Evaluation of IMERG Products on 5 September

The performance evaluation of the IMERG products on 5 September proves that the final run is the most reliable overall, particularly for high thresholds of rainfall. The late run has the best POD among the thresholds, reaching a value of 0.99 at 250 mm, but also shows considerably high FAR, increasing from 0.23 at 100 mm to 0.72 at 250 mm, making it susceptible to overestimation of extreme events (Figure 11). On the other hand, the final run also features the lowest FAR across most thresholds, with the sole increase being a

minimal one at 200 mm (0.25), thereby facilitating false alarm reduction. In addition, the final run realizes comparatively high CSI values, as it stands at 0.62 for 200 mm and 0.31 for 250 mm, and also preserves a more balanced PSS, with a peak of 0.61 at 200 mm. While the early and late runs show very good performance at low thresholds (50–150 mm) with POD values above 0.93, their accuracy and skill are lowered at high thresholds because of the increase in false alarms, leading to CSI values of 0.36 and 0.28 for 250 mm in the early and late runs, respectively. Hence, the final run is identified as the best product for extreme precipitation analysis with an optimum balance between event detection, false alarm ratios, and predictive skill overall.

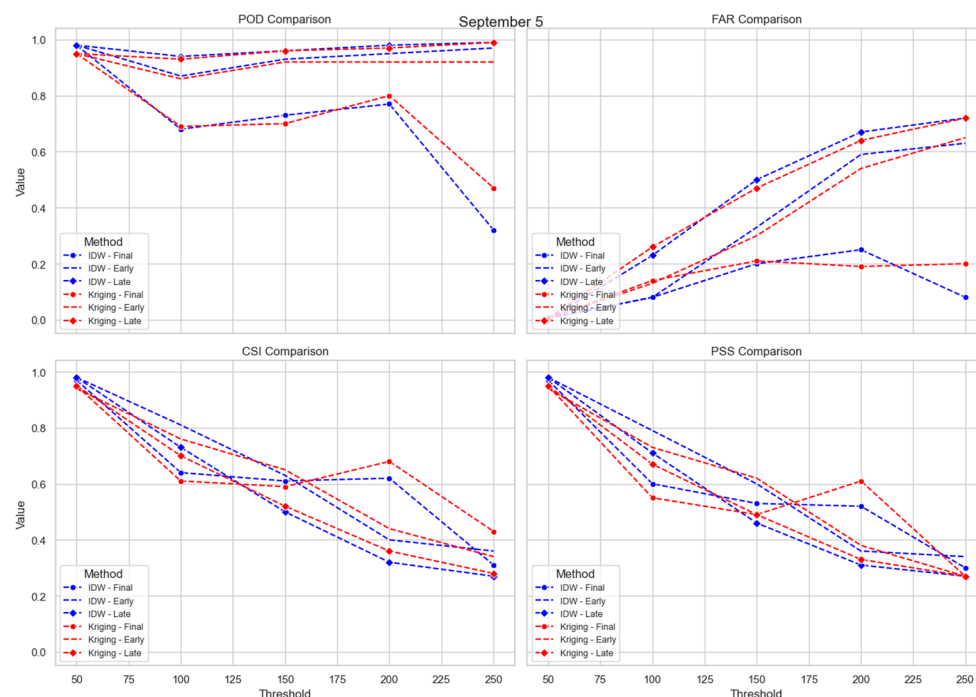


Figure 11. Performance indices of the total rainfall on 5 September IMERG products for different interpolation methods.

3.2.4. Spatial Distribution of Categorical Indices on 5 September

The verification scores of the 5 September event display consistent spatial patterns among the early run, final run, and late run outputs, especially with respect to changing rainfall thresholds from 50 mm to 400 mm (Figure 12). In all three runs, a distinct east–west gradient of performance is apparent, with greater POD and CSI values systematically found over the southeastern and eastern sectors of Thessaly, and lower values, particularly in FAR and PSS, more commonly appearing in the northwestern and central mountainous areas. In the early run, the southern and southeast plains have good detection skill, as indicated by high POD and moderate CSI values, especially at lower thresholds (50–100 mm). Conversely, the northwest has high FAR and lower POD, indicating overprediction or false negatives probably caused by poor station density or complicated terrain. The final run is characterized by an overall enhancement in the spatial coherence of verification scores. Though the eastern sector continues to perform well, particularly for 100 mm and 200 mm thresholds, the central–western zones experience marginal improvements in POD and PSS over the early run, reflecting improved interpolation-based correspondence with IMERG data in those sectors. In the late run, the spatial pattern is generally the same, although performance measures become increasingly localized and fragmented at the higher thresholds (300 mm and 400 mm). Notwithstanding, the southeastern sector still emerges as the most consistently high-performing area for all measures and approaches, while the northwest

persists as a low-performing area, with generally low CSI and high FAR. In general, the 5 September analysis highlights the persistent difficulty of precise rainfall reconstruction in northwestern Thessaly, whereas smoother interpolation techniques (such as IDW and kriging) exhibit solid performance in the eastern and southeastern areas, where station coverage and storm impact seem more pronounced.

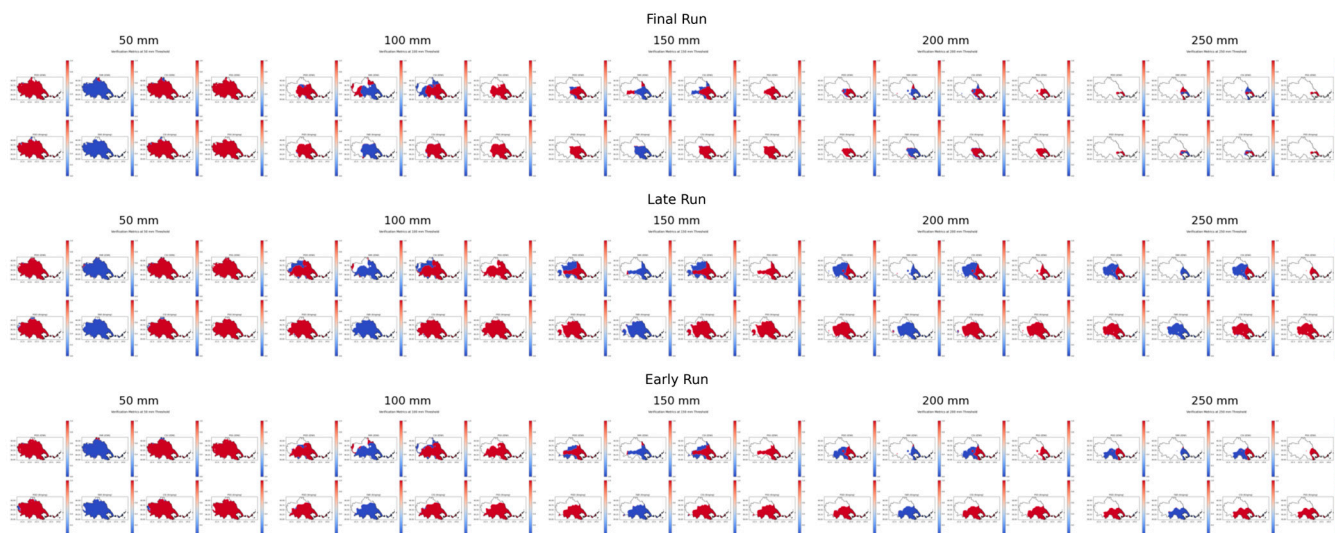


Figure 12. Spatial distribution of the performance indices of total rainfall on 5 September of IMERG products for different interpolation methods.

3.2.5. Overall Performance Evaluation of IMERG Products on 6 September

Figure 13 shows the fluctuations of indices' values between IDW and kriging for different thresholds. Overall, the final run provides satisfying results, especially at lower thresholds. POD ranges from 0.98 at 10 mm and 20 mm for both IDW and kriging, while keeping the FAR very low at 0.002–0.02 for IDW and 0 for kriging. As for heavy rainfall (100 mm), POD falls to 0.32 for IDW and 0.33 for kriging, despite constant CSI and PSS values of about 0.31–0.33. The early run shows the best detection at all thresholds, with a POD of 1 at 10 mm and 20 mm; its FAR goes up to 0.22 for IDW and 0.18 for kriging at 100 mm, which means lower accuracy at high thresholds (CSI equals 0.61 for IDW and 0.64 for kriging). Likewise, the late run displays a high POD (0.99–1) at lower thresholds but experiences an increased rate of false alarms, particularly at 100 mm (FAR = 0.33 for IDW and 0.28 for kriging). Despite this, it obtains a relatively better CSI (0.61–0.65) and PSS (0.56–0.6) compared to the early run. In summary, although the early and late runs are components in detecting events, they frequently exaggerate rainfall at extreme thresholds, whereas the final run provides a more balanced and reliable assessment for heavy precipitation events.

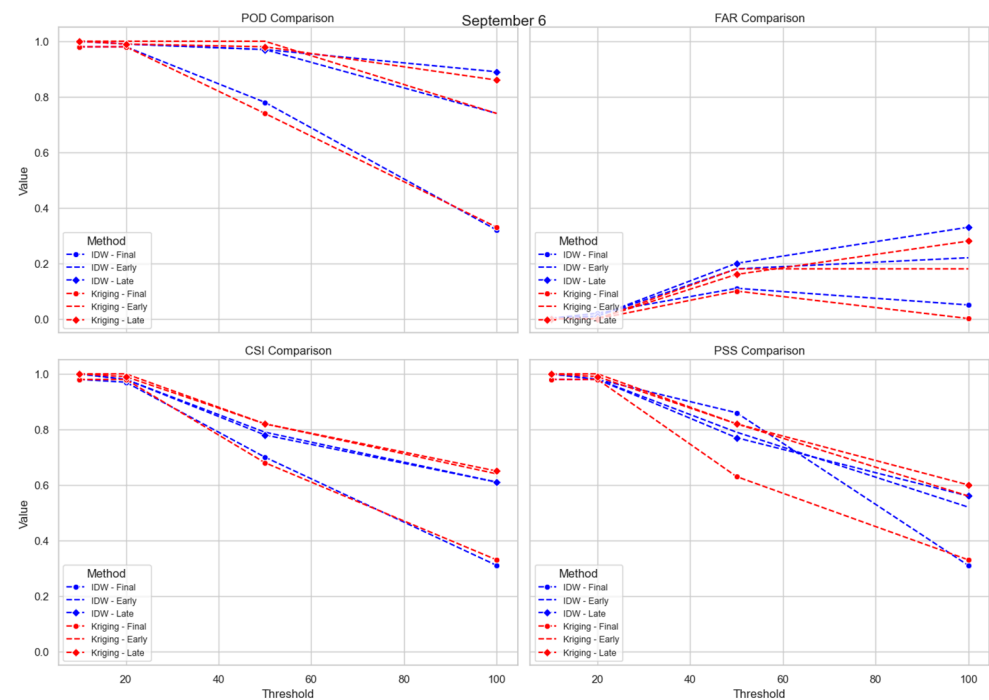


Figure 13. Performance indices of the total rainfall on 6 September IMERG products for different interpolation methods.

3.2.6. Spatial Distribution of Performance Metrics on 6 September

The spatial verification measures for the 6 September show coherent regional patterns across the early run, final run, and late run outputs, with metric performance varying by both rainfall thresholds (10–50 mm) and interpolation approaches (Figure 14). Throughout all runs, probability of detection (POD) is consistently high across southern and central regions of Thessaly, particularly at lower thresholds (10–30 mm). These regions are characterized by intense red shading, reflecting good detection of rainfall events. This pattern continues throughout interpolation approaches, although IDW and kriging produce marginally greater spatial coverage than more discrete methods. The false alarm ratio (FAR) rises substantially in northern and northeastern areas, particularly at higher thresholds (40–50 mm), where extensive blue-shaded areas develop. This implies that overestimation of rainfall events is more prevalent in those regions, perhaps due to gauge sparsity or complicated topography impacting interpolation accuracy. The critical success index (CSI) shows analogous spatial patterns to POD, with southern and southeastern areas recording the highest scores throughout all approaches. Nonetheless, CSI drops considerably in northwest and eastern mountainous regions at higher thresholds, reinforcing the challenge of accomplishing accurate detection at higher rainfall intensities. The Peirce skill score (PSS) also favors southern plains and eastern lowlands, where positive scores prevail. Northern and central Thessaly, on the other hand, witness considerable declines in skill (most evident at 40–50 mm) throughout all interpolation approaches. Of the methods, final run outputs have a tendency to portray more spatial coherence, whereas early run and late run outputs feature more fragmented patterns in higher-threshold scenarios. In conclusion, the 6 September spatial verification analysis indicates more robust model performance in southern and central regions of Thessaly, especially for lower thresholds. Yet, the northern and northeastern areas continue to be problematic, with increasing FAR and decreasing PSS for increasing thresholds, indicating spatially dependent limitations for all interpolation methods.

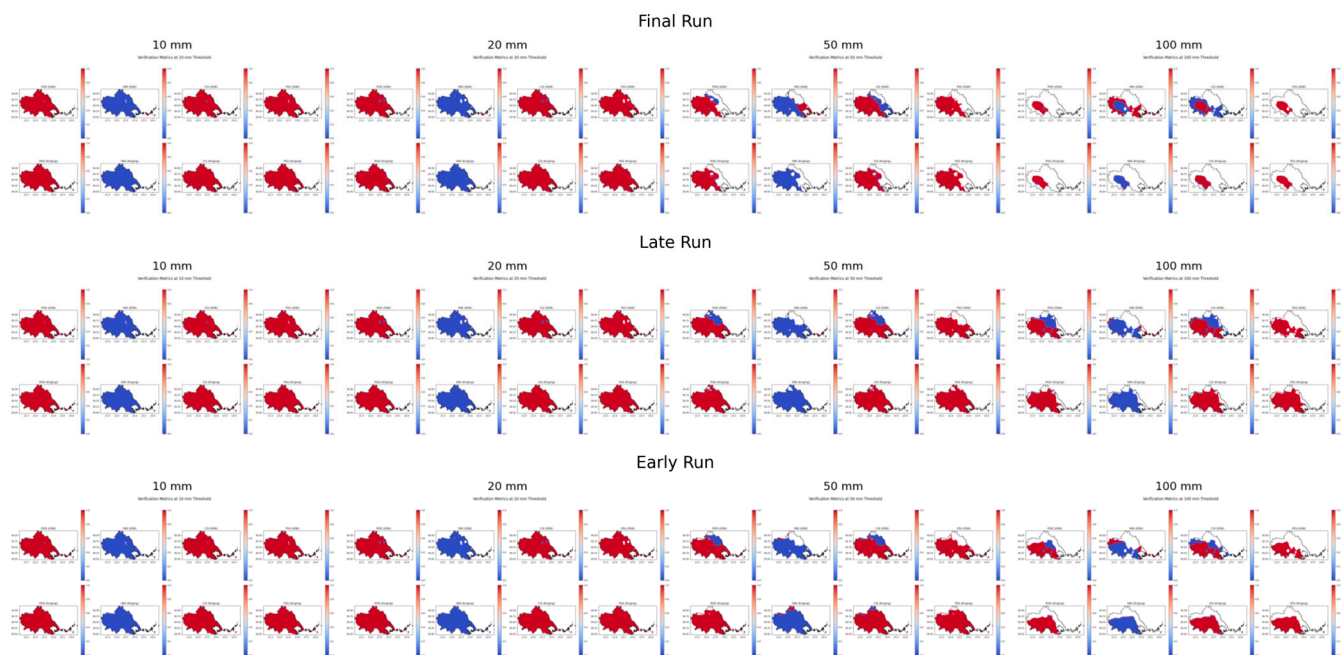


Figure 14. Spatial distribution of the performance indices of total rainfall on 6 September of IMERG products for different interpolation methods.

3.2.7. Overall Performance Evaluation of IMERG Products on 7 September

At 7 September, the final run considerably exemplifies the utmost level of reliability when tested, particularly when used in combination with the inverse distance weighting (IDW) approach, since it continuously registers high probability of detection (POD) values at low precipitation levels; that is, it scores a very impressive 0.98 at the 10 mm as well as the 20 mm levels, while at the same time maintaining the false alarm ratios (FAR) at the lowest levels (Figure 15). In contrast, an examination of the higher thresholds of 50 mm and 100 mm shows a considerable reduction in performance across methods, with kriging performing particularly poorly, culminating ultimately in complete failure at the 100 mm threshold, with a reported probability of detection (POD) of 0 and a false alarm rate (FAR) of 1. Although the early and late runs are performing well at low thresholds (10–20 mm), their performance deteriorates rapidly at larger thresholds (50–100 mm), mainly because of a large number of false alarms. These overestimations can affect operational decisions during extreme events. It is worth noting that although the late run demonstrates a slight enhancement of performance at lower thresholds compared to the early run, it should be understood that both runs exhibit very poor levels of skill when confronted with higher thresholds, reflecting the performance deficit indicated by the early run. As a conclusion, the final run derived from the IDW method is the most stable and reliable source of rainfall estimation, whereas the early and late runs tend to over-estimate the quantity of rainfall at the extremes, and this consequently results in an excessive count of false alarms that may influence decision-making procedures. The results underscore the critical importance of choosing appropriate interpolation techniques while recognizing their limitations for precipitation prediction across different scenarios.

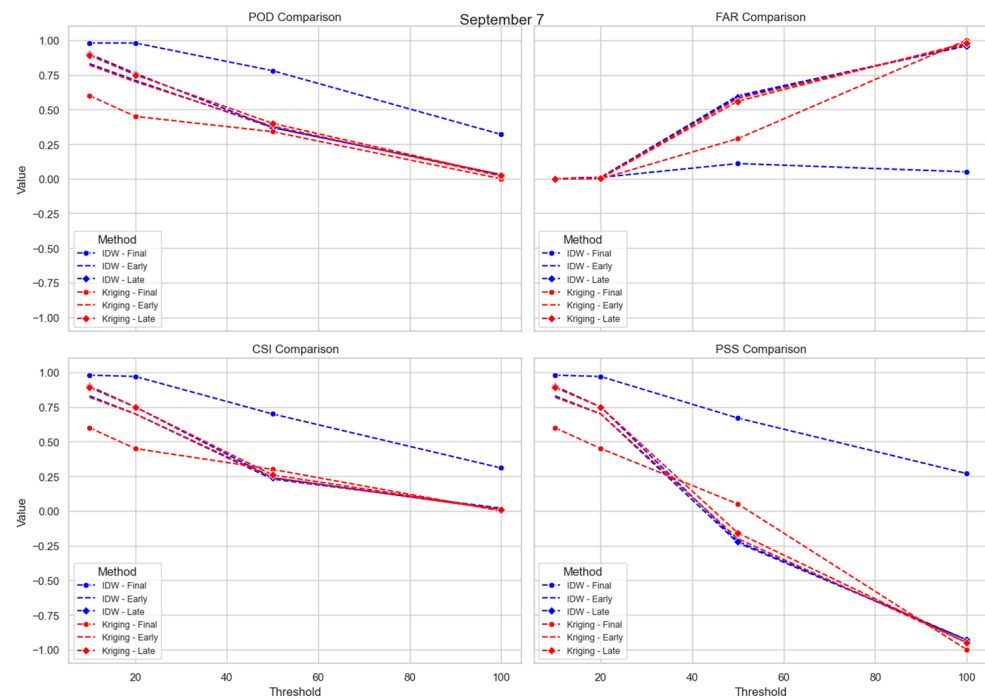


Figure 15. Performance indices of the total rainfall on 7 September IMERG products for different interpolation methods.

3.2.8. Spatial Distribution of Performance Indices on 7 September

On 7 September, the spatial patterns of verification measures over the early, final, and late runs exhibit strong and coherent patterns regarding precipitation thresholds and interpolation skill (Figure 16). In the early run, the probability of detection (POD) is high over most of the domain at low thresholds (10 mm and 20 mm), particularly in the southern and central areas, but progressively deteriorates in the northeast at higher thresholds (50 mm and 100 mm). The false alarm ratio (FAR) increases significantly in the northwest and central regions with rising thresholds, suggesting a tendency for overestimation in those areas. The critical success index (CSI) and Peirce skill score (PSS) reveal more pronounced performance in the southeastern sector, with decreasing scores in the northwest, particularly at the 50 mm and 100 mm levels. The final run is characterized by enhanced spatial coherence and modest improvements across all four measures, notably in the southern and southeastern areas, where lower FAR and elevated POD, CSI, and PSS values suggest improved calibration and overall event detection. The late run reinforces these patterns, with decreased FAR over the entire domain and considerably higher PSS and CSI in the central and eastern regions, even at high thresholds. Overall, across all runs and measures, performance is consistently stronger in the south and southeast, while northwestern areas remain troublesome for all interpolation approaches, especially as rainfall thresholds increase.

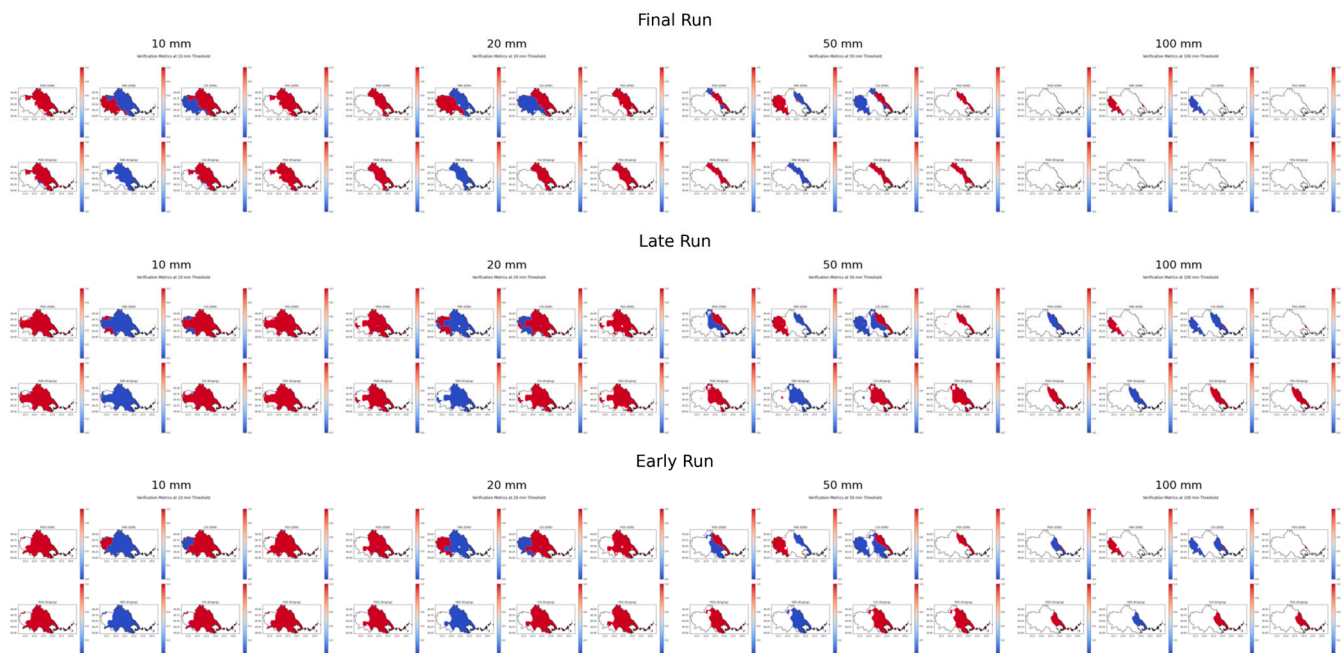


Figure 16. Spatial distribution of performance indices of the total rainfall from IMERG products on 7 September, using different interpolation methods.

4. Discussion

In this work, the assessment of the IMERG satellite products was analyzed against rain gauge data using statistical and categorical indices for Storm Daniel. According to the statistical parameters computed for accumulated rainfall (4–9 September) and daily estimates, the IMERG final run (IMERG-FR) tends to outperform the early (IMERG-ER) and late (IMERG-LR) runs, especially when using the kriging interpolation. At the total scale, IMERG-FR presents the highest correlation coefficient ($CC = 0.89$) and the lowest RMSE (133.11 mm) among the three products, indicating that it is more reliable and accurate (Figure 17). The bias of IMERG-FR is positive but much smaller than that of IMERG-LR, which greatly underestimates. However, IMERG-ER has the highest KGE (0.72), suggesting its overall better performance than the final run. On 4 September, IMERG-FR achieved its highest correlation coefficient ($CC = 0.9$) and consistently demonstrated lower RMSE values than IMERG-LR across all days. However, IMERG-ER has higher KGE values on some days, specifically 5 and 6 September, when it is better than IMERG-FR because of the bias correction and improved representation of variability. IMERG-LR tends to perform worse with greater RMSE and less KGE. This implies that it contains more errors and less accurately agrees with the observed rainfall. On the whole, IMERG-FR remains the most trustworthy for total rainfall estimation. Nonetheless, IMERG-ER performs better at the daily scale and total, particularly for KGE. However, our research solely focuses on IMERG products and reveals that IMERG-early run is better according to KGE and correlation coefficient (CC), especially during the whole event duration.

The spatial pattern of verification measures offers important information on the regional strengths and weaknesses of IMERG performance over Thessaly. Consistently across the various experiment runs and dates examined, the following spatial pattern was apparent: the southern and southeastern parts of the study area generally had higher values of probability of detection (POD) and Peirce skill score (PSS), especially at low thresholds (e.g., 10–50 mm), suggesting dependable detection and absence of bias for moderate precipitation events. In contrast, northwestern and mountainous areas often had higher false alarm ratios (FARs) and a lower critical success index (CSI), especially at high thresholds (greater

than 100 mm), pointing to a higher incidence of overestimation and poorer model capability in representing extreme events.

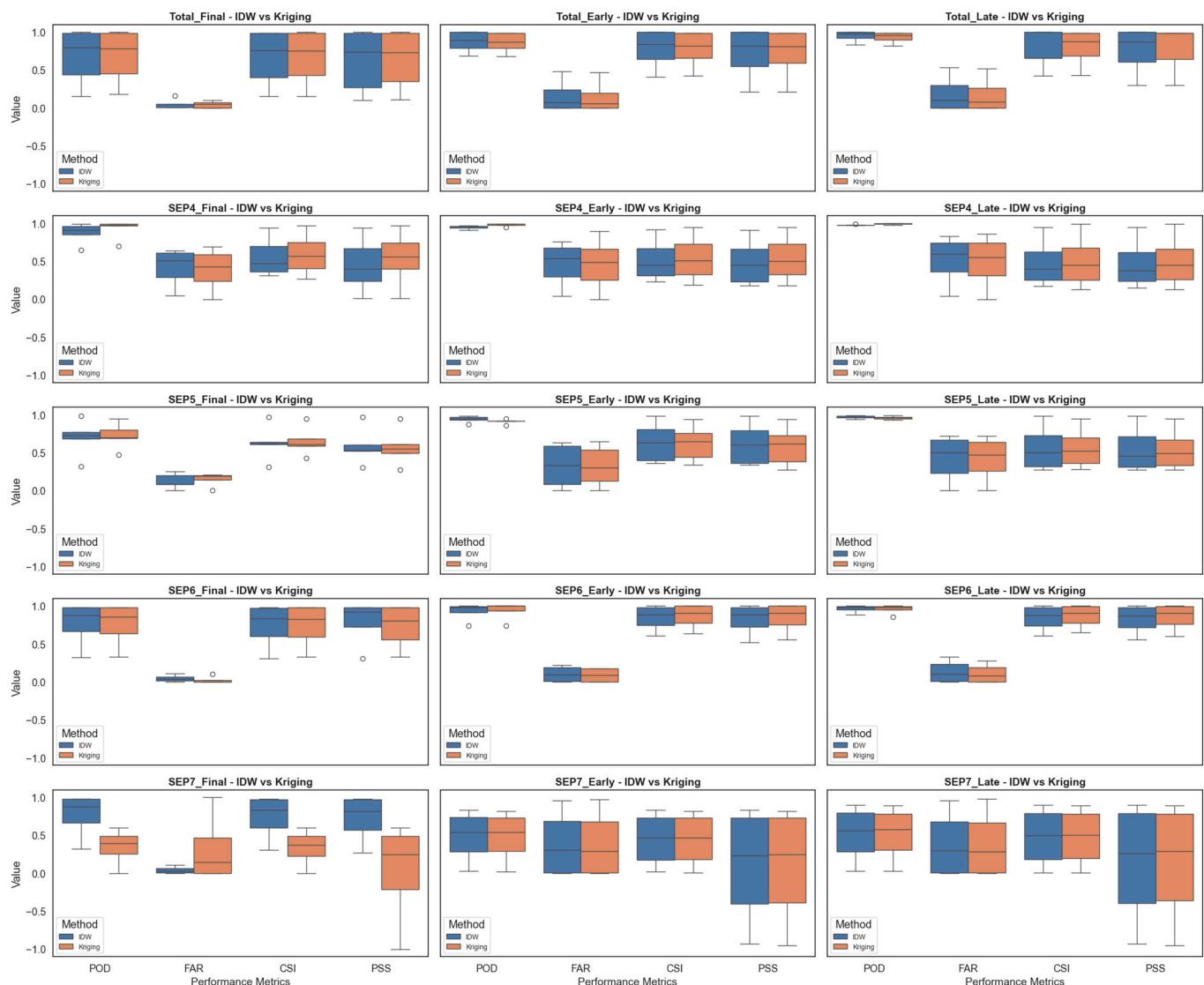


Figure 17. Boxplot of probabilistic indices on the total and daily level.

Interestingly, spatial variation was also seen between interpolation techniques. Kriging routinely exhibited smoother, more coherent spatial fields with marginally better performance at high thresholds (300–400 mm), whereas IDW had a tendency to create more localized strong and weak performance areas, indicative of its sensitivity to surrounding station values. Such spatial variations imply that regional topography, station density, and perhaps convective rainfall structures can impact IMERG detection reliability. Furthermore, temporal refinement enhanced spatial coherence: late runs consistently outperformed early runs, especially in the eastern and central plains, highlighting the value of using finalized satellite precipitation data in extreme event analyses.

Katsanos et al. also found that satellite-derived products underestimate precipitation, concurring with our evaluation that IMERG struggles to detect the maximum precipitation amounts [28]. As Katsanos illustrated how cloud convection is applicable in generating heavy precipitation and ought to utilize more than one dataset to achieve higher accuracy [28], our research presents a better picture of different IMERG products at day-to-day and event scales. In general, our findings support the main point of the two studies:

satellite-based rain observations are useful but are imperfect, especially in representing the distribution of extreme rain events in time and space. Similar research conducted by Yu [29] and Sakib [30] showed that IMERG products fail to measure extreme rainfall appropriately. The real-time products (IMERG-ER and IMERG-LR) are significantly in error, while IMERG-FR is a comparatively better choice but with some deficiencies. Sakib did mention, however, that IMERG-FR overestimates, especially inland. Yu's findings align with our results that the IMERG-LR product performed quite well and is a strong candidate for future precipitation evaluation [30]. Sakib's results show that IMERG accuracy varies by location, where it performs best in high-rainfall regions, as in our study area [30]. However, unlike Sakib, who found that IMERG-ER has a higher relative bias, our results show that IMERG-ER and IMERG-LR greatly underestimate intense rain [30]. The discrepancies between these studies suggest that IMERG performance is highly event-dependent, influenced by geographic and meteorological factors.

5. Conclusions

This research examined the performance of IMERG V06 precipitation datasets in quantitatively representing the extreme rain event of Medicane Daniel in 2023. Through comparative analysis of IMERG-ER, IMERG-LR, and IMERG-FR with rain gauge data obtained from in situ rain gauges, we ascertained their accuracy in estimating precipitation intensity and distribution. The research reveals that although IMERG products provide significant large-scale rainfall estimates, their performance is variable in various stages of storms and geographical locations. The results point to strengths and weaknesses of satellite-retrieved estimates of precipitation in the complex terrain of the Mediterranean region and thus underscore the importance of greater calibration and enhanced incorporation of ground observation. The main findings suggest that although the IMERG products provide valuable large-scale rainfall estimations, performance is extremely varied over storm periods and geographic areas, particularly over complex Mediterranean terrains. IMERG-FR was comparatively superior in accuracy, as found by previous studies [28,30], yet all the products fell short of the capture of extreme top rainfall intensities, underscoring perennial challenges with monitoring extreme precipitation from satellites. Spatial verification analysis showed significant geographical tendencies supplementing the general statistical assessment of IMERG data. Over the entire event and on a daily level, IMERG behaved more consistently in southern and eastern parts of Thessaly, while the performance decreased in northern and northwestern parts, particularly for higher rainfall intensities. These results highlight the spatially variable constraints of satellite precipitation products in heterogeneous terrain with diverse rainfall regimes.

For interpolation method choice, kriging exhibited marginally better spatial skill in reproducing extreme rainfall, especially at high-intensity events, but both kriging and IDW had decreased skill beyond 200 mm thresholds. The spatial patterns indicate the need to incorporate regional calibration methods or bias corrections to better enable IMERG's usage for hydrometeorological monitoring and prediction in complicated terrains. Finally, the late run products provided the most spatially consistent and skillful output, substantiating their utilization for retrospective rainfall verification and disaster response evaluation.

In spite of these constraints, IMERG data contribute essential inputs to risk management for climate, hydrological modeling, and early warning systems. Its near-real-time functionality (IMERG-ER/LR) facilitates early flood warnings and emergency response operations, while the research-quality IMERG-FR is used for post-event studies to enhance vulnerability assessments. Integration of IMERG data with hydrological models enhances stakeholders' capacity to simulate excessive runoff and flood events more realistically, especially in regions with sparse data availability and the absence of ground observations.

In addition, these datasets offer a baseline input for climate adaptation policy by enabling policymakers to locate areas of high risk and target resilience-strengthening interventions.

To address existing limitations, future studies should aim to (1) refine satellite rainfall algorithms to better capture instances of extreme precipitation, (2) apply machine learning methods to minimize biases in real-time products, and (3) blend multi-source datasets (e.g., satellite, radar, and gauges) to enhance spatial precision. As climate change is causing the frequency and intensity of hurricanes to rise, the creation of such tools will be critical to reducing risks and protecting vulnerable communities.

Author Contributions: Writing—original draft, E.L.; Writing—review & editing, A.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: These data were derived from the following resources available in the public domain: <https://gpm.nasa.gov/data-access/downloads/gpm> (accessed on 1 March 2025).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kouroutzoglou, J.; Samos, I.; Flocas, H.A.; Hatzaki, M.; Lamaris, C.; Mamara, A.; Emmannouil, A. Analysis of the Transition of an Explosive Cyclone to a Mediterranean Tropical-like Cyclone. *Atmosphere* **2021**, *12*, 1438. [CrossRef]
2. Lagouvardos, K.; Karagiannidis, A.; Dafis, S.; Kalimeris, A.; Kotroni, V. Ianos—A Hurricane in the Mediterranean. *Bull. Am. Meteorol. Soc.* **2022**, *103*, E1621–E1636. [CrossRef]
3. Miglietta, M.M. Mediterranean Tropical-Like Cyclones (Medicanes). *Atmosphere* **2019**, *10*, 206. [CrossRef]
4. Gutiérrez-Fernández, J.; Miglietta, M.M.; González-Alemán, J.J.; Gaertner, M.A. A New Refinement of Mediterranean Tropical-Like Cyclones Characteristics. *Geophys. Res. Lett.* **2024**, *51*, e2023GL106429. [CrossRef]
5. Vasiliades, L.; Farsiroto, E.; Psilovikos, A. An Integrated Hydrologic/Hydraulic Analysis of the Medicanes ‘Ianos’ Flood Event in Kalentzis River Basin, Greece. In Proceedings of the 7th IAHR Europe Congress: Innovative Water Management in a Changing Climate, Athens, Greece, 7–9 September 2022; pp. 235–236.
6. Mavroulis, S.; Mavrouli, M.; Lekkas, E.; Tsakris, A. Impact of the September 2023 Storm Daniel and Subsequent Flooding in Thessaly (Greece) on the Natural and Built Environment and on Infectious Disease Emergence. *Environments* **2024**, *11*, 163. [CrossRef]
7. Leivadiotis, E.; Kohnová, S.; Psilovikos, A. Evaluating flood events caused by Medicanes ‘Daniel’ in the Thessaly District (Central Greece) using remote sensing data and techniques. *Acta Hydrol. Slovaca* **2024**, *25*, 115–126. [CrossRef]
8. Leivadiotis, E.; Kohnová, S.; Psilovikos, A. IMERG-E and IMERG-L: A Comprehensive Evaluation of the Medicanes Daniel in Thessaly, Greece. Available online: <https://meetingorganizer.copernicus.org/EGU24/EGU24-8116.html> (accessed on 7 May 2025).
9. Mastrantonas, N.; Herrera-Lormendez, P.; Magnusson, L.; Pappenberger, F.; Matschullat, J. Extreme precipitation events in the Mediterranean: Spatiotemporal characteristics and connection to large-scale atmospheric flow patterns. *Int. J. Climatol.* **2021**, *41*, 2710–2728. [CrossRef]
10. Nastos, P.T.; Kapsomenakis, J.; Douvis, K.C. Analysis of precipitation extremes based on satellite and high-resolution gridded data set over Mediterranean basin. *Atmos. Res.* **2013**, *131*, 46–59. [CrossRef]
11. Hourngir, D.; Panegrossi, G.; Casella, D.; Sanò, P.; D’Adderio, L.P.; Liu, C. A 4-Year Climatological Analysis Based on GPM Observations of Deep Convective Events in the Mediterranean Region. *Remote. Sens.* **2021**, *13*, 1685. [CrossRef]
12. Retalis, A.; Katsanos, D.; Michaelides, S.; Tymvios, F. Evaluation of high-resolution satellite precipitation data over the Mediterranean Region. In *Precipitation Science*; Elsevier: Amsterdam, The Netherlands, 2022; pp. 159–175. [CrossRef]
13. Zhang, Y.; Ye, A.; Nguyen, P.; Analui, B.; Sorooshian, S.; Hsu, K. Error Characteristics and Scale Dependence of Current Satellite Precipitation Estimates Products in Hydrological Modeling. *Remote. Sens.* **2021**, *13*, 3061. [CrossRef]
14. Alhammoud, B.; Claud, C.; Funatsu, B.; Béranger, K.; Chaboureaud, J.-P. Patterns of Precipitation and Convection Occurrence over the Mediterranean Basin Derived from a Decade of Microwave Satellite Observations. *Atmosphere* **2014**, *5*, 370–398. [CrossRef]
15. Wisal, K.; Asif, K.; Afed Ullah, K.; Mujahid, K. Evaluation of hydrological modeling using climatic station and gridded precipitation dataset. *Mausam* **2021**, *71*, 717–728. [CrossRef]
16. Bournas, A.; Baltas, E. Development of a Storm-Tracking Algorithm for the Analysis of Radar Rainfall Patterns in Athens, Greece. *Water* **2024**, *16*, 2905. [CrossRef]

17. Spyrou, C.; Varlas, G.; Pappa, A.; Mentzafou, A.; Katsafados, P.; Papadopoulos, A.; Anagnostou, M.N.; Kalogiros, J. Implementation of a Nowcasting Hydrometeorological System for Studying Flash Flood Events: The Case of Mandra, Greece. *Remote. Sens.* **2020**, *12*, 2784. [[CrossRef](#)]
18. Gavahi, K.; Foroumandi, E.; Moradkhani, H. A deep learning-based framework for multi-source precipitation fusion. *Remote. Sens. Environ.* **2023**, *295*, 113723. [[CrossRef](#)]
19. Khan, R.S.; Bhuiyan, M.A.E. Artificial Intelligence-Based Techniques for Rainfall Estimation Integrating Multisource Precipitation Datasets. *Atmosphere* **2021**, *12*, 1239. [[CrossRef](#)]
20. Kang, J.; Wang, H.; Yuan, F.; Wang, Z.; Huang, J.; Qiu, T. Prediction of Precipitation Based on Recurrent Neural Networks in Jingdezhen, Jiangxi Province, China. *Atmosphere* **2020**, *11*, 246. [[CrossRef](#)]
21. Kofidou, M.; Stathopoulos, S.; Gemitzi, A. Review on spatial downscaling of satellite derived precipitation estimates. *Environ. Earth Sci.* **2023**, *82*, 424. [[CrossRef](#)]
22. Xiang, Y.; Chen, J.; Li, L.; Peng, T.; Yin, Z. Evaluation of Eight Global Precipitation Datasets in Hydrological Modeling. *Remote. Sens.* **2021**, *13*, 2831. [[CrossRef](#)]
23. Moazami, S.; Najafi, M.R. A comprehensive evaluation of GPM-IMERG V06 and MRMS with hourly ground-based precipitation observations across Canada. *J. Hydrol.* **2021**, *594*, 125929. [[CrossRef](#)]
24. Mahmoud, M.T.; Mohammed, S.A.; Hamouda, M.A.; Dal Maso, M.; Mohamed, M.M. Performance of the IMERG Precipitation Products over High-latitudes Region of Finland. *Remote. Sens.* **2021**, *13*, 2073. [[CrossRef](#)]
25. Huffman, G.J.; Bolvin, D.T.; Braithwaite, D.; Hsu, K.-L.; Joyce, R.J.; Kidd, C.; Nelkin, E.J.; Sorooshian, S.; Stocker, E.F.; Tan, J.; et al. Integrated Multi-satellite Retrievals for the Global Precipitation Measurement (GPM) Mission (IMERG). In *Satellite Precipitation Measurement*; Levizzani, V., Kidd, C., Kirschbaum, D.B., Kummerow, C.D., Nakamura, K., Turk, F.J., Eds.; Advances in Global Change Research; Springer International Publishing: Cham, Switzerland, 2020; Volume 67, pp. 343–353. [[CrossRef](#)]
26. Tan, J.; Huffman, G.J.; Bolvin, D.T.; Nelkin, E.J. IMERG V06: Changes to the Morphing Algorithm. *J. Atmos. Ocean. Technol.* **2019**, *36*, 2471–2482. [[CrossRef](#)]
27. Kolios, S.; Papavasileiou, N. Daily Rainfall Patterns During Storm ‘Daniel’ Based on Different Satellite Data. *Atmosphere* **2024**, *15*, 1277. [[CrossRef](#)]
28. Katsanos, D.; Retalis, A.; Kalogiros, J.; Psiloglou, B.E.; Roukounakis, N.; Anagnostou, M. Performance Evaluation of Satellite Precipitation Products During Extreme Events—The Case of the Medicane Daniel in Thessaly, Greece. *Remote. Sens.* **2024**, *16*, 4216. [[CrossRef](#)]
29. Yu, C.; Hu, D.; Di, Y.; Wang, Y. Performance evaluation of IMERG precipitation products during typhoon Lekima (2019). *J. Hydrol.* **2021**, *597*, 126307. [[CrossRef](#)]
30. Sakib, S.; Ghebreyesus, D.; Sharif, H.O. Performance Evaluation of IMERG GPM Products during Tropical Storm Imelda. *Atmosphere* **2021**, *12*, 687. [[CrossRef](#)]
31. Orengo, H.A.; Krahopoulou, A.; Garcia-Molsosa, A.; Palaiochoritis, K.; Stamati, A. Photogrammetric re-discovery of the hidden long-term landscapes of western Thessaly, central Greece. *J. Archaeol. Sci.* **2015**, *64*, 100–109. [[CrossRef](#)]
32. Skofronick-Jackson, G.; Kirschbaum, D.; Petersen, W.; Huffman, G.; Kidd, C.; Stocker, E.; Kakar, R. The Global Precipitation Measurement (GPM) mission’s scientific achievements and societal contributions: Reviewing four years of advanced rain and snow observations. *Q. J. R. Meteorol. Soc.* **2018**, *144*, 27–48. [[CrossRef](#)]
33. Jiang, L.; Bauer-Gottwein, P. How do GPM IMERG precipitation estimates perform as hydrological model forcing? Evaluation for 300 catchments across Mainland China. *J. Hydrol.* **2019**, *572*, 486–500. [[CrossRef](#)]
34. Dimitriou, E.; Efstratiadis, A.; Zotou, I.; Papadopoulos, A.; Iliopoulou, T.; Sakki, G.-K.; Mazi, K.; Rozos, E.; Koukouvinos, A.; Koussis, A.D.; et al. Post-Analysis of Daniel Extreme Flood Event in Thessaly, Central Greece: Practical Lessons and the Value of State-of-the-Art Water-Monitoring Networks. *Water* **2024**, *16*, 980. [[CrossRef](#)]
35. Prakash, S.; Mitra, A.K.; AghaKouchak, A.; Liu, Z.; Norouzi, H.; Pai, D.S. A preliminary assessment of GPM-based multi-satellite precipitation estimates over a monsoon dominated region. *J. Hydrol.* **2018**, *556*, 865–876. [[CrossRef](#)]
36. Gupta, H.V.; Kling, H.; Yilmaz, K.K.; Martinez, G.F. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.* **2009**, *377*, 80–91. [[CrossRef](#)]
37. Sentas, A.; Psilovikos, A.; Psilovikos, T. Statistical analysis and assessment of water quality parameters in Pagoneri, river Nestos. *Enterp.-Wide Risk Assess.* **2016**, *55*, 114–124.
38. Spiridis, A.; Konstantinidis, A.; Koutalou, V.; Perakis, K.; Psilovikos, A. Average hyper-annual rainfall as a regional variable in Central Macedonia: Use of a geostatistical approach to identify areas with drought trends. In Proceedings of the 7th IAHR Europe Congress: Innovative Water Management in a Changing Climate, Athens, Greece, 7–9 September 2022; pp. 406–409.

39. Charizopoulos, N.; Zagana, E.; Psilovikos, A. Assessment of natural and anthropogenic impacts in groundwater, utilizing multivariate statistical analysis and inverse distance weighted interpolation modeling: The case of a Scopia basin (Central Greece). *Environ. Earth Sci.* **2018**, *77*, 380. [[CrossRef](#)]
40. Karamoutsou, L.; Psilovikos, A. Deep Learning in Water Resources Management: The Case Study of Kastoria Lake in Greece. *Water* **2021**, *13*, 3364. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.