# Mapping the spatial variability of rainfall from a physiographic-based multilinear regression: model development and application to the Southwestern Iberian Peninsula

Verónica Ruiz-Ortiz[a*], Jorge M. G. P. Isidoro[b], Helena Maria Fernandez[c], Fernando M. Granja-Martins[c], Santiago García-López[d]

[a*]Department of Industrial and Civil Engineering, Algeciras School of Engineering and Technology, University of Cádiz, Algeciras, Spain; veronica.ruiz@uca.es; https://orcid.org/0000-0001-7638-1135 (corresponding author)

[b]Department of Civil Engineering, Institute of Engineering, University of Algarve, Portugal; Marine and Environmental Research Centre (MARE), Coimbra, Portugal; jisidoro@ualg.pt; http://orcid.org/0000-0002-6901-5652

[c]Department of Civil Engineering, Institute of Engineering, University of Algarve, Portugal; Research Centre for Tourism, Sustainability and Well-being (CinTurs), University of Algarve, Portugal; hfernand@ualg.pt; https://orcid.org/0000-0002-3677-1064

[c]Department of Civil Engineering, Institute of Engineering, University of Algarve, Portugal; Research Centre for Tourism, Sustainability and Well-being (CinTurs), University of Algarve, Portugal; fmmartin@ualg.pt; https://orcid.org/0000-0002-2709-804X

[d]Department of Earth Sciences, Faculty of Marine and Environmental Sciences, University of Cádiz, Puerto Real, Spain; santiago.garcia@uca.es; https://orcid.org/0000-0002-1245-7169

**Abstract**

A physiographic-based multilinear regression model supported by GIS was developed to estimate spatial rainfall variability in the Southwest Iberian Peninsula. The area study includes a wide diversity of landscape features and comprises four Portuguese regions and one Spanish province (totalising 28,860 km$^2$). The region suffers a very strong Mediterranean influence, with a major cleavage between winter and summer seasons. Thus, the analysis was carried out separately for the wet (October to March) and dry (April to September) semesters. From an initial set of 10 explanatory physiographic variables, five were selected to be used in the multilinear regression, as they allowed to generate models by map algebra that fitted well with the last 40-years of monthly rainfall data records. These records were obtained from 163 weather stations, filtered from an initial set of 230 (142 stations in Portugal and 88 in Spain). The correlation between the physiographic-based multilinear regression model and a model obtained by interpolation from rainfall historical data showed to be good or very good in approximately 75% of the area under study. Results show that physiographic-based models can be effectively used to estimate rainfall where there is a lack of raingauges, or to densify spatial resolution of rainfall between raingauges.

**Keywords:** Rainfall, Physiography, Multilinear regression, Interpolation, Map algebra, Iberian Peninsula

**Declaration of interest statement**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Acknowledgments**

This work is the result of a research stay of the first author in the University of Algarve, which was supported by the University of Cadiz (2021-025 / PU / PP-EST / MV). This work was supported by the FCT - Foundation for Science and Technology, strategic projects UIDB/04292/2020 granted to MARE and UIDB/04020/2020 granted to CinTurs.

**Data availability statement**

The authors confirm that the raw data supporting the findings of this study are available within the article. The raw data could be downloaded through the Portuguese and Spanish Official Networks of Hydrometeorological Records (Direção Regional de Agricultura e Pescas (DRAP); Junta de Andalucía (JA); Ministerio para la Transición Ecológica y el Reto Demográfico (MITECO); Sistema Nacional de Informação de Recursos Hídricos (SNIRH)).

https://www.drapc.gov.pt/base/agrometeorologia.htm

https://www.juntadeandalucia.es/agriculturaypesca/ifapa/riaweb/web/

https://www.miteco.gob.es/es/cartografia-y-sig/ide/descargas/otros/default.aspx

https://snirh.apambiente.pt/

# Mapping the spatial variability of rainfall from a physiographic-based multilinear regression: model development and application to the Southwestern Iberian Peninsula

## Introduction

Rainfall variability plays a fundamental role in all Earth surface processes, human activities, and their interactions. Water resources planning and management are intrinsically dependent on the spatial and temporal distribution of rainfall. From selecting a river cross-section for installing a major hydropower dam to deciding on how to better optimise water supply systems, rainfall variability is always a key-factor for decision making. Uncertainty on rainfall data has a strong influence in the performance of hydrological models (Barsi et al., 2014; Fraga et al., 2019; Montanari & Di Baldassarre, 2013; Waldron et al., 2020). Climate change is another example of a major source of uncertainty for hydrological modelling, as it can cause changes in rainfall patterns (Ghumman et al., 2013; Parveen & Sreekesh, 2018). As so, the quest for more accurate hydrological models is highly dependent on the knowledge and insight we gain on the structure and idiosyncrasies of rainfall. As Berndtsson and Niemczynowicz (1988) so well stated more than three decades ago in their study about spatial and temporal scales in rainfall analysis: “The future of rainfall analysis will continue to require more complex methodologies in order to match the new objectives due to an increasingly complex society”.

Physiography concerns the landforms present in a region. More specifically, physiography regards the study of the natural features of earth’s surface, as well as the study of the origin, evolution and processes that shape the landforms that characterise a region (Fehmi et al., 2014; Strahler, 1969). The association between physiographic features and earth-surface processes has been extensively studied over time. Examples of these since the beginning of this century are: the analysis of long-term hydrologic response of basins (Berger & Entekhabi, 2001), hydrologic response at different basin scales (Mohamoud, 2004), estimation of rainfall-runoff parameters (Hundecha et al., 2008), mapping stream water chemistry (Lyon et al., 2008), modelling monthly maximum stream temperatures (Guillemette et al., 2009), vulnerability to droughts (Pandey et al., 2010), assessing flood vulnerability (Shiau et al., 2011), estimating low-flow indices (Castiglioni et al., 2011), studying rainfall erosivity (Mello et al., 2013), estimating natural aquifer recharge (Ghiglieri et al., 2014), modelling moisture-landscape in soil (Kim, 2013), analysing hydrologic variability (Kuentz et al., 2017), assessing water balance (Abatzoglou & Ficklin, 2017), establishing the probability of the landslides (Fan et al., 2018), estimating flood frequency (Durocher et al., 2019), and mapping rainfall aggressiveness (Fernandez *et al.*, 2020).

Processes and methods to extrapolate point-source rainfall data acquired from raingauges to areas "within the influence range of a raingauge" have been deeply discussed in the literature for many years. Early studies were based on geometric processes, such as the widely used Thiessen Method where areas under the influence of different raingauges are arithmetically averaged (e.g., Chow *et al.*, 1988), or more complex polynomial surface fitting techniques (Edwards, 1973). From then on, researchers have used a variety of techniques, methodologies, and tools to improve the analysis and representation of rainfall's areal distribution. These include, *e.g.*, spatial interpolation methods, remote sensing retrieval, atmosphere reanalysis, or multi-source rainfall merging. For a detailed review on rainfall spatial estimations see Hu *et al*. (2019). Notwithstanding the specificities of each one of these analytical methods, they all are used to fill-in the blanks for a continuous representation of a naturally spatially distributed variable, *e.g.* rainfall intensity or rainfall depth.

A possible approach to attain this goal is to statistically correlate physiographic variables with the characteristics of rainfall, as the local physiography largely affects rainfall (Chandwani et al., 2015). Statistical methods have been widely and consistently used in physical geography for a long time (Harris & Jarvis, 2011; Unwin, 1977). In the particular case of rainfall modelling, Hu *et al.* (2019) states that introducing auxiliary (physiographic) information to augment rainfall spatial estimation accuracy depends on the strength of the correlation between the auxiliary variables and rainfall, and on the abundance of raingauges measurements. In fact, the use of these techniques and methods may be more relevant in areas with low gauge density, where rainfall is represented by too few raingauges, or even, in the limit, in ungauged areas. If a local rainfall pattern does not follow the large-scale pattern, such in an area shadowed from rain, and that particular location is not represented by a gauge, the fitted surface will not be able to represent that anomaly to the large-scale pattern. In these cases, statistics can help to improve the continuous representation of the spatial distribution of rainfall by incorporating physiographic features in the spatial interpolation of rainfall (Adhikary et al., 2017; Brown & Comrie, 2002; Goovaerts, 2000; Hevesi et al., 1992; Kyriakidis et al., 2001; Moliba Bankanza, 2014; Moral, 2010; Tobin et al., 2011).

This study aims to map the spatial variability of rainfall in the Southwest Iberian Peninsula based on physiographic variables. The physiography of this part of the Iberian Peninsula, which is under Mediterranean influence thus prone to torrential rainstorm events, shows a large diversity of features, *e.g.*, mountain ranges and valleys, plains, hills, and coastal lagoons. Furthermore, this area presents major asymmetries in the spatial distribution of raingauges. All this physiographic, climatic, and weather-observation scenario poses additional challenges for this type of analysis. From an initial set of 10 physiographic variables, five of these showed to correlate well with the rainfall records from 1980/81 to 2019/20 (40-years of records), thus being selected for the multilinear regression. Using only physiographical data, it was then possible to derive isohyet maps for the wet (October to March) and dry (April to September) semesters. This methodology contributes to a deeper insight on the complex relations between local physiography and spatial rainfall distribution.

## Materials and Methods

### Study area characterization

The study area, in the southwest of the Iberian Peninsula, includes three southern Portuguese regions and one Andalusian province in Spain. More precisely, the regions of Algarve, Baixo Alentejo, and Alentejo Litoral in Portugal, and the province of Huelva in Spain. The total extension of the area under study is 28,860 km$^2$ (Fig. 1).

**Fig. 1** Location of the study area within the Iberian Peninsula. The 163 weather stations with rainfall records selected to be used in the multilinear regression are highlighted

The physiography of this area closely relates to the structure and lithological nature of the different geological units of the region. Especially, with regard to the distribution and orientation of the drainage networks and the mountain ranges. In this context, three lithological assemblages can be distinguished in the study area (Vera, 2004): (i) Units of the Hesperian Massif, which extend over most of the area under study; these are ancient rocks (Proterozoic terminal to the Carboniferous), intensely deformed as a consequence of the Hercynian orogeny. Within it, in the northern sector, one can find the so-called Ossa-Morena Zone, which presents a higher degree of metamorphism with large recumbent folds and ductile thrusts of NW-SE orientation and abundant magmatism. The most important mountains in the study area (*Sierra de Aracena*) are in this domain, in the north of the province of Huelva, with a maximum altitude of 1055 m and slopes up to 37º. In the southern sector of the Hesperian Massif, separated from the northern sector by a great tectonic accident, is the so-called South Portuguese Zone, which main distinguishing feature is the existence of a submarine volcano-sedimentary complex that underwent hydrothermal alteration, giving rise to one of the most important deposits of polymetallic sulphides in the world (Iberian Pyrite Belt). In this domain, within the Algarve region, *Serra de Monchique* (maximum altitude 908 m, volcanic origin) and *Serra do Caldeirão* (maximum altitude 589 m) are the most important orographic structures; (ii) units of the Mesozoic cover mainly constituted by carbonates, emerging in a narrow coastal strip in the Algarve region, and giving rise to elevated plateau morphologies several tens of meters above the sea level; and (iii) the more recent tertiary basins, which consists of sedimentary formations of marine, littoral, and fluvial origin, scarcely deformed, which give rise to very soft reliefs associated with the filling of the basins of the Tajo-Sado river (to NW) and the Guadalquivir river (to SE).

**Fig. 2** Lithological, geological and tectonic characteristics of the study area

The climate of the area under study is Mediterranean, characterised by hot and dry summers in the coastal area (subtype Csa, according to the Köppen-Geiger climate classification), with cool and dry summers in the hilly areas (subtype Csb) (Beck et al., 2018; Kottek et al., 2006; Peel et al., 2007). The mean annual temperature is 18°C with monthly average temperature ranging from -4°C in the winter to 44 °C in the summer, and the mean annual rainfall is about 521 mm (Pulido-Calvo et al., 2020) with a major spatiotemporal irregularity. On average, about 42% of the annual rainfall occurs in the three-month winter season, mainly caused by frontal and orographic rainfall events. Maximum and minimum annual rainfall values of 1397 mm and 264 mm are associated , respectively, with the mountain and the coastal areas (Neves et al., 2020).

Agriculture, livestock, tourism/ecotourism, mining, and civil construction are the most important economic activities. Many of these are conditioned by the irregularity and seasonality of trainfall, as agriculture depends on large volumes of water for irrigation in the dry season, and tourism is mostly concentrated in the summer (with a strong demand by international and domestic tourists for the coastal areas). The most eastern area is setting above a formation with the highest concentration of massive sulphides in the Earth's crust (Almodóvar et al., 2019). As such, mining is fundamental for the region's economy, despite the subsequent environmental degradation, namely by acid runoff through the mine tailings, that seriously affects the water reservoirs for public water supply and irrigation (Ruiz-Ortiz et al., 2021).

The study area also comprises eight natural spaces of important ecological and landscape values, classified by the European Union network's Natura 2000 (EC, 2008). Several protected species of fauna and flora can be found in this area. Examples are the Iberian lynx (Linx pardina), imperial eagle (Aquila heliaca), Eurasian black vulture (Aegypius monachus), or the holm oak (Quercus rotundifolia) among others, representing a good ecological indicator. Good quality water resources are thus essential for maintaining the ecological state of these protected areas, as well as for the economic activities that are necessary for human development.

Rainfall data

The historical period under analysis comprises 40 hydrologic years, ranging from 1980/81 to 2019/20. Monthly rainfall data was initially obtained from 230 raingauges of Portuguese and Spanish weather stations, 142 in Portugal and 88 in Spain. All these stations are distributed, although irregularly, throughout the study area and are under the Portuguese and Spanish Official Networks of Hydrometeorological Records *(Direção Regional de Agricultura e Pescas (DRAP); Junta de Andalucía (JA); Ministerio para la Transición Ecológica y el Reto Demográfico (MITECO); Sistema Nacional de Informação de Recursos Hídricos* (SNIRH)).

Since the rainfall records showed many data gaps and different raingauges had different recording periods, it was necessary to homogenise the datasets. Firstly, raingauges with less than 180 months of data, the equivalent to 15 years of records, were discarded (65 of 230 stations) following the criteria used by Akbar *et al.* (2018) and Zhang *et al.* (2019) and two more stations were removed due to measurements errors in raw data. Secondly, the historical records of each raingauge were correlated with the records from all the other rainguages by double mass analysis to assure homogeneity and robustness. Finally, the rainfall series were systematically completed through the bivariate regression model CORMUL (González-Hidalgo et al., 2002; Squintu et al., 2019) implemented in the software CHAC (CEDEX, 2013). The series completion considered a previous monthly stationarisation, using a prioritisation exponent of 0.1 and setting up a prioritisation threshold of 0.7, higher than the minimum value of multiple correlation coefficients typically used in hydrology (CEDEX, 2013; Penagos Cruz, 2014). According to the aforementioned criteria used by other authors in hydrology studies (Zazo et al., 2020), a total of 163 rainfall stations were considered and completed for this study (see Fig. 1).

Physiographic explanatory variables

Based on background knowledge of the physical geography of the study area, the following physiographic explanatory variables were initially selected for this analysis: elevation (H), slope (S), curvature (C), aspect defined by its sine (SINA) and cosine (COSA), distance to coast (DC), geographic coordinates (X, Y), Normalised Difference Vegetation Index in the wet (NDVI_W) and dry (NDVI_D) periods, and radiation in the wet (RW) and dry (RD) periods. The wet and dry periods correspond, respectively, to the semesters of October to March and of April to September. Firstly, a Digital Elevation Model (DEM) of the study area with a resolution of 30 m was obtained from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM) Version 3 (National Administration Space Aeronautics (NASA) and Ministry of Economy Trade and Industry (METI)). The ASTER mission has a stereoscopic capability to acquire stereo image data with a base-to-height ratio of 0.6. The image's spatial horizontal resolution is 15 m, with a 60×60 km ground area for each scene. Automatic photogrammetric processing of data collected by the satellite is carried out, generating a DEM for all the emerged lands that is distributed freely. Studies to validate and characterise the ASTER GDEM confirm that accuracies for this global product are 20 m and 30 m at a 95% confidence level, respectively, for vertical and horizontal data (*e.g.*, Zhao *et al.*, 2010).

The altitude in meters (H) of the weather stations were obtained from the DEM, and then compared/validated against the information of the Official Network of Hydrometeorological Records. Regarding the planimetry, after a meticulous analysis to detect errors, some notorious differences were detected. These differences were due to some inaccuracies in the Official Networks of Hydrometeorological Records data, and so the coordinates of these stations were corrected.

The slope (S), curvature (C), and aspect (SINA and COSA) physiographic models were generated by fitting a quadratic surface to the DEM for the selected kernel size (further detailed in this section) and taking the appropriate derivatives according to Wood (1996), using the software ENVI 4.3. The slope and curvature of the surface determines the morphometric features. S is measured in degrees, with 0° as a reference for a horizontal plane. The cross-sectional curvature, intersecting with the plane of the slope's normal and perpendicular aspect directions, are measures of the surface curvature in the across slope directions. Convex and concave surfaces result, respectively, in positive and negative values. Aspect identifies the downslope direction of the maximum gradient of elevation from each cell to its neighbours. The aspect value of each cell in the output raster indicates the surface's azimuth, and it is measured clockwise in degrees from 0º (north) to 360º (north). In this study, two preferential directions (N-S and W-E) have been identified. According to this, SINA and COSA are defined, respectively, regarding the directions W-E and N-S. The resulting normalised SINA and COSA values ranges between -1 to 1.

The distance to coast (DC) corresponds to the minimum planimetric distance from each raingauge to the coastline and was obtained directly from the Geographic Information System (GIS). Geographic coordinates correspond to X and Y coordinates in the UTM projection, zone 29 in the European Terrestrial Reference System 1989 (ETRS89).

The NDVI model was obtained through Google Earth Engine (GEE). GEE is a multi-petabyte analysis-ready data catalogue co-located with a high-performance, intrinsically parallel, computation service. It is accessed and controlled through an internet-accessible Application Programming Interface (API) and an associated web-based interactive development environment (IDE), enabling rapid prototyping and visualisation of results (Gorelick et al., 2017). NDVI was obtained by using Landsat-5 images of the years 1984 to 2012 acquired from the Thematic Mapper (TM) sensor, and Landsat-8 images from the years 2013 to 2020 acquired from the Operational Land Imager (OLI) sensor, both with a spatial resolution of 30 m. These images present a cloudiness of less than 22%. NDVI index represents the contrast of the two-band characteristics of a multispectral raster dataset: the absorption of chlorophyll pigment in the red band (RED), and the high reflectivity of plant tissue in the near-infrared (NIR) band. NDVI ranges from -1 to 1 and was calculated using Equation (1):

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{1}$$

Where NIR is reflectance in band 4 (0.76 - 0.90 µm) for the TM sensor and band 5 (0.85 - 0.88 µm) for the OLI sensor, and RED is reflectance in band 3 (0.63 - 0.69 µm) for the TM sensor and band 4 (0.64 - 0.67 µm) for the OLI sensor.

The solar radiation models were derived based on the hemispherical viewshed algorithm developed by Rich *et al.* (1994) and Fu and Rich (2002), using the Spatial Analyst tools of ArcGis 10. Solar radiation (Wh/m$^2$) corresponds to the total solar radiation accumulated for the wet and dry periods of the year 2019/2020, without considering the atmospheric effect.

Three products with different kernel sizes were generated for the slope, curvature, aspect, NDVI, and radiation models. Kernel sizes of 3, 9, and 15 were used, and the multi-scale physiographic information was extracted for the location of each of the 165 raingauges. The use of different kernel sizes had two goals in mind: first, to analyse the influence of cell size on the correlation of rainfall with the physiographic variables and, second, to smooth out the imprecision of the raw data.

Multilinear regression

Multilinear regression was used to explore the relationship between rainfall (in the wet and dry periods) and the explanatory variables (obtained with the three kernels), thus resulting in six equations. Multilinear regression has been used by other authors for this specific purpose (*e.g.*, Kumari *et al.*, 2017; Segarra *et al.*, 2020). This tool has also been used for several other purposes in hydrological studies, such as analysing soil erosion (H. M. Fernandez et al., 2016), producing rainfall aggressiveness maps (H. Fernandez et al., 2020), estimating runoff (Abatzoglou & Ficklin, 2017) or evaporation (Singh et al., 2021), or evaluating surface water budgets (Narbondo et al., 2020). The general equation for a multilinear regression is presented in Equation (2).

$$y_i = \beta_0 + \sum_{1}^{k} \beta_k x_k + \varepsilon_i \tag{2}$$

Where $y_i$ is the dependent variable (rainfall), $\beta_0$ is the intercept parameter, $\beta_k$ is the regression coefficient of the independent variable $x_k$ (*e.g.*, slope, curvature, …) and $\varepsilon_i$ is the random error.

The p-values of the intercept parameter and of the regression coefficients were verified for a confidence level of 95%. The model adjustment ($R^2$), the residuals validation (normality, homogeneity of variance, independence, and Cook's distance), and the Akaike Information Criterion (AIC) were calculated to validate the multilinear regressions. Different combinations of explanatory variables were tested preliminary. From the resulting equations in this preliminary analysis, two equations were selected for the wet period and four for the dry period, all of these with $R^2$ values between 0.675 and 0.783. This procedure revealed two raingauges with major errors in the raw data, which were eliminated from the dataset.

Rainfall Models

The Map Algebra tool of ArcGis 10.6 was used to generate rainfall models based on the six selected equations above-mentioned. The regression models' results were then compared against the observed average rainfall maps from the 1980-2020 period. The latter were generated by Inverse Distance Weighting (IDW) interpolation for a power of 2. Residual and difference maps were also produced, thus allowing an easier assessment of the models quality. Fig. 2 conceptualises the methodology followed in this study.

**Fig. 3** Flowchart of the methodology followed to generate the multilinear rainfall regression models based on physiographic data

## Results

Intermediate results were obtained before generating the final rainfall regression models. For the location of the 163 raingauges selected in the study area, *i.e.*, those complying with the criteria of gaps<180 and ρ>0.7, 28 explanatory variables were extracted through the physiographic models, considering the different kernel sizes (Fig. 3). Annex 1 summarises the most important datasets used in this study, for all the selected raingauges.

**Fig. 4** Explanatory (physiographic) variables used in the multilinear regression: a) elevation; b) slope (from kernel 15); c) curvature (from kernel 3); d) coordinate X; e) coordinate Y; f) distance to coast; g) aspect (from kernel 3); h) NDVI (from kernel 3) only for the wet period model, and i) radiation (from kernel 15) only for the dry period model. Note: Aspect has been divided into SINA and COSA

Observed rainfall in the wet and dry periods was correlated with each of the explanatory variables. Since three different kernel sizes were used and two periods were analysed, this resulted in six correlation matrices. After finding the best correlation, 10 explanatory variables were selected for each period. The resulting lowest correlation of rainfall with the explanatory variables is -0.064 (in the wet period) and the highest is 0.768 (in the dry period) (Table 1).

**Table 1** Correlation matrix between explanatory variables: a) wet period (W); b) dry period (D). The numbers after S, C, SINA, COSA, NDVI_W, R_W, NDVI_D, and R_D indicate the size of the kernel that yields the best correlation

| a) | P_W | H | S15 | C3 | SINA9 | COSA3 | DC | X | Y | NDVI_W3 | R_W15 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| P_W | 1 | 0.543 | 0.296 | 0.084 | -0.176 | 0.101 | -0.083 | 0.146 | -0.195 | -0.064 | 0.107 |
| H | | 1 | 0.380 | 0.090 | -0.089 | 0.014 | 0.597 | 0.316 | 0.322 | -0.028 | 0.391 |
| S15 | | | 1 | 0.048 | -0.048 | 0.089 | 0.133 | 0.035 | 0.027 | 0.090 | -0.248 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| C3 | | | | 1 | 0.027 | -0.084 | 0.121 | -0.012 | 0.073 | -0.042 | 0.313 |
| SINA9 | | | | | 1 | -0.023 | 0.003 | 0.120 | 0.011 | 0.006 | -0.023 |
| COSA3 | | | | | | 1 | -0.053 | 0.031 | -0.014 | -0.024 | -0.590 |
| DC | | | | | | | 1 | 0.364 | 0.760 | -0.001 | 0.406 |
| X | | | | | | | | 1 | 0.061 | -0.304 | 0.345 |
| Y | | | | | | | | | 1 | -0.019 | 0.252 |
| NDVI_W3 | | | | | | | | | | 1 | -0.051 |
| R_W15 | | | | | | | | | | | 1 |

| b) | P_D | H | S15 | C3 | SINA3 | COSA9 | DC | X | Y | NDVI_D15 | R_D15 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| P_D | 1 | $\boxed{0.768}$ | 0.330 | 0.135 | -0.117 | 0.113 | 0.476 | 0.345 | 0.363 | 0.222 | 0.464 |
| H | | 1 | 0.380 | 0.090 | -0.114 | 0.081 | 0.597 | 0.316 | 0.322 | 0.183 | 0.626 |
| S15 | | | 1 | 0.048 | -0.029 | 0.134 | 0.133 | 0.035 | 0.027 | 0.316 | -0.222 |
| C3 | | | | 1 | -0.018 | 0.023 | 0.121 | -0.012 | 0.073 | -0.028 | 0.327 |
| SINA3 | | | | | 1 | -0.034 | 0.003 | 0.120 | 0.011 | -0.043 | -0.015 |
| COSA9 | | | | | | 1 | -0.053 | 0.031 | -0.014 | 0.209 | -0.326 |
| DC | | | | | | | 1 | 0.364 | 0.760 | -0.147 | 0.526 |
| X | | | | | | | | 1 | 0.061 | -0.286 | 0.390 |
| Y | | | | | | | | | 1 | -0.045 | 0.320 |
| NDVI_D15 | | | | | | | | | | 1 | -0.173 |
| R_D15 | | | | | | | | | | | 1 |

The multilinear regression models selected are presented in Equations (3) and (4), respectively, for the wet and dry periods. These two equations were obtained for a 95% confidence level, and selected after evaluating $R^2$, p-value, and the AIC (Wet period: $R^2$=0.692, p-value=$2.2\times10^{-16}$, AIC=721.856; Dry period: $R^2$=0.783, p-value=$2.2\times10^{-16}$, AIC=324.801).

$$P\_W = 47.090 + (9.909 * 10^{-2} * H) + (1.414 * 10^{-4} * X) - (3.535 * 10^{-4} * DC) + (47.660 * NDVI_{W3}) - (3.363 * 10^{-4} * X * NDVI_{W3}) \tag{3}$$

$$P\_D = 5.849 + (2.58 * 10^{-2} * H) + (6.799 * 10^{-5} * X) + (4.625 * 10^{-5} * Y) - (3.625 * 10^{-5} * DC) + (29.640 * NDVI_{D15}) - (1.594 * 10^{-4} * X * NDVI_{D15}) \tag{4}$$

With *P_W* and *P_D* in mm, and *H*, *X*, *Y*, and *DC* in m. *NDVI_W3* and *NDVI_D15* are dimensionless.

Fig. 4 shows the evaluation of residuals for the multilinear regression according to the homogeneity of variance, normality, independence, and Cook’s distance. The four assumptions for validation of residuals are fulfilled for the wet and dry models. In both these models, the homogeneity of variance and the independence present a random distribution. Q-Q plot shows a distribution of residuals close to a normal distribution. Cook’s distance is always <0.5, except for station 152 in the dry period model which is slightly over 1.0; thus, no sample influences the residuals (Cook & Weisberg, 1982).

**Fig. 5** Evaluation of the residuals according to homogeneity of variance, normality, independence, and Cook’s distance: a) wet period and b) dry period, both for the years 1980-2020

The multilinear regression was validated, and the explanatory variables interpolated for the study area resulting in two rainfall models (wet and dry periods). Simultaneously, the historical rainfall data was interpolated by IDW. Table 2 shows the main statistics for the rainfall multilinear regression models, the IDW-interpolated historical rainfall models, and the differences between the regression and interpolated models, for both periods.

**Table 2** Summary of statistics from the models obtained by multilinear regression, IDW-interpolation, and the differences between both (units are in mm)

| | | MIN | MAX | MEAN | STANT.DEV |
|---|---|---|---|---|---|
| Model Regression | Wet period | 37.93 | 123.17 | 68.41 | 11.14 |
| Model IDW | | 45.38 | 136.26 | 69.73 | 13.73 |
| Difference IDW-Regression | | -25.26 | 32.53 | 1.30 | 9.46 |
| Model Regression | Dry | 11.69 | 40.72 | 22.94 | 4.01 |
| Model IDW | | 14.34 | 43.55 | 23.06 | 4.72 |

| | | | | |
|---|---|---|---|---|
| Difference IDW-Regression | -9.79 | 8.78 | 0.12 | 2.64 |

The rainfall models obtained by regression (from the physiographic variables) and by IDW interpolation (from the historical records), as well as their differences, are summarised in Fig. 5 and 6. In the wet period (Fig. 5), a strong similarity between these two models is clearly visible. Monthly rainfall ranges from 38 to 123 mm in the regression model, and from 45 to 136 mm in the IDW model, while the average rainfall is 68 and 69 mm, respectively (Table 2). Rainfall occurs mostly in the mountainous areas, while on the river valleys drops up to 60%. Three positive rainfall anomalies are clearly detectable, corresponding to the most important orographic structures of the study area and where the altitude is maximum (*Sierra de Aracena*, *Serra de Monchique*, and *Serra do Caldeirão*). The maximum difference between both models is 32 mm. However, in 88% of the study area (25,397 km$^2$) this difference is less than 15 mm, *i.e.*, approximately 10% of the maximum rainfall obtained by multilinear regression (123.17 mm).

**Fig. 6** Monthly average rainfall in the wet (October-March) period from 1980 to 2020: a) Multilinear regression model (R); b) IDW interpolation model; c) Differences between both models (IDW-R)

In the dry period (Fig. 6), monthly rainfall ranges from 11 to 40 mm in the regression model, and 14 to 43 mm in the IDW model. Average rainfall in both models is approximately 23 mm (Table 2). Similarly to the wet period rainfall is higher in the mountainous areas; however, the minimum values are found near the coast where rainfall drops around 40%. The maximum difference between both models is 9 mm, *i.e.*, 20% of the maximum rainfall obtained by multilinear regression in this period (40.72 mm). Nonetheless, in 90% of the study area (25,974 km$^2$) this difference is less than 4.5 mm.

**Fig. 7** Monthly average rainfall in the dry (April-September) period from 1980 to 2020: a) Multilinear regression model (R); b) IDW interpolation model; c) Differences between both models (IDW-R)

## Discussion

Mapping of the spatial variability of rainfall from a set of physiographic variables and using multilinear regression was tested in a large area of 28,860 km$^2$ of the southern Iberian Peninsula. The area under study shows an important variety of physiographic features (e.g., mountains, coast, great floodplains...). The density of raingauges across the study area varies significantly. From these raingauges 40 years (1980-2022) of rainfall records were selected, a longer period than the minimum stablished by WMO (2020) for computing climate data (30 years). The analysis of the spatial variability of rainfall was separated in two periods, the wet (October to March) and the dry (April to September) semesters. These periods were defined after previous studies such as Feidas et al. (2013) that modelled and mapped rainfall in Greece, based on 18 topographic and geographic parameters, by seasons (spring, summer, autumn and winter) and annually. One of the conclusions of this work was that regression models do not provide a good correlation between rainfall and the explanatory variables for the transition seasons (spring and autumn). Thus, in this study, we opted to model for the wet and dry semesters.

In the course of the proposed methodological development some limitations must be considered. Regarding the raw rainfall data, according to the Official Networks of Hydrometeorological, some stations were set in unreasonable locations, such as over dam reservoirs and the sea, wherefore their coordinates were corrected. Moreover, two stations were discarded after the analysis of the residuals, which showed systematic errors in the raw data. A thorough assessment of the consistency of the raw data and a meticulous analysis to find errors were carried out, to improve the quality of the regression model, i.e., a model that is more representative of the reality. On the other hand, some explanatory variables shown errors in particular locations, such as the variable NDVI that returned negative values in water bodies like reservoirs or large rivers. This induced a localised sharp decrease of rainfall in the model, which was overcome by applying a low-pass filter to smooth it out. Moreover, when the orography is highly irregular the regression model also suffers from grater inaccuracy. This was tackled through the assessment of how different kernel sizes in the physiographic models can lead to more accurate results, a novel aspect that is approach not included in any work related to this subject, as far the authors of this study can acknowledge. This assessment was carried out using basic statistics (mean, maximum, minimum, and standard deviation). After several tests, three kernel sizes were selected (kernel 3, 9, and 15, which is equivalent to calculation windows of 90, 270 and 450 m in real size). The maximum kernel size was conditioned by the available computational processing capacity. Kernel 3 and kernel 9 models led to important differences, whilst differences between kernel 9 and kernel 15 showed to be minimal. The explanatory variable most affected by the kernel size was curvature, with differences up to 80% in the correlation with rainfall when considering the different kernel sizes. These differences ranged from 10 to 55% for the remaining explanatory variables. Larger kernel sizes generally lead to stronger correlations as irregularities and extreme values were smoothed.

In this study, the explanatory variables with a stronger correlation with rainfall, thus selected for the the multilinear regression models, were elevation (H), distance to coast (DC), coordinates X and Y, and NDVI. Coordinate Y was only used in the wet semester model. Table 3 shows the relative importance of these variables to the regression models. The remaining physiographic variables, i.e., slope (S), curvature (C), aspect (SINA and COSA), and radiation (R) have been shown very little influence on the spatial distribution of rainfall, what is important in itself to be highlighted here.

**Table 3** Influence of physiographic variables in the multilinear regression

| | H | X | Y | DC | NDVI |
|---|---|---|---|---|---|
| WET | 0.9767 | 0.5294 | -- | -0.6524 | 0.3130 |
| DRY | 0.7256 | 0.7265 | 0.3204 | -0.1909 | 0.4527 |

In the wet period rainfall is mainly affected by altitude (H). Higher rainfall values are concentrated in the hilly areas of *Aracena* and *Monchique* mountain ranges, respectively located in the northeast and in the southwest of the study area. Excluding these areas with higher elevation, rainfall variability is mostly dependent of the distance to coast (DC). This variable presents an inverse variation with rainfall, i.e., the greater the distance from the coast, the less rainfall occurs in the north central zone, which corresponds to the alluvial plains of the *Guadiana* River. The geographical longitude (variable X) has a medium influence on rainfall, which is partially justified by the orography, as the main orographic structure within the study area (*Aracena* mountain range) is located to the east. On the other hand, the geographical latitude (variable Y) does not influence the variability of rainfall in the wet period. This is justified as rainfall events in the wet period are mostly cyclonic, originated by Atlantic frontal systems moving from west to east. Finally, NDVI shows a moderate correspondence with rainfall due to anthropogenic influence, as large areas of the studied zone are occupied by farming greenhouses.

In the dry period, the explanatory variables with stronger correlation to rainfall are coordinate X and elevation (H), both presenting a similar weight in the multilinear regression. As explained before, this happens since both variables are related to orography as the higher relief is in the northeast of the study area. Moreover, in dry periods, rainfall is mostly convective in the areas with a smooth orography such as alluvial plains. In these areas, warm and moist air promote the formation of Cumulonimbus clouds, leading to localised rainstorm events. This also explain the influence of coordinate Y, although with a weaker correlation. Finally, NDVI has a higher influence in the dry period when compared to the wet period, due to the irregularity of water distribution that lead to a localised scarcity. As expected they present a positive correlation. Also, rainfall events during the driest and warmest periods are sometimes linked to a particular synoptic phenomenon where a small-scale storm located over the Gulf of Cádiz forces the atmospheric circulation from south to north due to its counterclockwise rotation. This phenomenon happens less frequently during the wet period.

Stablishing comparisons from studies in other regions of the Iberian Peninsula and in the rest of the world, Marquínez et al. (2003) developed a multiple linear regression model of rainfall for the mountains regions of Cantabria (Spain). This model was developed using data collected between 1966 to 1990 from 117 weather stations, and used five topographic descriptors as independent variables: elevation, distance from the coastline, distance from the west, and a measurement of elevation and slope averages into homogeneous areas. Rainfall was found to increase mostly with elevation, as was also observed in the present study, although the authors noticed a sharper rainfall gradient on elevations over 1000 m. They also identified peculiar rainfall patterns linked to variables distance from the west and distance to the sea (coordinates X and DC in the present study), which were related to the locations of the dominant frontal systems. Feidas et al. (2013), noticed that western continental Greece receives higher depths of rainfall than the eastern part of the mainland country, as illustrated by the positive correlation with the distance to Aegean Sea. This is also due to areas with higher elevation, that show a positive and stronger correlation with rainfall, mainly in the winter. Moreover, while the NDVI was found to be strongly and positively correlated with rainfall all over the year due to the dependence of vegetation vigour on the total amount of rainfall in an area, the variable elevation was excluding in the summer model, with a high importance in the winter model.

The relation between DC, coordinate X and H and rainfall has been identified in the present study. This relation is particularly noticeable in the wet season, while in the dry season H shows a lower weight in the multilinear regression. Other authors also found similar correlations. Jin et al. (2016) used multiple linear regression to study the annual average rainfall distribution in the middle reaches of the Yellow River basin, in the north of China. Records from 432 weather stations were correlated with explanatory variables such as NDVI, solar radiation, slope, and aspect. Rainfall showed to have the strongest positive correlation with NDVI, followed by solar radiation, DEM, and slope, all with a significance level of 0.01. This is justified by the direct relation between rainfall depth and NDVI, thereby indicating that vegetation growth in the region is closely linked to rainfall. However, unlike the two previously cited studies, in this region NDVI shows a moderate relation with rainfall due to the anthropogenic influence. Finally, Brown and Corine (2002) developed a statistical modelling technique based on five topographic explanatory variables to estimate the mean winter rainfall in New Mexico (USA) for the period of 1961-1990. Whilst the resulting model justified 63% of the spatial variance during cross-validation, the authors concluded including physiographic parameters such as distance to coast or other proximity variables could improve the model results, as it was incorporated in this study.

Rainfall models based on physiographic variables provided a good accuracy (Portalés et al., 2010). Having this in mind, the multilinear regression presents advantages over classical methodologies. For example, the classical interpolation is limited by the density and spacing of raingauges (Bárdossy & Pegram, 2013; Hurtado et al., 2021), whilst multilinear regression based on physiographic variables can provide a stronger spatial continuity and wider coverage with a good accuracy. Nonetheless, an important prediction error in interpolation can appears at the study area boundaries, known as ‘edge effects’. These effects can be mitigated by taking into account rainfall data from raingauges outside of the study area but close to its boundaries (Borges et al., 2016). However, this supposes an important processing of extra information that it is not required in statistical methods. Other limitations of classic interpolation of rainfall are (Bárdossy & Pegram, 2013; Newman et al., 2015): a) the necessity of specific handling of zeros; b) interpolation quality has to be assessed (squared errors and possible bias); c) uncertainty at larger spatial scales is of great importance; and d) rainfall is strongly influenced by topography.

In this study, rainfall maps generated by, IDW interpolation (from historical records) and multilinear regression (from physiographic variables) were compared. Fig. 8 shows the differences between the IDW and the multilinear regression models. In approximately 45% of the study area, i.e., 13,000 km$^2$, the adjustment between both models is very good, showing differences below 15% (<5mm for the wet period and <1.5mm for the dry period). This strong goodness-of-fit occur in areas with homogeneous orography, but not necessarily where more data is available, thus showing the suitability of this type of physiographic-based multilinear rainfall regression model when the number of raingauges is too low or has an irregular spatial distribution. In almost 30% of the area under study (8,650 $km^2$) a good correspondence between both models can be observed, with differences of rainfall ranging from 15 to 30% (5-10 mm for the wet period and 1.5-3.0 mm for the dry period). 15% of the study area (4,300 $km^2$) presents a reasonable matching between the IDW and physiographic-based models, with differences of 10-15 mm for the wet period and 3.0-4.5 mm for the dry period (30-45% of the maximum difference). Only 10% (2,880 $km^2$) of the study area shows a low adjustment between both models. This lack of adjustment is more marked in the hilly areas, where raw data do not have a normal distribution and is not stationary, and so the classical methods of interpolation are not too efficient (Borges et al., 2016).

**Fig. 8** Differences between IDW-Regression models by area

This research presents new insights on rainfall modelling based on physiographic variables. The methodology used in this study was applied to the South of the Iberian Peninsula; however, it can be reproduced in other locations with similar climatic and physiographic features. The analysis was divided into wet and dry semesters to consider, among other factors, the anomalies in rainfall patterns mainly caused by frontal systems entering from the south in cold periods and from the west in warm periods. A meticulous completion of the historical series has been carried out, which is not visible in all the works related to this subject (Feidas et al., 2013; H. Fernandez et al., 2020; Jin et al., 2016), but will lead to more accurate results. Moreover, this study included the analysis of how cell size (kernel size) on influences the correlation of rainfall and physiographic variables, and it helps to smooth out imprecisions of raw data which, as far the authors know, it hasn't been studied yet.

## Conclusions

This study presents a methodology to estimate the spatial variability of rainfall based on physiographic variables. From an initial set of 10 physiographic variables, five of these showed a strong correlation with 40-years of historical rainfall records (1980-2020). The five explanatory variables selected to generate the regression model were: elevation (H), distance to coast (DC), geographic coordinates (X, Y), and the Normalized Difference Vegetation Index in the wet (NDVI_W) and dry (NDVI_D) periods. The rainfall historical records were obtained from 163 weather stations, selected from an initial set of 230, after discarding the ones with large gaps in the data series, weak homogeneity and robustness, or erroneous raw data.

The physiographic-based model was developed and applied to a large area of the southwestern Iberian Peninsula, comprising the Portuguese regions of Algarve, Baixo Alentejo, and Alentejo Litoral, as well as the Spanish province of Huelva, totalising 28,860 km$^2$. This region is characterised by a strong seasonality with relatively mild winters and very warm summers, typical of Mediterranean regions. To better simulate and represent this seasonality, two regression equations were generated, specifically for the wet ($R^2$=0.692) and for the dry ($R^2$=0.783) semesters, respectively from April to September, and from October to March. The rainfall spatial distribution maps generated from these two equations were compared against maps generated by interpolating the historical rainfall records. From this comparison, and for the area under study, it became possible to conclude that:

- A meticulous and thorough analysis to detect errors in the raw data is fundamental to attain a quality regression model.

- In the wet period, elevation (H) and distance to coast (DC) can explain a substantial part of rainfall depth, and DC shows an inverse relation with rainfall;

- In the dry period, coordinate X shows the heaviest influence on rainfall depth followed by altitude (H), partially justified by the occurrence of convective rainstorms in the dry semester;

- In areas with a smooth orography, a very good accuracy has been found between classical interpolation and multilinear regression results;

- Generally, when the orography is highly irregular, the regression models' accuracy diminishes. This limitation can be minimised by testing different kernel sizes for some explanatory variable models;

- Multilinear regression is a good tool to estimate rainfall in scarcely gauged areas or with an irregular distribution of raingauges; moreover, it can mitigate some undesirable issues related to interpolation such as the edge effect;

- In hilly areas, the proposed methodology presents a lower adjustment when compared to areas with a smooth orography. However, in these conditions, the reliability is better than other classic methodologies such as interpolation by IDW.

Physiographic-based models can be successfully used to estimate rainfall in areas with a sparse distribution of raingauges (e.g. for environmental studies in remote areas) or to refine rainfall data resolution for local-scale analysis. A deeper knowledge about the relation between rainfall patterns and physiographic variables would allow to continue with futures research directions in this area such as defining the variability of the drought in the area using indexes like the SPI (Vélez-Nicolás et al., 2022) or quantifying runoff

(Zazo et al., 2020). Both are of special relevance in this zone where the importance of irrigated agriculture and the significant water stress requires better-informed decision making and the elaboration of early warning systems. Moreover, this information can aid in the design of decision support systems (Ruiz-Ortiz et al., 2019), in the area to analyse current water management and propose several strategies under different rainfall scenarios.

## References

Abatzoglou, J. T., & Ficklin, D. L. (2017). Climatic and physiographic controls of spatial variability in surface water balance over the contiguous United States using the Budyko relationship. *Water Resources Research*, *53*(9), 7630–7643. https://doi.org/10.1002/2017WR020843

Adhikary, S. K., Muttil, N., & Yilmaz, A. G. (2017). Cokriging for enhanced spatial interpolation of rainfall in two Australian catchments. *Hydrological Processes*, *31*(12), 2143–2161. https://doi.org/10.1002/hyp.11163

Akbar, R., Short Gianotti, D., McColl, K. A., Haghighi, E., Salvucci, G. D., & Entekhabi, D. (2018). Hydrological Storage Length Scales Represented by Remote Sensing Estimates of Soil Moisture and Precipitation. *Water Resources Research*, *54*(3), 1476–1492. https://doi.org/10.1002/2017WR021508

Almodóvar, G. R., Yesares, L., Sáez, R., Toscano, M., González, F., & Pons, J. M. (2019). Massive sulfide ores in the iberian pyrite belt: Mineralogical and textural evolution. *Minerals*, *9*(11), 653. https://doi.org/10.3390/min9110653

Bárdossy, A., & Pegram, G. (2013). Interpolation of precipitation under topographic influence at different time scales. *Water Resources Research*, *49*(8), 4545–4565. https://doi.org/10.1002/WRCR.20307

Barsi, J. A., Schott, J. R., Hook, S. J., Raqueno, N. G., Markham, B. L., & Radocinski, R. G. (2014). Landsat-8 thermal infrared sensor (TIRS) vicarious radiometric calibration. *Remote Sensing*, *6*(11), 11607–11626. https://doi.org/10.3390/rs61111607

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future köppen-geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*(1), 1–12. https://doi.org/10.1038/sdata.2018.214

Berger, K. P., & Entekhabi, D. (2001). Basin hydrologic response relations to distributed physiographic descriptors and climate. *Journal of Hydrology*, *247*(3–4), 169–182. https://doi.org/10.1016/S0022-1694(01)00383-3

Berndtsson, R., & Niemczynowicz, J. (1988). Spatial and temporal scales in rainfall analysis - Some aspects and future perspectives. *Journal of Hydrology*, *100*(1–3), 293–313. https://doi.org/10.1016/0022-1694(88)90189-8

Borges, P. de A., Franke, J., da Anunciação, Y. M. T., Weiss, H., & Bernhofer, C. (2016). Comparison of spatial interpolation methods for the estimation of precipitation distribution in Distrito Federal, Brazil. *Theoretical and Applied Climatology*, *123*(1–2), 335–348. https://doi.org/10.1007/s00704-014-1359-9

Brown, D. P., & Comrie, A. C. (2002). Spatial modeling of winter temperature and precipitation in Arizona and New Mexico, USA. *Climate Research*, *22*(2), 115–128. https://doi.org/10.3354/cr022115

Castiglioni, S., Castellarin, A., Montanari, A., Skøien, J. O., Laaha, G., & Blöschl, G. (2011). Smooth regional estimation of low-flow indices: Physiographical space based interpolation and top-kriging. *Hydrology and Earth System Sciences*, *15*(3), 715–727. https://doi.org/10.5194/hess-15-715-2011

CEDEX (Centro de Estudios y experimentación de obras públicas). (2013). *Cálculo hidrometeorológico de aportaciones y crecidas. Manual CHAC* (5.06 beta1). Ministerio de Transportes, movilidad y Agenda Urbana. https://ceh.cedex.es/chac/

Chandwani, V., Vyas, S. K., Agrawal, V., & Sharma, G. (2015). Soft Computing Approach for Rainfall-runoff Modelling: A Review. *Aquatic Procedia*, *4*, 1054–1061. https://doi.org/10.1016/j.aqpro.2015.02.133

Chow, V. Te, Maidment, D. R., & Mays, L. W. (1988). *Applied Hydrology* (N. Y. McGraw-Hill Book Company (ed.)). McGraw. http://ponce.sdsu.edu/Applied_Hydrology_Chow_1988.pdf

Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression* (C. and H. New York (ed.); Retrieved). Chapman and Hall. https://hdl.handle.net/11299/37076

Direção Regional de Agricultura e Pescas do Centro (DRAP). (n.d.). *Agrometeorologia*. Retrieved March 11, 2021, from https://www.drapc.gov.pt/base/agrometeorologia.htm

Durocher, M., Burn, D. H., Mostofi Zadeh, S., & Ashkar, F. (2019). Estimating flood quantiles at ungauged sites using nonparametric regression methods with spatial components. *Hydrological Sciences Journal*, *64*(9), 1056–1070. https://doi.org/10.1080/02626667.2019.1620952

EC (European Commission). (2008). *Natura 2000 - Protecting Europe´s biodiversity* (Susanne Wegefelt, Nature & Biodiversity Unit, & D. Environment (eds.); Oxford, UK). https://doi.org/10.2779/45963

Edwards, K. A. (1973). Estimating areal rainfall by fitting surfaces to irregularly spaced data. In W. M. O. WMO (Ed.), *Proc. of the International Symposium on the Distribution of Precipitation in Mountainous Areas* (pp. 565–587). http://hydrologie.org/redbooks/a106/iahs_106_V1_0000.pdf

Fan, J. C., Chang, S. C., Liao, K. W., Guo, J. J., Liu, C. H., Chang, Y. C., Huang, C. L., & Yang, C. H. (2018). The impact of physiographic factors upon the probability of slides occurrence: a case study from the Kaoping River Basin, Taiwan. *Journal of the Chinese Institute of Engineers, Transactions of the Chinese Institute of Engineers,Series A*, *41*(5), 419–429. https://doi.org/10.1080/02533839.2018.1482236

Fehmi, J. S., Niu, G. Y., Scott, R. L., & Mathias, A. (2014). Evaluating the effect of rainfall variability on vegetation establishment in a semidesert grassland. *Environmental Monitoring and Assessment*, *186*(1), 395–406. https://doi.org/10.1007/S10661-013-3384-Z

Feidas, H., Karagiannidis, A., Keppas, S., Vaitis, M., Kontos, T., Zanis, P., Melas, D., & Anadranistakis, E. (2013). Modeling and mapping temperature and precipitation climate data in Greece using topographical and geographical parameters. *Theoretical and Applied Climatology*, *118*(1–2), 133–146. https://doi.org/10.1007/S00704-013-1052-4

Fernandez, H. M., Martins, F. M. G., Isidoro, J. M. G. P., Zavala, L., & Jordán, A. (2016). Soil erosion, Serra de Grândola (Portugal). *Journal of Maps*, *12*(5), 1138–1142. https://doi.org/10.1080/17445647.2015.1135829

Fernandez, H., Martins, F., & Isidoro, J. M. G. P. (2020). Mapping rainfall aggressiveness from physiographical data: application to the Grândola Mountain Range (Alentejo, Portugal).

*Physical Geography*, *41*(5), 451–466. https://doi.org/10.1080/02723646.2019.1674557

Fraga, I., Cea, L., & Puertas, J. (2019). Effect of rainfall uncertainty on the performance of physically based rainfall–runoff models. *Hydrological Processes*, *33*(1), 160–173. https://doi.org/10.1002/hyp.13319

Fu, P., & Rich, P. M. (2002). A geometric solar radiation model with applications in agriculture and forestry. *Computers and Electronics in Agriculture*, *37*(1–3), 25–35. https://doi.org/10.1016/S0168-1699(02)00115-1

Ghiglieri, G., Carletti, A., & Pittalis, D. (2014). Runoff coefficient and average yearly natural aquifer recharge assessment by physiography-based indirect methods for the island of Sardinia (Italy) and its NW area (Nurra). *Journal of Hydrology*, *519*(PB), 1779–1791. https://doi.org/10.1016/j.jhydrol.2014.09.054

Ghumman, A. R., Hassan, I., Khan, Q. U. Z., & Kamal, M. A. (2013). Investigation of impact of environmental changes on precipitation pattern of Pakistan. *Environmental Monitoring and Assessment*, *185*(6), 4897–4905. https://doi.org/10.1007/S10661-012-2911-7

González-Hidalgo, J. C., de Luis Arrillaga, M., Stepánek, P., Raventós Bonvehí, J., & Cuadrats Prats, J. M. (2002). Reconstrucción, estabilidad y proceso de homogeneizado de series de precipitación en ambientes de elevada variabilidad pluvial. In S. M. V. y M. A. S. J. M. Cuadrat (Ed.), *VII Reunión Nacional de Climatología: La información climática como herramienta de gestión ambiental* (pp. 47–58). Universidad de Zaragoza.

Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, *228*(1–2), 113–129. https://doi.org/10.1016/S0022-1694(00)00144-X

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Guillemette, N., St-Hilaire, A., Ouarda, T. B. M. J., Bergeron, N., Robichaud, É., & Bilodeau, L. (2009). Feasibility study of a geostatistical modelling of monthly maximum stream temperatures in a multivariate space. *Journal of Hydrology*, *364*(1–2), 1–12. https://doi.org/10.1016/j.jhydrol.2008.10.002

Harris, R., & Jarvis, C. (2011). *Statistics for Geography and Environmental Science* (Routledge (ed.); 1st Editio). Taylor & Francis. https://www.routledge.com/Statistics-for-Geography-and-Environmental-Science/Harris-Jarvis/p/book/9780131789333

Hevesi, J. A., Istok, J. D., & Flint, A. L. (1992). Precipitation Estimation in Mountainous Terrain Using Multivariate Geostatistics. Part I: Structural Analysis. *Journal of Applied Meteorology and Climatology*, *31*(7), 661–676. https://doi.org/https://doi.org/10.1175/1520-0450(1992)031<0661:PEIMTU>2.0.CO;2

Hu, Q., Li, Z., Wang, L., Huang, Y., Wang, Y., & Li, L. (2019). Rainfall spatial estimations: A review from spatial interpolation to multi-source data merging. In *Water (Switzerland)* (Vol. 11, Issue 3, p. 579). MDPI AG. https://doi.org/10.3390/w11030579

Hundecha, Y., Ouarda, T. B. M. J., & Bárdossy, A. (2008). Regional estimation of parameters of a rainfall-runoff model at ungauged watersheds using the “spatial” structures of the parameters within a canonical physiographic-climatic space. *Water Resources Research*, *44*(1), 1427. https://doi.org/10.1029/2006WR005439

Hurtado, S. I., Zaninelli, P. G., Agosta, E. A., & Ricetti, L. (2021). Infilling methods for monthly precipitation records with poor station network density in Subtropical Argentina. *Atmospheric Research*, *254*, 105482. https://doi.org/10.1016/J.ATMOSRES.2021.105482

Jin, Q., Zhang, J., Shi, M., & Huang, J. (2016). Estimating loess plateau average annual precipitation with multiple linear regression kriging and geographically weighted regression kriging. *Water (Switzerland)*, *8*(6), 266. https://doi.org/10.3390/W8060266

Junta de Andalucía (JA). (n.d.). *Red de Información Agroclimática de Andalucía (RIA) | Instituto de Investigación y Formación Agraria y Pesquera (IFAPA)*. Retrieved March 11, 2021, from https://www.juntadeandalucia.es/agriculturaypesca/ifapa/riaweb/web/

Kim, D. (2013). Incorporation of multi-scale spatial autocorrelation in soil moisture-landscape modeling. *Physical Geography*, *34*(6), 441–455. https://doi.org/10.1080/02723646.2013.857267

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World Map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, *15*(3), 259–263. https://doi.org/10.1127/0941-2948/2006/0130

Kuentz, A., Arheimer, B., Hundecha, Y., & Wagener, T. (2017). Understanding hydrologic variability across Europe through catchment classification. *Hydrology and Earth System Sciences*, *21*(6), 2863–2879. https://doi.org/10.5194/hess-21-2863-2017

Kumari, M., Singh, C. K., & Basistha, A. (2017). Clustering Data and Incorporating Topographical Variables for Improving Spatial Interpolation of Rainfall in Mountainous Region. *Water Resources Management*, *31*(1), 425–442. https://doi.org/10.1007/s11269-016-1534-0

Kyriakidis, P. C., Kim, J., & Miller, N. L. (2001). Geostatistical Mapping of Precipitation from Rain Gauge Data Using Atmospheric and Terrain Characteristics. *Journal of Applied Meteorology and Climatology*, *40*(11), 1855–1877. https://doi.org/https://doi.org/10.1175/1520-0450(2001)040<1855:GMOPFR>2.0.CO;2

Lyon, S. W., Seibert, J., Lembo, A. J., Steenhuis, T. S., & Walter, M. T. (2008). Incorporating landscape characteristics in a distance metric for interpolating between observations of stream water chemistry. *Hydrology and Earth System Sciences*, *12*(5), 1229–1239. https://doi.org/10.5194/hess-12-1229-2008

Marquínez, J., Lastra, J., & García, P. (2003). Estimation models for precipitation in mountainous regions: The use of GIS and multivariate analysis. *Journal of Hydrology*, *270*(1–2), 1–11. https://doi.org/10.1016/S0022-1694(02)00110-5

Mello, C. R., Viola, M. R., Beskow, S., & Norton, L. D. (2013). Multivariate models for annual rainfall erosivity in Brazil. *Geoderma*, *202–203*, 88–102. https://doi.org/10.1016/j.geoderma.2013.03.009

Ministerio para la Transición Ecológica y el Reto Demográfico (MITECO). (n.d.). *Estaciones Climatológicas de la Agencia Estatal de Meteorología (AEMET)*. Retrieved March 11, 2021, from https://www.miteco.gob.es/es/cartografia-y-sig/ide/descargas/otros/default.aspx

Mohamoud, Y. (2004). Comparison of hydrologic responses at different watershed scales. *Office of Research and Development, United States Environmental Protection Agency*, 1–81.

Moliba Bankanza, J. C. (2014). Spatial modeling of summer precipitation over the Czech Republic using physiographic variables. *Geographical Research*, *52*(1), 85–105. https://doi.org/10.1111/1745-5871.12041

Montanari, A., & Di Baldassarre, G. (2013). Data errors and hydrological modelling: The role of model structure to propagate observation uncertainty. *Advances in Water Resources*, *51*, 498–504. https://doi.org/10.1016/j.advwatres.2012.09.007

Moral, F. J. (2010). Comparison of different geostatistical approaches to map climate variables: Application to precipitation. *International Journal of Climatology*, *30*(4), 620–631. https://doi.org/10.1002/joc.1913

Narbondo, S., Gorgoglione, A., Crisci, M., & Chreties, C. (2020). Enhancing physical similarity approach to predict runoff in ungauged watersheds in sub-tropical regions. *Water (Switzerland)*, *12*(2). https://doi.org/10.3390/w12020528

National Administration Space Aeronautics (NASA), & Ministry of Economy Trade and Industry (METI). (n.d.). *Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM)*. Retrieved April 23, 2021, from https://gdemdl.aster.jspacesystems.or.jp/index_en.html

Neves, M. C., Nunes, L. M., & Monteiro, J. P. (2020). Evaluation of GRACE data for water resource management in Iberia: a case study of groundwater storage monitoring in the Algarve region. *Journal of Hydrology: Regional Studies*, *32*, 100734. https://doi.org/10.1016/j.ejrh.2020.100734

Newman, A. J., Clark, M. P., Craig, J., Nijssen, B., Wood, A., Gutmann, E., Mizukami, N., Brekke, L., & Arnold, J. R. (2015). Gridded Ensemble Precipitation and Temperature Estimates for the Contiguous United States. *Journal of Hydrometeorology*, *16*(6), 2481–2500. https://doi.org/10.1175/JHM-D-15-0026.1

Pandey, R. P., Pandey, A., Galkate, R. V, Byun, H.-R., Bimal, ·, Mal, C., Pandey, R. P., Pandey, A., Galkate, R. V, Byun, H.-R., & Mal, B. C. (2010). Integrating Hydro-Meteorological and Physiographic Factors for Assessment of Vulnerability to Drought. *Water Resour Manage*, *24,* 4199–4217. https://doi.org/10.1007/s11269-010-9653-5

Parveen, U., & Sreekesh, S. (2018). Physiographic Influence on Rainfall Variability: A Case Study of Upper Ganga Basin. In E. E. and D. R. R. Climate Change (Ed.), *Sustainable Development Goals Series* (pp. 59–73). Springer, Cham. https://doi.org/10.1007/978-3-319-56469-2_4

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, *11*(5), 1633–1644. https://doi.org/10.5194/hess-11-1633-2007

Penagos Cruz, G. (2014). Variables hidrometereológicas asociadas al cambio climático en Girardot y la Región del Alto Magdalena. *Ambiente y Desarrollo*, *18*(35), 134. https://doi.org/10.11144/Javeriana.AyD18-35.vhac

Portalés, C., Boronat, N., Pardo-Pascual, J. E., & Balaguer-Beser, A. (2010). Seasonal precipitation interpolation at the Valencia region with multivariate methods using geographic and topographic information. *International Journal of Climatology*, *30*(10), 1547–1563. https://doi.org/10.1002/JOC.1988

Pulido-Calvo, I., Gutiérrez-Estrada, J. C., & Sanz-Fernández, V. (2020). Drought and Ecological Flows in the Lower Guadiana River Basin (Southwest Iberian Peninsula). *Water*, *12*(3), 677. https://doi.org/10.3390/w12030677

Rich, P. M., Hetrick, W. A., Saving, S. C., & Dubayah, R. O. (1994). USING VIEWSHED MODELS TO CALCULATE INTERCEPTED SOLAR RADIATION: APPLICATIONS IN ECOLOGY. *American Society for Photogrammetry and Remote Sensing Technical Papers*, 524–529.

Ruiz-Ortiz, V., Garciá-López, S., Solera, A., & Paredes, J. (2019). Contribution of decision support systems to water management improvement in basins with high evaporation in Mediterranean climates. *Hydrology Research*, *50*(4), 1020–1036. https://doi.org/10.2166/nh.2019.014

Ruiz-Ortiz, V., García-López, S., Vélez-Nicolás, M., Sánchez-Bellón, Á., de Villar, A. C., & Contreras, F. (2021). Learning from hydrological and hydrogeological problems in civil engineering. Study of reservoirs in Andalusia, Spain. *Engineering Geology*, *282*, 105916. https://doi.org/10.1016/j.enggeo.2020.105916

Segarra, J., González-Torralba, J., Aranjuelo, Í., Araus, J. L., & Kefauver, S. C. (2020). Estimating wheat grain yield using Sentinel-2 imagery and exploring topographic features and rainfall effects on wheat performance in Navarre, Spain. *Remote Sensing*, *12*(14). https://doi.org/10.3390/rs12142278

Shiau, J.-T., Chen, C.-N., & Tsai, C.-T. (2011). Physiographic Drainage-Inundation Model Based Flooding Vulnerability Assessment. *Water Resources Management*, *26*(5), 1307–1323. https://doi.org/10.1007/S11269-011-9960-5

Singh, A., Singh, R. M., Senthil Kumar, A. R., Kumar, A., Hanwat, S., & Tripathi, V. K. (2021). Evaluation of soft computing and regression-based techniques for the estimation of evaporation. *Journal of Water and Climate Change*, *12*(1), 32–43. https://doi.org/10.2166/wcc.2019.101

Sistema Nacional de Informação de Recursos Hídricos (SNIRH). (n.d.). *Redes de Monitorização*. Retrieved March 11, 2021, from https://snirh.apambiente.pt/

Squintu, A. A., van der Schrier, G., Brugnara, Y., & Klein Tank, A. (2019). Homogenization of daily temperature series in the European Climate Assessment & Dataset. *International Journal of Climatology*, *39*(3), 1243–1261. https://doi.org/10.1002/joc.5874

Strahler, A. N. (1969). Physical Geography. *John Wiley*. https://scholar.google.es/scholar?hl=es&as_sdt=0%2C5&q=Strahler%2C+A.N.+%281969%29.+Physical+Geography.+John+Wiley+%26+Sons%2C+New+York%2C+732+pp&btnG=

Tobin, C., Nicotina, L., Parlange, M. B., Berne, A., & Rinaldo, A. (2011). Improved interpolation of meteorological forcings for hydrologic applications in a Swiss Alpine region. *Journal of Hydrology*, *401*(1–2), 77–89. https://doi.org/10.1016/j.jhydrol.2011.02.010

Unwin, D. J. (1977). Statistical methods in physical geography: *Progress in Physical Geoprаphy*, *1*(2), 185–221. https://doi.org/10.1177/030913337700100201

Vélez-Nicolás, M., García-López, S., Ruiz-Ortiz, V., Zazo, S., & Molina, J. L. (2022). *Precipitation Variability and Drought Assessment Using the SPI: Application to Long-Term Series in the Strait of Gibraltar Area*. https://doi.org/10.3390/w14060884

Vera, J. A. (2004). *Geología de España* (J. A. Vera (ed.); SGE-IGME). https://books.google.pt/books?hl=es&lr=&id=n1SO6IjVhZEC&oi=fnd&pg=PA1&dq=Vera+libro+geología&ots=7GeUv01izE&sig=S-utdKDNq9YpHesSHL1Vb3nDePI&redir_esc=y#v=onepage&q=Vera libro geología&f=false

Waldron, B., Gui, D., Liu, Y., Feng, L., & Dai, H. (2020). Assessing water distribution and agricultural expansion in the Cele Oasis, China. *Environmental Monitoring and Assessment*, *192*(5). https://doi.org/10.1007/S10661-020-8233-2

WMO (World Meteorological Organization). (2020). Climate Data and Monitoring. *Tenth Seminar for Homogenization and Quality Control in Climtologicla Database and Fith Conference on Spatial Interpolation Techniques in Climatology and Meteorology*.

Wood, J. (1996). The geomorphological characterisation of Digital Elevation Models. In *Ph. D. Thesis, University of Leicester, Department of Geography, Leicester, UK.*

Zazo, S., Molina, J.-L., Ruiz-Ortiz, V., Vélez-Nicolás, M., & García-López, S. (2020). Modeling River Runoff Temporal Behavior through a Hybrid Causal–Hydrological (HCH) Method. *Water*, *12*(11), 3137. https://doi.org/10.3390/w12113137

Zhang, C., Zhou, X., & Lei, W. (2019). Necessary length of daily precipitation time series for different entropy measures. *Earth Science Informatics*, *12*(4), 475–487. https://doi.org/10.1007/s12145-019-00392-1

Zhao, G., Xue, H., & Ling, F. (2010). Assessment of ASTER GDEM Performance by Comparing with SRTM and ICESat/GLAS Data in Central China. In Y. Liu & A. Chen

(Eds.), *18th International Conference on Geoinformatics*. IEEE GRSS; Geograph Soc China.

Figure2 Click here to access/download;Figure;Fig2.tif

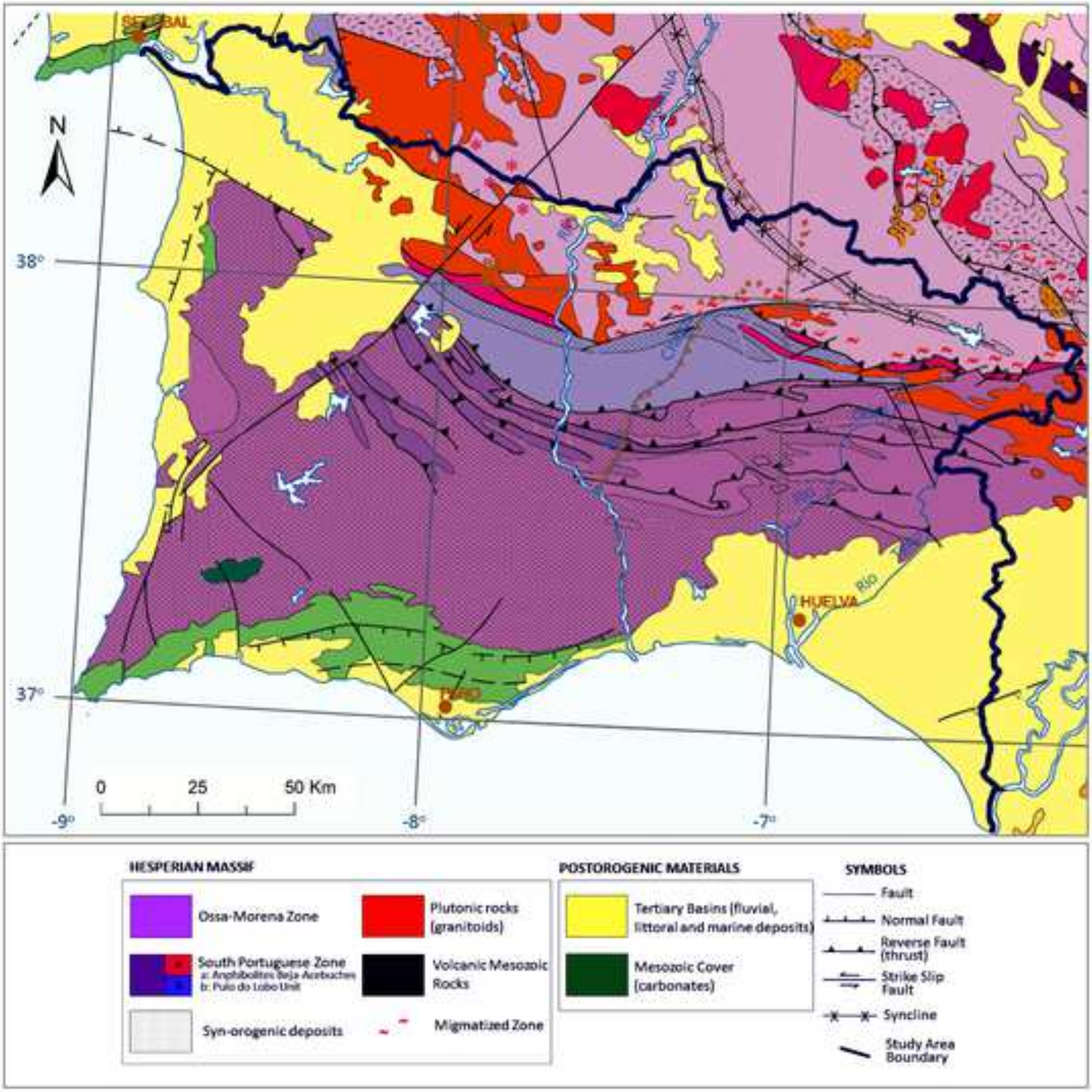


Figure3 

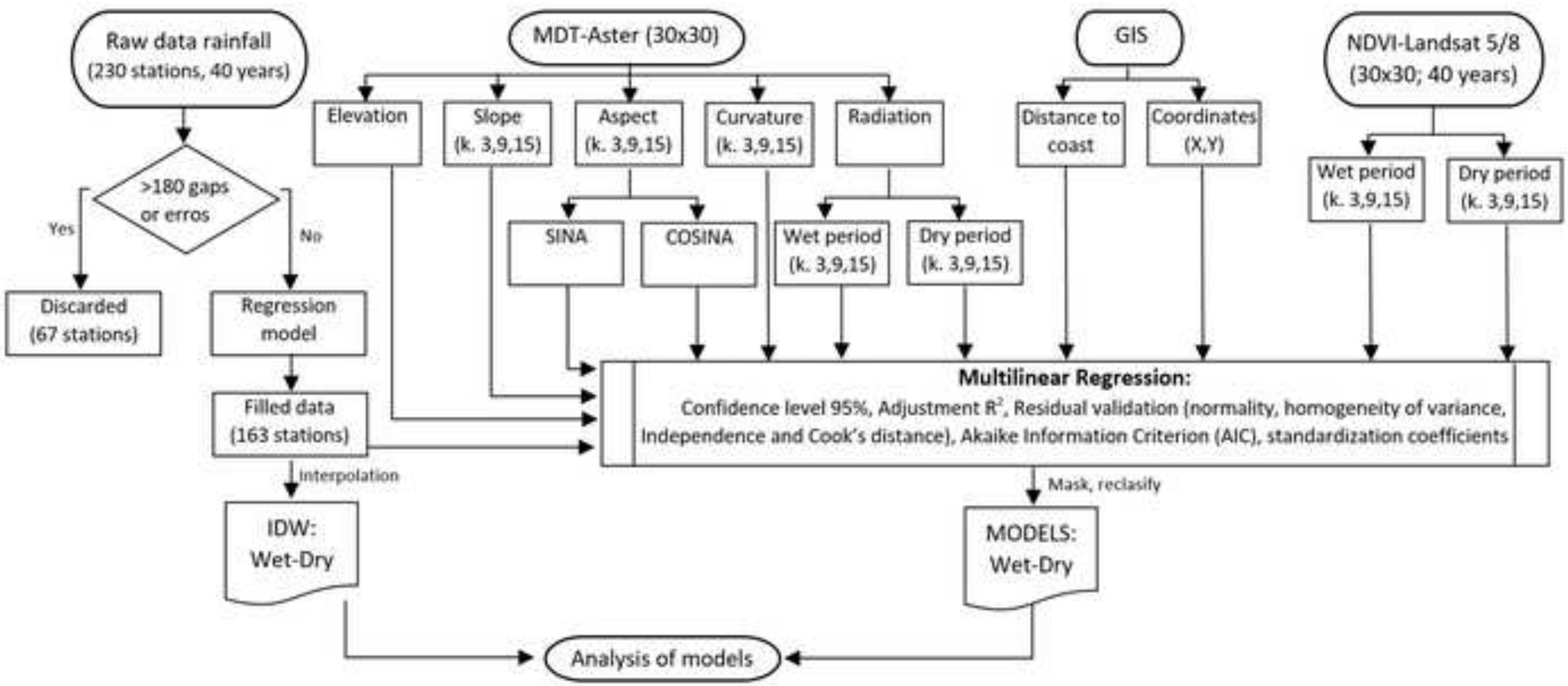

Raw data rainfall
(230 stations, 40 years)
>180 gaps
or erros
Yes
No
Discarded
(67 stations)
Regression
model
Filled data
(163 stations)
Interpolation
IDW:
Wet-Dry
MDT-Aster (30x30)
Elevation
Slope
(k. 3,9,15)
Aspect
(k. 3,9,15)
Curvature
(k. 3,9,15)
Radiation
SINA
COSINA
Wet period
(k. 3,9,15)
Dry period
(k. 3,9,15)
GIS
Distance to
coast
Coordinates
(X,Y)
NDVI-Landsat 5/8
(30x30; 40 years)
Wet period
(k. 3,9,15)
Dry period
(k. 3,9,15)
Multilinear Regression:
Confidence level 95%, Adjustment R², Residual validation (normality, homogeneity of variance,
Independence and Cook's distance), Akaike Information Criterion (AIC), standardization coefficients
Mask, reclasify
MODELS:
Wet-Dry
Analysis of models

Figure4

Click here to access/download;Figure;Fig4.tif

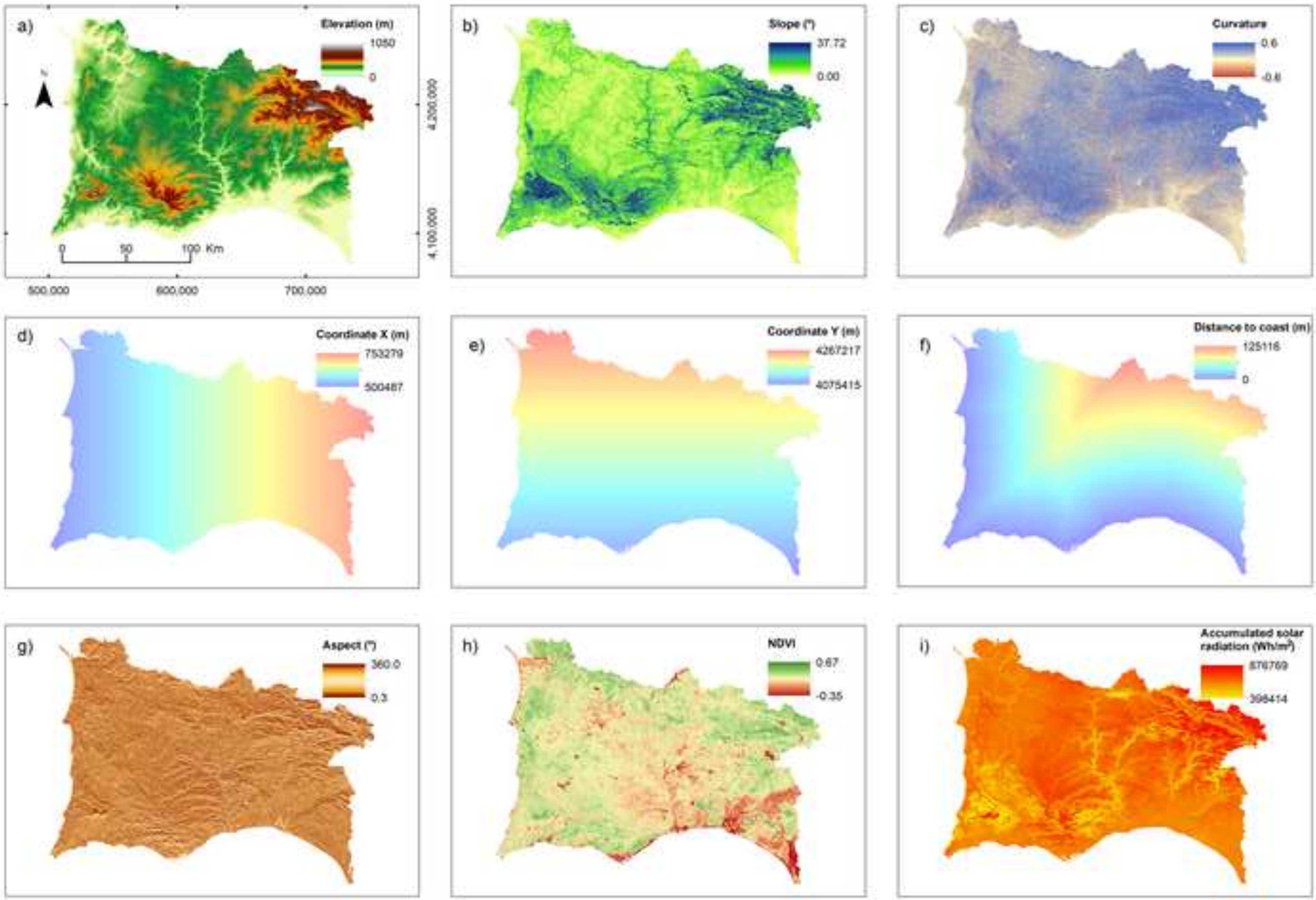


Figure5

Click here to access/download;Figure;Fig5.tif

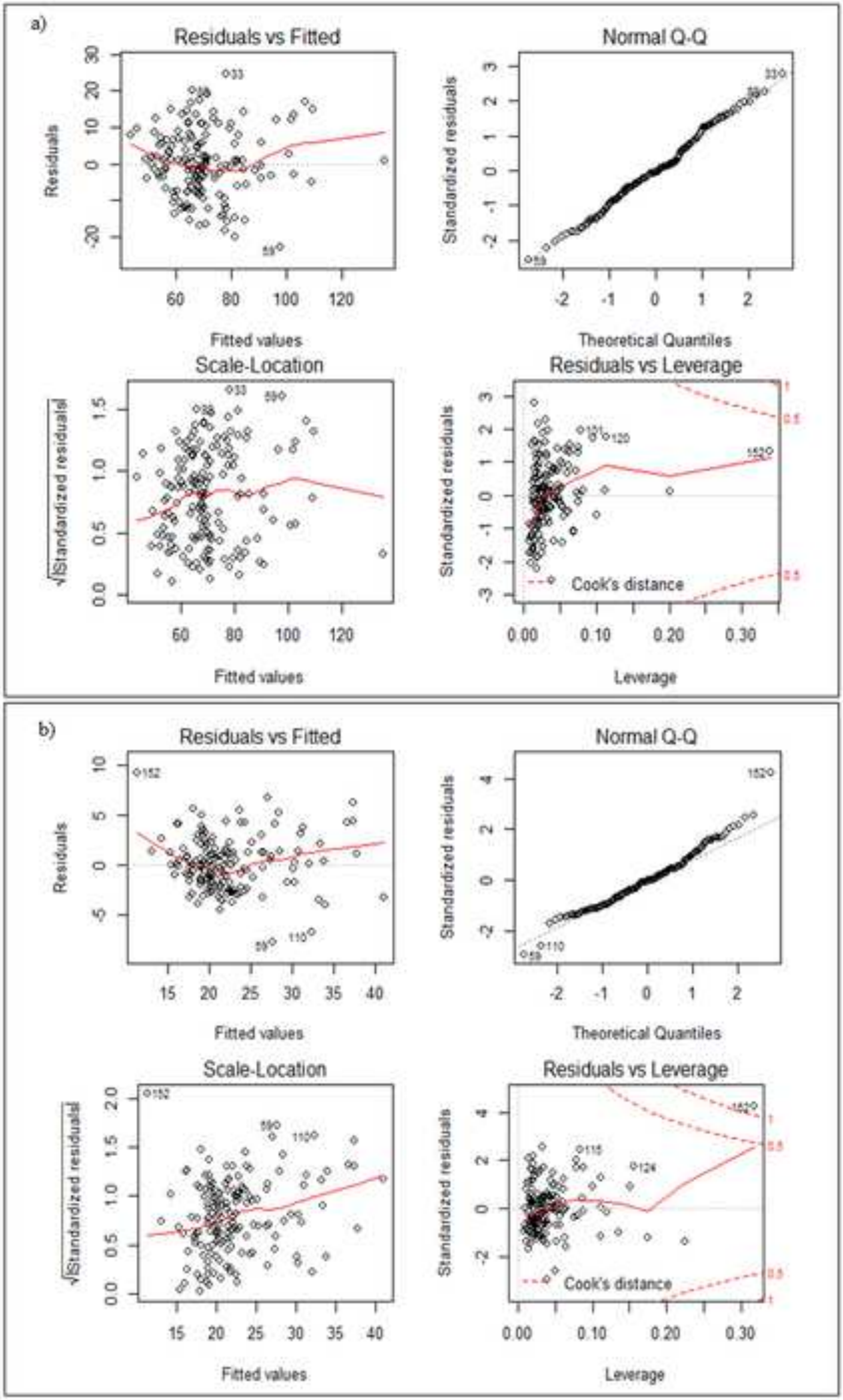
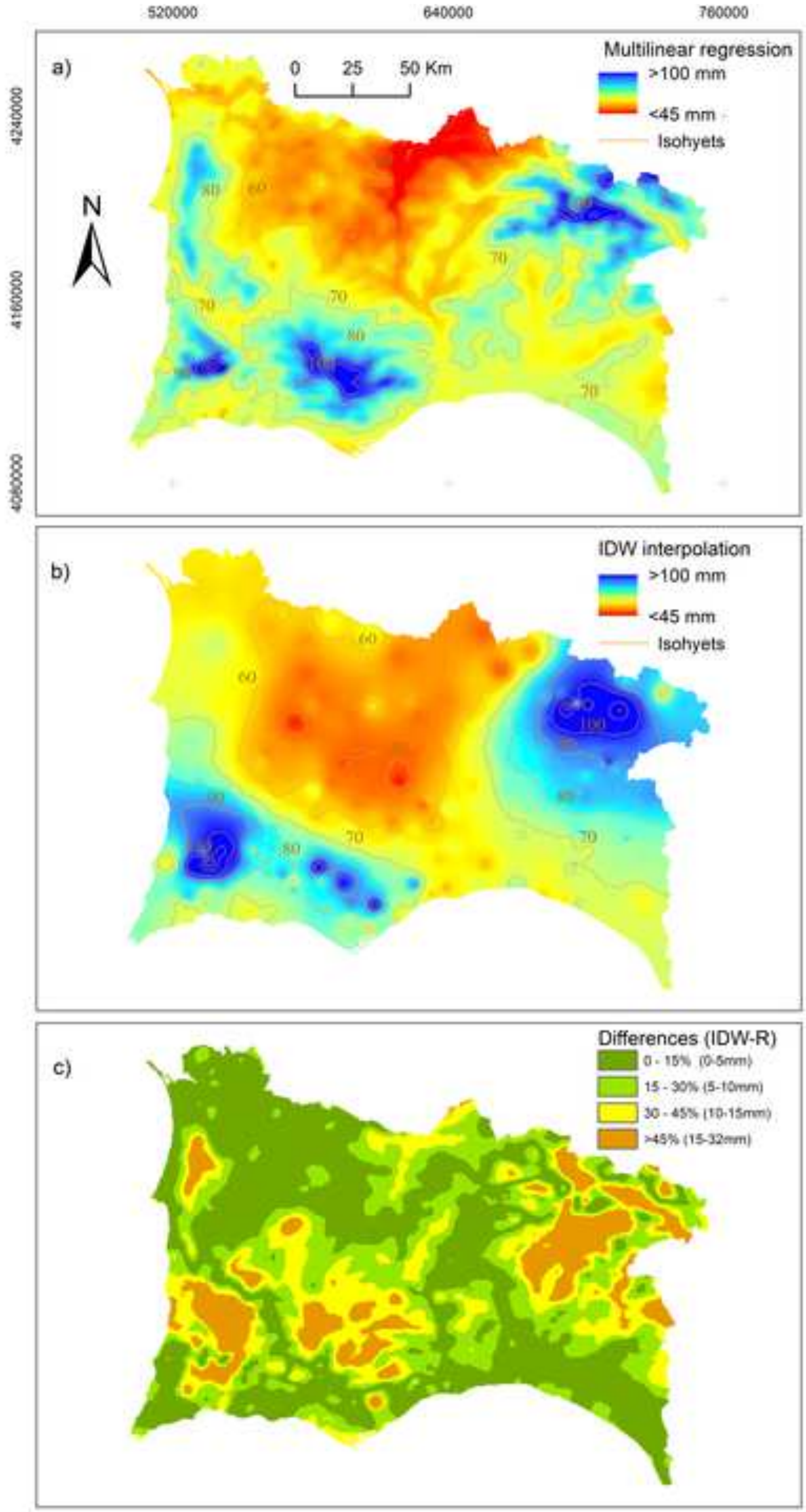


Figure6




520000
640000
760000
4240000
4160000
4080000
a)
0 25 50 Km
N
Multilinear regression
>100 mm
<45 mm
Isohyets
b)
IDW interpolation
>100 mm
<45 mm
Isohyets
c)
Differences (IDW-R)
0 - 15% (0-5mm)
15 - 30% (5-10mm)
30 - 45% (10-15mm)
>45% (15-32mm)

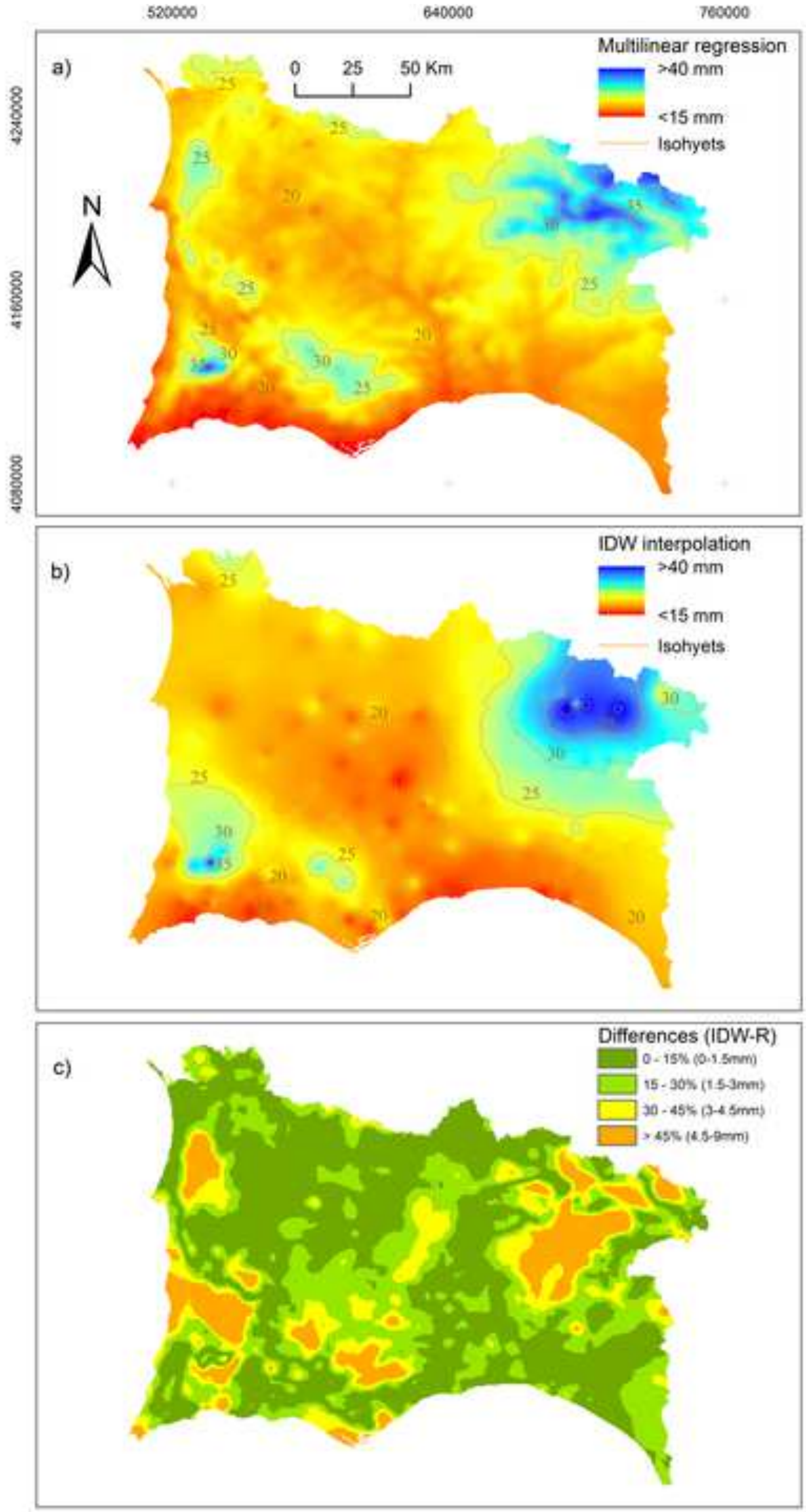
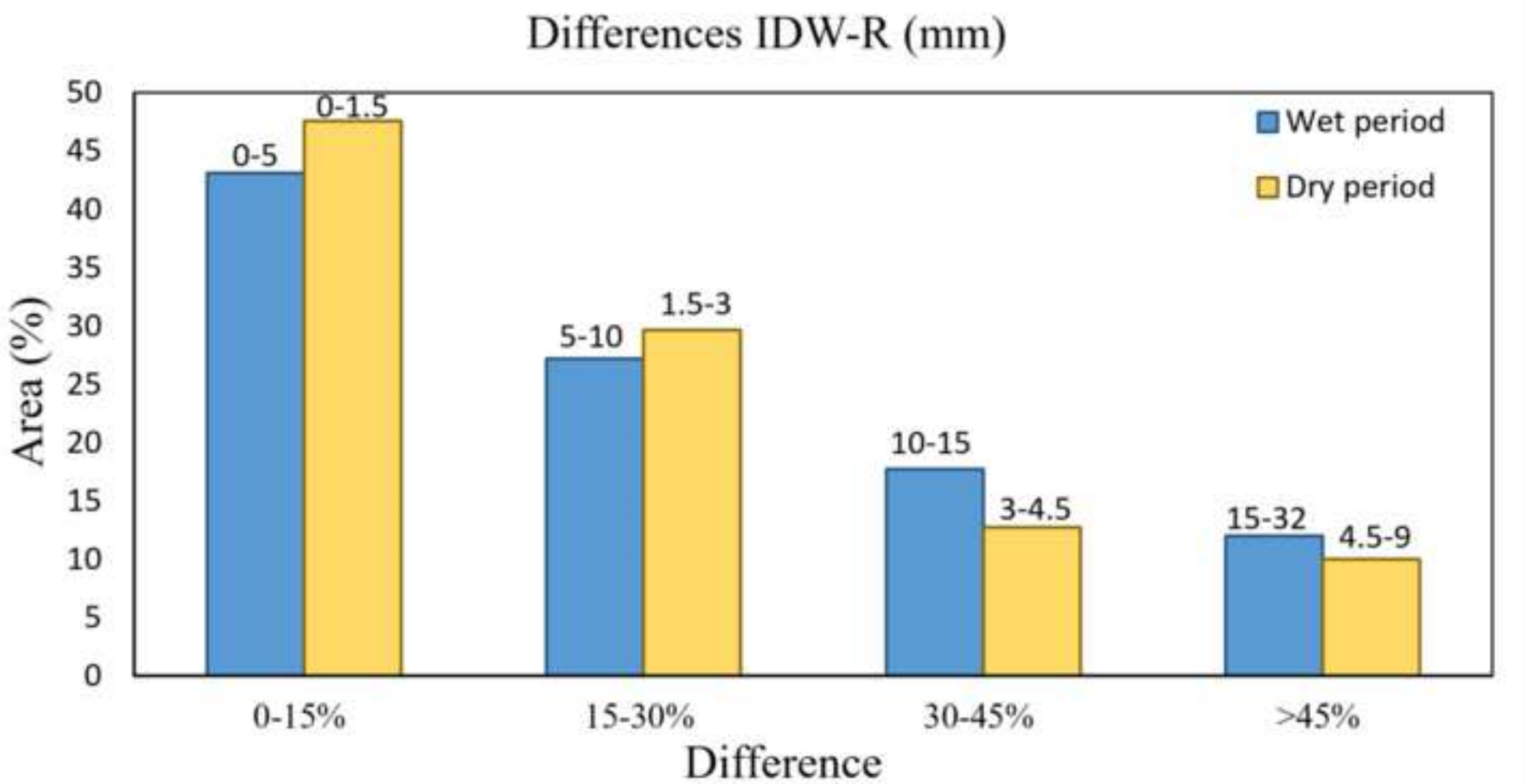

520000
640000
760000
4240000
4160000
4080000
a)
Multilinear regression
>40 mm
<15 mm
Isohyets
0 25 50 Km
N
b)
IDW interpolation
>40 mm
<15 mm
Isohyets
c)
Differences (IDW-R)
0 - 15% (0-1.5mm)
15 - 30% (1.5-3mm)
30 - 45% (3-4.5mm)
> 45% (4.5-9mm)

Figure8




Differences IDW-R (mm)
Area (%)
50
45
40
35
30
25
20
15
10
5
0
0-5
0-1.5
5-10
1.5-3
10-15
3-4.5
15-32
4.5-9
Wet period
Dry period
0-15%
15-30%
30-45%
>45%
Difference