A new approach to the Openness index for landform characterisation

F. Alonso-Sarría[a], F. Gomariz-Castillo[d], F. Cánovas-García[b]

[a]*Instituto del Agua y Medio Ambiente, Universidad de Murcia. Edificio D, Campus de Espinardo, s/n, 30001 Murcia, Spain*
[b]*Unidad Predepartamental de Ingeniera Civil, Universidad Politcnica de Cartagena, Paseo Alfonso XIII, 52. 30203 Cartagena, Spain*
[c]*Instituto Euromediterráneo del Agua, Campus de Espinardo, s/n, 30001 Murcia, Spain*
[d]*Instituto del Agua y Medio Ambiente, Universidad de Murcia. Edificio D, Campus de Espinardo, s/n, 30001 Murcia, Spain*

Abstract

Openness is a multi-scale geomorphometric feature that has not been widely used despite its potential. The original approach, that averages zenith and nadir angles in the eight main compass directions, is modified to take into account openness in all directions available; in addition, openness is calculated in different directions and different scales. An statistical analysis and Random Forest classification are carried out to check if the modifications introduced provide significantly different results than the original approach. In addition, it was tested if the multi-scale and multi-direction openness provide relevant and complementary information to total openness. The results show that the original algorithm produces biased openness estimations, systematically higher. In addition, multi-scale and multi-direction openness features produce more accurate random forest classifications. Accuracy increases from 0.62 when using total openness to 0.66 when using the multi-scale approach, 0.73 when using multi-direction approach and 0.75 when both are used.

Keywords: Geomorphometric Features, Openness, ANOVA, GRASS GIS

1. Introduction

Geomorphometry is an interdisciplinary field that encompasses Earth sciences, mathematics and computer science (Pike et al., 2009). Its objective is the quantitative characterisation of land surface topography (Pike, 1995,

2000; Pike et al., 2009; Hengl and Reuter, 2009; Wilson, 2012) in order to relate it with processes acting on Earth surface, as topography is one of the most important factors controlling several environmental (atmospheric, hydrologic, geomorphic, biogeographic or ecologic) processes (Wilson, 2012). One of the main objectives of modern geomorphometry is the definition and extraction of quantitative geomorphometric features (GF) from a Digital Elevation Model (DEM) of a given resolution (cell size), used as a model of the continuous terrain surface, and the segmentation of the landscape into spatial entities also using a DEM as initial information (Wilson and Bishop, 2013).

Several GFs have been proposed to characterise the multi-scale morphological attributes of the topography, and several approaches have been developed to classify such parameters. One of the most accepted divides them into local and regional parameters (Wilson, 2012). The calculation of regional parameters involves a number of cells that has to be determined by the same algorithm that performs the calculation, for instance, the contributing area to a cell. Local parameter values are supposed to be representative of points in space but are calculated from the cells surrounding the cell that corresponds to the analysed point. Most of the local GFs are terrain derivatives, both first derivatives (slope and aspect) and second derivatives (curvatures) (Evans and Minár, 2011; Evans, 2013; Krebs et al., 2015). Recently, the use of third derivatives has been introduced (Florinsky, 2009; Minár et al., 2013). The algorithms to calculate first and second derivatives (Evans, 1972; Zevenbergen and Thorne, 1987) used subsets of 3x3 cells around the cell of interest, although with different approaches. This dependence on algorithms and resolution, and the fact that the terrain surface is not truly differentiable, lead Florinsky (1998) or Wilson (2012) to suggest that local GFs should be considered as mathematical variables rather than real-world values.

Wood (1996) proposed an algorithm to calculate such derivatives using larger window sizes taking into account that different processes might work at different scales and the interesting scale is not necessarily the 3x3 window. The resulting algorithm was implemented as a GRASS module (r.param.scale). The neighbouring cells used in the calculation are the ones enclosed in a square or circle centred in the analysed point and with a side length or diameter given by the user as a parameter. These GFs should be called focal instead of local following the concept of focal function in raster GIS (De Meers, 2002). A different type of focal parameters are related to surface roughness estimated using the standard deviation or coefficient of

variation of elevation in some circle or window (Wilson, 2012). Other approaches include Hobson (1972), Iwahashi and Kamiya (1995), Riley et al. (1999) or the use of semivariograms or fractal analysis (Olaya, 2009). In addition, there are GFs based on the vertical position of the cell in relation to the neighbouring cells, for instance the Topographic Position Index (De Reu et al., 2013) or the percentile of the cell height in relation to the heights of the surrounding cells. Roughness and terrain position might also be calculated with different windows sizes, and so, at different scales.

Moore et al. (1993) and Wilson and Gallant (2000) distinguished among primary and secondary GFs. The former, local or regional, are calculated directly from the DEM and the later are obtained by operating with the primary GFs, for instance Topographic Wetness Index (Beven et al., 1995) or the LS parameter in the Universal Soil Loss Equation (Moore et al., 1991).

Despite this multiplication of GFs and scales, one of the objectives of geomorphometry is to find a minimal set of GFs, as universal as possible, that convey the maximum information on the topography and that are also related with geomorphological processes (Wood, 1996). Gallant and Dowling (2003) encouraged the definition of multi-scale GFs. Yokoyama et al. (2002) proposed a multi-scale GF called openness, defined as the average of the zenith (or nadir) angles along the main eight compass directions (N, NE, E, SE, S, SW, W, NW) measured from a cell using a radial distance L. A positive openness is defined using the zenith angles and a negative openness using the nadir angles. The zenith angles are calculated as $\phi = 90 - \beta$ where β is the larger slope among the origin cell and all the cells in a given compass direction and within a distance lower than the radial distance, whose heights are higher than the origin cell. On the other hand, the nadir angle is calculated as $\psi = 90 + \delta$ where δ is the larger slope between the origin cell and all the cells in a given compass direction and within a distance lower than the radial distance, whose height is lower than the origin cell (Figure 1). Jasiewicz and Stepinski (2013) calculated negative openness as $\psi = 90 - \delta$, this way both indices (positive and negative openness) have the same range, and it is always possible to obtain one of the nadir approaches subtracting the other to 180. From Figure 1, it is clear that neither positive nor negative openness can increase when the radial distance increases. The same is true for their averages as every cell crossed by the eight main directions are analysed only once in each different window size. Yokoyama et al. (2002) claimed that openness expresses a topographic attribute that is possibly not captured by other combinations of GFs, although they acknowledged that this had to

be tested. Later, Prima et al. (2006) used openness and slope to classify landforms using a supervised approach based on the Jeffries-Matusita and Mahalanobis distances, and Bishop et al. (2010) used openness to study mountain topographic evolution in the central Karakorum.

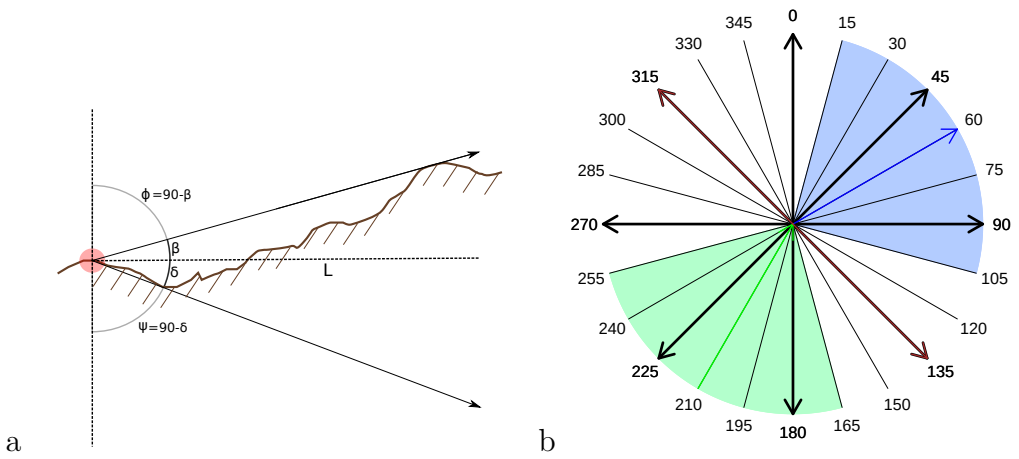

Figure 1: a) Calculation of the zenith (ϕ) and nadir (ψ) angles for a given radial distance (L). β is the maximum slope to a surrounding cell with higher height than the origin cell, δ is the maximum slope from a surrounding cell with lower height than the origin cell. Negative openness was calculated as $\psi = 90 - \delta$, as in Jasiewicz and Stepinski (2013). b) Comparison of the O8 (Yokoyama et al., 2002) approach and the OT approach presented in this paper. The thicker black lines show the 8 main directions and the thinner black lines show all the possible directions in a window of 7x7.

Although Yokoyama et al. (2002) presented several examples of how openness measured in the eight main directions could be used as a geometric signature to classify other landforms, their main goal was to recognise volcanic domes and craters, so the averaging approach was appropriate enough. However, we think that averaging the eight directions is less useful when trying to detect anisotropic landforms as those produced by fluvial processes. Another problem, from our point of view, is that the use of just eight directions implies an important loss of information when the radial distance increases. The number of perimeter cell as a function of the window size is $4 \cdot (ws-1)$, so the proportion of lost information is $(ws-3)/(ws-1)$, reaching, for example, a value of 80% for a radial distance of 5 cells (ws=11).

Jasiewicz and Stepinski (2013) used openness to develop a novel landform classification approach based on the number of directions with positive ($\phi >$

ψ), negative ($\phi < \psi$) or neutral ($\phi = \psi$) openness using a threshold to define the maximum difference among both angles for which the openness is considered neutral. The authors claim that this approach is multi-scale, meaning that using a large enough radial distance, all the lower distance openness values are integrated and the result is equivalent to using in each cell the appropriate window size. The results shown are quite impressive in terms of the convergence of the resulting map when L increases whereas similar classification approaches based on curvatures (Dikau et al., 1995; Wood, 1996) change substantially with window size leading eventually to a flat class map for very large window sizes.

However, we think this approach still has the problem of the loss of information when using eight directions; in fact, Jasiewicz and Stepinski (2013) recommend L=20 cells (ws=41), that means a loss of information of 95 %. Jasiewicz and Stepinski (2013) stated that openness values converge to a single value when increasing window size; however, we think that the way convergence is reached might provide important information when classifying landforms. In addition, the approach of Jasiewicz and Stepinski (2013) consider the patterns of positive, negative and neutral openness as rotation invariant, that means that it does not consider any principal direction in the surroundings of the cell. We think that such a direction exists: the flow direction, and should be taken into account, especially in landscapes dominated by fluvial processes.

The objective of this paper is to 1) modify the original openness approach by using all the directions available for a given window size (OT), 2) calculate 3 complementary openness measurements: openness in the downslope direction (ODN), upslope direction (OUP) and transverse directions (OBN), 3) analyze how these measurements change when the radial distance increase in comparison with the original Yokoyama et al. (2002) approach, 4) compare these openness measurements with the original approach (O8), 5) relate these changes to the main landforms present in the study area, and 6) write the code as a GRASS v. 7 module available to anyone interested.

We will test three null hypothesis:

1. O8 values are not significantly different from OT values. The alternative hypothesis is that the values differ when the window size is greater than 3.
2. Information provided by openness measurements with different window sizes is summarised by a proper window size in such a way that

openness values calculated with smaller window sizes do not provide additional useful information. The alternative hypothesis is that openness calculated at different scales provide useful information about the landforms.

3. ODN, OUP and OBN are not significantly different to OT and do not provide additional information to OT. The alternative hypothesis is that the values are different and convey useful information for landform classification.

2. Material and methods

2.1. The openness algorithm

The original Yokoyama et al. (2002) algorithm averages zenith and nadir angles in the eight main compass directions to compute positive and negative (respectively) openness (Figure 1 b). We call these measurements as O8P and O8N. Our algorithm calculates positive (OTP) and negative (OTN) total openness averaging zenith and nadir angles in all compass directions available for a given window size. Figure 1 (a) shows, as an example, the eight main directions and the 24 available directions when using a 7x7 window.

In addition, the downslope direction for the analysed cell is calculated using a drainage layer; in our case, this layer was calculated with the GRASS module r.watershed; the calculation procedure is illustrated in Figure 2 (a). The flow direction is followed until the border of the window is reached, the downslope direction is calculated as the direction of a line joining the centres of the analysed cell and the output cell in the border of the cell. The upslope direction is more complex as several candidate cells might appear. The procedure is similar, but following the cells upslope until the border of the window. Among all the cells contributing to the analysed cell, an upslope cell is identified as the border cell with the lowest slope (following the flow path) respect to the analysed cell. If the border of the window is not reached, the lowest slope criterion is applied to all the contributing cells. Finally, if there is no contributing cell, the upslope direction is considered to be the same as the downslope direction but in opposite sense. Once downslope and upslope directions are obtained, the transverse directions are obtained as those that halve the two angles formed by the downslope and upslope directions. Figure 2 (a) shows this procedure for the case shown in figure 1 (a) in which the downslope directions is a_d=60^{o}E, the upslope direction is a_u=210 ^{o}E) and the corresponding values of a_1 and a_2 are 135 and 315 ^{o}E

respectively. To calculate OUP, ODN and OBN, a weighted average of nadir and zenith angles is obtained with the weights calculated as:

$$wDN(a) = max(0, cos(2 \cdot \pi \cdot (a - a_d)/180) \quad (1)$$

$$wUP(a) = max(0, cos(2 \cdot \pi \cdot (a - a_u)/180) \quad (2)$$

$$wBN(a) = max(max(0, cos(2 \cdot \pi \cdot (a - a_1)/180), max(0, cos(2 \cdot \pi \cdot (a - a_2)/180)) \quad (3)$$

where $wDN(a)$ is the weight for direction a when calculating downslope openness, $wUP(a)$ is the weight for direction a when calculating upslope openness, $wBN(a)$ is the weight for direction a when calculating transverse openness, a_d is the downslope direction, a_u is the upslope direction, and a_1 and a_2 are the transverse directions. Finally, the weights are normalised by dividing them by the sum of their values. For example, for downslope openness weights it would be:

$$wDN(a) = \frac{wDN(a)}{\sum_{i=1}^{360} wDN(i)} \quad (4)$$

Figure 1 (b) shows the weights to calculate ODN, OUP and OBN corresponding to the directions shown in Figure 2 (a).

2.2. Testing the hypothesis

To check the first of the three aforementioned hypothesis: O8 is not significantly different to OT when estimating openness at different scales, a Wilcoxon test was performed for each landform and window size to determine if the average differences are significantly different to zero. Wilcoxon test was preferred to the Student's t test because in all landforms, both openness values are not Gaussian and the variances are homogeneous. We checked these properties using the Shapiro-Wilk and the Leveene tests respectively.

To check the second hypothesis: that a single openness value calculated with a large enough window size integrates the information of the smaller window sizes, we did an ANOVA in each landform comparing the openness calculated with different window sizes. Once the existence of an effect was discovered using ANOVA, a Tukey-Kramer contrast was carried out to identify significant differences among factors; this contrast is based on a Student's t:

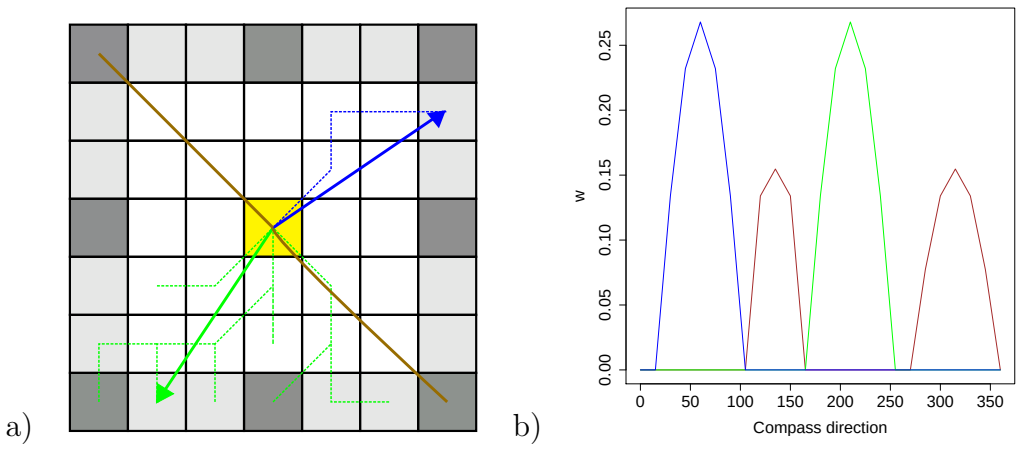

Figure 2: Determination of the downslope (blue lines), upslope (gree lines) and transverse (brown) directions (a) and associated weights (b) in the same window presented in 1 (a).

$$t = \frac{|X_i - X_j|}{SE_{ij}} \tag{5}$$

where SE_{ij} is a pooled estimation of the standard error of the means obtained from the covariance matrix of the openness values.

The reason why we did not use Wilcoxon test is that when comparing openness values for different window sizes, variances are not homogeneous. So we used ANOVA and tried to correct for heteroscedasticity using a heteroscedasticity-consistent covariance matrix of the parameters (HC3). With this methodology, heteroscedasticity effects are avoided even if its form is not known (Long and Ervin, 2000; Zeileis, 2004). HC estimators use residuals ($\hat{e}_i^2$) to estimate the covariance matrix:

$$Var[\hat{\beta}] = HC3 = (X'X)^{-1}X'\hat{\Omega}_3 X(X'X)^{-1} \tag{6}$$

$$\hat{\Omega}_3 = diag\left\{ \frac{\hat{e_1^2}}{(1-h_{11})^2}, ..., \frac{\hat{e_i^2}}{(1-h_{ii})^2} \right\} \tag{7}$$

where h_{ii} is the ii element in the matrix $(X'X)^{-1}X'$.

With this methodology, a more robust covariance matrix estimator for the Tukey-Kramer contrast can be computed, especially when the number

of cases is small. This procedure allows obtaining groups of window sizes with significantly equal openness values. If all the scales belong to the same group, we can consider that the window size does not affect the openness value.

For each landform, a set of characteristic window sizes were identified. When there is only one group, there is not scale difference and ws=3 is considered representative. If two groups appear, ws=3 is considered the characteristic window size for the most detailed scale, and the first window sizes that belong to the second group but not to the first group is considered the characteristic window size for the smallest scale. When several groups appear, the procedure continues until a set of representative scales is obtained for each landform. However, when several window sizes could be used for a given group, the objective is to use the same window size in all landforms for the sake of parsimony. Finally, a set of all characteristic window sizes is obtained for each openness measure.

Additionally, we classified the landforms using Random Forest (Breiman, 2001). This is one of the most used classification algorithms. It is an ensemble classifier based on decision trees that uses bagging and random feature selection to reduce correlation among trees, giving more sense to the whole ensemble learning concept (James et al., 2013).

The number of classification trees (ntree) and the number of features randomly chosen to split each node (mtry) are the parameters that the user must decide or optimise; however, the method is not very sensitive to them, and the default value (ntree=500 and mtry=$\sqrt{p}$, where p is the number of features) generally give highly accurate results (Liaw and Wiener, 2002; Cánovas-García and Alonso-Sarría, 2015; Belgiu and Drăgu, 2016). A larger description of random forest classification can be found in James et al. (2013).

We used two different random forest models:

- Simple Model (O): Just OTP and OTN with ws=101 are used as predictors
- multi-scale Model (MSO): OTP and OTN in all the relevant window sizes are used as predictors

If the second model achieves a significantly higher accuracy than the first one, we could conclude that the multi-scale approach provides relevant information to the classification algorithm.

In relation with the third hypothesis, we did the same multi-scale analysis for ODN, OUP and OBN and classify a third Random Forest model with OTP, OTN, ODNP, ODNN, OUPP, OUPN, OBNP and OBNN only with a window size of 101 cells. This multi-direction model was labelled as MDO. Once again, a significantly higher accuracy with the MDO would mean that the multi-direction approach provides relevant information to the classification algorithm. Finally a multi-scale and multidimension RF model was calibrated using the eight openness measurements in their characteristic window sizes. This last RF model was labelled MSMDO and will test the interaction of both approaches.

Because of the random component in RF, each time the algorithm is used with the same data-set the results will differ. We used this fact to obtain a distribution of accuracy values after 100 executions of the four classification model. An ANOVA using HC3 parameters was used to identify when differences in accuracy and in the omission error of each class were significantly different.

The objective of the RF classification is not to obtain an accurate classification, but to determine if the new openness measurements provide substantial new information that can increase accuracy and decrease omission errors of the different classes. RF is especially useful in this respect as its low parameter sensitivity simplify the classification process.

All the statistical analysis were performed in R using the randomForest package (Liaw and Wiener, 2002) for the RF classification and the car (Fox and Weisberg, 2011), multcomp (Hothorn et al., 2008), effects (Fox, 2003) and PMCMR (Pohlert, 2014) packages for the heteroscedasticity-consistent ANOVA.

3. Study case: The Guadalentín basin

3.1. Study area

The Guadalentín basin (Figure 3) is a large (3300 km^2) basin, located in the southwest of the Region of Murcia, whose head reaches the northeastern part of Andalucia. It is the main tributary of the River Segura. All the area is inside the Betic ranges and shows a great landform and lithological variety, with rock formations from the Paleozoic to the Quaternary.

The most relevant structural characteristic in the area is a large tectonic depression bounded to the north by the Lorca-Alhama fault and to the south by the Palomares and North-Carrascoy faults. These faults have produced

the sinking of the valley and the raising of the surrounding reliefs. Silva (2014) summarises the geology and geomorphology of the depression. The North fault has greater intensity than the South, so the northern reliefs are higher and have isolated two sedimentary basins (Lorca and Mula) separated by the Sierra Espua range. These basins were filled during the Neogen (Tortonian and Messinian) by loams, clays, gypsum, sandstones and conglomerates (López Bermúdez and Romero Díaz, 2009).

During the Pliocene period, several N-S faults intersected the aforementioned tectonic accidents and opened the previously endorheic sedimentary basins. As a result, the baseline levels changed and new equilibrium profiles were developed in the pre-existing river network, new channels developed in the sedimentary basins that would be responsible for their partial emptying. As a consequence, large alluvial fans were formed in the Guadalentín basin, at the foot of the reliefs. The coalescence of such alluvial fans produced also glacis. Torrential systems appeared disconnected from the base level and are spread divergently over the alluvial fans. The upper Guadalentín river drained the Lorca basin, whereas the Librilla and Algeciras ramblas drained part of the Mula basin towards the Guadalentín. Calmel-Avila (2002) points out that changes in the rainfall regime between the ages of Bronze and iron intensified the process.

The rapid emptying of the Lorca and Mula basins, due to the poor mechanical resistance of their sedimentary materials, led to extensive badland landscapes (López Bermúdez and Romero Díaz, 1989). The results are basins with rounded headwaters and very dense and torrential drainage networks, short steep slopes and low bifurcation coefficients that allow the accelerated and superimposed arrival of flows from different sectors of the headland (López Bermúdez et al., 2002).

In summary, the study area is dominated by fluvial processes and is an appropriate landscape to test if openness calculated in different directions might improve the average openness estimation. We selected eight different landforms, representative of the study area, to analyse the results of the different openness algorithms:

- Plains: Wide flat extensions independently of extension and altitude. In the study area, they correspond mainly with the alluvial valley of the main channel.

- Alluvial fans: Detritic deposits located in significant slope breaks. They have a slightly positive profile and transverse curvatures. In plant, they

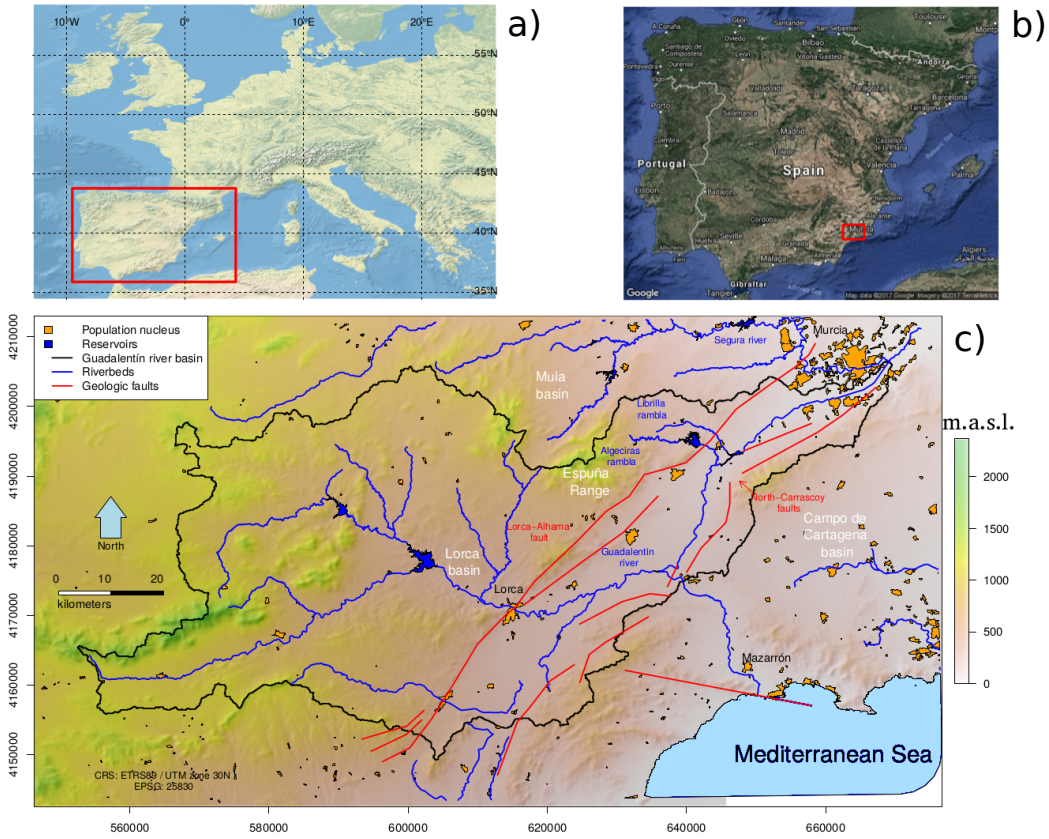

Figure 3: Location of the study area. a) Location of Iberian Peninsula in Europe, b) location of the study area in the Iberian Peninsula and c) map of the Guadalentin river basin.

are circular or fan-shaped, and they also have a cone-shaped section. That is the reason why they are also named debris cones.

- Glacis: Surfaces with a gentle slope that fringe mountain areas. In this case, they are accumulation glacis mainly generated by the successive alluvial fans. Their profile section is straight or gently concave.

- V shaped streams: Erosion is mainly vertical and the consequent incision generates strong slopes in both banks, they are mainly located in the upper basin.

- U shaped streams (ramblas): Mainly located in the central depression. Erosion is mainly lateral.
- Ridges: Elevated heights originated by morphological units such as slope or interfluvial dividers.
- Slopes: More or less regular in terms of the degree of inclination, their origin is not linked to sedimentary but to tectonic or erosive processes.
- Badland areas: Located in easily eroded rocks (marls or clay), usually in arid or semi-arid climatic conditions; they are generated by the erosive power of concentrated water.

3.2. Data used

We used the official MDE produced by the Spanish Geographical National Institute (IGN) with a resolution of 5 m. It is derived from LiDAR points with a density of approximately 0.5 points/m^2, meaning that the value in each cell is an average of approximately 12.5 values. Vertical accuracy of the points is RMSE$<$ 0.2 m (IGN, 2017). If this value is taken as a standard deviation, the RMSE of the cells can be computed as $0.2/\sqrt{12.5} = 0.056$ m.

Sixty training points were obtained for each of the aforementioned classes, except for the badland areas where only 27 were obtained. The 0.5 m resolution images of the Spanish National Plan of Airborne Orthophotography (PNOA), the contour lines extracted from the DEM, and field work for the most difficult points were used as a reference to support the identification of the different landforms in randomly selected zooms in the study area.

The identification of some classes has proved to be difficult, particularly due to scale problems and the similarities of some of the classes. In addition, the pixel size is a constraint that does not allow to select landforms smaller than a certain size. That is the reason why other relevant classes as fluvial terraces were not included. Special attention was paid when distinguishing between "U" and "V" shaped channels, as "U" shaped channels with a width lower than 5 m would not be recognised by the classification algorithms as their bed will not be properly characterized by the features taken into account. The definition of alluvial fans and glacis is also difficult as the latter are usually formed by a succession of fans that in some cases retain their two-dimensional contour.

Ten openness related GFs were calculated using windows of 3, 5, 7, 9, 15, 25, 37, 51, 63, 75, 87 and 101 cells. As the pixel size is 5 m, the shorter

window is equivalent to 225 m^2 and the larger window is equivalent to an area of approximately 0.25 km^2. The ten GFs are:

- O8P: Positive openness as defined in Yokoyama et al. (2002).
- O8N: Negative openness as defined in Yokoyama et al. (2002).
- OTP: Total positive openness calculated averaging zenith angles in all directions.
- OTN: Total negative openness calculated averaging nadir angles in all directions.
- ODNP: Downslope positive openness calculated with a weighted average of zenith angles in directions near the downslope direction.
- ODNN: Downslope negative openness calculated with a weighted average of nadir angles in directions near the downslope direction.
- OUPP: Upslope positive openness calculated with a weighted average of zenith angles in directions near the upslope direction.
- OUPN: Upslope negative openness calculated with a weighted average of nadir angles in directions near the upslope direction.
- OBNP: Transverse positive openness calculated with a weighted average of zenith angles in directions near the transverse directions.
- OBNN: Transverse negative openness calculated with a weighted average of nadir angles in directions near the transverse directions.

3.3. Results and discussion

The first result of this research is a GRASS v.7.2 module, called r.openness, available in github (https://github.com/pacoalonso/r.openness). It allows to calculate both O8 and OT (positive and negative), and, if the user provides a flow direction layer, ODN, OUP and OBN positive and negative openness. It includes the possibility of using a circular window instead of a square window, although only the later was used in this research. Figure 4 shows, as an example, the openness map obtained for the Guadalentn basin with a window size of 5 cells.

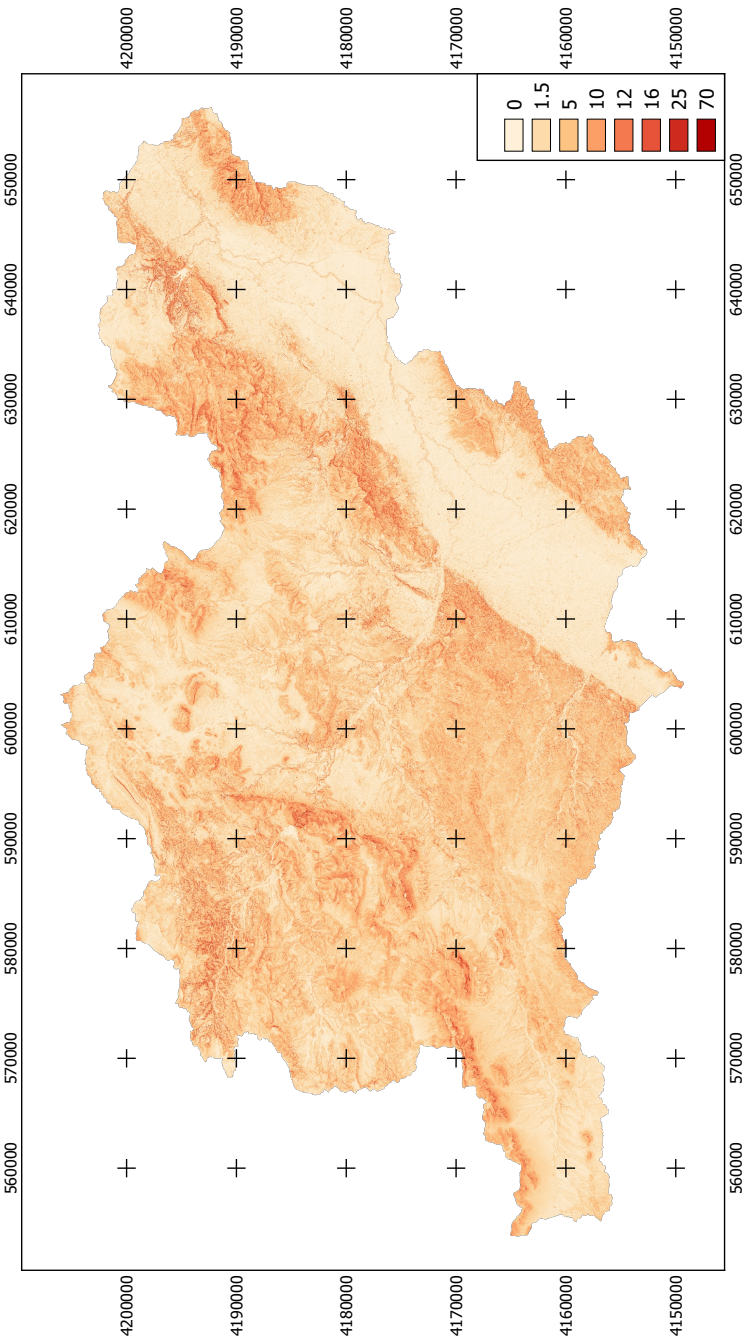

Figure 4: Positive openness map with a window size of 5 cells.

Figure 5 shows the distribution of differences among O8 and OT (both for positive and negative openness) for the eight landforms analysed in the Guadalentn basin and for the 12 window sizes. It is clear that there is a positive bias due to the positive skewness of slope. The bias seems fairly constant along the window sizes but the dispersion tends to increase with window size. In relation to the magnitude of the differences, it depends on the landform and is, logically, related to the magnitude of the slopes characteristic of each landform. Table 1 shows the p-values per landform and scale of the Kruskal Wallis rank test. Most of the p-values indicate that the average differences are significantly larger than zero and, consequently, that O8 underestimates openness. Only the positive openness in some window sizes are not significantly different for badlands and U-shaped streams. Anyway, figure 5 shows the higher variance in the differences for such window sizes, meaning high differences, both for positive and negative openness when they are estimated using the O8 approach. In summary, we should reject our first null hypothesis and accept that O8 overestimates openness.

Positive	5	7	9	15	25	37	51	63	75	87	101
Alluvial fans	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Badlands	0.0003	**0.0654**	0.0182	**0.3484**	**0.0934**	0.0290	0.0386	0.0386	0.0443	**0.0739**	**0.0577**
U streams	0.0028	0.0360	0.0393	**0.4088**	0.0466	0.0140	0.0022	0.0030	0.0080	0.0075	0.0012
V streams	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ridges	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Glacis	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Slopes	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Flat areas	0.0125	0.0018	0.0121	0.0007	0.0009	0.0037	0.0034	0.0025	0.0030	0.0053	0.0099
Negative	5	7	9	15	25	37	51	63	75	87	101
Alluvial fans	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Badlands	0.0002	0.0003	0.0002	0.0021	0.0004	0.0021	0.0017	0.0013	0.0010	0.0010	0.0010
U streams	0.0005	0.0007	0.0010	0.0003	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
V streams	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ridges	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Glacis	0.0069	0.0000	0.0005	0.0001	0.0001	0.0000	0.0001	0.0005	0.0009	0.0001	0.0003
Slopes	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Flat areas	0.0001	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 1: P-values of the Kruskal-Wallis test to test the hypothesis that openness estimated as O8 is not different to openness estimated as OTT. P-values larger than 0.05 are highlighted in bold.

Figure 6 shows the results of the ANOVA and window size comparisons for total, positive and negative, openness carried out to test our second hypothesis. It shows average openness, the dispersion estimated using the heteroscedasticity-consistent covariance matrix, and the groups of window sizes with significantly equal openness. These results show differences among landforms: badlands, ridges and flat areas do not show scale differences in positive openness, whereas badlands, both streams and flat areas do not show scale differences in negative openness. The rest of the cases show different

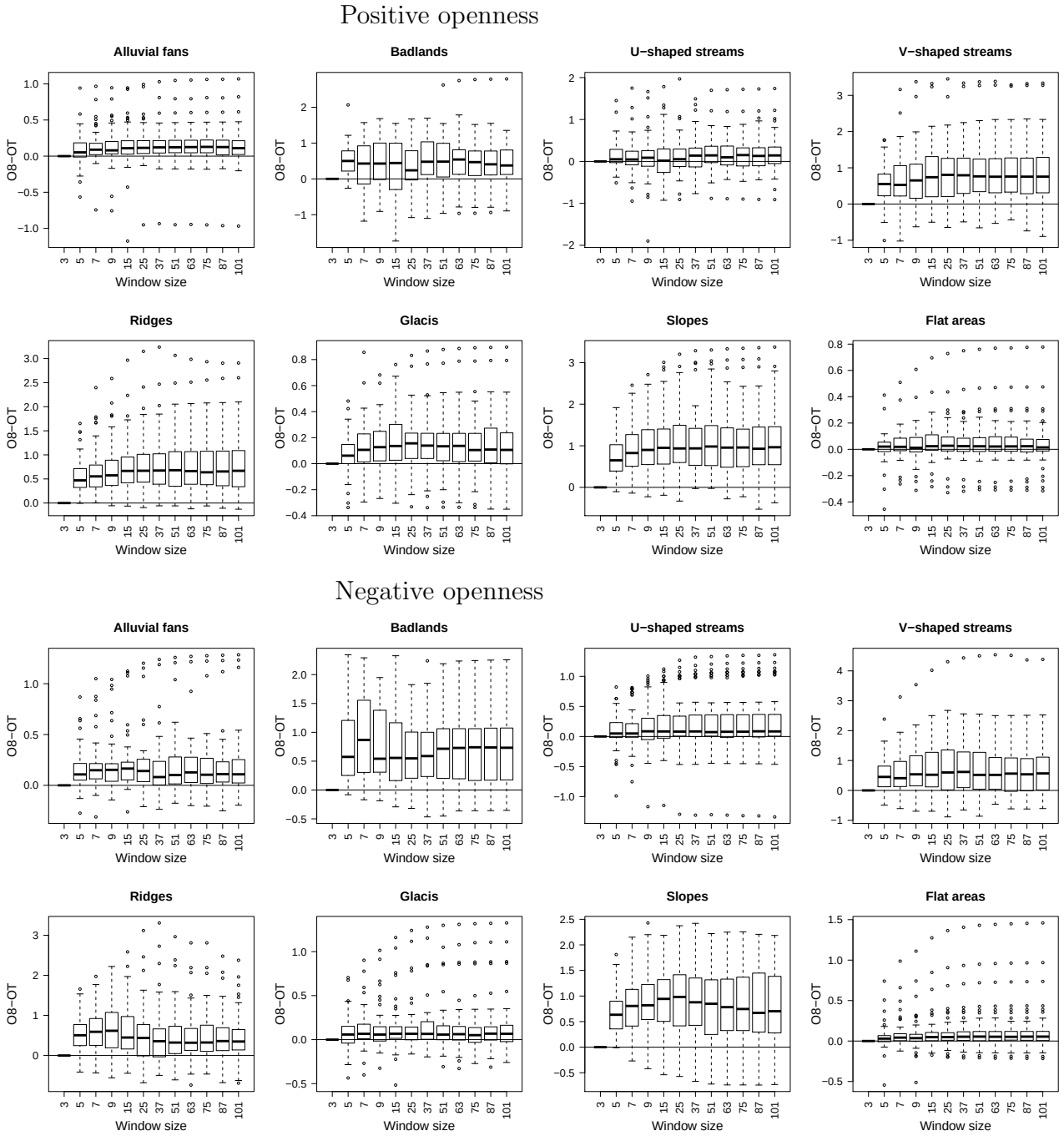

Figure 5: Difference among O8 and OT positive (top) and negative (bottom) openness in real landforms in the Guadalentin basin.

behaviours at different window sizes. The same approach was carried out for downslope, upslope and transverse openness. Table 2 summarizes the resulting characteristic scales for all openness GFs, landforms and window sizes.

Landform	OT	OTN	ODNP	ODNN	OUPP	OUPN	OBNP	OBNN
Alluvial fans	3, 75	3, 37	3	3	3	3	3	3, 37
Badlands	3	3	3, 75	3	3	3	3	3
U streams	3, 15, 75	3	3, 15, 75	3	3, 63	3	3, 101	3
V streams	3, 7, 75	3	3, 5, 9, 75	3	3, 63	3	3	3
Ridges	3	3, 5, 7, 37, 87	3	3	3	3,15, 37	3, 101	3, 15, 37
Glacis	3, 75	3, 87	3, 75	3	3	3	3, 101	3
Slopes	3, 15	3, 9, 87	3	3	3, 63	3, 87	3, 101	3, 25
Flat areas	3	3	3, 75	3	3	3	3, 25, 101	3
Total	3, 7, 15, 75	3, 5, 7, 37, 87	3, 5, 9, 15, 75	3	3, 63	3, 15, 87	3, 101	3, 15, 25, 37

Table 2: Relevant scales for each openness type and window size.

Figure 7 summarises the results of the Random Forest classifications. Both the multi-scale and the multi-direction approaches increase significantly accuracy. Omission errors increase in alluvial fans and badlands when the multi-scale approach is used, and there is no difference in flat areas. The reduction in omission errors is especially important in U-shaped streams.

The multi-direction approach decreases omission errors significantly in all landforms, especially in V-shaped streams and ridges. The use of both approaches increases accuracy an reduce omission errors in most of the landforms, although increases it in badlands, V-shaped streams and flat areas. In fact, the interaction of both approaches significantly decrease all accuracy indicators, although this decrease is lower than the improvement provided in an additive way by both approaches.

Although obtaining an accurate classification is not the main goal of this research, it is noteworthy that the omission errors are quite low in most of the landforms except in glacis and alluvial fans, that usually are confused with each other, and for badlands, openness does not seem to be an appropriate GFs for this landform.

4. Conclusions

The O8 algorithm produces a bias that overestimates openness, both positive and negative in almost all cases. The reason is that slope is a positively skewed feature and, in consequence, a small sample will underestimate its mean. This fact tends to increase with window size.

Openness measures computed with different window sizes are significantly different. It is then possible to identify characteristic windows sizes whose

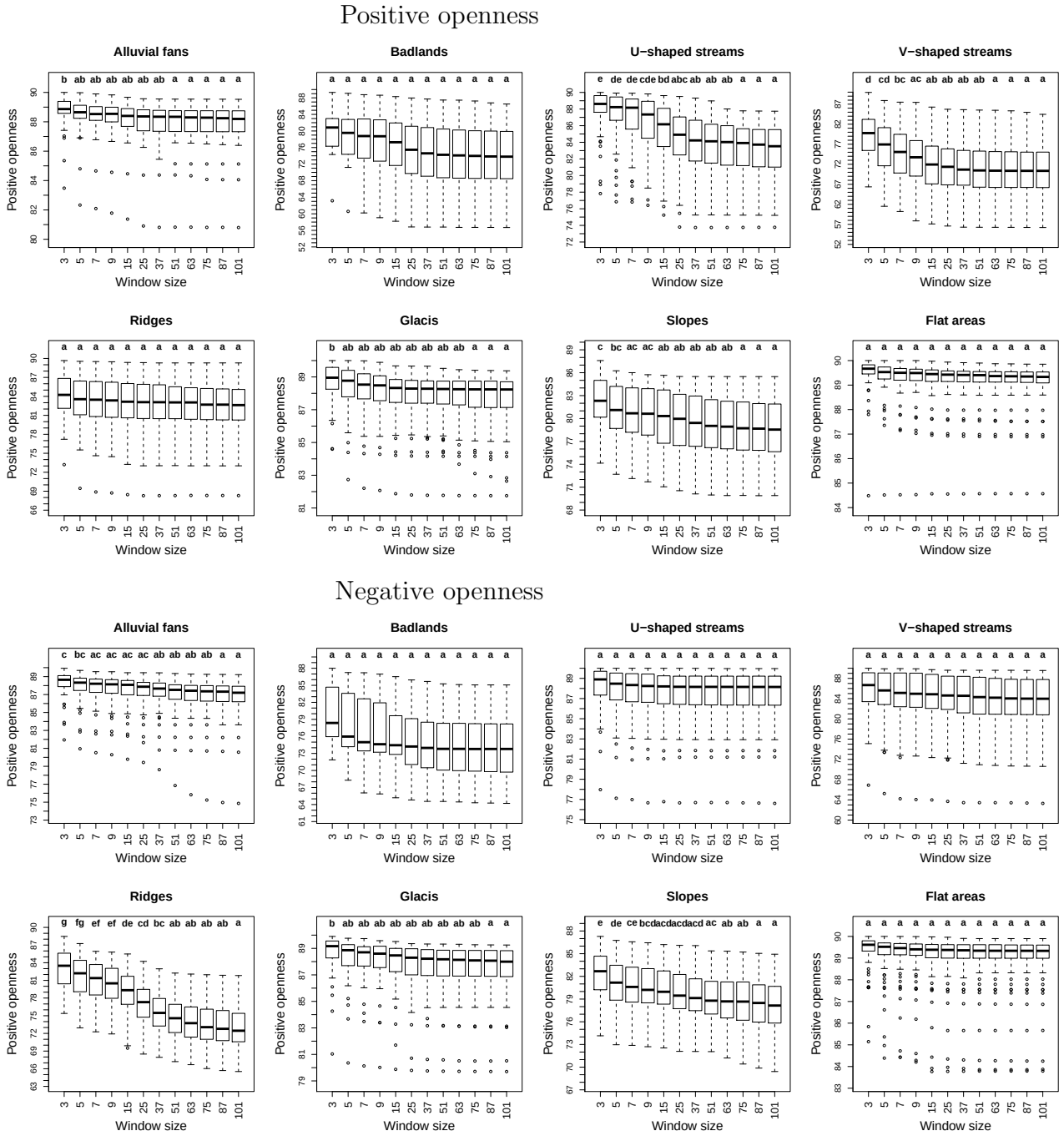

Figure 6: Scale groups total in Positive (OTP) and negative (OTN) openness in real landforms.

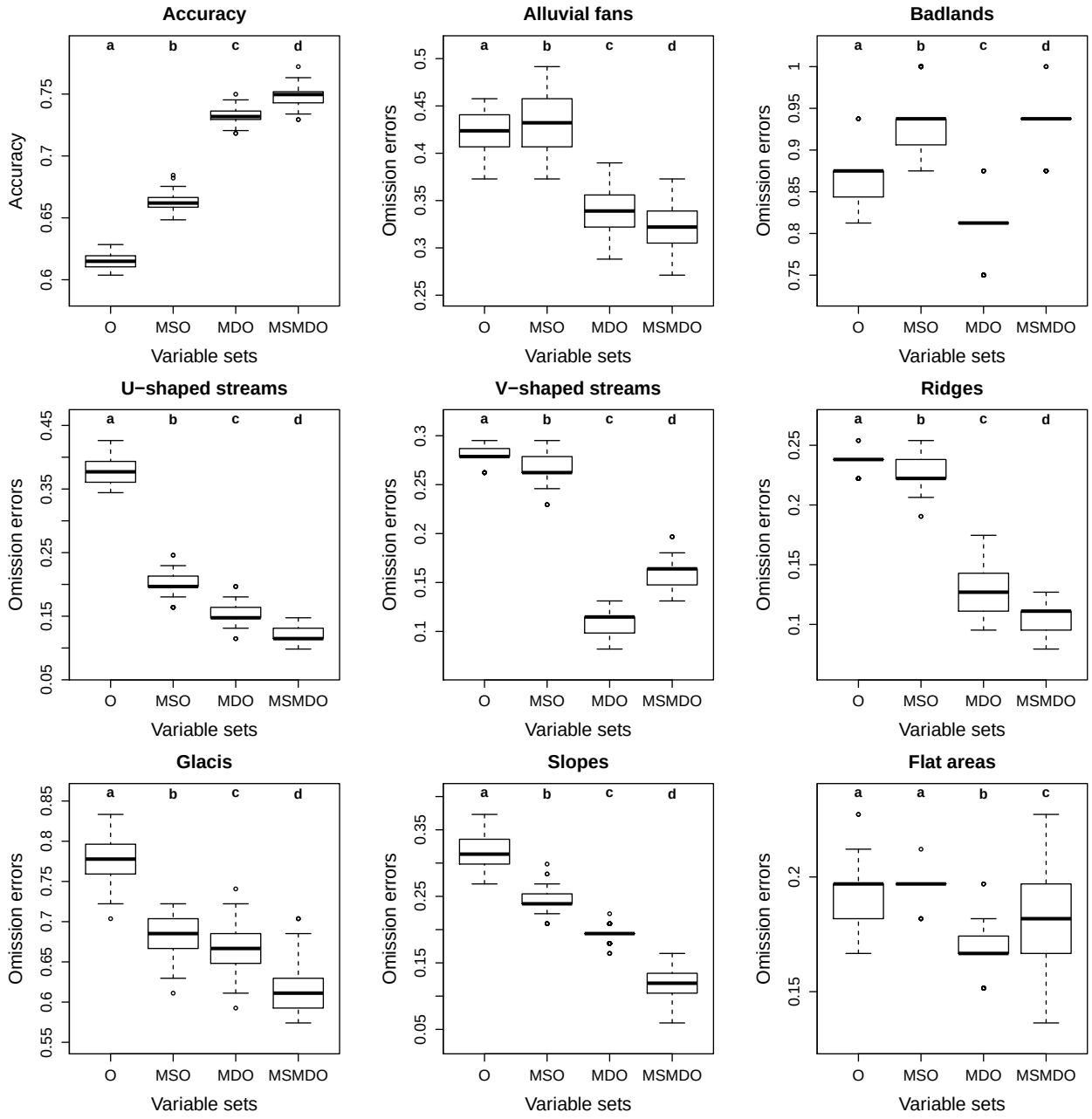

Figure 7: Accuracies and omission errors using different openness GFs combinations. O: Total positive and negative openness with ws=101, MSO: Total positive and negative openness with all characteristic window sizes, MDO: Total, downslope, upslope and transverse openness with ws=101, MSMDO: Total, downslope, upslope and transverse openness with all characteristic window sizes.

use in landform classification increases accuracy significantly. This multi-scale approach is especially useful when classifying U-shaped channels.

The multi-direction approach also contributes to an increase in accuracy in general but in this case it also contributes to a decrease in omission errors in every landform.

Although the interaction of both approaches have a negative effect on accuracy, the additive effect of both approaches contribute to an increase in accuracy, although increase omission errors in badlands, V-shaped streams and flat areas.

5. Acknowledgement

This work partly resulted from a post-doctoral contract included in the Programa Saavedra Fajardo (20023/SF/16) funded by Consejera de Educacin y Universidades de la Comunidad Autnoma de la Regin de Murcia through Fundacin Sneca-Agencia de Ciencia y Tecnologa de la Regin de Murcia.

References

Belgiu, M., Drăgu, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing 114, 24 – 31.

Beven, K., Lamb, R., Quinn, P., Romanowicz, R., Freer, J., 1995. Topmodel. In: Singh, V. P. (Ed.), Computer Models of Watershed Hydrology. Water Resour. Publ., pp. 627–668.

Bishop, M., Bush, A., Copland, L., Kamp, U., Owen, L., Seong, Y., Shroder, J. J., 2010. Climate change and mountain topographic evolution in the central karakorum, pakistan. Annals of the Association of American Geographers 100 (4), 772–793.

Breiman, L., 2001. Random forests. Machine Learning 45 (1), 5–32.

Calmel-Avila, M., 2002. The librilla rambla, an example of morphogenetic crisis in the holocene (Murcia, spain). Quaternary International 93-94.

Cánovas-García, F., Alonso-Sarría, F., 2015. Optimal Combination of Classification Algorithms and Feature Ranking Methods for Object-Based Classification of Submeter Resolution Z/I-Imaging DMC Imagery. Remote Sensing 7, 4651–4677.

De Meers, M., 2002. GIS Modelling in Raster. Wiley.

De Reu, J., Bourgeois, J., Bats, M., Zwertvaegher, A., Gelorini, V., De Smedt, P., Chu, W., Antrop, M., De Maeyer, P., Finke, P., Van Meirvenne, M., Verniers, J., Cromb, P., 2013. Application of the topographic position index to heterogeneous landscapes. Geomorphology 186, 39–49.

Dikau, R., Brabb, E. E., Mark, R. K., Pike, R. J., 1995. Morphometric Landform Analysis of New Mexico. Zeitschrift für Geomorphologie Supplementband 101, 109–126.

Evans, I., 1972. General geomorphometry, derivatives of altitude, and descriptive statistics. In: Chorley, R. (Ed.), Spatial analysis in Geomorphology. Harper & Row, pp. 17–90.

Evans, I., 2013. Land surface derivatives: history, calculation and further development. In: Geomorphometry.org/2013.

Evans, I., Minár, J., 2011. A classification of geomorphometric variables. In: Geomorphometry.org/2011.

Florinsky, I., 2009. Computation of the third order curvature derivatives from a digital elevation model. International Journal of Geographical Information Science I 23 (2), 213–231.

Florinsky, I. V., 1998. Accuracy of local topographic variables derived from digital elevation models. International Journal of Geographical Information Science 12 (1), 47–62.

Fox, J., 2003. Effect displays in R for generalised linear models. Journal of Statistical Software 8 (15), 1–27.
URL `http://www.jstatsoft.org/v08/i15/`

Fox, J., Weisberg, S., 2011. An R Companion to Applied Regression, 2nd Edition. Sage, Thousand Oaks CA.
URL `http://socserv.socsci.mcmaster.ca/jfox/Books/Companion`

Gallant, J., Dowling, T., 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. Water Resources Research 39 (12), 14.

Hengl, T., Reuter, H., 2009. Geomorphometry. Concepts, software and applications. Elsevier.

Hobson, R. D., 1972. Surface roughness in topography: quantitative approach. In: Chorley, R. J. (Ed.), Spatial Analysis in Geomorphology. Methuen, Londres, pp. 221–245.

Hothorn, T., Bretz, F., Westfall, P., 2008. Simultaneous inference in general parametric models. Biometrical Journal 50 (3), 346–363.

IGN, 2017. Pnoa lidar.
URL `http://pnoa.ign.es/presentacion`

Iwahashi, J., Kamiya, I., 1995. Landform classification using digital elevation model by the skills of image processing–mainly using the digital national land information. Geoinformatics 6 (2), 97–108.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics). Springer.

Jasiewicz, J., Stepinski, T., 2013. Geomorphons – a pattern recognition approach to classification and mapping of landforms. Geomorphology 182, 147–156.

Krebs, P., Markus Stocker, M., Pezzatti, G., Conedera, M., 2015. An alternative approach to transverse and profile terrain curvature. International Journal of Geographical Information Science, 29 (4), 643–666.

Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. R News 2 (3), 18–22.

Long, J., Ervin, L., 2000. Using heteroscedasticity consistent standard errors in the linear regression model. Am. Stat. 54, 217–224.

López Bermúdez, F., Barber, G., Alonso Sarria, F., Belmonte Serrato, F., 2002. Natural resources in the Guadalentín Basin (South-east Spain): water as a key factor. John Wiley & Sons, Ltd., pp. 233–245.

López Bermúdez, F., Romero Díaz, A., 1989. Piping erosion and badland development in south-east spain. Catena Supplement 14, 59–73.

López Bermúdez, F., Romero Díaz, A., 2009. Soil erosion and desertification in neogene-quaternary basins of the murcia region.

Minár, J., Jenčo, M., Evans, I., Minár, J., Kadlec, M., Krcho, J., Pacina, J., Burian, L., Benová, A., 2013. Third-order geomorphometric variables (derivatives): definition, computation and utilization of changes of curvatures. International Journal of Geographical Information Science 27 (7), 1381–1402.

Moore, I., Gessler, P., Nielsen, G., Peterson, G., 1993. Soil atribute prediction using terrain analysis. Soil Sci. Soil Am. J. 57, 443–452.

Moore, I., Grayson, R., Ladson, A., 1991. Digital terrain modeling: a review of hydrological, geomorphological, and biological applications. Hydrological Processes 5, 3–30.

Olaya, V., 2009. Basic land surface parameters. In: Hengl, T., Reuter, H. (Eds.), Geomorphometry. Concepts, Software, Applications. Elsevier, pp. 141–170.

Pike, R., 1995. Geomorphometry - progress, practice and prospect. Zeitschrift fr Geomorphologie Supplementband 101, 221–238.

Pike, R. J., 2000. Geomorphometry –- diversity in quantitative surface analysis. Progress in Physical Geography 24 (1), 1–20.

Pike, R. J., Evans, I. S., Hengl, T., 2009. Geomorphometry: A brief guide. In: Hengl, T., Reuter, H. I. (Eds.), Geomorphometry - Concepts, Software, Applications. Vol. 33 of Developments in Soil Science. Elsevier, pp. 3 – 30.

Pohlert, T., 2014. The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR). R package.
URL `http://CRAN.R-project.org/package=PMCMR`

Prima, O. D. A., Echigo, A., Yokoyama, R., Yoshida, T., 2006. Supervised landform classification of northeast Honshu from dem-derived thematic maps. Geomorphology 78 (3-4), 373 – 386.

Riley, S., De Gloria, S., Elliot, R., 1999. A terrain ruggedness index that quantifies topographic heterogeneity. Intermountain Journal of Sciences 5 (1-4), 23–27.

Silva, P., 2014. The Guadalentín Tectonic Depression, Betic Cordillera, Murcia. Springer.

Wilson, J., 2012. Digital terrain modeling. Geomorphology, 107–121.

Wilson, J., Bishop, M., 2013. Geomorphometry. In: Shroder, J., Bishop, M. (Eds.), Treatise on Geomorphology. Vol. 3. Academic Press, pp. 162–186.

Wilson, J. P., Gallant, J. C. (Eds.), 2000. Terrain Analysis: Principles and Applications. John Wiley & Sons, Inc.

Wood, J., 1996. The geomorphological characterization of digital elevation models. Ph.D. thesis, Department of Geography. University of Leicester.

Yokoyama, R., Shlrasawa, M., Pike, R., 2002. Visualizing Topography by Openness: A New Application of Image Processing to Digital Elevation Models. Photogrammetric Engineering & Remote Sensing 68 (3), 257–265.

Zeileis, A., 2004. Econometric computing with hc and hac covariance matrix estimators. Journal of Statistical Software 11 (10), 1–17.

Zevenbergen, L. W., Thorne, C. R., 1987. Quantitative analysis of land surface topography. Earth Surface Processes and Landforms 12, 47–56.