Elsevier Editorial System(tm) for Geomorphology

Manuscript Draft

Manuscript Number: GEOMOR-8878R1

Title: Predicting gully occurrence at watershed scale: comparing topographic indices and multivariate statistical models

Article Type: Research Paper

Keywords: Gully erosion susceptibility; Topographic indices; Multivariate Adaptive Regression Splines (MARS); Logistic Regression (LR)

Corresponding Author: Professor Christian Conoscenti, Ph.D.

Corresponding Author's Institution: University of Palermo, Italy

First Author: Christian Conoscenti, Ph.D.

Order of Authors: Christian Conoscenti, Ph.D.; Edoardo Rotigliano, Ph.D.

Abstract: In this study, the ability of five topographic indices to predict the gully trajectories observed in two adjacent watersheds located in Sicily (Italy) was evaluated. Two of these indices, named MSPI and MTWI, as far as we know, have never been employed to this aim. They were obtained by multiplying the stream power index (SPI) and the topographic wetness index (TWI), respectively, by the convergence index (CI). The predictive ability of the topographic indices was measured by using both cut-off independent (AUC: area under the receiver operating characteristic curve) and dependent statistics (Cohen's kappa index κ, sensitivity, specificity). These statistics were calculated also for 100 MARS (multivariate adaptive regression splines) and 100 LR (logistic regression) model runs, which used as predictors the topographic variables (i.e. contributing area, slope steepness, plan curvature and convergence index) combined into the five indices. Performance statistics of both topographic indices and statistical models were calculated using 100 random samples of 2 m grid cells, which were extracted only from flow concentration lines. This was done in order to focus the validation process on where gully erosion is more likely to occur. MSPI achieved the best predictive skill (AUC > 0.93; κ > 0.71) among the topographic indices and exhibited similar and better accuracy than local (i.e. trained and validated in the same watershed) and transferred (i.e. trained in one watershed and tested in the other one) LR models, respectively. On the other hand, MSPI performed similarly to transferred MARS runs (AUC > 0.92; κ > 0.71) but slightly worse than local MARS runs (AUC > 0.95; κ > 0.77). Based on the results of this experiment, it can be inferred that (i) including CI helps in detecting hollow areas where gullies are more likely to occur and (ii) MPSI can be a valid alternative to a data driven approach for mapping gully erosion susceptibility in areas where a gully inventory is not available, which is necessary to calibrate statistical models.

Research Data Related to this Submission

There are no linked research data sets for this submission. The following reason is given:

Data will be made available on request

Dear Editor, we have revised our work and manuscript in order to address comments and suggestions of the three reviewers.

---RESPONSES TO COMMENTS OF REVIEWER #1---

GENERAL COMMENT

This is a well-written article with clearly defined objectives, methods, results, and discussion. The study presents statistical and geospatial evaluation of the predictive performance of five topographic index models and two statistical modeling approaches. In addition to three standard TI models, two modified models were introduced by adding a convergence index CI to SPI and TWI models. This addition seemed to improve the gully predictions by adding sort of a "weighting" factor to each cell that would enforce channel incision. While this modification did not improve the TWI model's prediction, it drastically enhanced the prediction ability of the SPI model. The TWI model has a logarithmic form while the SPI model is a simple two factor product. The two statistical models (LR and MARS) unsurprisingly outperformed the TI models due to their extensive calibration on topographic attributes prior to their application to the watershed. It was interesting to learn that the thresholds for topographic index models and occurrence statistics were found to be in the range of the ones reported for Kansas, USA. With different soil types, land management, rainfall characteristics, etc, in two regions of the world, basic topographic features (slope, contributing area, curvature, convergence) appeared to yield similar results. Gullies provide sources of extensive erosion and their placement in the watershed is not yet fully understood, a development of predictive techniques is of paramount importance for watershed management and planning. I recommend the article to be published after minor comments below are addressed.

RESPONSE

Dear Reviewer, the authors wish to thank you for having read carefully our work and for having appreciated, as well as for your valuable suggestions. We have modified our manuscript to address your suggestions.

COMMENT #1

L. 148: "combine two or more primary topographic attributes". Which ones? Recommend providing examples.

RESPONSE #1

We added: “including contributing area, slope steepness, plan curvature and convergence index agree”. These are the primary topographic attributes which were combined.

COMMENT #2

L. 149: Any topographic index modelling heavily relies on the quality of DEM. A more specific description of the locally-developed Lidar-based DEM would be appreciated, the URL refers to a site in Italian.

RESPONSE #2

We agree. We added more details, providing information about vertical accuracy of the DEM. Moreover, we provided a new URL (WMS server) which can be used to load the DEM in a GIS software.

COMMENT #3

L. 197: "where CI (m) is the convergence index (Köthe et al., 1996)." Although the basic description of the CI index is presented, the mathematical definition remains unclear. The citation refers to publication in German with no English translation. Since this index provides a significant improvement to the SPI model, providing mathematical and possibly graphical representation would be very helpful.

RESPONSE #3

Thanks for this comment. We definitely agree. We added more details about how CI is calculated. We hope that now is more clear.

COMMENT #4
Table 1 and Table 2. The values in the tables have 4 or 3 decimal places. Recommend maybe rounding up to the same number, say 3?
RESPONSE #4
We agree, 4 decimal places are too much. We rounded up the values of Table 1 to 3 decimal places. We also rounded to 1 or 2 decimal places large numbers of Table 2.

---RESPONSES TO COMMENTS OF REVIEWER #2---

GENERAL COMMENT
Dear Editor
I have read this article very carefully. Unfortunately I have the following concerns. I am not satisfy with present form of this article. I am not positive for my decision.
RESPONSE
Dear Reviewer, the authors wish to thank you for having read carefully our work and for having provided very useful suggestions and comments which helped us to improve our manuscript. We have modified our manuscript to address your suggestions/comments.

COMMENT #1
Highlights (for review)
In the first highlight, please remove word of "we".
RESPONSE #1
Ok, done. We rephrased the point to: "The ability of five topographic indices to predict gullies was evaluated"

COMMENT #2
Headers
Headers aren't according to reference format, for example" ABSTRACT" isn't correct, please replace with "Abstract"
RESPONSE #2
Ok, done. We changed the title of the abstract paragraph according to the suggestion.

COMMENT #3
In general, please remove word of "we" from all of text. This isn't suitable for international publications.
RESPONSE #3
The active voice we + verb was changed to the passive voice.

COMMENT #4
Line 16: five or four indices?
RESPONSE #4
(Line 16) - We changed line 16 and we hope that now is more clear. The topographic indices are five. These are made by different combinations of four primary topographic attributes (i.e. contributing area, slope steepness, plan curvature and convergence index).

COMMENT #5
In general, abstract is so vague… it is important to clear this part very carefully.
RESPONSE #5
We changed some parts of the abstract and we hope that now is less vague. Indeed, it provides information about the indices and the statistical models employed to predict the gullies, the validation strategy, the metrics employed to measure the predictive performance, the results and the conclusions.

COMMENT #6
1. Introduction
Line 42: "Gullying" isn't a correct word. Pease edit it.

Lines 113-116: these paragraph isn't suitable for introduction.

RESPONSE #6
(Line 42) - We are sorry but the term "gullying" is often used as a synonymous of "gully erosion" in a number of very important publications. For example: "Gully Erosion: Procedures to Adopt When Modelling Soil Erosion in Landscapes Affected by Gullying" (Poesen, Torri, Van Walleghem, 2011) or "Badlands and gullying" (Howard, 2009).

(Lines 113-116) – We agree. The paragraph was moved to the methods section (lines 291-294).

COMMENT #7
2.1. Study area and gully inventory
Lines 135-137: authors just used from GE images? Any field surveys?

Line 139: please introduce source a 2-m raster DEM.
RESPONSE #7
(Lines 135-137) - Actually, the gully inventory was prepared by analyzing the GE image dated 3 May 2015, but also field surveys to check the inventory were performed. Now we specify it in the text. However, we specified that most of the gully channels in the drainage basins are ephemeral and are usually filled in by tillage within few months after their initiation. Thus not all the mapped gullies were visible in the field or in more recent GE images.

(Line 139) - We added more details about the DEM and a new URL (WMS server) which can be used to load the DEM in a GIS software. (Lines 116-117)

COMMENT #8
2.2. Topographic indices
Eq. 2, PLANC is plan curvature… please write plan curvature in text.

Eqs. 4 and 5: I am not satisfy from these two indices because they are using from CI and this index used from AS and S… So, I think authors used from double AS and S. Is this correct?
RESPONSE #8
(Eq. 2) – *PLANC* is explained just below the equation, as done for *As* (specific contributing area) and *S* (slope gradient). Moreover, *PLANC* is employed in many other papers to refer to "plan curvature".

(Eqs. 4 and 5) – Now we explained better how *CI* is calculated (Lines 190-195). This attribute is calculated from slope aspect (not from *As* and/or *S*).

COMMENT #9
2.3. Statistical modelling
It is better to authors separate description of models from multicollineairty test. In general, I think methodology needs to write better than previous to remove some vague.

Line 228: how many gully locations do you find in two watersheds? Authors written 1928 and 717 cells. How many gully locations?
RESPONSE #9
In the 2.3 section, we explain that the topographic attributes As, S, PLANC and CI, were used as independent variables of MARS and LR models. Since these statistical techniques require absence of multicollinearity, we simply verified that there was no strong relationship between the topographic attributes employed as predictors. We do not consider this as a result of the research, that's why we reported it in the methods section.

(Line 228) – Thanks for this comment, we realized that we did not provide this important information. In the revised version of the manuscript, we specify in line 141 that "The inventory includes 115 gullies (83 in W1, 32 in W2)".

COMMENT #10
2.5. Gully prediction maps
Line 290: the four ensemble statistical models? Ensemble models?
In general, I think it is important to add a flowchart of used methodology. Really, methodology is written difficult to understand it.

RESPONSE #10
(Line 290) – Dear Reviewer, thanks for this comment. We followed your suggestion and added a flow chart (new Fig. 1) to schematically explain the methodology. We hope that it helps to understand the following steps: 1) random selection of 100 calibration samples for W1 and 100 for W2, each including 25% of the gully pixels and the same number of non-event pixels; 2) Each of this sample was used to calibrate one LR and one MARS model, thus we have calibrated 100 LR and 100 MARS models; 3) random selection of 100 validation samples for W1 and 100 for W2 (each including 25% of the gully pixels and the same number of non-event pixels); 4) for both LR and MARS, calculation of gully probability for the pixels of each validation sample by averaging the score provided by the 100 model runs. Thus we have for one validation pixel 100 probability values, which were averaged to provide one value of gully probability. As explained in line 262 (and in the flow-chart), the "LR and MARS ensemble models" are prepared by averaging the score of the 100 model runs.
We report here the following text taken from Kotu and Deshpande, 2015 (citation added to the manuscript), which explain well what an ensemble model is: "Ensemble modeling is a process where multiple diverse models are created to predict an outcome, either by using many different modeling algorithms or using different training data sets. The ensemble model then aggregates the prediction of each base model and results in once final prediction for the unseen data." https://www.sciencedirect.com/topics/computer-science/ensemble-modeling
In our case, the ensemble model is prepared "using different training data sets" and calculating the average value of probability.

COMMENT #11
3. Results
This part started by sub-header "predictive performance measured…"
I think author at first have to write about gully prediction maps and then add validation and other things.
RESPONSE #11
Dear Reviewer, we first present the results of the validation process because these serve as premise to understand the reliability of the models which were applied to all the study area to prepare the gully erosion susceptibility maps.

COMMENT #12
4. Discussion
This part written the same with introduction, indeed it isn't a discussion (Lines 374-390).

I don't know why authors presented text without and with figures… any reason?
RESPONSE #12
Dear Reviewer, we added this part because we think that a comparison with the results found in other areas by applying the same indices could be very useful (as also highlighted by Reviewer #1).

We believe that the figure showing the kernel density plots of *CI* and *PLANC* is useful in the discussion section, because it supports the hypothesis that the contribution of *CI* in increasing the ability to discriminate between non-gully and gully cells is higher than that provided by *PLANC*.

COMMENT #13
Figures:
What is your reason for adding Fig. 2? You don't use from these factors in your analysis.

In Fig. 3, contour lines are 10-m, but in Fig. 4 they are 2-m. which one?

I cant understand Fig. 7. Please present gully erosion map for each watershed separately.
RESPONSE #13
Dear Reviewer, we believe that the maps shown in Fig. 2 (Fig. 3 in the revised version) could be useful for the reader. Indeed, the elevation map (DEM) is used to derive the topographic variables; slope angle is included in the topographic indices and is used as predictor variable of LR and MARS models; Lithology and Soil use maps (as well as elevation and slope) may help in understanding the geomorphological setting of the area.

We used different contour intervals because Fig. 3 (Fig. 4 in the revised version) shows the entire area whereas Fig. 4 (Fig. 5 in the revised version) shows at larger scale a small portion of the area to highlight the correspondence between flow lines and gully trajectories.

Fig. 7 (Fig. 8 in the revised version) shows two small portions of the catchments W1 and W2, corresponding to the GE views shown in Fig. 3 (Fig. 4 in the revised version).

---RESPONSES TO COMMENTS OF REVIEWER #3---

GENERAL COMMENT
I read the manuscript carefully. I found it interesting and practical. However, there are some unclear points and drawbacks in the manuscript.
I have provided some minor comments to improve the manuscript.
RESPONSE
Dear Reviewer, the authors wish to thank you for having read carefully our manuscript and for having provided very useful suggestions. We have modified our manuscript in order to address your comments.

COMMENT #1
It's interesting that you have used two indices (MSPI and MTWI) for the first time in this field of study. I'm not sure that they have developed by authors or they are previously available. If the later is correct, please add their original citations. In addition, these indices are the main part of the study and readers expect to get some information regarding them. Please describe these indices in detail and say how they can reflect gully erosion (direct and indirect impacts).
RESPONSE #1
Thanks for this comment. As far as we know, the two indices, which we called MSPI and MTWI, have never been proposed or used before. We added more details about how MSPI and MTWI are calculated and about the expected relationship with the spatial distribution of the gullies.

COMMENT #2

Literature review should be improved. There are some studies that used data-mining and machine learning models for gully-erosion susceptibility mapping. Please consider them in the introduction and discussion sections.

RESPONSE #2

Dear reviewer, in the revised version of our manuscript we consider also the contribution of other recent studies which employed data-mining and machine learning models to map gully-erosion susceptibility.

COMMENT #3

I agree with this sentence in the manuscript "MPSI can be a valid alternative to a data driven approach for mapping gully erosion susceptibility in areas where a gully inventory is not available" but authors have improperly proposed this point in the Highlight section. Data-mining and machine learning always perform better than a single index when a gully erosion inventory is available.

RESPONSE #3

We agree, thus we changed the highlight point to: "*MSPI* can be an alternative to a data-driven approach if gullies are not yet mapped"

- The ability of five topographic indices to predict gullies was evaluated
- Two of these indices, named *MSPI* and *MTWI*, have never been used to predict gullies
- Among the indices tested, *MSPI* (= *SPI* • *CI*) exhibited the best accuracy
- The convergence index (*CI*) helps in detecting where a gully is more likely to occur
- *MSPI* can be an alternative to a data-driven approach if gullies are not yet mapped

Predicting gully occurrence at watershed scale: comparing topographic indices and multivariate statistical models

Christian Conoscenti, Edoardo Rotigliano

A B S T R A C T

In this study, the ability of five topographic indices to predict the gully trajectories observed in two adjacent watersheds located in Sicily (Italy) was evaluated. Two of these indices, named *MSPI* and *MTWI*, as far as we know, have never been employed to this aim. They were obtained by multiplying the stream power index (*SPI*) and the topographic wetness index (*TWI*), respectively, by the convergence index (*CI*). The predictive ability of the topographic indices was measured by using both cut-off independent (*AUC*: area under the receiver operating characteristic curve) and dependent statistics (Cohen's kappa index κ, sensitivity, specificity). These statistics were calculated also for 100 MARS (multivariate adaptive regression splines) and 100 LR (logistic regression) model runs, which used as predictors the topographic variables (i.e. contributing area, slope steepness, plan curvature and convergence index) combined into the five indices. Performance statistics of both topographic indices and statistical models were calculated using 100 random samples of 2 m grid cells, which were extracted only from flow concentration lines. This was done in order to focus the validation process on where gully erosion is more likely to occur. *MSPI* achieved the best predictive skill (*AUC* > 0.93; κ > 0.71) among the topographic indices and exhibited similar and better accuracy than local (i.e. trained and validated in the same watershed) and transferred (i.e. trained in one watershed and tested in the other one) LR models, respectively. On the other hand, *MSPI* performed similarly to transferred MARS runs (*AUC* > 0.92; κ > 0.71) but slightly worse than local MARS runs (*AUC* > 0.95; κ > 0.77). Based on the results of this experiment, it can be inferred that (i) including *CI* helps in detecting hollow areas where gullies are more likely to occur and (ii) *MPSI* can be a valid alternative to a data driven approach for mapping gully erosion susceptibility in areas where a gully inventory is not available, which is necessary to calibrate statistical models.

Predicting gully occurrence at watershed scale: comparing topographic indices and multivariate statistical models

Christian Conoscenti [a],*, Edoardo Rotigliano [a]

[a] Department of Earth and Marine Sciences (DISTEM), University of Palermo, Via Archirafi 22, 90123 Palermo, Italy

~~A B S T R A C T~~

Abstract

In this study,~~ we evaluated~~ the ability of five topographic indices to predict the gully trajectories observed in two adjacent watersheds located in Sicily (Italy~~).~~) was evaluated. Two of these indices, named *MSPI* and *MTWI*, as far as we know, have never been employed to this aim. They were obtained by multiplying the stream power index (*SPI*) and the topographic wetness index (*TWI*), respectively, by the convergence index (*CI*). The predictive ability of the topographic indices was measured by using both cut-off independent (*AUC*: area under the receiver operating characteristic curve) and dependent statistics (Cohen's kappa index κ, sensitivity, specificity). These statistics were calculated also for 100 MARS (multivariate adaptive regression splines) and 100 LR (logistic regression) model runs, which used as predictors the topographic variables ~~combined in the five indices~~ (i.e. contributing area, slope steepness, plan curvature and convergence index~~).~~) combined into the five indices. Performance statistics of both topographic indices and statistical models were calculated using 100 random samples of 2 m grid cells, which were extracted only from flow concentration lines. This was done in order to focus the validation process on where gully erosion is more likely to occur. *MSPI* achieved the best predictive skill (*AUC* > 0.93; κ > 0.71) among the

Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1, Border: : (No border), Pattern: Clear (White)
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1, Italian (Italy)
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1, Italian (Italy)
Formatted: Font color: Text 1

topographic indices and exhibited similar and better accuracy than local (i.e. trained and validated in the same watershed) and transferred (i.e. trained in one watershed and tested in the other one) LR models, respectively. On the other hand, *MSPI* performed similarly to transferred MARS runs (*AUC* > 0.92; κ > 0.71) but slightly worse than local MARS runs (*AUC* > 0.95; κ > 0.77). Based on the results of this experiment, ~~we infer~~it can be inferred that (i) including *CI* helps in detecting hollow areas where gullies are more likely to occur and (ii) *MPSI* can be a valid alternative to a data driven approach for mapping gully erosion susceptibility in areas where a gully inventory is not available, which is necessary to calibrate statistical models.

Keywords: Gully erosion susceptibility; Topographic indices; Multivariate Adaptive Regression Splines (MARS); Logistic Regression (LR); Geographic Information System (GIS)

* Corresponding author. Tel.: +39 09123864670; fax: +39 0916169908. E-mail address: christian.conoscenti@unipa.it (C. Conoscenti).

Formatted: Font color: Text 1, Italian (Italy)

Formatted: Font color: Text 1

Formatted: Font color: Text 1, Italian (Italy)

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1, Pattern: Clear

Formatted: Font color: Text 1

1. Introduction

Gully erosion causes land degradation in a wide range of environmental conditions. The development of gullies in agricultural watersheds may induce high soil loss and reduction of water availability, leading to a significant decrease of soil quality and crop yield. Moreover, gully channels hamper the trafficability of the fields causing extra damages and costs to farmers (Poesen et al., 2003, 2011).

Gullying is a threshold phenomenon that is mainly controlled by rainfall, topography, soil, lithology and land use. Gullies occur only after a threshold of runoff erosivity and soil erodibility is exceeded. In addition to rainfall, runoff erosive power depends on topography which regulates discharge, concentration and velocity of overland flow (e.g., Moore et al., 1988; Desmet et al., 1999; Poesen et al., 2003; Valentin et al., 2005; Gómez-Gutiérrez et al., 2009a; Daggupati et al., 2013; Conoscenti et al., 2013). Morphology, density and development of gullies in a given landscape is also significantly controlled by parent material (Oostwoud Wijdenes et al., 2000; Vandekerckhove et al., 2001; Poesen et al., 2011). Furthermore, gully occurrence is controlled by resistance of soil, which is influenced by soil properties such as texture, bulk density, moisture conditions, organic matter content (Poesen et al., 2003). Soil erosion susceptibility is also related to crop type and stage, as well as tillage direction and conservation practices (Parker et al., 2007). Also, several studies have reported triggering of gullies or increasing of gully erosion rates as being caused by land use changes, intensification of farming activities and overgrazing (Poesen et al., 2003; Valentin et al., 2005; Zucca et al., 2006; Gómez-Gutiérrez et al., 2009b).

Planning of gully erosion control in agricultural watersheds requires either quantifying soil loss and predicting gully location. Several process-based models have been developed to quantify gully erosion (e.g., CREAMS, Knisel, 1980; EGEM, Merkel et al., 1988; GLEAMS, Knisel, 1993; Sidorchuk, 1999; REGEM, Gordon et al., 2007). However, these models require physical input variables that are difficult to measure at the watershed scale. Soil loss due to gully erosion can be also evaluated by using empirical models which are based on relationships established between

Formatted: Font color: Text 1, Italian (Italy)

Formatted: Font color: Text 1

Formatted: Font color: Text 1, Italian (Italy)

Formatted: Font color: Text 1

volume and length of the gully channels (e.g., Nachtergaele et al., 2001; Capra and Scicolone, 2002; Capra et al., 2005; Caraballo-Arias et al., 2014, 2015).

Formatted: Font color: Text 1, Italian (Italy)

Formatted: Font color: Text 1

Prediction of gully location can be achieved by identifying a topographic threshold that has to be exceeded for a gully to form. A number of studies have proposed topographic threshold lines defined on a log-log plot of local slope gradient (*S*) versus upslope contributing area (*A*) measured at gully heads (e.g., Patton and Schumm, 1975; Montgomery and Dietrich, 1992; Nachtergaele et al., 2001b; Zucca et al., 2006; Nazari Samani et al., 2009). Both these topographic attributes are indeed widely considered to play the role of controlling factors in the gully formation process as they act as proxies for flow velocity and discharge, respectively. The approach based on *S*–*A* threshold lines assumes that for a given *A*, a critical *S* exists above which runoff erosivity is large enough to produce gully erosion. The *S*–*A* threshold can be used to predict gullies by classifying a study area into non-event positions (below the threshold line) and event positions (on or above the threshold line). However, this approach tends to overestimate the likelihood of gully occurrence (Svoray et al., 2012; Gómez-Gutiérrez et al., 2015), providing a high number of false positives (i.e. non-gullied positions classified as gullied).

Furthermore, several topographic indices have been employed to predict gully location (e.g., Thorne et al., 1986; Moore et al., 1988; Vandaele et al., 1996; Desmet et al., 1999). These models rely on the assumption that gully formation depends on a combination of primary topographic attributes (Wilson and Gallant, 2000) which reflect erosivity of concentrated overland flow; gully erosion occurs when the topographic index exceeds a critical threshold value. Daggupati et al. (2013), Sekaluvu et al. (2015) and Sheshukov et al. (2018) have compared the ability to discriminate between gullied and non-gullied areas of several topographic indices, which were applied using different thresholds. Their analyses revealed that gully predictions were not accurate without identifying an optimal threshold through local calibration. Indeed, they have observed that a low threshold causes high number of false positives whereas a high threshold produces high number of false negatives (i.e. gullied sites predicted as non-gullied).

Formatted: Font color: Text 1, Italian (Italy)

Formatted: Font color: Text 1

Recently, accurate predictions of gully locations have been achieved by using statistical modeling and data mining techniques such as logistic regression, classification and regression trees, multivariate adaptive regression splines, stochastic gradient treeboost, artificial neural network, random forest, maximum entropy, etc. ~~(e.g., Meyer and Martínez-Casasnovas, 1999; Gómez-Gutiérrez et al., 2009c; Eustace et al., 2011; Svoray et al., 2012; Conoscenti et al., 2014, 2018; Dewitte et al., 2015; Angileri et al., 2016; Pourghasemi et al., 2017; Garosi et al., 2018, 2019).~~(e.g., Meyer and Martínez-Casasnovas, 1999; Gómez-Gutiérrez et al., 2009c; Eustace et al., 2011; Svoray et al., 2012; Conoscenti et al., 2014, 2018; Dewitte et al., 2015; Angileri et al., 2016; Pourghasemi et al., 2017; Rahmati et al., 2016, 2017a, 2017b; Garosi et al., 2018, 2019; Azareh et al., 2019; Choubin et al., 2019; Javidan et al., 2019). These techniques are able to analyze and model the relationships between gully locations and spatial variability of a set of environmental predictors related to topography, land use, parent materials and soils. Based on the identified statistical relationships, these techniques allow for calculating a probability of gully occurrence that ranges from 0 to 1, for each position (usually grid cell) in a given area. However, an important drawback in these procedures, which are data-driven, is that they generate prediction images which efficiently explain the gully distribution in the study area but tend to fail when exported to other areas, even if located at a close distance (Conoscenti et al., 2018).

This study focuses on investigating the topographic control of gully erosion caused by concentrated overland flow at watershed scale. The experiment was carried out in two small agricultural watersheds located in Sicily (Italy). The main goal of the study was to evaluate and compare the ability to predict the location of gullies achieved by using a set of topographic indices, which includes three indices previously proposed for predicting gully location and two modified versions of them. Predictive models of gully occurrence were prepared also by using logistic regression (LR; Hosmer and Lemeshow, 2000) and multivariate adaptive regression splines (MARS; Friedman, 1991), two statistical modeling techniques which have been successfully used to this aim in previous studies (e.g., Vanwalleghem et al., 2008; Gómez-Gutiérrez et al., 2009c; Svoray et al.,

Formatted: Font color: Text 1, Italian (Italy)

Formatted: Font color: Text 1

2012; Conoscenti et al., 2014, 2018; Dewitte et al., 2015). To further assess the ability to predict gully occurrence provided by the five topographic indices, their accuracy was compared with that achieved by LR and MARS models.

2. Materials and Methods

~~The statistical analyses~~In this study, the topographic analysis was carried out using a LiDAR-derived 2×2 m Digital Elevation Model (DEM; Regione Siciliana, 2010), with vertical accuracy of 0.1–0.2 m. The GIS calculations were performed using SAGA-GIS software (Conrad et al., 2015). The calibration of MARS and LR and the validation of both topographic indices and statistical models were performed using the R software (R Core Team, 2017) with the packages “raster~~”~~” (Hijmans, 2017), “usdm” (Naimi, 2015), “splitstackshape” (Mahto, 2018), “pROC” (Robin et al., 2011), “ROCR” (Sing et al., 2005), “caret” (Wing and Kuhn, 2018) and “earth” (Milborrow, 2018). The flow-chart of Fig. 1 shows a schematic overview of the methodology, which is described in detail in the following sections.

3. ~~Materials and Methods~~

3.1. Study area and gully inventory

The experiment was carried out in two adjacent agricultural watersheds located in central-western Sicily (Fig. ~~1~~2), approximately 35 km south-east of the city of Palermo. The westernmost watershed (W1) drains an area of 621.7 ha whereas the easternmost one (W2) covers 901.4 ha. The study area experiences a typical Mediterranean climate with an average annual rainfall of 711 mm (time interval: 2002–2017; Camporeale rainfall station; Regione Siciliana – SIAS - Servizio Informativo Agrometeorologico Siciliano), with a minimum in July (5.6 mm) and a maximum in December (88.7 mm). Topography of the two investigated watersheds is slightly different (Fig. ~~2a~~3a–b):

Formatted: Font color: Text 1, Italian (Italy)
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1

elevation ranges from 185 to 576 m a.s.l. in W1 (mean = 303 m) and from 209 to 571 m a.s.l. in W2 (mean = 345 m), whereas average slope gradient is 10.1° (SD = 5.0°) and 9.7° (SD = 6.9°), respectively. Soils are mostly regosols and vertisols with fine-medium texture (Fierotti, 1988). Lithologies are mainly eluvial-colluvial deposits, sands of the Late Miocene Terravecchia Fm., clays of the Middle-Late Miocene Castellana Sicula Fm., silty-clays and sandy-silts of the Terravecchia Fm. (Fig. ~~2c~~3c). Primary land covers are arable lands (mainly cereal fields) and vineyards, which occupy 92% of W1 and 80% of W2 (Fig. ~~2d~~3d).

Both watersheds are affected by gully erosion which increases soil loss, causes landscape dissection and hampers the movement of farm machines. Most of the gully channels in the drainage basins are ephemeral and are usually filled in by tillage within few months after their initiation. Conoscenti et al. (2018) created a gully inventory of the watersheds by analyzing a Google Earth image acquired on 3 May 2015 (Fig. ~~3).~~4) and by carrying out field surveys. As their objective was to model gully erosion due to overland flow concentration, the inventory includes only gullies located on concentrated flow pathways. The latter were extracted from ~~a 2 m raster Digital Elevation Model (DEM; Regione Siciliana, 2010)~~the DEM, by calculating for each cell the value of upstream contributing area. To ensure consistency between mapped gullies and contributing area, gully trajectories have been slightly modified in order to exactly match flow pathways and to ensure that contributing area increases along each gully from head to mouth (Fig. ~~4~~5). The inventory includes 115 gullies (83 in W1, 32 in W2) and reveals that gully erosion is more severe in W1 (gully density = 0.73 km^{-1}) than in W2 (0.18 km^{-1}). Gullies mostly occur on eluvial-colluvial deposits and clays. As regards land cover, arable lands host most of the gully trajectories.

3.2. Topographic indices

In this experiment, ~~we assessed~~ the ability to predict gully location of five topographic indices was assessed. These indices, which combine two or more primary topographic attributes (Wilson and Gallant, 2000~~). These attributes~~), including contributing area, slope steepness, plan curvature and convergence index, were calculated for each grid cell of ~~a LiDAR-derived 2×2 m~~the DEM ~~(Regione~~

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

~~Siciliana, 2010),~~, by using terrain analysis tools of SAGA-GIS software (Conrad et al., 2015).

Three topographic indices adopted here, namely stream power index (*SPI*), compound topographic index (*CTI*) and topographic wetness index (*TWI*), have been employed in previous studies to predict location of ephemeral gullies in cultivated watersheds (e.g., Vandaele et al., 1996; Parker et al., 2007; Daggupati et al., 2013, 2014; Sekaluvu et al., 2015; Sekaluvu and Sheshukov, 2016; Sheshukov et al., 2018).

The *SPI* (Moore et al., 1988, 1991) is a measure of erosive power of concentrated runoff and is calculated as:

$$SPI = A_s \bullet S \tag{1}$$

where A_s ($m^2\ m^{-1}$) is the specific contributing area and S ($m\ m^{-1}$) is the local slope gradient. A_s and S are employed as surrogates for flow discharge and velocity. A_s was extracted from upslope contributing area (A), which in turn was calculated by applying the single flow direction (also referred to as D8) algorithm (O'Callaghan and Mark, 1984), after filling sinks in the DEM. To obtain A_s, A has to be divided by the contour width within the pixel (Desmet and Govers, 1996). As the contour width can be set to the average of the grid cell width (i.e., 2.0 m) and the grid cell diagonal (i.e., 2.8 m), A_s was calculated dividing A by 2.4.

The *CTI* (Thorne et al., 1986) is defined as:

$$CTI = As \bullet S \bullet PLANC \tag{2}$$

where *PLANC* (m/100 m) is the curvature of the contour line (Hengl and Reuter, 2008). *PLANC* is a measure of local flow convergence and divergence and thus reflects the degree of concentration of the runoff. *CTI* is employed in the USDA Agricultural Non-Point Source (AGNPS) modelling

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

system (Bingner and Theurer, 2001) to identify potential ephemeral gully locations throughout a watershed (Parker et al., 2007; Momm et al., 2012, 2013).

TWI (Moore et al., 1988; 1991) is a measure of soil saturation and is calculated as:

$$TWI = \ln (A_s / S) \quad (3)$$

As *TWI* reflects zones of saturation in a watershed, it could also be an index of the potential location of ephemeral gullies. Indeed, gully heads often form where soils become very wet and lose their strength (Moore et al., 1988).

In addition to *SPI*, *CTI* and *TWI*, ~~we explored~~ the ability to predict gully locations of other two topographic indices was explored. These indices are modified versions of *SPI* and *CTI* and are calculated as:

$$MSPI = A_s \bullet S \bullet CI \quad (4)$$

$$MTWI = \ln (A_s / S) \bullet CI \quad (5)$$

~~where *CI* (m) is the convergence index (Köthe et al., 1996). *CI* measures to what extent neighboring cells point to the center cell and is calculated by setting a search radius. Differently from *PLANC*, which depends on local morphology, *CI* describes the general shape of the landscape up to a scale that depends from the set search radius. In this experiment, the *CI* value of each cell was calculated by averaging the values obtained by varying the search radius from 1 to 10 cells. As, *PLANC* and *CI* calculated by SAGA-GIS have negative values on concavities (e.g. valley bottoms) and positive values on convexities (e.g. ridges), a change in the sign of both parameters was performed before using them to calculate the topographic indices employed to predict gully location.~~

where *CI* is the convergence index (Köthe et al., 1996; Kiss, 2004; Thommeret et al., 2010). *CI*

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

measures to what extent neighboring cells point to the center cell. *CI* is calculated as the average difference between actual aspect of surrounding cells within a moving square or circular window and the direction to the center cell, minus 90 degrees. The value ranges from -90 degrees (max convergence) by 0 (planar slopes) to 90 degrees (max divergence). *CI* provided by SAGA-GIS is normalized between -100 and 100. Differently from *PLANC*, which depends on local morphology, *CI* describes the general shape of the landscape up to a scale that depends from the size of the moving window. In this experiment, the *CI* value of each cell was calculated by averaging the values obtained varying the search radius of a circular moving window from 1 to 10 cells. As *PLANC* and *CI* calculated by SAGA-GIS have negative values on concavities (e.g. valley bottoms) and positive values on convexities (e.g. ridges), a change in the sign of both parameters was performed before using them to calculate the topographic indices employed to predict gully location.

MSPI and *MTWI* could help in predicting gully occurrence as they estimate runoff erosive power and potential soil saturation, respectively, and incorporate a weighting factor which reflects flow convergence/divergence (i.e. *CI*).

3.3. Statistical modelling

In our experiment, the location of the gullies was also predicted by employing two statistical techniques, namely logistic regression (LR; Hosmer and Lemeshow, 2000) and multivariate adaptive regression splines (MARS; Friedman, 1991).

LR is a generalized linear model with a logistic link function. LR is among the most common statistical technique for prediction of gully occurrence (e.g., Meyer and Martínez-Casasnovas, 1999; Lucà et al., 2011; Conoscenti et al., 2014; Dewitte et al., 2015; Selkimäki and González-Olabarria, 2016). Conversely, MARS has been employed only recently to model gully erosion (Gómez-Gutiérrez et al., 2009a, 2009c, 2015; Arabameri et al., 2018; Garosi et al., 2018; Conoscenti et al., 2018). LR and MARS enable modelling of relationships between continuous and/or categorical independent variables and a dichotomous dependent variable (i.e. event or non-event). Both

Formatted: Font color: Text 1

Formatted: Font color: Text 1, Italian (Italy)

Formatted: Font color: Text 1

techniques consist of an additive combination of terms. LR has a linear structure with constant coefficients across the entire range of the predictor variables. Conversely, MARS uses piece-wise linear regressions with breaks at the knots to describe non-linear relationships between event occurrence and predictors. To reduce the complexity of the models, ~~we prepared~~ MARS models were prepared with terms made of single predictors ~~whereas,~~; as regards LR models, ~~we adopted~~ a bilateral stepwise strategy ~~that~~, which selects only the most significant predictors, was adopted. Please refer to Hosmer and Lemeshow (2000) and Friedman (1991) for further details about LR and MARS, respectively.

LR and MARS models were prepared by using as predictor variables the primary topographic attributes *S*, A_s, *PLANC* and *CI*. Since both the employed statistical techniques require absence of multicollinearity, the degree of correlation among these four variables was evaluated before running the models. To this aim, ~~we used~~ the variance inflation factor (*VIF*) (Jebur et al., 2014; Heckmann et al., 2014; Bui et al., 2015; Conoscenti et al., 2016; Cama et al., 2017; Rotigliano et al., 2019; Vargas-Cuervo et al., 2019), was employed. The results, which were interpreted according to the “rule of 10~~””~~”, revealed absence of strong correlations among the predictor variables (*VIF* range: 1.0 – 1.1).

Calibration of the statistical models was carried out separately in W1 and W2, where 100 learning samples were prepared by randomly selecting the 25% of the total number of event pixels and the same number of non-event pixels. This percentage was chosen in order to achieve a compromise between the attempt to minimize the effects of spatial autocorrelation and the effort to obtain robust models, by using a sufficiently large number of cases. Since 1928 and 717 gully cells were identified in W1 and W2, respectively, the W1 learning samples include 964 pixels (i.e. 482 non-event + 482 event cells, the latter corresponding to 25% of 1928) whereas 358 pixels (i.e. 179 non-event + 179 event cells, the latter corresponding to 25% of 717) form the W2 samples. The learning samples were employed to perform 100 LR and 100 MARS model runs in each of the watersheds. Hereafter, MARS1 and LR1 are used to indicate model runs calibrated in W1 whereas MARS2 and

Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1

LR2 indicate model runs calibrated in W2.

3.4. Validation strategy

The ability to predict gully occurrence of topographic indices and statistical models was measured on a network of flow lines which were identified separately in W1 and W2 by using two different thresholds of contributing area. The thresholds were set equal to the minimum ~~*A*~~A_s of W1 and W2 gully cells, respectively, after discarding values below the 1st percentile which were regarded as outliers. By using this approach, ~~we measured and compared~~ the predictive performance of topographic indices and statistical models ~~focusing~~was measured where drainage area is sufficient to trigger gully erosion, given the rainfall, soil, bedrock and land use characteristics which caused gullying in our study watersheds.

One hundred validation samples were prepared by randomly selecting pixels from flow lines of both W1 and W2. Like the calibration samples, also the validation samples include the 25% of the gully cells and a same number of non-gully cells. The value of the topographic indices was used directly as a score to predict the distribution of gully cells. As regards statistical modelling, the probability of gully occurrence was calculated from LR and MARS ensemble models~~, which were prepared by averaging the score of the 100 model runs.~~ (Kotu and Deshpande, 2015), which were prepared by averaging the score of the 100 model runs. This procedure was applied in order to generate a more stable performance of the models and to mitigate the effects of prevalence (i.e. different proportion of event/non-event cells in the study area) (Svoray et al., 2012). ~~We measured the~~The predictive performance of both “local” (i.e. calibrated and validated in the same watershed) and “transferred” (i.e. calibrated in one watershed and validated in the other one) statistical models was measured.

The accuracy of the topographic indices and statistical models was assessed by plotting for each validation sample the receiver operating characteristic (ROC) curve (e.g., Lasko et al., 2005; Brenning, 2005; Frattini et al., 2010; Cama et al., 2015, 2016) and by calculating the area under the ROC curve (*AUC*). ROC curve analysis is a cut-off independent technique for assessing the

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

performance of predictive models, which plots all possible values of sensitivity (i.e. true positive rate, *TPR*) against the corresponding value of 1-specificity (i.e. false positive rate, *FPR*). The ideal predictive model achieves an *AUC* value close to 1, whereas a value close to 0.5 reveals inaccuracy in the model (Nandi and Shakoor, 2009). In this experiment, accuracy of the models was interpreted as acceptable, excellent or outstanding if *AUC* values were higher than 0.7, 0.8 and 0.9, respectively (Hosmer and Lemeshow, 2000). In both W1 and W2, a group of 100 ROC curves and related *AUC* values, was obtained (one for each validation sample) for each topographic index and statistical model. Comparisons between *AUC* groups were performed by using box plots and the Wilcoxon signed-rank test, setting the level of significance at 0.01.

Furthermore, the predictive ability of topographic indices and statistical models was evaluated by using cut-off dependent performance metrics such as Cohen's kappa index (Cohen, 1960; Landis and Koch, 1977; Monserud and Leemans, 1992; Geissen et al., 2007; Frattini et al., 2010; Sterlacchini et al., 2011), sensitivity (or *TPR*) and specificity (i.e. true negative rate, *TNR*). The Cohen's kappa index (κ) reflects the degree of agreement between prediction and observation and is calculated as:

$$\kappa = P_{obs} - P_{exp} / (1 - P_{exp}) \quad (6)$$

where P_{obs} and P_{exp} are the observed and the expected proportion of agreement, respectively. κ values were interpreted according to Monserud and Leemans (1992), which evaluated the agreement between model prediction and observation as: 1.00, perfect; 0.85–0.99, excellent; 0.70–0.85, very good; 0.55–0.70, good; 0.40–0.55, fair; 0.20–0.40, poor; 0.05–0.20, very poor; <0.05, null.

~~To calculate κ, *TPR* and *FPR*, we first prepared the average ROC curve from each group of 100 validation ROC curves. We then identified the optimal cut-off values of these curves by using the Youden's index (*J*) (Youden, 1950; Angileri et al., 2016; Cama et al., 2017; Rotigliano et al., 2019),~~

Formatted: Font color: Text 1, Italian (Italy)

Formatted: Font color: Text 1

Formatted: Font color: Text 1, Italian (Italy)

Formatted: Font color: Text 1

Formatted: Font color: Text 1, Italian (Italy)

Formatted: Font color: Text 1

~~which corresponds to the threshold that maximizes the sum of sensitivity and specificity. Then, by using *J* as threshold (*T*) to classify the grid pixels as not susceptible (score < *T*) or as susceptible (score > *T*) to gully erosion, we prepared the contingency tables for each topographic index and ensemble statistical model.~~

Firstly, to calculate κ, *TPR* and *FPR*, the average ROC curve from each group of 100 validation ROC curves was prepared. Then, the optimal cut-off values of these curves were identified by using the Youden's index (*J*) (Youden, 1950; Angileri et al., 2016; Cama et al., 2017; Rotigliano et al., 2019), which corresponds to the threshold that maximizes the sum of sensitivity and specificity. Then, by using *J* as threshold (*T*) to classify the grid pixels as not susceptible (score < *T*) or as susceptible (score > *T*) to gully erosion, the contingency tables were prepared for each topographic index and ensemble statistical model.

3.5. Gully prediction maps

A gully susceptibility map of the study area was obtained from each of the topographic indices and the four ensemble statistical models which were prepared by averaging the score of 100 MARS and LR model runs. Susceptibility to gully erosion was then classified into four levels according to thresholds that were calculated separately in W1 and W2 by using the steps described below, which were repeated for each topographic index and ensemble statistical model. First, *J* was used to separate the pixels of the 100 validation samples into a low susceptibility dataset (score < *J*) and a high susceptible dataset (score > *J*). Then, ~~we prepared~~ the average ROC curve and ~~calculated~~ the Youden index were calculated for both the low susceptibility dataset (J_{low}) and ~~for~~ the high susceptibility dataset (J_{high}). Finally, ~~we identified~~ the following four levels of susceptibility to gully erosion were identified: i) low (score ≤ J_{low}); ii) moderate (J_{low} < score ≤ *J*); iii) high (*J* < score ≤ J_{high}); iv) very high (score > J_{high}).

4. Results

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

4.1. Predictive performance measured by using a cut-off independent statistic

The ability of the topographic indices and statistical models to discriminate between gully and non-gully cells of the validation samples is graphically represented by the box plots of Fig. ~~5~~6. Each box plot reveals the variability of a group of 100 *AUC* values by indicating their quartiles, as well as the lowest and the highest data still within 1.5 interquartile range of the lower quartile and of the upper quartile, respectively. Furthermore, descriptive statistics such as mean and standard deviation of each *AUC* group are reported in Table 1.

Formatted: Font color: Text 1

The *AUC* values reflect excellent (*AUC* > 0.8) to outstanding (*AUC* > 0.9) discrimination ability of indices and models applied to predict gullies occurred in the studied watersheds. However, significant differences of accuracy can be detected.

MSPI performed clearly better than the other indices in both watersheds. In W1, only *SPI* achieved a similar performance but still significantly lower than that obtained from *MSPI*. In W2, *SPI* performed better than *TWI* but not significantly different from *CTI* and *MTWI*. *TWI* performed better than its modified version (i.e. *MTWI*) in W1, whereas the opposite was observed in W2.

As regards statistical models, MARS performed better than LR in both watersheds. Accuracy of MARS and LR is significantly different even in W1, where *AUC* values appear quite similar. A not significant difference was observed only in W1 between local (i.e. trained in W1) LR and transferred (i.e. trained in W2) MARS models (*p*-value = 0.284). In W1, both MARS and LR local models (i.e. MARS1 and LR1) exhibited higher accuracy than transferred models (i.e. MARS2 and LR2). On the other hand, a not significant difference of *AUC* was observed in W2 between local and transferred LR models (*p*-value = 0.5221).

The *AUC* values and the Wilcoxon signed-rank test revealed an overall better predictive performance of the statistical models with respect to the topographic indices, with the exception of *MSPI*. The latter indeed achieved outstanding accuracy in both watersheds. In W1, *MSPI* exhibited the same accuracy of transferred MARS and local LR runs and better predictive ability than transferred LR runs. In W2, *MSPI* achieved higher accuracy than both local and transferred LR runs and the same accuracy of MARS1. Only local MARS models performed significantly better than MSPI.

4.2. Predictive performance measured by using cut-off dependent statistics

Fig. ~~6~~7 shows the average ROC curves obtained from the validation of the topographic indices and statistical models in W1 and W2. These curves were employed to calculate the optimal cut-off (*T*) that maximizes the sum of sensitivity and specificity and which graphically corresponds to the maximum distance to the diagonal lines plotted in Fig. ~~6~~8. The value of *T*, as well as those of kappa index (κ), *TPR* and *TNR* are reported in Table 2. Kappa values obtained for the five topographic indices vary from 0.625 to 0.795 indicating a good ($\kappa > 0.55$) to very good ($\kappa > 0.70$) ability to discriminate between event and non-event pixels. As revealed by *AUC* values, the kappa index also demonstrated that *MSPI* achieved the best predictive skill in both watersheds. *SPI* reached a κ value close that of *MSPI* in W1. Conversely, *SPI* accuracy appears similar to that of *TWI* and *MTWI* in W2, where *CTI* achieved the second best κ value. As regards sensitivity and specificity, *MSPI* obtained the highest values in W1 whereas in W2 a slightly higher *TPR* and *TNR* was observed for *MTWI* and *TWI*, respectively.

Kappa index revealed approximately the same difference of performance between MARS and LR models which is highlighted by the *AUC* values. Indeed, MARS achieved higher κ values in both watersheds, with more enhanced difference of accuracy occurring in W2, where LR models are below the threshold indicating very good performance ($\kappa > 0.7$). The difference of performance observed in W1 appears related more to a difference in specificity than in sensitivity, which is very similar for MARS and LR models. On the other hand, in W2, MARS runs exhibit higher values of both *TPR* and *TNR*, whereas only transferred models show a similar sensitivity.

Kappa, *TPR* and *TNR* confirm that *MSPI* achieves approximately the same accuracy of MARS runs. Furthermore, these statistics reveal that *MSPI* outperforms both LR local and transferred models which in turn show better discrimination ability when compared to the other topographic indices, with the exception of *SPI*, in W1, and *CTI*, in W2.

Formatted: Font color: Text 1, Italian (Italy)

Formatted: Font color: Text 1, Italian (Italy)

Formatted: Font color: Text 1, Italian (Italy)

Formatted: Font color: Text 1, Italian (Italy)

Formatted: Font color: Text 1, Italian (Italy)

Formatted: Font color: Text 1

Formatted: Font color: Text 1, Italian (Italy)

Formatted: Font color: Text 1

Formatted: Font color: Text 1, Italian (Italy)

Formatted: Font color: Text 1

4.3. Gully prediction maps

Fig. ~~7~~8 shows the gully prediction maps for the sectors of W1 and W2 highlighted in Fig.~~3~~ 4, obtained from the topographic indices and the ensemble statistical models. To aid the assessment of the maps, Fig. ~~8~~9 plots the relative frequency distributions of non-event and event pixels across the susceptibility levels. The gully erosion susceptibility maps show very low probability of gully occurrence in most part of the study area, with the exception of few flow lines where susceptibility level is from moderate to very high. Only maps derived from *MTWI* and LR, especially in W2, show slightly larger sectors with moderate to high probability of gully occurrence. This is confirmed by the bar plots of Fig. ~~8~~9, which reveal that non-event cells occur with a frequency higher than 5% only over moderate probability levels of *MTWI* maps and of LR maps of W2. On the other hand, although their very low frequency, high and very high susceptibility levels of all the maps host most of the gully pixels. In particular, the maps derived from *SPI*, *MSPI* and MARS1 ensemble model, achieve the highest percentage of gully pixels within the very high level of susceptibility (Fig. ~~8~~9).

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

5. Discussion

The results of our experiment showed that the spatial distribution of gullies can be effectively predicted by using either topographic indices or statistical models.

Both cut-off independent and dependent performance metrics revealed that, among the employed topographic indices, the best accuracy in predicting gully occurrence is achieved by *MSPI* whereas *MTWI* exhibited similar or worse performance than *SPI*, *CTI* and *TWI*. The ability of the latter indices to discriminate between gully and non-pixels was evaluated and compared, by identifying optimal thresholds and by calculating the κ index, in three previous studies (Daggupati et al., 2013; Sekaluvu et al., 2015; Sheshukov et al., 2018) performed in Kansas. Daggupati et al. (2013) estimated the thresholds of 30 – 50, 62, and 12, respectively, for *SPI*, *CTI* and *TWI*. Sekaluvu et al. (2015) and Sheshukov et al. (2018) report that the critical thresholds required by *CTI* to best predict

Formatted: Font color: Text 1, Italian (Italy)

Formatted: Font color: Text 1

the gullies of two watersheds of central Kansas are equal to 79.4 and 25.1. These values are relatively similar to the *CTI* thresholds estimated by Daggupati et al. (2013) and those calculated in our experiment (52.8 and 24.3). As regards *SPI*, the thresholds found in our study (270.9 and 127.0) are of the same order of magnitude of those calculated by Sekaluvu et al. (2015) and Sheshukov et al. (2018) (501.2 and 158.5), but higher than the values reported by Daggupati et al. (2013). Furthermore, the *TWI* critical thresholds estimated in our experiment (9.7 and 9.4) are quite similar to those calculated for the Kansas areas (12.0 – 18.2).

By applying the thresholds cited above, Daggupati et al. (2013) found a poor predictive performance of *CTI* and *TWI* but a fair agreement between observed gullies and prediction obtained using *SPI* (κ: 0.40 – 0.55). This is in accordance with what ~~we~~ observed in W1 but not in W2, where *CTI* achieved a higher κ value than *SPI* and *TWI*. A similar result is reported by Sekaluvu et al. (2015) and Sheshukov et al. (2018), who observed a better accuracy of *CTI*, which achieved a κ value of 0.29 and 0.32 in two watersheds of central Kansas. However, it is worth noting that the range of κ obtained in our experiment for *SPI*, *CTI* and *TWI* is quite higher (0.63 – 0.77) than the values calculated in Kansas. This could be explained by considering that the trajectory of our gullies was adjusted to fit lines of flow concentration extracted from the DEM. This procedure indeed prevents gullies to intersect cells with very low or null drainage area, which can be caused by mapping errors or inadequate DEM resolution, and thus may yield a stronger positive relationship between gully occurrence and contributing area. Furthermore, the higher values of κ achieved by topographic indices in predicting our gullies can be also explained by considering that validation in this experiment was performed at the pixel scale while a sub-watershed scale was employed in the studies performed in Kansas (Daggupati et al., 2013; Sekaluvu et al., 2015; Sheshukov et al., 2018).

To explain the better accuracy of *MSPI* with respect to the other indices, ~~we hypothesize~~it can be hypothesized that adding *CI* to the *SPI* formula helps in detecting areas of enhanced flow concentration and, thus, in identifying cells which are likely to host a gully. Moreover, since *MSPI*

Formatted: Font color: Text 1, Italian (Italy)
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1, Italian (Italy)
Formatted: Font color: Text 1
Formatted: Font color: Text 1, Italian (Italy)
Formatted: Font color: Text 1
Formatted: Font color: Text 1, Italian (Italy)
Formatted: Font color: Text 1
Formatted: Font color: Text 1, Italian (Italy)
Formatted: Font color: Text 1
Formatted: Font color: Text 1

performs clearly better than *CTI* in both investigated watersheds, ~~we infer~~it can be inferred that the contribution of *CI* in increasing the ability to discriminate between non-gully and gully cells is higher than that provided by *PLANC*. This hypothesis is corroborated by the frequency distributions of *CI* and *PLANC* measured on gully and non-gully cells, which are revealed by the kernel density plots of Fig. ~~9~~10. These plots show that *CI* distributions measured along gully trajectories are clearly different from those calculated for non-event cells, whereas no such difference can be observed for *PLANC*. Furthermore, *PLANC* does not improve appreciably the predictive ability of *CTI* with respect to *SPI*; indeed, *SPI* achieves higher *AUC* values in both studied watersheds and higher κ value in W1. On the other hand, *CI* did not improve the predictive skill of *TWI*, as *MTWI* performed better than *TWI* only in W2.

As regards statistical modelling of gully occurrence, validation performed in our study area revealed a better predictive skill of MARS with respect to LR. This results is in line with other studies, like that of Garosi et al. (2018), which also found a better performance of MARS (*AUC*: 74.5–90.2) with respect to LR (*AUC*: 66.4–85.6) in predicting gully erosion susceptibility in Iran. MARS provided slightly better accuracy also in another Sicilian watershed (Gómez-Gutiérrez et al., 2015), where LR has been previously employed to predict the same gully inventory (Conoscenti et al., 2014). Also Rahmati et al. (2019) observed better accuracy of MARS in predicting the same gully inventory of this study, although performing validation on pixels selected from the entire watersheds and employing a quite larger number of predictors, which include land use and bedrock. The better performance of MARS was somewhat expected given the widely accepted assumption that gullying is a threshold phenomenon and the ability of MARS to model non-linear relationships between event occurrence and predictor variables. Indeed, MARS is able to identify, across the range of the predictors, different linear functions separated by knots which may correspond to potential thresholds for gully initiation.

AUC and κ values revealed that, in our study area, statistical models predict the occurrence of gullies with better accuracy than topographic indices, with the exception of *MSPI*. The latter

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1, Italian (Italy)

Formatted: Font color: Text 1

Formatted: Font color: Text 1, Italian (Italy)

Formatted: Font color: Text 1

Formatted: Font color: Text 1, Italian (Italy)

Formatted: Font color: Text 1

exhibited indeed similar or better predictive performance than local LR models and transferred LR and MARS models, whereas only local MARS2 model runs achieved better accuracy. Due to their data-driven nature, a better fit of MARS and LR to the observed gully data was expected prior to performing the experiment. Coefficients of local MARS and LR equations were indeed calculated on the basis of the observed spatial distribution of gullies within the training areas. Also transferred models, although calibrated in one watershed and validated in the other one, were expected to achieve better accuracy than topographic indices, due to the closeness of the two areas and their similar environmental conditions. Therefore, the difference in performance observed between *MSPI* and the transferred statistical models suggests that where an inventory of gullies is not available, reliable maps of gully erosion susceptibility can be prepared by using *MSPI*. This holds in particular if only topographic data is available at high resolution. Indeed, it is worth considering that predictive ability of multivariate statistical models can be improved by including variables reflecting, at high resolution, land use, soil and bedrock characteristics.

The gully erosion prediction maps derived from both topographic indices and ensemble statistical models exhibit an optimal distribution of the susceptibility levels in relation to gullies location. Indeed, at least 89% of observed non-gully cells fall within the lowest susceptibility level whereas between 53% (*CTI* map in W2) and 71% (*SPI* map in W1) of gully cells intersect the highest class of gully occurrence probability. ~~We infer that, in~~In addition to the reliability of the employed indices and models, it can be inferred that the large agreement observed between prediction maps and gully spatial distribution is due to the method employed to identify the susceptibility classes, which was based on the Youden's index (*J*).

Formatted: Font color: Text 1

Formatted: Font color: Text 1

6. Concluding remarks

In this experiment, ~~we evaluated~~ the ability of a set of five topographic indices to predict the spatial distribution of the gullies observed in two adjacent watersheds located in Sicily (Italy~~).~~) was evaluated. Two of these indices, named *MSPI* and *MTWI*, as far as we know, have never been employed to this aim; they were obtained by multiplying the stream power index (*SPI*) and the

Formatted: Font color: Text 1

Formatted: Font color: Text 1

topographic wetness index (*TWI*), respectively, by the convergence index (*CI*). The predictive ability of the topographic indices was measured by using both cut-off independent and dependent statistics and compared to the performance of multivariate statistical models, which use as predictors the same topographic variables of the five indices (i.e. contributing area, slope steepness, plan curvature and convergence index).

The validation results revealed that topographic indices and statistical models achieved excellent to outstanding accuracy in predicting the spatial distribution of the gullies observed in our study area. Statistical models performed better than topographic indices with the exception of *MPSI*. Since the proposed index showed the best predictive performance among the topographic indices, ~~we infer~~it can be inferred that the inclusion of *CI* helps in detecting hollow areas where gullies are more likely to occur. Furthermore, *MSPI* exhibited similar or better predictive skill than transferred statistical models (i.e. models calibrated in one watershed and validated in the other one). This suggests that *MPSI* can be a valid alternative to a data driven approach for identifying potential gully locations in areas where a gully inventory is not available, which is necessary to calibrate statistical models.

Acknowledgments

This research was developed in the framework of the project FLUMEN (project number: 318969), funded by the EU (call identifier: FP7-PEOPLE-2012-IRSES), scientific responsible for the University of Palermo: Prof. Dr. C. Conoscenti. C. Conoscenti and E. Rotigliano have commonly shared the whole research phases. This manuscript benefited greatly from the kind suggestions and comments of the three anonymous reviewers.

References

Angileri, S.E., Conoscenti, C., Hochschild, V., Märker, M., Rotigliano, E., Agnesi, V., 2016. Water erosion susceptibility mapping by applying stochastic gradient treeboost to the imera Meridionale River basin (Sicily, Italy). Geomorphology 262, 61–76. https://doi.org/10.1016/j.geomorph.2016.03.018

Arabameri, A., Pradhan, B., Pourghasemi, H.R., Rezaei, K., Kerle, N., 2018. Spatial Modelling of Gully Erosion Using GIS and R Programing: A Comparison among Three Data Mining Algorithms. Appl. Sci. 8, 1369. https://doi.org/10.3390/app8081369

Azareh, A., Rahmati, O., Rafiei-Sardooi, E., Sankey, J.B., Lee, S., Shahabi, H., Ahmad, B. Bin, 2019. Modelling gully-erosion susceptibility in a semi-arid region, Iran: Investigation of applicability of certainty factor and maximum entropy models. Sci. Total Environ. 655, 684–696. https://doi.org/10.1016/j.scitotenv.2018.11.235

Bingner, R.L., Theurer, F.D., 2001. AGNPS 98: A Suite of water quality models for watershed use, in: Proceedings of the Sediment: Monitoring, Modeling, and Managing, 7th Federal Interagency Sedimentation Conference. Reno, NV, 25-29 March 2001.

Brenning, A., 2005. Spatial prediction models for landslide hazards: review, comparison and evaluation. Nat. Hazards Earth Syst. Sci. 5, 853–862. https://doi.org/10.5194/nhess-5-853-2005

Bui, D.T., Tuan, T.A., Klempe, H., Pradhan, B., Revhaug, I., 2015. Spatial prediction models for shallow landslide hazards : a comparative assessment of the efficacy of support vector machines , artificial neural networks , kernel logistic regression , and logistic model tree. Landslides. https://doi.org/10.1007/s10346-015-0557-6

Cama, M., Lombardo, L., Conoscenti, C., Agnesi, V., Rotigliano, E., 2015. Predicting storm-triggered debris flow events: application to the 2009 Ionian Peloritan disaster (Sicily, Italy). Nat. Hazards Earth Syst. Sci. 15, 1785–1806. https://doi.org/10.5194/nhess-15-1785-2015

Cama, M., Conoscenti, C., Lombardo, L., Rotigliano, E., 2016. Exploring relationships between grid cell size and accuracy for debris-flow susceptibility models: a test in the Giampilieri catchment (Sicily, Italy). Environ. Earth Sci. 75. https://doi.org/10.1007/s12665-015-5047-6

Cama, M., Lombardo, L., Conoscenti, C., Rotigliano, E., 2017. Improving transferability strategies for debris flow susceptibility assessment: Application to the Saponara and Itala catchments (Messina, Italy). Geomorphology 288, 52–65. https://doi.org/10.1016/j.geomorph.2017.03.025

Capra, A., Mazzara, L.M., Scicolone, B., 2005. Application of the EGEM model to predict ephemeral gully erosion in Sicily, Italy. Catena 59, 133-146 ST-Application of the EGEM model to pre. https://doi.org/10.1016/j.catena.2004.07.001

Capra, A., Scicolone, B., 2002. Ephemeral gully erosion in a wheat-cultivated area in Sicily (Italy). Biosyst. Eng. 83, 119–126. https://doi.org/10.1006/bioe.2002.0092

Caraballo-Arias, N.A., Conoscenti, C., Di Stefano, C., Ferro, V., 2015. A new empirical model for estimating calanchi Erosion in Sicily, Italy. Geomorphology 231, 292–300. https://doi.org/10.1016/j.geomorph.2014.12.017

Caraballo-Arias, N.A., Conoscenti, C., Di Stefano, C., Ferro, V., 2014. Testing GIS-morphometric analysis of some Sicilian badlands. CATENA 113, 370–376. ~~https://doi.org/10.1016/j.catena.2013.08.021~~https://doi.org/10.1016/j.catena.2013.08.021

Choubin B., Rahmati O., Tahmasebipour N., Feizizadeh B., Pourghasemi H.R. (2019) Application of Fuzzy Analytical Network Process Model for Analyzing the Gully Erosion Susceptibility. In: Pourghasemi H., Rossi M. (eds) Natural Hazards GIS-Based Spatial Modeling Using Data Mining Techniques. Advances in Natural and Technological Hazards Research, vol 48. Springer, Cham

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 37–46.

Conoscenti, C., Agnesi, V., Angileri, S., Cappadonia, C., Rotigliano, E., Märker, M., 2013. A GIS-based approach for gully erosion susceptibility modelling: a test in Sicily, Italy. Environ. Earth Sci. 70, 1179–1195. https://doi.org/10.1007/s12665-012-2205-y

Conoscenti, C., Agnesi, V., Cama, M., Caraballo-Arias, N.A., Rotigliano, E., 2018. Assessment of Gully Erosion Susceptibility Using Multivariate Adaptive Regression Splines and Accounting for Terrain Connectivity. L. Degrad. Dev. 29, 724–736. https://doi.org/10.1002/ldr.2772

Conoscenti, C., Angileri, S., Cappadonia, C., Rotigliano, E., Agnesi, V., Märker, M., 2014. Gully erosion susceptibility assessment by means of GIS-based logistic regression: a case of Sicily (Italy). Geomorphology 204, 399–411. https://doi.org/10.1016/j.geomorph.2013.08.021

Conoscenti, C., Rotigliano, E., Cama, M., Caraballo-Arias, N.A., Lombardo, L., Agnesi, V., 2016. Exploring the effect of absence selection on landslide susceptibility models: A case study in Sicily, Italy. Geomorphology 261, 222–235. https://doi.org/10.1016/j.geomorph.2016.03.006

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geosci. Model Dev. 8, 1991–2007. https://doi.org/10.5194/gmd-8-1991-2015

Daggupati, P., Douglas-Mankin, K.R., Sheshukov, A.Y., 2013. Predicting ephemeral gully location and length using topographic index models. Trans. ASABE 56, 1427–1440. https://doi.org/10.13031/trans.56.10087

Daggupati, P., Sheshukov, A.Y., Douglas-Mankin, K.R., 2014. Evaluating ephemeral gullies with a process-based topographic index model. CATENA 113, 177–186. https://doi.org/10.1016/j.catena.2013.10.005

Desmet, P.J.J., Govers, G., 1996. Comparison of routing algorithms for digital elevation models and their implications for predicting ephemeral gullies. Int. J. Geogr. Inf. Syst. 10, 311–331. https://doi.org/10.1080/02693799608902081

Desmet, P.J.J., Poesen, J., Govers, G., Vandaele, K., 1999. Importance of slope gradient and contributing area for optimal prediction of the initiation and trajectory of ephemeral gullies. Catena 37, 377–392. https://doi.org/10.1016/S0341-8162(99)00027-2

Dewitte, O., Daoudi, M., Bosco, C., Van Den Eeckhaut, M., 2015. Predicting the susceptibility to gully initiation in data-poor regions. Geomorphology 228, 101–115. https://doi.org/10.1016/j.geomorph.2014.08.010

Eustace, A.H., Pringle, M.J., Denham, R.J., 2011. A risk map for gully locations in central Queensland, Australia. Eur. J. Soil Sci. 62, 431–441. https://doi.org/10.1111/j.1365-2389.2011.01375.x

Fierotti, G., 1988. Carta dei Suoli della Sicilia. Istituto di Agronomia, Università di Palermo e Regione Sicilia, Assessorato Territorio ed Ambiente, Palermo.

Frattini, P., Crosta, G., Carrara, A., 2010. Techniques for evaluating the performance of landslide susceptibility models. Eng. Geol. 111, 62–72. https://doi.org/10.1016/j.enggeo.2009.12.004

Friedman, J.H., 1991. Multivariate adaptive regression splines. Ann. Stat. 19, 1–141.

Garosi, Y., Sheklabadi, M., Porghasemi, H.R., Besalatpour, A.A., Conoscenti, C., Van Oost, K., 2018. Comparison of differences in resolution and sources of controlling factors for gully erosion susceptibility mapping. Geoderma 330, 65–78. https://doi.org/10.1016/j.geoderma.2018.05.027

Garosi, Y., Sheklabadi, M., Conoscenti, C., Pourghasemi, H.R., Van Oost, K., 2019. Assessing the performance of GIS-based machine learning models with different accuracy measures for determining susceptibility to gully erosion. Sci. Total Environ. 664, 1117–1132. https://doi.org/10.1016/j.scitotenv.2019.02.093

Geissen, V., Kampichler, C., López-de Llergo-Juárez, J.J., Galindo-Acántara, A., 2007. Superficial and subterranean soil erosion in Tabasco, tropical Mexico: Development of a decision tree modeling approach. Geoderma 139, 277–287. https://doi.org/10.1016/j.geoderma.2007.01.002

Gómez-Gutiérrez, Á., Conoscenti, C., Angileri, S.E., Rotigliano, E., Schnabel, S., 2015. Using topographical attributes to evaluate gully erosion proneness (susceptibility) in two mediterranean basins: advantages and limitations. Nat. Hazards 79, 291–314. https://doi.org/10.1007/s11069-015-1703-0

Gómez-Gutiérrez, Á., Schnabel, S., Felicísimo, Á.M., 2009a. Modelling the occurrence of gullies in rangelands of southwest Spain. Earth Surf. Process. Landforms 34, 1894–1902. https://doi.org/10.1002/esp1881

Gómez-Gutiérrez, Á., Schnabel, S., Lavado Contador, F., 2009b. Gully erosion , land use and topographical thresholds during the last 60 years in a small rangeland catchment in SW Spain. L. Degrad. Dev. 20, 535–550. https://doi.org/10.1002/ldr

Gómez-Gutiérrez, Á., Schnabel, S., Lavado Contador, F., 2009c. Using and comparing two nonparametric methods (CART and MARS) to model the potential distribution of gullies. Ecol. Modell. 220, 3630–3637. https://doi.org/10.1016/j.ecolmodel.2009.06.020

Gordon, L.M., Bennett, S.J., Bingner, R.L., Theurer, F.D., Alonso, C. V, 2007. Simulating ephemeral gully erosion in AnnAGNPS. Trans. ASABE.

Heckmann, T., Gegg, K., Gegg, a., Becht, M., 2014. Sample size matters: investigating the effect of sample size on a logistic regression susceptibility model for debris flows. Nat. Hazards Earth Syst. Sci. 14, 259–278. https://doi.org/10.5194/nhess-14-259-2014

Hengl, T., Reuter, H.I., 2008. Geomorphometry: Concepts, Software, Applications. Elsevier, Amsterdam.

Hijmans, R.J., 2017. raster: Geographic Data Analysis and Modeling [WWW Document]. URL https://cran.r-project.org/package=raster

Hosmer, D.W., Lemeshow, S., 2000. Applied logistic regression, Wiley Series in Probability and Statistics, Wiley series in probability and statistics: Texts and references section. Wiley. ~~https://doi.org/10.1198/tech.2002.s650~~https://doi.org/10.1198/tech.2002.s650

Javidan, N., Kavian, A., Pourghasemi, H.R., Conoscenti, C., Jafarian, Z., 2019. Gully Erosion Susceptibility Mapping Using Multivariate Adaptive Regression Splines—Replications and Sample Size Scenarios. Water 11, 2319, 1–21. https://doi.org/10.3390/w11112319

Jebur, M.N., Pradhan, B., Tehrany, M.S., 2014. Optimization of landslide conditioning factors using very high-resolution airborne laser scanning (LiDAR) data at catchment scale. Remote Sens. Environ. 152, 150–165. https://doi.org/10.1016/j.rse.2014.05.013

Kiss, R., 2004. Determination of drainage network in digital elevation models, utilities and limitations. J. Hungarian Geomathematics 2, 16–29.

Knisel, W.G., 1993. GLEAMS: Groundwater loading effects of agricultural management systems. Publication No. 5. Tifton, Ga.: University of Georgia, Coastal Plains Experiment Station.

Knisel, W.G., 1980. CREAMS: A field scale model for chemicals, runoff and erosion from agricultural management systems. US Dep. Agric. Conserv. Res. Rep. 26, 474–485.

Köthe, R., Gehrt, E., Böhner, J., 1996. Automatische Reliefanalyse für geowissenschaftliche Anwendungen—derzeitiger Stand undWeiterentwicklungen des Programms SARA. Arbeitshefte Geol. 1, 31–37.

Kotu, V., Deshpande, B., 2015. Chapter 2 - Data Mining Process, in: Kotu, V., Deshpande, B. (Eds.), Predictive Analytics and Data Mining. Morgan Kaufmann, Boston, pp. 17–36. https://doi.org/https://doi.org/10.1016/B978-0-12-801460-8.00002-1

Landis, J.R., Koch, G.G., 1977. The Measurement of Observer Agreement for Categorical Data. Biometrics 33, 159–174. https://doi.org/10.2307/2529310

Lasko, T.A., Bhagwat, J.G., Zou, K.H., Ohno-Machado, L., 2005. The use of receiver operating characteristic curves in biomedical informatics. J. Biomed. Inform. 38, 404–415. https://doi.org/10.1016/j.jbi.2005.02.008

Lucà, F., Conforti, M., Robustelli, G., 2011. Comparison of GIS-based gullying susceptibility mapping using bivariate and multivariate statistics: Northern Calabria, South Italy. Geomorphology 134, 297–308. https://doi.org/10.1016/j.geomorph.2011.07.006

Mahto, A., 2018. splitstackshape: Stack and Reshape Datasets After Splitting Concatenated Values [WWW Document]. URL https://cran.r-project.org/package=splitstackshape

Merkel, W.H., Woodward, D.E., Clarke, C.D., 1988. Ephemeral gully erosion model (EGEM). Agric. For. Rangel. Hydrol. Am. Soc. Agric. Eng. Publ. 07–88, 315–323.

Meyer, A., Martínez-Casasnovas, J.A., 1999. Prediction of existing gully erosion in vineyard parcels of the NE Spain: a logistic modelling approach. Soil Tillage Res. 50, 319–331. https://doi.org/10.1016/S0167-1987(99)00020-3

Milborrow, S., 2018. earth: Multivariate Adaptive Regression Splines [WWW Document]. URL https://cran.r-project.org/package=earth

Momm, H.G., Bingner, R.L., Wells, R.R., Wilcox, D., 2012. Agnps GIS-based tool for watershed-scale identification and mapping of cropland potential ephemeral gullies. Appl. Eng. Agric. 28, 17–29.

Momm, H.G., Bingner, R.L., Wells, R.R.W., Rigby, J.R., Dabney, S.M., 2013. Effect of Topographic Characteristics on Compound Topographic Index for Identification of Gully Channel Initiation Locations, in: Transactions of the ASABE. pp. 523–537. https://doi.org/10.13031/2013.42673

Monserud, R.A., Leemans, R., 1992. Comparing global vegetation maps with the Kappa statistic. Ecol. Modell. 62, 275–293. https://doi.org/https://doi.org/10.1016/0304-3800(92)90003-W

Montgomery, D.R., Dietrich, W.E., 1992. Channel initiation and the problem of landscape scale. Science (80-.). 255, 826–830. https://doi.org/10.1126/science.255.5046.826

Moore, I.D., Burch, G.J., Mackenzie, D.H., 1988. Topographic effects on the distribution of surface soil water and the location of ephemeral gullies. Trans. ASAE 32 32, 1098–1107.

Moore, I.D., Grayson, R.B., Ladson, A.R., 1991. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrol. Process. 4, 3–30.

Nachtergaele, J., Poesen, J., Vandekerckhove, L., Oostwoud Wijdenes, D., Roxo, M., 2001a. Testing the Ephemeral Gully Erosion Model (EGEM) in Mediterranean environments, in: Stott, D.E., Mohtar, R.H., Steinhardt, G.C. (Eds.), Sustaining the Global Farm – Selected Papers from the 10th International Soil Conservation Organization Meeting, May 24-29, 1999, West Lafayette, IN. pp. 1024–1028.

Nachtergaele, J., Poesen, J., Vandekerckhove, L., Oostwoud Wijdenes, D., Roxo, M., 2001b. Testing the Ephemeral Gully Erosion Model (EGEM) for two Mediterranean environments. Earth Surf. Process. Landforms 26, 17–30. https://doi.org/10.1002/1096-9837(200101)26:1<17::AID-ESP149>3.0.CO;2-7

Naimi, B., 2015. Uncertainty analysis for species distribution models. R Software Package.

Nandi, A., Shakoor, A., 2009. A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses. Eng. Geol. 110, 11–20. https://doi.org/10.1016/j.enggeo.2009.10.001

Nazari Samani, A., Ahmadi, H., Jafari, M., Boggs, G., Ghoddousi, J., Malekian, A., 2009. Geomorphic threshold conditions for gully erosion in Southwestern Iran (Boushehr-Samal watershed). J. Asian Earth Sci. 35, 180–189. https://doi.org/10.1016/j.jseaes.2009.02.004

O'Callaghan, J.F., Mark, D.., 1984. The extraction of drainage networks from digital elevation data. Comput. Vision, Graph. Image Process. 28, 323–344. https://doi.org/http://dx.doi.org/10.1016/S0734-189X(84)80011-0

Oostwoud Wijdenes, D.J., Poesen, J., Vandekerckhove, L., Ghesquiere, M., 2000. Spatial distribution of gully head activity and sediment supply along an ephemeral channel in a Mediterranean environment. Catena 39, 147–167. https://doi.org/10.1016/S0341-8162(99)00092-2

Parker, C., Thorne, C., Bingner, R., Wells, R., Wilcox, D., 2007. Automated Mapping of Potential For Ephemeral Gully Formation in Agricultural Watersheds, in: Proceedings of the 2nd Joint Federal Interagency Conference, June 27-July 1, 2010, Las Vegas, Nevada. Las Vegas, NV, pp. 1–12.

Patton, P.C., Schumm, S.A., 1975. Gully erosion, northwestern Colorado: a threshold phenomenon. Geology 31, 187–199.

Poesen, J., Nachtergaele, J., Verstraeten, G., Valentin, C., 2003. Gully erosion and environmental change: importance and research needs. Catena 50, 91–133. https://doi.org/10.1016/S0341-8162(02)00143-1

Poesen, J., Torri, D., Van Walleghem, T., 2011. Gully Erosion: Procedures to Adopt When Modelling Soil Erosion in Landscapes Affected by Gullying, in: Morgan, P.C., Nearing, M.A. (Eds.), Handbook of Erosion Modelling. John Wiley & Sons, Ltd, Chichester, UK, pp. 360–386. https://doi.org/10.1002/9781444328455.ch19

Pourghasemi, H.R., Yousefi, S., Kornejady, A., Cerdà, A., 2017. Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling. Sci. Total Environ. 609, 764–775. https://doi.org/10.1016/j.scitotenv.2017.07.198

R Core Team, 2017. R: A Language and Environment for Statistical Computing.

Rahmati, O., ~~Kornejady~~Haghizadeh, A., ~~Samadi, M., Deo, R.C., Conoscenti, C., Lombardo, L., Dayal, K., Taghizadeh-Mehrjardi, R.,~~ Pourghasemi, H.R., ~~Kumar, S., Bui, D.T., 2019. PMT: New analytical framework for automated evaluation~~Noormohamadi, F., 2016. Gully erosion susceptibility mapping: the role of ~~geo-environmental modelling approaches. Sci. Total Environ. 664, 296–311.~~GIS-based bivariate statistical models and their comparison. Nat. Hazards 82, 1–28. https://doi.org/10.~~1016/j.scitotenv.2019.02.017~~1007/s11069-016-2239-7

Rahmati, O., Tahmasebipour, N., Haghizadeh, A., Pourghasemi, H.R., Feizizadeh, B., 2017. Evaluating the influence of geo-environmental factors on gully erosion in a semi-arid region of Iran: An integrated framework. Sci. Total Environ. 579, 913–927. https://doi.org/10.1016/j.scitotenv.2016.10.176

Rahmati, O., Tahmasebipour, N., Haghizadeh, A., Pourghasemi, H.R., Feizizadeh, B., 2017. Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion. Geomorphology 298, 118–137. https://doi.org/10.1016/j.geomorph.2017.09.006

Rahmati, O., Kornejady, A., Samadi, M., Deo, R.C., Conoscenti, C., Lombardo, L., Dayal, K., Taghizadeh-Mehrjardi, R., Pourghasemi, H.R., Kumar, S., Bui, D.T., 2019. PMT: New analytical framework for automated evaluation of geo-environmental modelling approaches. Sci. Total Environ. 664, 296–311.

https://doi.org/10.1016/j.scitotenv.2019.02.017
Regione Siciliana, 2010. Modello digitale del terreno (MDT) 2m x 2m Regione Siciliana - ATA 2007-2008 [WWW Document]. URL http://~~www~~map.sitr.regione.sicilia.it/gis/services/DEM/DEM_2x2_2007_2008_GB/MapServer/WMSServer
Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., Müller, M., 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12, 77.
Rotigliano, E., Martinello, C., Hernandéz, M.A., Agnesi, V., Conoscenti, C., 2019. Predicting the landslides triggered by the 2009 96E/Ida tropical storms in the Ilopango caldera area (El Salvador, CA): optimizing MARS-based model building and validation strategies. Environ. Earth Sci. 78, 210. https://doi.org/10.1007/s12665-019-8214-3
Sekaluvu, L., Sheshukov, A.Y., 2016. Topographical thresholds for ephemeral gully identification in watersheds with highly disturbed terrain, in: 2016 American Society of Agricultural and Biological Engineers Annual International Meeting, ASABE 2016. https://doi.org/10.13031/aim.20162461481
Sekaluvu, L., Sheshukov, A.Y., Hutchinson, S.L., 2015. Accuracy of topographic index models at prediction of ephemeral gullies, in: American Society of Agricultural and Biological Engineers Annual International Meeting 2015. pp. 4432–4439.
Selkimäki, M., González-Olabarria, J.R., 2016. Assessing Gully Erosion Occurrence in Forest Lands in Catalonia (Spain). L. Degrad. Dev. https://doi.org/10.1002/ldr.2533
Sheshukov, A.Y., Sekaluvu, L., Hutchinson, S.L., 2018. Accuracy of topographic index models at identifying ephemeral gully trajectories on agricultural fields. Geomorphology 306, 224–234. https://doi.org/10.1016/j.geomorph.2018.01.026
Sidorchuk, A., 1999. Dynamic and static models of gully erosion. Catena 37, 401–414. https://doi.org/10.1016/S0341-8162(99)00029-6
Sing, T., Sander, O., Beerenwinkel, N., Lengauer, T., 2005. ROCR: visualizing classifier performance in R. Bioinformatics 21, 7881.
Sterlacchini, S., Ballabio, C., Blahut, J., Masetti, M., Sorichetta, a., 2011. Spatial agreement of predicted patterns in landslide susceptibility maps. Geomorphology 125, 51–61. https://doi.org/10.1016/j.geomorph.2010.09.004
Svoray, T., Michailov, E., Cohen, A., Rokah, L., Sturm, A., 2012. Predicting gully initiation: comparing data mining techniques, analytical hierarchy processes and the topographic threshold. Earth Surf. Process. Landforms 37, 607–619. https://doi.org/10.1002/esp.2273
Thommeret, N., Bailly, J.S., Puech, C., 2010. Extraction of thalweg networks from DTMs: application to badlands. Hydrol. Earth Syst. Sci. 14, 1527–1536. https://doi.org/10.5194/hess-14-1527-2010
Thorne, C.R., Zevenbergen, L.W., Grissinger, E.H., Murphey, J.B., 1986. Ephemeral gullies as source of sediments, in: Proceedings of the 4th Federal Interagency Sedimentation Conference. March 24–27, 1986, Las Vegas, NV, pp. 3,152-3,161.
Valentin, C., Poesen, J., Li, Y., 2005. Gully erosion: Impacts, factors and control. Catena 63, 132–153. https://doi.org/10.1016/j.catena.2005.06.001
Vandaele, K., Poesen, J., Marques da Silva, J.R., Desmet, P., 1996. Rates and predictability of ephemeral gully erosion in two contrasting environments. Geomorphol. Reli. Proc., Environ. 2, 83–96.
Vandekerckhove, L., Poesen, J., Oostwoud Wijdenes, D., Gyssels, G., 2001. Short-term bank gully retreat rates in Mediterranean environments. Catena 44, 133–161. https://doi.org/10.1016/S0341-8162(00)00152-1
Vanwalleghem, T., Van Den Eeckhaut, M., Poesen, J., Govers, G., Deckers, J., 2008. Spatial analysis of factors controlling the presence of closed depressions and gullies under forest: Application of rare event logistic regression. Geomorphology 95, 504–517. https://doi.org/10.1016/j.geomorph.2007.07.003
Vargas-Cuervo, G., Rotigliano, E., Conoscenti, C., 2019. Prediction of debris-avalanches and -flows triggered by a tropical storm by using a stochastic approach: An application to the events occurred in Mocoa (Colombia) on 1 April 2017. Geomorphology 339, 31–43. https://doi.org/https://doi.org/10.1016/j.geomorph.2019.04.023
Wilson, J.P., Gallant, J.C., 2000. Terrain Analysis: Principles and Applications. Wiley & Sons, Inc., Canada.
Wing, J., Kuhn, M., 2018. caret: Classification and Regression Training [WWW Document]. URL https://cran.r-project.org/package=caret
Youden, W.J., 1950. Index for rating diagnostic tests. Cancer 32–35.
Zucca, C., Canu, A., Della Peruta, R., 2006. Effects of land use and landscape on spatial distribution and morphological features of gullies in an agropastoral area in Sardinia (Italy). Catena 68, 87–95. https://doi.org/10.1016/j.catena.2006.03.015

Formatted: No underline, Font color: Text 1
Formatted: No underline, Font color: Text 1
Formatted: Font color: Text 1
Formatted
Formatted: Font color: Text 1
Formatted
Formatted: Font color: Text 1
Formatted
Formatted: Font color: Text 1
Formatted
Formatted: Font color: Text 1
Formatted
Formatted: Font color: Text 1
Formatted
Formatted: Font color: Text 1
Formatted
Formatted: Font color: Text 1
Formatted
Formatted: Font color: Text 1
Formatted
Formatted: Font color: Text 1
Formatted
Formatted: Font color: Text 1
Formatted
Formatted: Font color: Text 1
Formatted
Formatted: Font color: Text 1
Formatted
Formatted: Font color: Text 1
Formatted
Formatted: Font color: Text 1
Formatted
Formatted: Font color: Text 1
Formatted
Formatted: Font color: Text 1
Formatted
Formatted: Font color: Text 1
Formatted
Formatted: Font color: Text 1
Formatted
Formatted: Font color: Text 1

CAPTIONS

Fig. 1. ~~Location (a) and topographic map (b) of the watersheds W1 and W2~~Flow-chart of methodology.

~~Fig. 2~~Fig. 2. Location (a) and topographic map (b) of the watersheds W1 and W2.

Fig. 3. Elevation (a), slope steepness (b), lithology (c) and land cover (d) maps of the watersheds W1 and W2.

Fig. ~~3~~4. Gully maps of the watersheds W1 and W2 and Google Earth views of two gully-prone sectors of the study area.

Fig. ~~4~~5. An example showing correspondence between gullies and flow pathways.

Fig. ~~5~~6. Box plots showing the variability of the 100 *AUC* values calculated in W1 and W2 for the topographic indices and local and transferred statistical models.

Fig. ~~6~~7. Average ROC curves obtained from the validation of the topographic indices and statistical models in W1 and W2.

Fig. ~~7~~8. Gully erosion susceptibility maps for the sectors of W1 and W2 highlighted in Fig.3. First and third columns show maps calculated from the topographic indices. Second and fourth columns show maps calculated from local and transferred statistical models. White pixels were not investigated because they intersect anthropogenic features (i.e. urban areas, artificial lakes or roads) or fall within a 10 m buffer around river channels.

Fig. ~~8~~9. Relative frequency distributions of non-event and event pixels across the susceptibility levels of the gully erosion susceptibility maps.

Fig. ~~9~~10. Kernel density plots of *CI* and *PLANC* calculated for gully and non-gully cells of the watersheds W1 and W2.

Table 1. Mean and standard deviation of the 100 *AUC* values calculated for the topographic indices and local and transferred statistical models.

Table 2. Cut-off (*T*) dependent statistics calculated in W1 and W2 for the topographic indices and local and transferred statistical models.

Predicting gully occurrence at watershed scale: comparing topographic indices and multivariate statistical models

Christian Conoscenti [a,]*, Edoardo Rotigliano [a]

[a] Department of Earth and Marine Sciences (DISTEM), University of Palermo, Via Archirafi 22, 90123 Palermo, Italy

Abstract

In this study, the ability of five topographic indices to predict the gully trajectories observed in two adjacent watersheds located in Sicily (Italy) was evaluated. Two of these indices, named *MSPI* and *MTWI*, as far as we know, have never been employed to this aim. They were obtained by multiplying the stream power index (*SPI*) and the topographic wetness index (*TWI*), respectively, by the convergence index (*CI*). The predictive ability of the topographic indices was measured by using both cut-off independent (*AUC*: area under the receiver operating characteristic curve) and dependent statistics (Cohen's kappa index κ, sensitivity, specificity). These statistics were calculated also for 100 MARS (multivariate adaptive regression splines) and 100 LR (logistic regression) model runs, which used as predictors the topographic variables (i.e. contributing area, slope steepness, plan curvature and convergence index) combined into the five indices. Performance statistics of both topographic indices and statistical models were calculated using 100 random samples of 2 m grid cells, which were extracted only from flow concentration lines. This was done in order to focus the validation process on where gully erosion is more likely to occur. *MSPI* achieved the best predictive skill (*AUC* > 0.93; κ > 0.71) among the topographic indices and exhibited similar and better accuracy than local (i.e. trained and validated in the same watershed) and transferred (i.e. trained in one watershed and tested in the other one) LR models, respectively. On the other hand, *MSPI* performed similarly to transferred MARS runs (*AUC* > 0.92; κ > 0.71) but slightly worse than local MARS runs (*AUC* > 0.95; κ > 0.77). Based on the results of this experiment, it can be inferred that (i) including *CI* helps in detecting hollow areas where gullies are more likely to occur and (ii) *MPSI* can be a valid alternative to a data driven approach for mapping gully erosion susceptibility in areas where a gully inventory is not available, which is necessary to calibrate statistical models.

Keywords: Gully erosion susceptibility; Topographic indices; Multivariate Adaptive Regression Splines (MARS); Logistic Regression (LR); Geographic Information System (GIS)

* Corresponding author. Tel.: +39 09123864670; fax: +39 0916169908. E-mail address: christian.conoscenti@unipa.it (C. Conoscenti).

1. Introduction

Gully erosion causes land degradation in a wide range of environmental conditions. The development of gullies in agricultural watersheds may induce high soil loss and reduction of water availability, leading to a significant decrease of soil quality and crop yield. Moreover, gully channels hamper the trafficability of the fields causing extra damages and costs to farmers (Poesen et al., 2003, 2011).

Gullying is a threshold phenomenon that is mainly controlled by rainfall, topography, soil, lithology and land use. Gullies occur only after a threshold of runoff erosivity and soil erodibility is exceeded. In addition to rainfall, runoff erosive power depends on topography which regulates discharge, concentration and velocity of overland flow (e.g., Moore et al., 1988; Desmet et al., 1999; Poesen et al., 2003; Valentin et al., 2005; Gómez-Gutiérrez et al., 2009a; Daggupati et al., 2013; Conoscenti et al., 2013). Morphology, density and development of gullies in a given landscape is also significantly controlled by parent material (Oostwoud Wijdenes et al., 2000; Vandekerckhove et al., 2001; Poesen et al., 2011). Furthermore, gully occurrence is controlled by resistance of soil, which is influenced by soil properties such as texture, bulk density, moisture conditions, organic matter content (Poesen et al., 2003). Soil erosion susceptibility is also related to crop type and stage, as well as tillage direction and conservation practices (Parker et al., 2007). Also, several studies have reported triggering of gullies or increasing of gully erosion rates as being caused by land use changes, intensification of farming activities and overgrazing (Poesen et al., 2003; Valentin et al., 2005; Zucca et al., 2006; Gómez-Gutiérrez et al., 2009b).

Planning of gully erosion control in agricultural watersheds requires either quantifying soil loss and predicting gully location. Several process-based models have been developed to quantify gully erosion (e.g., CREAMS, Knisel, 1980; EGEM, Merkel et al., 1988; GLEAMS, Knisel, 1993; Sidorchuk, 1999; REGEM, Gordon et al., 2007). However, these models require physical input variables that are difficult to measure at the watershed scale. Soil loss due to gully erosion can be also evaluated by using empirical models which are based on relationships established between volume and length of the gully channels (e.g., Nachtergaele et al., 2001; Capra and Scicolone, 2002; Capra et al., 2005; Caraballo-Arias et al., 2014, 2015).

Prediction of gully location can be achieved by identifying a topographic threshold that has to be exceeded for a gully to form. A number of studies have proposed topographic threshold lines defined on a log-log plot of local slope gradient (*S*) versus upslope contributing area (*A*) measured at gully heads (e.g., Patton and Schumm, 1975; Montgomery and Dietrich, 1992; Nachtergaele et al., 2001b; Zucca et al., 2006; Nazari Samani et al., 2009). Both these topographic attributes are indeed widely considered to play the role of controlling factors in the gully formation process as they act as proxies for flow velocity and discharge, respectively. The approach based on *S*–*A* threshold lines assumes that for a given *A*, a critical *S* exists above which runoff erosivity is large enough to produce gully erosion. The *S*–*A* threshold can be used to predict gullies by classifying a study area into non-event positions (below the threshold line) and event positions (on or above the threshold line). However, this approach tends to overestimate the likelihood of gully occurrence (Svoray et al., 2012; Gómez-Gutiérrez et al., 2015), providing a high number of false positives (i.e. non-gullied positions classified as gullied).

Furthermore, several topographic indices have been employed to predict gully location (e.g., Thorne et al., 1986; Moore et al., 1988; Vandaele et al., 1996; Desmet et al., 1999). These models rely on the assumption that gully formation depends on a combination of primary topographic attributes (Wilson and Gallant, 2000) which reflect erosivity of concentrated overland flow; gully erosion occurs when the topographic index exceeds a critical threshold value. Daggupati et al. (2013), Sekaluvu et al. (2015) and Sheshukov et al. (2018) have compared the ability to discriminate between gullied and non-gullied areas of several topographic indices, which were applied using different thresholds. Their analyses revealed that gully predictions were not accurate without identifying an optimal threshold through local calibration. Indeed, they have observed that a low threshold causes high number of false positives whereas a high threshold produces high number of false negatives (i.e. gullied sites predicted as non-gullied).

Recently, accurate predictions of gully locations have been achieved by using statistical modeling and data mining techniques such as logistic regression, classification and regression trees, multivariate adaptive regression splines, stochastic gradient treeboost, artificial neural network, random forest, maximum entropy, etc. (e.g., Meyer and Martínez-Casasnovas, 1999; Gómez-Gutiérrez et al., 2009c; Eustace et al., 2011; Svoray et al., 2012; Conoscenti et al., 2014, 2018; Dewitte et al., 2015; Angileri et al., 2016; Pourghasemi et al., 2017; Rahmati et al., 2016, 2017a, 2017b; Garosi et al., 2018, 2019; Azareh et al., 2019; Choubin et al., 2019; Javidan et al., 2019). These techniques are able to analyze and model the relationships between gully locations and spatial variability of a set of environmental predictors related to topography, land use, parent materials and soils. Based on the identified statistical relationships, these techniques allow for calculating a probability of gully occurrence that ranges from 0 to 1, for each position (usually grid cell) in a given area. However, an important drawback in these procedures, which are data-driven, is that they generate prediction images which efficiently explain the gully distribution in the study area but tend to fail when exported to other areas, even if located at a close distance (Conoscenti et al., 2018).

This study focuses on investigating the topographic control of gully erosion caused by concentrated overland flow at watershed scale. The experiment was carried out in two small agricultural watersheds located in Sicily (Italy). The main goal of the study was to evaluate and compare the ability to predict the location of gullies achieved by using a set of topographic indices, which includes three indices previously proposed for predicting gully location and two modified versions of them. Predictive models of gully occurrence were prepared also by using logistic regression (LR; Hosmer and Lemeshow, 2000) and multivariate adaptive regression splines (MARS; Friedman, 1991), two statistical modeling techniques which have been successfully used to this aim in previous studies (e.g., Vanwalleghem et al., 2008; Gómez-Gutiérrez et al., 2009c; Svoray et al., 2012; Conoscenti et al., 2014, 2018; Dewitte et al., 2015). To further assess the ability to predict gully occurrence provided by the five topographic indices, their accuracy was compared with that achieved by LR and MARS models.

2. Materials and Methods

In this study, the topographic analysis was carried out using a LiDAR-derived 2×2 m Digital Elevation Model (DEM; Regione Siciliana, 2010), with vertical accuracy of 0.1–0.2 m. The GIS calculations were performed using SAGA-GIS software (Conrad et al., 2015).

The calibration of MARS and LR and the validation of both topographic indices and statistical models were performed using the R software (R Core Team, 2017) with the packages “raster” (Hijmans, 2017), “usdm” (Naimi, 2015), “splitstackshape” (Mahto, 2018), “pROC” (Robin et al., 2011), “ROCR” (Sing et al., 2005), “caret” (Wing and Kuhn, 2018) and “earth” (Milborrow, 2018). The flow-chart of Fig. 1 shows a schematic overview of the methodology, which is described in detail in the following sections.

2.1. Study area and gully inventory

The experiment was carried out in two adjacent agricultural watersheds located in central-western Sicily (Fig. 2), approximately 35 km south-east of the city of Palermo. The westernmost watershed (W1) drains an area of 621.7 ha whereas the easternmost one (W2) covers 901.4 ha. The study area experiences a typical Mediterranean climate with an average annual rainfall of 711 mm (time interval: 2002–2017; Camporeale rainfall station; Regione Siciliana – SIAS - Servizio Informativo Agrometeorologico Siciliano), with a minimum in July (5.6 mm) and a maximum in December (88.7 mm). Topography of the two investigated watersheds is slightly different (Fig. 3a–b): elevation ranges from 185 to 576 m a.s.l. in W1 (mean = 303 m) and from 209 to 571 m a.s.l. in W2 (mean = 345 m), whereas average slope gradient is 10.1° (SD = 5.0°) and 9.7° (SD = 6.9°), respectively. Soils are mostly regosols and vertisols with fine-medium texture (Fierotti, 1988). Lithologies are mainly eluvial-colluvial deposits, sands of the Late Miocene Terravecchia Fm., clays of the Middle-Late Miocene Castellana Sicula Fm., silty-clays and sandy-silts of the Terravecchia Fm. (Fig. 3c). Primary land covers are arable lands (mainly cereal fields) and vineyards, which occupy 92% of W1 and 80% of W2 (Fig. 3d).

Both watersheds are affected by gully erosion which increases soil loss, causes landscape dissection and hampers the movement of farm machines. Most of the gully channels in the drainage basins are ephemeral and are usually filled in by tillage within few months after their initiation. Conoscenti et al. (2018) created a gully inventory of the watersheds by analyzing a Google Earth image acquired on 3 May 2015 (Fig. 4) and by carrying out field surveys. As their objective was to model gully erosion due to overland flow concentration, the inventory includes only gullies located on concentrated flow pathways. The latter were extracted from the DEM, by calculating for each cell the value of upstream contributing area. To ensure consistency between mapped gullies and contributing area, gully trajectories have been slightly modified in order to exactly match flow pathways and to ensure that contributing area increases along each gully from head to mouth (Fig. 5). The inventory includes 115 gullies (83 in W1, 32 in W2) and reveals that gully erosion is more severe in W1 (gully density = 0.73 km^{-1}) than in W2 (0.18 km^{-1}). Gullies mostly occur on eluvial-colluvial deposits and clays. As regards land cover, arable lands host most of the gully trajectories.

2.2. Topographic indices

In this experiment, the ability to predict gully location of five topographic indices was assessed. These indices, which combine two or more primary topographic attributes (Wilson and Gallant, 2000), including contributing area, slope steepness, plan curvature and convergence index, were calculated for each grid cell of the DEM, by using terrain analysis tools of SAGA-GIS software (Conrad et al., 2015).

Three topographic indices adopted here, namely stream power index (*SPI*), compound topographic index (*CTI*) and topographic wetness index (*TWI*), have been employed in previous studies to predict location of ephemeral gullies in cultivated watersheds (e.g., Vandaele et al., 1996; Parker et al., 2007; Daggupati et al., 2013, 2014; Sekaluvu et al., 2015; Sekaluvu and Sheshukov, 2016; Sheshukov et al., 2018).

The *SPI* (Moore et al., 1988, 1991) is a measure of erosive power of concentrated runoff and is calculated as:

$$SPI = A_s \bullet S \qquad (1)$$

where A_s (m^2 m^{-1}) is the specific contributing area and S (m m^{-1}) is the local slope gradient. A_s and S are employed as surrogates for flow discharge and velocity. A_s was extracted from upslope contributing area (A), which in turn was calculated by applying the single flow direction (also referred to as D8) algorithm (O'Callaghan and Mark, 1984), after filling sinks in the DEM. To obtain A_s, A has to be divided by the contour width within the pixel (Desmet and Govers, 1996). As the contour width can be set to the average of the grid cell width (i.e., 2.0 m) and the grid cell diagonal (i.e., 2.8 m), A_s was calculated dividing A by 2.4.

The *CTI* (Thorne et al., 1986) is defined as:

$$CTI = As \bullet S \bullet PLANC \qquad (2)$$

where *PLANC* (m/100 m) is the curvature of the contour line (Hengl and Reuter, 2008). *PLANC* is a measure of local flow convergence and divergence and thus reflects the degree of concentration of the runoff. *CTI* is employed in the USDA Agricultural Non-Point Source (AGNPS) modelling system (Bingner and Theurer, 2001) to identify potential ephemeral gully locations throughout a watershed (Parker et al., 2007; Momm et al., 2012, 2013).

TWI (Moore et al., 1988; 1991) is a measure of soil saturation and is calculated as:

$$TWI = \ln{(A_s / S)} \qquad (3)$$

As *TWI* reflects zones of saturation in a watershed, it could also be an index of the potential location of ephemeral gullies. Indeed, gully heads often form where soils become very wet and lose their strength (Moore et al., 1988).

In addition to *SPI*, *CTI* and *TWI*, the ability to predict gully locations of other two topographic indices was explored. These indices are modified versions of *SPI* and *CTI* and are calculated as:

$$MSPI = A_s \bullet S \bullet CI \tag{4}$$

$$MTWI = \ln (A_s / S) \bullet CI \tag{5}$$

where *CI* is the convergence index (Köthe et al., 1996; Kiss, 2004; Thommeret et al., 2010). *CI* measures to what extent neighboring cells point to the center cell. *CI* is calculated as the average difference between actual aspect of surrounding cells within a moving square or circular window and the direction to the center cell, minus 90 degrees. The value ranges from -90 degrees (max convergence) by 0 (planar slopes) to 90 degrees (max divergence). *CI* provided by SAGA-GIS is normalized between -100 and 100. Differently from *PLANC*, which depends on local morphology, *CI* describes the general shape of the landscape up to a scale that depends from the size of the moving window. In this experiment, the *CI* value of each cell was calculated by averaging the values obtained varying the search radius of a circular moving window from 1 to 10 cells. As *PLANC* and *CI* calculated by SAGA-GIS have negative values on concavities (e.g. valley bottoms) and positive values on convexities (e.g. ridges), a change in the sign of both parameters was performed before using them to calculate the topographic indices employed to predict gully location.

MSPI and *MTWI* could help in predicting gully occurrence as they estimate runoff erosive power and potential soil saturation, respectively, and incorporate a weighting factor which reflects flow convergence/divergence (i.e. *CI*).

2.3. Statistical modelling

In our experiment, the location of the gullies was also predicted by employing two statistical techniques, namely logistic regression (LR; Hosmer and Lemeshow, 2000) and multivariate adaptive regression splines (MARS; Friedman, 1991).

LR is a generalized linear model with a logistic link function. LR is among the most common statistical technique for prediction of gully occurrence (e.g., Meyer and Martínez-Casasnovas, 1999; Lucà et al., 2011; Conoscenti et al., 2014; Dewitte et al., 2015; Selkimäki and González-Olabarria, 2016). Conversely, MARS has been employed only recently to model gully erosion (Gómez-Gutiérrez et al., 2009a, 2009c, 2015; Arabameri et al., 2018; Garosi et al., 2018; Conoscenti et al., 2018). LR and MARS enable modelling of relationships between continuous and/or categorical independent variables and a dichotomous dependent variable (i.e. event or non-event). Both techniques consist of an additive combination of terms. LR has a linear structure with constant coefficients across the entire range of the predictor variables. Conversely, MARS uses piece-wise linear regressions with breaks at the knots to describe non-linear relationships between event occurrence and predictors. To reduce the complexity of the models, MARS models were prepared with terms made of single predictors; as regards LR models, a bilateral stepwise strategy, which selects only the most significant predictors, was adopted. Please refer to Hosmer and Lemeshow (2000) and Friedman (1991) for further details about LR and MARS, respectively.

LR and MARS models were prepared by using as predictor variables the primary topographic attributes *S*, A_s, *PLANC* and *CI*. Since both the employed statistical techniques require absence of multicollinearity, the degree of correlation among these four variables was evaluated before running the models. To this aim, the variance inflation factor (*VIF*) (Jebur et al., 2014; Heckmann et al., 2014; Bui et al., 2015; Conoscenti et al., 2016; Cama et al., 2017; Rotigliano et al., 2019; Vargas-Cuervo et al., 2019), was employed. The results, which were interpreted according to the "rule of 10", revealed absence of strong correlations among the predictor variables (*VIF* range: 1.0 – 1.1).

Calibration of the statistical models was carried out separately in W1 and W2, where 100 learning samples were prepared by randomly selecting the 25% of the total number of event pixels and the same number of non-event pixels. This percentage was chosen in order to achieve a compromise between the attempt to minimize the effects of spatial autocorrelation and the effort to obtain robust models, by using a sufficiently large number of cases. Since 1928 and 717 gully cells were identified in W1 and W2, respectively, the W1 learning samples include 964 pixels (i.e. 482 non-event + 482 event cells, the latter corresponding to 25% of 1928) whereas 358 pixels (i.e. 179 non-event + 179 event cells, the latter corresponding to 25% of 717) form the W2 samples. The learning samples were employed to perform 100 LR and 100 MARS model runs in each of the watersheds. Hereafter, MARS1 and LR1 are used to indicate model runs calibrated in W1 whereas MARS2 and LR2 indicate model runs calibrated in W2.

2.4. Validation strategy

The ability to predict gully occurrence of topographic indices and statistical models was measured on a network of flow lines which were identified separately in W1 and W2 by using two different thresholds of contributing area. The thresholds were set equal to the minimum A_s of W1 and W2 gully cells, respectively, after discarding values below the 1st percentile which were regarded as outliers. By using this approach, the predictive performance of topographic indices and statistical models was measured where drainage area is sufficient to trigger gully erosion, given the rainfall, soil, bedrock and land use characteristics which caused gullying in our study watersheds.

One hundred validation samples were prepared by randomly selecting pixels from flow lines of both W1 and W2. Like the calibration samples, also the validation samples include the 25% of the gully cells and a same number of non-gully cells. The value of the topographic indices was used directly as a score to predict the distribution of gully cells. As regards statistical modelling, the probability of gully occurrence was calculated from LR and MARS ensemble models (Kotu and Deshpande, 2015), which were prepared by averaging the score of the 100 model runs. This procedure was applied in order to generate a more stable performance of the models and to mitigate the effects of prevalence (i.e. different proportion of event/non-event cells in the study area) (Svoray et al., 2012). The predictive performance of both “local” (i.e. calibrated and validated in the same watershed) and “transferred” (i.e. calibrated in one watershed and validated in the other one) statistical models was measured.

The accuracy of the topographic indices and statistical models was assessed by plotting for each validation sample the receiver operating characteristic (ROC) curve (e.g., Lasko et al., 2005; Brenning, 2005; Frattini et al., 2010; Cama et al., 2015, 2016) and by calculating the area under the ROC curve (*AUC*). ROC curve analysis is a cut-off independent technique for assessing the performance of predictive models, which plots all possible values of sensitivity (i.e. true positive rate, *TPR*) against the corresponding value of 1-specificity (i.e. false positive rate, *FPR*). The ideal predictive model achieves an *AUC* value close to 1, whereas a value close to 0.5 reveals inaccuracy in the model (Nandi and Shakoor, 2009). In this experiment, accuracy of the models was interpreted as acceptable, excellent or outstanding if *AUC* values were higher than 0.7, 0.8 and 0.9, respectively (Hosmer and Lemeshow, 2000). In both W1 and W2, a group of 100 ROC curves and related *AUC* values, was obtained (one for each validation sample) for each topographic index and statistical model. Comparisons between *AUC* groups were performed by using box plots and the Wilcoxon signed-rank test, setting the level of significance at 0.01.

Furthermore, the predictive ability of topographic indices and statistical models was evaluated by using cut-off dependent performance metrics such as Cohen’s kappa index (Cohen, 1960; Landis and Koch, 1977; Monserud and Leemans, 1992; Geissen et al., 2007; Frattini et al., 2010; Sterlacchini et al., 2011), sensitivity (or *TPR*) and specificity (i.e. true negative rate, *TNR*). The Cohen’s kappa index (κ) reflects the degree of agreement between prediction and observation and is calculated as:

$$\kappa = P_{\text{obs}} - P_{\text{exp}} / (1 - P_{\text{exp}}) \qquad (6)$$

where P_{obs} and P_{exp} are the observed and the expected proportion of agreement, respectively. κ values were interpreted according to Monserud and Leemans (1992), which evaluated the agreement between model prediction and observation as: 1.00, perfect; 0.85–0.99, excellent; 0.70–0.85, very good; 0.55–0.70, good; 0.40–0.55, fair; 0.20–0.40, poor; 0.05–0.20, very poor; <0.05, null.

Firstly, to calculate κ, *TPR* and *FPR*, the average ROC curve from each group of 100 validation ROC curves was prepared. Then, the optimal cut-off values of these curves were identified by using the Youden's index (J) (Youden, 1950; Angileri et al., 2016; Cama et al., 2017; Rotigliano et al., 2019), which corresponds to the threshold that maximizes the sum of sensitivity and specificity. Then, by using J as threshold (T) to classify the grid pixels as not susceptible (score < T) or as susceptible (score > T) to gully erosion, the contingency tables were prepared for each topographic index and ensemble statistical model.

2.5. Gully prediction maps

A gully susceptibility map of the study area was obtained from each of the topographic indices and the four ensemble statistical models which were prepared by averaging the score of 100 MARS and LR model runs. Susceptibility to gully erosion was then classified into four levels according to thresholds that were calculated separately in W1 and W2 by using the steps described below, which were repeated for each topographic index and ensemble statistical model. First, J was used to separate the pixels of the 100 validation samples into a low susceptibility dataset (score < J) and a high susceptible dataset (score > J). Then, the average ROC curve and the Youden index were calculated for both the low susceptibility dataset (J_{low}) and the high susceptibility dataset (J_{high}). Finally, the following four levels of susceptibility to gully erosion were identified: i) low (score ≤ J_{low}); ii) moderate (J_{low} < score ≤ J); iii) high (J < score ≤ J_{high}); iv) very high (score > J_{high}).

3. Results

3.1. Predictive performance measured by using a cut-off independent statistic

The ability of the topographic indices and statistical models to discriminate between gully and non-gully cells of the validation samples is graphically represented by the box plots of Fig. 6. Each box plot reveals the variability of a group of 100 *AUC* values by indicating their quartiles, as well as the lowest and the highest data still within 1.5 interquartile range of the lower quartile and of the upper quartile, respectively. Furthermore, descriptive statistics such as mean and standard deviation of each *AUC* group are reported in Table 1.

The *AUC* values reflect excellent (*AUC* > 0.8) to outstanding (*AUC* > 0.9) discrimination ability of indices and models applied to predict gullies occurred in the studied watersheds. However, significant differences of accuracy can be detected.

MSPI performed clearly better than the other indices in both watersheds. In W1, only *SPI* achieved a similar performance but still significantly lower than that obtained from *MSPI*. In W2, *SPI* performed better than *TWI* but not significantly different from *CTI* and *MTWI*. *TWI* performed better than its modified version (i.e. *MTWI*) in W1, whereas the opposite was observed in W2.

As regards statistical models, MARS performed better than LR in both watersheds. Accuracy of MARS and LR is significantly different even in W1, where *AUC* values appear quite similar. A not significant difference was observed only in W1 between local (i.e. trained in W1) LR and transferred (i.e. trained in W2) MARS models (*p*-value = 0.284). In W1, both MARS and LR local models (i.e. MARS1 and LR1) exhibited higher accuracy than transferred models (i.e. MARS2 and LR2). On the other hand, a not significant difference of *AUC* was observed in W2 between local and transferred LR models (*p*-value = 0.5221).

The *AUC* values and the Wilcoxon signed-rank test revealed an overall better predictive performance of the statistical models with respect to the topographic indices, with the exception of *MSPI*. The latter indeed achieved outstanding accuracy in both watersheds. In W1, *MSPI* exhibited the same accuracy of transferred MARS and local LR runs and better predictive ability than transferred LR runs. In W2, *MSPI* achieved higher accuracy than both local and transferred LR runs and the same accuracy of MARS1. Only local MARS models performed significantly better than MSPI.

3.2. Predictive performance measured by using cut-off dependent statistics

Fig. 7 shows the average ROC curves obtained from the validation of the topographic indices and statistical models in W1 and W2. These curves were employed to calculate the optimal cut-off (T) that maximizes the sum of sensitivity and specificity and which graphically corresponds to the maximum distance to the diagonal lines plotted in Fig. 8. The value of T, as well as those of kappa index (κ), *TPR* and *TNR* are reported in Table 2. Kappa values obtained for the five topographic indices vary from 0.625 to 0.795 indicating a good (κ > 0.55) to very good (κ > 0.70) ability to discriminate between event and non-event pixels. As revealed by *AUC* values, the kappa index also demonstrated that *MSPI* achieved the best predictive skill in both watersheds. *SPI* reached a κ value close that of *MSPI* in W1. Conversely, *SPI* accuracy appears similar to that of *TWI* and *MTWI* in W2, where *CTI* achieved the second best κ value. As regards sensitivity and specificity, *MSPI* obtained the highest values in W1 whereas in W2 a slightly higher *TPR* and *TNR* was observed for *MTWI* and *TWI*, respectively.

Kappa index revealed approximately the same difference of performance between MARS and LR models which is highlighted by the *AUC* values. Indeed, MARS achieved higher κ values in both watersheds, with more enhanced difference of accuracy occurring in W2, where LR models are below the threshold indicating very good performance (κ > 0.7). The difference of performance observed in W1 appears related more to a difference in specificity than in sensitivity, which is very similar for MARS and LR models. On the other hand, in W2, MARS runs exhibit higher values of both *TPR* and *TNR*, whereas only transferred models show a similar sensitivity.

Kappa, *TPR* and *TNR* confirm that *MSPI* achieves approximately the same accuracy of MARS runs. Furthermore, these statistics reveal that *MSPI* outperforms both LR local and transferred models which in turn show better discrimination ability when compared to the other topographic indices, with the exception of *SPI*, in W1, and *CTI*, in W2.

3.3. Gully prediction maps

Fig. 8 shows the gully prediction maps for the sectors of W1 and W2 highlighted in Fig. 4, obtained from the topographic indices and the ensemble statistical models. To aid the assessment of the maps, Fig. 9 plots the relative frequency distributions of non-event and event pixels across the susceptibility levels. The gully erosion susceptibility maps show very low probability of gully occurrence in most part of the study area, with the exception of few flow lines where susceptibility level is from moderate to very high. Only maps derived from *MTWI* and LR, especially in W2, show slightly larger sectors with moderate to high probability of gully occurrence. This is confirmed by the bar plots of Fig. 9, which reveal that non-event cells occur with a frequency higher than 5% only over moderate probability levels of *MTWI* maps and of LR maps of W2. On the other hand, although their very low frequency, high and very high susceptibility levels of all the maps host most of the gully pixels. In particular, the maps derived from *SPI*, *MSPI* and MARS1 ensemble model, achieve the highest percentage of gully pixels within the very high level of susceptibility (Fig. 9).

4. Discussion

The results of our experiment showed that the spatial distribution of gullies can be effectively predicted by using either topographic indices or statistical models.

Both cut-off independent and dependent performance metrics revealed that, among the employed topographic indices, the best accuracy in predicting gully occurrence is achieved by *MSPI* whereas *MTWI* exhibited similar or worse performance than *SPI*, *CTI* and *TWI*. The ability of the latter indices to discriminate between gully and non-pixels was evaluated and compared, by identifying optimal thresholds and by calculating the κ index, in three previous studies (Daggupati et al., 2013; Sekaluvu et al., 2015; Sheshukov et al., 2018) performed in Kansas. Daggupati et al. (2013) estimated the thresholds of 30 – 50, 62, and 12, respectively, for *SPI*, *CTI* and *TWI*. Sekaluvu et al. (2015) and Sheshukov et al. (2018) report that the critical thresholds required by *CTI* to best predict the gullies of two watersheds of central Kansas are equal to 79.4 and 25.1. These values are relatively similar to the *CTI* thresholds estimated by Daggupati et al. (2013) and those calculated in our experiment (52.8 and 24.3). As regards *SPI*, the thresholds found in our study (270.9 and 127.0) are of the same order of magnitude of those calculated by Sekaluvu et al. (2015) and Sheshukov et al. (2018) (501.2 and 158.5), but higher than the values reported by Daggupati et al. (2013). Furthermore, the *TWI* critical thresholds estimated in our experiment (9.7 and 9.4) are quite similar to those calculated for the Kansas areas (12.0 – 18.2).

By applying the thresholds cited above, Daggupati et al. (2013) found a poor predictive performance of *CTI* and *TWI* but a fair agreement between observed gullies and prediction obtained using *SPI* (κ: 0.40 – 0.55). This is in accordance with what observed in W1 but not in W2, where *CTI* achieved a higher κ value than *SPI* and *TWI*. A similar result is reported by Sekaluvu et al. (2015) and Sheshukov et al. (2018), who observed a better accuracy of *CTI*, which achieved a κ value of 0.29 and 0.32 in two watersheds of central Kansas. However, it is worth noting that the range of κ obtained in our experiment for *SPI*, *CTI* and *TWI* is quite higher (0.63 – 0.77) than the values calculated in Kansas. This could be explained by considering that the trajectory of our gullies was adjusted to fit lines of flow concentration extracted from the DEM. This procedure indeed prevents gullies to intersect cells with very low or null drainage area, which can be caused by mapping errors or inadequate DEM resolution, and thus may yield a stronger positive relationship between gully occurrence and contributing area. Furthermore, the higher values of κ achieved by topographic indices in predicting our gullies can be also explained by considering that validation in this experiment was performed at the pixel scale while a sub-watershed scale was employed in the studies performed in Kansas (Daggupati et al., 2013; Sekaluvu et al., 2015; Sheshukov et al., 2018).

To explain the better accuracy of *MSPI* with respect to the other indices, it can be hypothesized that adding *CI* to the *SPI* formula helps in detecting areas of enhanced flow concentration and, thus, in identifying cells which are likely to host a gully. Moreover, since *MSPI* performs clearly better than

CTI in both investigated watersheds, it can be inferred that the contribution of *CI* in increasing the ability to discriminate between non-gully and gully cells is higher than that provided by *PLANC*. This hypothesis is corroborated by the frequency distributions of *CI* and *PLANC* measured on gully and non-gully cells, which are revealed by the kernel density plots of Fig. 10. These plots show that *CI* distributions measured along gully trajectories are clearly different from those calculated for non-event cells, whereas no such difference can be observed for *PLANC*. Furthermore, *PLANC* does not improve appreciably the predictive ability of *CTI* with respect to *SPI*; indeed, *SPI* achieves higher *AUC* values in both studied watersheds and higher κ value in W1. On the other hand, *CI* did not improve the predictive skill of *TWI*, as *MTWI* performed better than *TWI* only in W2.

As regards statistical modelling of gully occurrence, validation performed in our study area revealed a better predictive skill of MARS with respect to LR. This results is in line with other studies, like that of Garosi et al. (2018), which also found a better performance of MARS (*AUC*: 74.5–90.2) with respect to LR (*AUC*: 66.4–85.6) in predicting gully erosion susceptibility in Iran. MARS provided slightly better accuracy also in another Sicilian watershed (Gómez-Gutiérrez et al., 2015), where LR has been previously employed to predict the same gully inventory (Conoscenti et al., 2014). Also Rahmati et al. (2019) observed better accuracy of MARS in predicting the same gully inventory of this study, although performing validation on pixels selected from the entire watersheds and employing a quite larger number of predictors, which include land use and bedrock. The better performance of MARS was somewhat expected given the widely accepted assumption that gullying is a threshold phenomenon and the ability of MARS to model non-linear relationships between event occurrence and predictor variables. Indeed, MARS is able to identify, across the range of the predictors, different linear functions separated by knots which may correspond to potential thresholds for gully initiation.

AUC and κ values revealed that, in our study area, statistical models predict the occurrence of gullies with better accuracy than topographic indices, with the exception of *MSPI*. The latter exhibited indeed similar or better predictive performance than local LR models and transferred LR and MARS models, whereas only local MARS2 model runs achieved better accuracy. Due to their data-driven nature, a better fit of MARS and LR to the observed gully data was expected prior to performing the experiment. Coefficients of local MARS and LR equations were indeed calculated on the basis of the observed spatial distribution of gullies within the training areas. Also transferred models, although calibrated in one watershed and validated in the other one, were expected to achieve better accuracy than topographic indices, due to the closeness of the two areas and their similar environmental conditions. Therefore, the difference in performance observed between *MSPI* and the transferred statistical models suggests that where an inventory of gullies is not available, reliable maps of gully erosion susceptibility can be prepared by using *MSPI*. This holds in particular if only topographic data is available at high resolution. Indeed, it is worth considering that predictive ability of multivariate statistical models can be improved by including variables reflecting, at high resolution, land use, soil and bedrock characteristics.

The gully erosion prediction maps derived from both topographic indices and ensemble statistical models exhibit an optimal distribution of the susceptibility levels in relation to gullies location. Indeed, at least 89% of observed non-gully cells fall within the lowest susceptibility level whereas between 53% (*CTI* map in W2) and 71% (*SPI* map in W1) of gully cells intersect the highest class of gully occurrence probability. In addition to the reliability of the employed indices and models, it can be inferred that the large agreement observed between prediction maps and gully spatial distribution is due to the method employed to identify the susceptibility classes, which was based on the Youden's index (*J*).

5. Concluding remarks

In this experiment, the ability of a set of five topographic indices to predict the spatial distribution of the gullies observed in two adjacent watersheds located in Sicily (Italy) was evaluated. Two of these indices, named *MSPI* and *MTWI*, as far as we know, have never been employed to this aim; they were obtained by multiplying the stream power index (*SPI*) and the topographic wetness index (*TWI*), respectively, by the convergence index (*CI*). The predictive ability of the topographic indices was measured by using both cut-off independent and dependent statistics and compared to the performance of multivariate statistical models, which use as predictors the same topographic variables of the five indices (i.e. contributing area, slope steepness, plan curvature and convergence index).

The validation results revealed that topographic indices and statistical models achieved excellent to outstanding accuracy in predicting the spatial distribution of the gullies observed in our study area. Statistical models performed better than topographic indices with the exception of *MPSI*. Since the proposed index showed the best predictive performance among the topographic indices, it can be inferred that the inclusion of *CI* helps in detecting hollow areas where gullies are more likely to occur. Furthermore, *MSPI* exhibited similar or better predictive skill than transferred statistical models (i.e. models calibrated in one watershed and validated in the other one). This suggests that *MPSI* can be a valid alternative to a data driven approach for identifying potential gully locations in areas where a gully inventory is not available, which is necessary to calibrate statistical models.

Acknowledgments

This research was developed in the framework of the project FLUMEN (project number: 318969), funded by the EU (call identifier: FP7-PEOPLE-2012-IRSES), scientific responsible for the University of Palermo: Prof. Dr. C. Conoscenti. C. Conoscenti and E. Rotigliano have commonly shared the whole research phases. This manuscript benefited greatly from the kind suggestions and comments of the three anonymous reviewers.

References

Angileri, S.E., Conoscenti, C., Hochschild, V., Märker, M., Rotigliano, E., Agnesi, V., 2016. Water erosion susceptibility mapping by applying stochastic gradient treeboost to the imera Meridionale River basin (Sicily, Italy). Geomorphology 262, 61–76. https://doi.org/10.1016/j.geomorph.2016.03.018

Arabameri, A., Pradhan, B., Pourghasemi, H.R., Rezaei, K., Kerle, N., 2018. Spatial Modelling of Gully Erosion Using GIS and R Programing: A Comparison among Three Data Mining Algorithms. Appl. Sci. 8, 1369. https://doi.org/10.3390/app8081369

Azareh, A., Rahmati, O., Rafiei-Sardooi, E., Sankey, J.B., Lee, S., Shahabi, H., Ahmad, B. Bin, 2019. Modelling gully-erosion susceptibility in a semi-arid region, Iran: Investigation of applicability of certainty factor and maximum entropy models. Sci. Total Environ. 655, 684–696. https://doi.org/10.1016/j.scitotenv.2018.11.235

Bingner, R.L., Theurer, F.D., 2001. AGNPS 98: A Suite of water quality models for watershed use, in: Proceedings of the Sediment: Monitoring, Modeling, and Managing, 7th Federal Interagency Sedimentation Conference. Reno, NV, 25-29 March 2001.

Brenning, A., 2005. Spatial prediction models for landslide hazards: review, comparison and evaluation. Nat. Hazards Earth Syst. Sci. 5, 853–862. https://doi.org/10.5194/nhess-5-853-2005

Bui, D.T., Tuan, T.A., Klempe, H., Pradhan, B., Revhaug, I., 2015. Spatial prediction models for shallow landslide hazards : a comparative assessment of the efficacy of support vector machines , artificial neural networks , kernel logistic regression , and logistic model tree. Landslides. https://doi.org/10.1007/s10346-015-0557-6

Cama, M., Lombardo, L., Conoscenti, C., Agnesi, V., Rotigliano, E., 2015. Predicting storm-triggered debris flow events: application to the 2009 Ionian Peloritan disaster (Sicily, Italy). Nat. Hazards Earth Syst. Sci. 15, 1785–1806. https://doi.org/10.5194/nhess-15-1785-2015

Cama, M., Conoscenti, C., Lombardo, L., Rotigliano, E., 2016. Exploring relationships between grid cell size and accuracy for debris-flow susceptibility models: a test in the Giampilieri catchment (Sicily, Italy). Environ. Earth Sci. 75. https://doi.org/10.1007/s12665-015-5047-6

Cama, M., Lombardo, L., Conoscenti, C., Rotigliano, E., 2017. Improving transferability strategies for debris flow susceptibility assessment: Application to the Saponara and Itala catchments (Messina, Italy). Geomorphology 288, 52–65. https://doi.org/10.1016/j.geomorph.2017.03.025

Capra, A., Mazzara, L.M., Scicolone, B., 2005. Application of the EGEM model to predict ephemeral gully erosion in Sicily, Italy. Catena 59, 133-146 ST-Application of the EGEM model to pre. https://doi.org/10.1016/j.catena.2004.07.001

Capra, A., Scicolone, B., 2002. Ephemeral gully erosion in a wheat-cultivated area in Sicily (Italy). Biosyst. Eng. 83, 119–126. https://doi.org/10.1006/bioe.2002.0092

Caraballo-Arias, N.A., Conoscenti, C., Di Stefano, C., Ferro, V., 2015. A new empirical model for estimating calanchi Erosion in Sicily, Italy. Geomorphology 231, 292–300. https://doi.org/10.1016/j.geomorph.2014.12.017

Caraballo-Arias, N.A., Conoscenti, C., Di Stefano, C., Ferro, V., 2014. Testing GIS-morphometric analysis of some Sicilian badlands. CATENA 113, 370–376. https://doi.org/10.1016/j.catena.2013.08.021

Choubin B., Rahmati O., Tahmasebipour N., Feizizadeh B., Pourghasemi H.R. (2019) Application of Fuzzy Analytical Network Process Model for Analyzing the Gully Erosion Susceptibility. In: Pourghasemi H., Rossi M. (eds) Natural Hazards GIS-Based Spatial Modeling Using Data Mining Techniques. Advances in Natural and Technological Hazards Research, vol 48. Springer, Cham

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 37–46.

Conoscenti, C., Agnesi, V., Angileri, S., Cappadonia, C., Rotigliano, E., Märker, M., 2013. A GIS-based approach for gully erosion susceptibility modelling: a test in Sicily, Italy. Environ. Earth Sci. 70, 1179–1195. https://doi.org/10.1007/s12665-012-2205-y

Conoscenti, C., Agnesi, V., Cama, M., Caraballo-Arias, N.A., Rotigliano, E., 2018. Assessment of Gully Erosion Susceptibility Using Multivariate Adaptive Regression Splines and Accounting for Terrain Connectivity. L. Degrad. Dev. 29, 724–736. https://doi.org/10.1002/ldr.2772

Conoscenti, C., Angileri, S., Cappadonia, C., Rotigliano, E., Agnesi, V., Märker, M., 2014. Gully erosion susceptibility assessment by means of GIS-based logistic regression: a case of Sicily (Italy). Geomorphology 204, 399–411. https://doi.org/10.1016/j.geomorph.2013.08.021

Conoscenti, C., Rotigliano, E., Cama, M., Caraballo-Arias, N.A., Lombardo, L., Agnesi, V., 2016. Exploring the effect of absence selection on landslide susceptibility models: A case study in Sicily, Italy. Geomorphology 261, 222–235. https://doi.org/10.1016/j.geomorph.2016.03.006

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geosci. Model Dev. 8, 1991–2007. https://doi.org/10.5194/gmd-8-1991-2015

Daggupati, P., Douglas-Mankin, K.R., Sheshukov, A.Y., 2013. Predicting ephemeral gully location and length using topographic index models. Trans. ASABE 56, 1427–1440. https://doi.org/10.13031/trans.56.10087

Daggupati, P., Sheshukov, A.Y., Douglas-Mankin, K.R., 2014. Evaluating ephemeral gullies with a process-based topographic index model. CATENA 113, 177–186. https://doi.org/10.1016/j.catena.2013.10.005

Desmet, P.J.J., Govers, G., 1996. Comparison of routing algorithms for digital elevation models and their implications for predicting ephemeral gullies. Int. J. Geogr. Inf. Syst. 10, 311–331. https://doi.org/10.1080/02693799608902081

Desmet, P.J.J., Poesen, J., Govers, G., Vandaele, K., 1999. Importance of slope gradient and contributing area for optimal prediction of the initiation and trajectory of ephemeral gullies. Catena 37, 377–392. https://doi.org/10.1016/S0341-8162(99)00027-2

Dewitte, O., Daoudi, M., Bosco, C., Van Den Eeckhaut, M., 2015. Predicting the susceptibility to gully initiation in data-poor regions. Geomorphology 228, 101–115. https://doi.org/10.1016/j.geomorph.2014.08.010

Eustace, A.H., Pringle, M.J., Denham, R.J., 2011. A risk map for gully locations in central Queensland, Australia. Eur. J. Soil Sci. 62, 431–441. https://doi.org/10.1111/j.1365-2389.2011.01375.x

Fierotti, G., 1988. Carta dei Suoli della Sicilia. Istituto di Agronomia, Università di Palermo e Regione Sicilia, Assessorato Territorio ed Ambiente, Palermo.

Frattini, P., Crosta, G., Carrara, A., 2010. Techniques for evaluating the performance of landslide susceptibility models. Eng. Geol. 111, 62–72. https://doi.org/10.1016/j.enggeo.2009.12.004

Friedman, J.H., 1991. Multivariate adaptive regression splines. Ann. Stat. 19, 1–141.

Garosi, Y., Sheklabadi, M., Porghasemi, H.R., Besalatpour, A.A., Conoscenti, C., Van Oost, K., 2018. Comparison of differences in resolution and sources of controlling factors for gully erosion susceptibility mapping. Geoderma 330, 65–78. https://doi.org/10.1016/j.geoderma.2018.05.027

Garosi, Y., Sheklabadi, M., Conoscenti, C., Pourghasemi, H.R., Van Oost, K., 2019. Assessing the performance of GIS-based machine learning models with different accuracy measures for determining susceptibility to gully erosion. Sci. Total Environ. 664, 1117–1132. https://doi.org/10.1016/j.scitotenv.2019.02.093

Geissen, V., Kampichler, C., López-de Llergo-Juárez, J.J., Galindo-Acántara, A., 2007. Superficial and subterranean soil erosion in Tabasco, tropical Mexico: Development of a decision tree modeling approach. Geoderma 139, 277–287. https://doi.org/10.1016/j.geoderma.2007.01.002

Gómez-Gutiérrez, Á., Conoscenti, C., Angileri, S.E., Rotigliano, E., Schnabel, S., 2015. Using topographical attributes to evaluate gully erosion proneness (susceptibility) in two mediterranean basins: advantages and limitations. Nat. Hazards 79, 291–314. https://doi.org/10.1007/s11069-015-1703-0

Gómez-Gutiérrez, Á., Schnabel, S., Felicísimo, Á.M., 2009a. Modelling the occurrence of gullies in rangelands of southwest Spain. Earth Surf. Process. Landforms 34, 1894–1902. https://doi.org/10.1002/esp1881

Gómez-Gutiérrez, Á., Schnabel, S., Lavado Contador, F., 2009b. Gully erosion , land use and topographical thresholds during the last 60 years in a small rangeland catchment in SW Spain. L. Degrad. Dev. 20, 535–550. https://doi.org/10.1002/ldr

Gómez-Gutiérrez, Á., Schnabel, S., Lavado Contador, F., 2009c. Using and comparing two nonparametric methods (CART and MARS) to model the potential distribution of gullies. Ecol. Modell. 220, 3630–3637. https://doi.org/10.1016/j.ecolmodel.2009.06.020

Gordon, L.M., Bennett, S.J., Bingner, R.L., Theurer, F.D., Alonso, C. V, 2007. Simulating ephemeral gully erosion in AnnAGNPS. Trans. ASABE.

Heckmann, T., Gegg, K., Gegg, a., Becht, M., 2014. Sample size matters: investigating the effect of sample size on a logistic regression susceptibility model for debris flows. Nat. Hazards Earth Syst. Sci. 14, 259–278. https://doi.org/10.5194/nhess-14-259-2014

Hengl, T., Reuter, H.I., 2008. Geomorphometry: Concepts, Software, Applications. Elsevier, Amsterdam.

Hijmans, R.J., 2017. raster: Geographic Data Analysis and Modeling [WWW Document]. URL https://cran.r-project.org/package=raster

Hosmer, D.W., Lemeshow, S., 2000. Applied logistic regression, Wiley Series in Probability and Statistics, Wiley series in probability and statistics: Texts and references section. Wiley. https://doi.org/10.1198/tech.2002.s650

Javidan, N., Kavian, A., Pourghasemi, H.R., Conoscenti, C., Jafarian, Z., 2019. Gully Erosion Susceptibility Mapping Using Multivariate Adaptive Regression Splines—Replications and Sample Size Scenarios. Water 11, 2319, 1–21. https://doi.org/10.3390/w11112319

Jebur, M.N., Pradhan, B., Tehrany, M.S., 2014. Optimization of landslide conditioning factors using very high-resolution airborne laser scanning (LiDAR) data at catchment scale. Remote Sens. Environ. 152, 150–165. https://doi.org/10.1016/j.rse.2014.05.013

Kiss, R., 2004. Determination of drainage network in digital elevation models, utilities and limitations. J. Hungarian Geomathematics 2, 16–29.

Knisel, W.G., 1993. GLEAMS: Groundwater loading effects of agricultural management systems. Publication No. 5. Tifton, Ga.: University of Georgia, Coastal Plains Experiment Station.

Knisel, W.G., 1980. CREAMS: A field scale model for chemicals, runoff and erosion from agricultural management systems. US Dep. Agric. Conserv. Res. Rep. 26, 474–485.

Köthe, R., Gehrt, E., Böhner, J., 1996. Automatische Reliefanalyse für geowissenschaftliche Anwendungen—derzeitiger Stand undWeiterentwicklungen des Programms SARA. Arbeitshefte Geol. 1, 31–37.

Kotu, V., Deshpande, B., 2015. Chapter 2 - Data Mining Process, in: Kotu, V., Deshpande, B. (Eds.), Predictive Analytics and Data Mining. Morgan Kaufmann, Boston, pp. 17–36. https://doi.org/https://doi.org/10.1016/B978-0-12-801460-8.00002-1

Landis, J.R., Koch, G.G., 1977. The Measurement of Observer Agreement for Categorical Data. Biometrics 33, 159–174. https://doi.org/10.2307/2529310

Lasko, T.A., Bhagwat, J.G., Zou, K.H., Ohno-Machado, L., 2005. The use of receiver operating characteristic curves in biomedical informatics. J. Biomed. Inform. 38, 404–415. https://doi.org/10.1016/j.jbi.2005.02.008

Lucà, F., Conforti, M., Robustelli, G., 2011. Comparison of GIS-based gullying susceptibility mapping using bivariate and multivariate statistics: Northern Calabria, South Italy. Geomorphology 134, 297–308. https://doi.org/10.1016/j.geomorph.2011.07.006

Mahto, A., 2018. splitstackshape: Stack and Reshape Datasets After Splitting Concatenated Values [WWW Document]. URL https://cran.r-project.org/package=splitstackshape

Merkel, W.H., Woodward, D.E., Clarke, C.D., 1988. Ephemeral gully erosion model (EGEM). Agric. For. Rangel. Hydrol. Am. Soc. Agric. Eng. Publ. 07–88, 315–323.

Meyer, A., Martínez-Casasnovas, J.A., 1999. Prediction of existing gully erosion in vineyard parcels of the NE Spain: a logistic modelling approach. Soil Tillage Res. 50, 319–331. https://doi.org/10.1016/S0167-1987(99)00020-3

Milborrow, S., 2018. earth: Multivariate Adaptive Regression Splines [WWW Document]. URL https://cran.r-project.org/package=earth

Momm, H.G., Bingner, R.L., Wells, R.R., Wilcox, D., 2012. Agnps GIS-based tool for watershed-scale identification and mapping of cropland potential ephemeral gullies. Appl. Eng. Agric. 28, 17–29.

Momm, H.G., Bingner, R.L., Wells, R.R.W., Rigby, J.R., Dabney, S.M., 2013. Effect of Topographic Characteristics on Compound Topographic Index for Identification of Gully Channel Initiation Locations, in: Transactions of the ASABE. pp. 523–537. https://doi.org/10.13031/2013.42673

Monserud, R.A., Leemans, R., 1992. Comparing global vegetation maps with the Kappa statistic. Ecol. Modell. 62, 275–293. https://doi.org/https://doi.org/10.1016/0304-3800(92)90003-W

Montgomery, D.R., Dietrich, W.E., 1992. Channel initiation and the problem of landscape scale. Science (80-.). 255, 826–830. https://doi.org/10.1126/science.255.5046.826

Moore, I.D., Burch, G.J., Mackenzie, D.H., 1988. Topographic effects on the distribution of surface soil water and the location of ephemeral gullies. Trans. ASAE 32 32, 1098–1107.

Moore, I.D., Grayson, R.B., Ladson, A.R., 1991. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrol. Process. 4, 3–30.

Nachtergaele, J., Poesen, J., Vandekerckhove, L., Oostwoud Wijdenes, D., Roxo, M., 2001a. Testing the Ephemeral Gully Erosion Model (EGEM) in Mediterranean environments, in: Stott, D.E., Mohtar, R.H., Steinhardt, G.C. (Eds.), Sustaining the Global Farm – Selected Papers from the 10th International Soil Conservation Organization Meeting, May 24-29, 1999, West Lafayette, IN. pp. 1024–1028.

Nachtergaele, J., Poesen, J., Vandekerckhove, L., Oostwoud Wijdenes, D., Roxo, M., 2001b. Testing the Ephemeral Gully Erosion Model (EGEM) for two Mediterranean environments. Earth Surf. Process. Landforms 26, 17–30. https://doi.org/10.1002/1096-9837(200101)26:1<17::AID-ESP149>3.0.CO;2-7

Naimi, B., 2015. Uncertainty analysis for species distribution models. R Software Package.

Nandi, A., Shakoor, A., 2009. A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses. Eng. Geol. 110, 11–20. https://doi.org/10.1016/j.enggeo.2009.10.001

Nazari Samani, A., Ahmadi, H., Jafari, M., Boggs, G., Ghoddousi, J., Malekian, A., 2009. Geomorphic threshold conditions for gully erosion in Southwestern Iran (Boushehr-Samal watershed). J. Asian Earth Sci. 35, 180–189. https://doi.org/10.1016/j.jseaes.2009.02.004

O'Callaghan, J.F., Mark, D.., 1984. The extraction of drainage networks from digital elevation data. Comput. Vision, Graph. Image Process. 28, 323–344. https://doi.org/http://dx.doi.org/10.1016/S0734-189X(84)80011-0

Oostwoud Wijdenes, D.J., Poesen, J., Vandekerckhove, L., Ghesquiere, M., 2000. Spatial distribution of gully head activity and sediment supply along an ephemeral channel in a Mediterranean environment. Catena 39, 147–167. https://doi.org/10.1016/S0341-8162(99)00092-2

Parker, C., Thorne, C., Bingner, R., Wells, R., Wilcox, D., 2007. Automated Mapping of Potential For Ephemeral Gully Formation in Agricultural Watersheds, in: Proceedings of the 2nd Joint Federal Interagency Conference, June 27-July 1, 2010, Las Vegas, Nevada. Las Vegas, NV, pp. 1–12.

Patton, P.C., Schumm, S.A., 1975. Gully erosion, northwestern Colorado: a threshold phenomenon. Geology 31, 187–199.

Poesen, J., Nachtergaele, J., Verstraeten, G., Valentin, C., 2003. Gully erosion and environmental change: importance and research needs. Catena 50, 91–133. https://doi.org/10.1016/S0341-8162(02)00143-1

Poesen, J., Torri, D., Van Walleghem, T., 2011. Gully Erosion: Procedures to Adopt When Modelling Soil Erosion in Landscapes Affected by Gullying, in: Morgan, P.C., Nearing, M.A. (Eds.), Handbook of Erosion Modelling. John Wiley & Sons, Ltd, Chichester, UK, pp. 360–386. https://doi.org/10.1002/9781444328455.ch19

Pourghasemi, H.R., Yousefi, S., Kornejady, A., Cerdà, A., 2017. Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling. Sci. Total Environ. 609, 764–775. https://doi.org/10.1016/j.scitotenv.2017.07.198

R Core Team, 2017. R: A Language and Environment for Statistical Computing.

Rahmati, O., Haghizadeh, A., Pourghasemi, H.R., Noormohamadi, F., 2016. Gully erosion susceptibility mapping: the role of GIS-based bivariate statistical models and their comparison. Nat. Hazards 82, 1–28. https://doi.org/10.1007/s11069-016-2239-7

Rahmati, O., Tahmasebipour, N., Haghizadeh, A., Pourghasemi, H.R., Feizizadeh, B., 2017. Evaluating the influence of geo-environmental factors on gully erosion in a semi-arid region of Iran: An integrated framework. Sci. Total Environ. 579, 913–927. https://doi.org/10.1016/j.scitotenv.2016.10.176

Rahmati, O., Tahmasebipour, N., Haghizadeh, A., Pourghasemi, H.R., Feizizadeh, B., 2017. Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion. Geomorphology 298, 118–137. https://doi.org/10.1016/j.geomorph.2017.09.006

Rahmati, O., Kornejady, A., Samadi, M., Deo, R.C., Conoscenti, C., Lombardo, L., Dayal, K., Taghizadeh-Mehrjardi, R., Pourghasemi, H.R., Kumar, S., Bui, D.T., 2019. PMT: New analytical framework for automated evaluation of geo-environmental modelling approaches. Sci. Total Environ. 664, 296–311. https://doi.org/10.1016/j.scitotenv.2019.02.017

Regione Siciliana, 2010. Modello digitale del terreno (MDT) 2m x 2m Regione Siciliana - ATA 2007-2008 [WWW Document]. URL http://map.sitr.regione.sicilia.it/gis/services/DEM/DEM_2x2_2007_2008_GB/MapServer/WMSServer

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., Müller, M., 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12, 77.

Rotigliano, E., Martinello, C., Hernandéz, M.A., Agnesi, V., Conoscenti, C., 2019. Predicting the landslides triggered

by the 2009 96E/Ida tropical storms in the Ilopango caldera area (El Salvador, CA): optimizing MARS-based model building and validation strategies. Environ. Earth Sci. 78, 210. https://doi.org/10.1007/s12665-019-8214-3
Sekaluvu, L., Sheshukov, A.Y., 2016. Topographical thresholds for ephemeral gully identification in watersheds with highly disturbed terrain, in: 2016 American Society of Agricultural and Biological Engineers Annual International Meeting, ASABE 2016. https://doi.org/10.13031/aim.20162461481
Sekaluvu, L., Sheshukov, A.Y., Hutchinson, S.L., 2015. Accuracy of topographic index models at prediction of ephemeral gullies, in: American Society of Agricultural and Biological Engineers Annual International Meeting 2015. pp. 4432–4439.
Selkimäki, M., González-Olabarria, J.R., 2016. Assessing Gully Erosion Occurrence in Forest Lands in Catalonia (Spain). L. Degrad. Dev. https://doi.org/10.1002/ldr.2533
Sheshukov, A.Y., Sekaluvu, L., Hutchinson, S.L., 2018. Accuracy of topographic index models at identifying ephemeral gully trajectories on agricultural fields. Geomorphology 306, 224–234. https://doi.org/10.1016/j.geomorph.2018.01.026
Sidorchuk, A., 1999. Dynamic and static models of gully erosion. Catena 37, 401–414. https://doi.org/10.1016/S0341-8162(99)00029-6
Sing, T., Sander, O., Beerenwinkel, N., Lengauer, T., 2005. ROCR: visualizing classifier performance in R. Bioinformatics 21, 7881.
Sterlacchini, S., Ballabio, C., Blahut, J., Masetti, M., Sorichetta, a., 2011. Spatial agreement of predicted patterns in landslide susceptibility maps. Geomorphology 125, 51–61. https://doi.org/10.1016/j.geomorph.2010.09.004
Svoray, T., Michailov, E., Cohen, A., Rokah, L., Sturm, A., 2012. Predicting gully initiation: comparing data mining techniques, analytical hierarchy processes and the topographic threshold. Earth Surf. Process. Landforms 37, 607–619. https://doi.org/10.1002/esp.2273
Thommeret, N., Bailly, J.S., Puech, C., 2010. Extraction of thalweg networks from DTMs: application to badlands. Hydrol. Earth Syst. Sci. 14, 1527–1536. https://doi.org/10.5194/hess-14-1527-2010
Thorne, C.R., Zevenbergen, L.W., Grissinger, E.H., Murphey, J.B., 1986. Ephemeral gullies as source of sediments, in: Proceedings of the 4th Federal Interagency Sedimentation Conference. March 24–27, 1986, Las Vegas, NV, pp. 3,152-3,161.
Valentin, C., Poesen, J., Li, Y., 2005. Gully erosion: Impacts, factors and control. Catena 63, 132–153. https://doi.org/10.1016/j.catena.2005.06.001
Vandaele, K., Poesen, J., Marques da Silva, J.R., Desmet, P., 1996. Rates and predictability of ephemeral gully erosion in two contrasting environments. Geomorphol. Reli. Proc., Environ. 2, 83–96.
Vandekerckhove, L., Poesen, J., Oostwoud Wijdenes, D., Gyssels, G., 2001. Short-term bank gully retreat rates in Mediterranean environments. Catena 44, 133–161. https://doi.org/10.1016/S0341-8162(00)00152-1
Vanwalleghem, T., Van Den Eeckhaut, M., Poesen, J., Govers, G., Deckers, J., 2008. Spatial analysis of factors controlling the presence of closed depressions and gullies under forest: Application of rare event logistic regression. Geomorphology 95, 504–517. https://doi.org/10.1016/j.geomorph.2007.07.003
Vargas-Cuervo, G., Rotigliano, E., Conoscenti, C., 2019. Prediction of debris-avalanches and -flows triggered by a tropical storm by using a stochastic approach: An application to the events occurred in Mocoa (Colombia) on 1 April 2017. Geomorphology 339, 31–43. https://doi.org/https://doi.org/10.1016/j.geomorph.2019.04.023
Wilson, J.P., Gallant, J.C., 2000. Terrain Analysis: Principles and Applications. Wiley & Sons, Inc., Canada.
Wing, J., Kuhn, M., 2018. caret: Classification and Regression Training [WWW Document]. URL https://cran.r-project.org/package=caret
Youden, W.J., 1950. Index for rating diagnostic tests. Cancer 32–35.
Zucca, C., Canu, A., Della Peruta, R., 2006. Effects of land use and landscape on spatial distribution and morphological features of gullies in an agropastoral area in Sardinia (Italy). Catena 68, 87–95. https://doi.org/10.1016/j.catena.2006.03.015

CAPTIONS

Fig. 1. Flow-chart of methodology.

Fig. 2. Location (a) and topographic map (b) of the watersheds W1 and W2.

Fig. 3. Elevation (a), slope steepness (b), lithology (c) and land cover (d) maps of the watersheds W1 and W2.

Fig. 4. Gully maps of the watersheds W1 and W2 and Google Earth views of two gully-prone sectors of the study area.

Fig. 5. An example showing correspondence between gullies and flow pathways.

Fig. 6. Box plots showing the variability of the 100 *AUC* values calculated in W1 and W2 for the topographic indices and local and transferred statistical models.

Fig. 7. Average ROC curves obtained from the validation of the topographic indices and statistical models in W1 and W2.

Fig. 8. Gully erosion susceptibility maps for the sectors of W1 and W2 highlighted in Fig.3. First and third columns show maps calculated from the topographic indices. Second and fourth columns show maps calculated from local and transferred statistical models. White pixels were not investigated because they intersect anthropogenic features (i.e. urban areas, artificial lakes or roads) or fall within a 10 m buffer around river channels.

Fig. 9. Relative frequency distributions of non-event and event pixels across the susceptibility levels of the gully erosion susceptibility maps.

Fig. 10. Kernel density plots of *CI* and *PLANC* calculated for gully and non-gully cells of the watersheds W1 and W2.

Table 1. Mean and standard deviation of the 100 *AUC* values calculated for the topographic indices and local and transferred statistical models.

Table 2. Cut-off (*T*) dependent statistics calculated in W1 and W2 for the topographic indices and local and transferred statistical models.

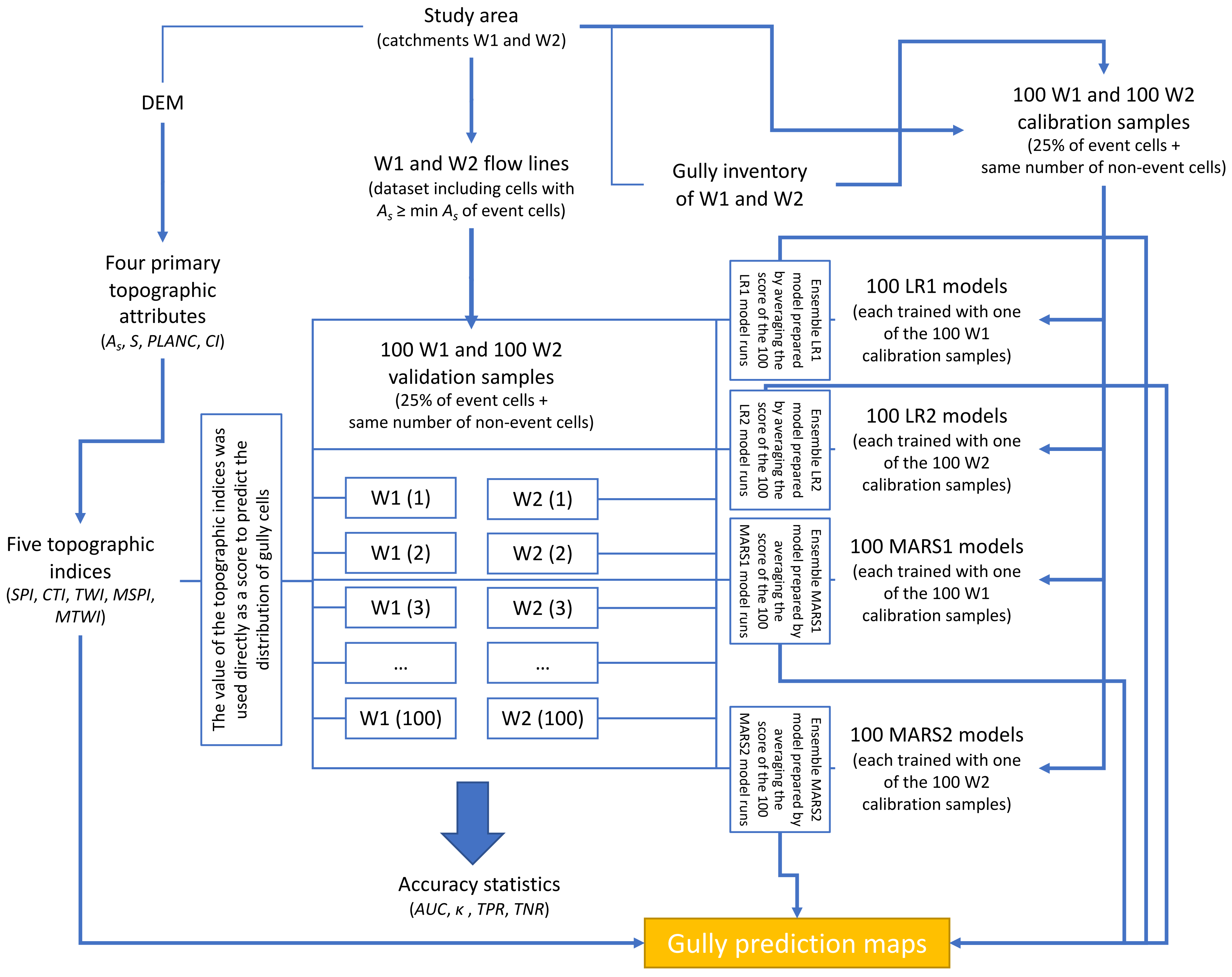
Tables

Table 1. Mean and standard deviation of the 100 *AUC* values calculated for the topographic indices and local and transferred statistical models.

		MARS1	MARS2	LR1	LR2	SPI	CTI	TWI	MSPI	MTWI
W1	Mean	0.961	0.953	0.952	0.943	0.945	0.902	0.926	0.953	0.913
	Std. Dev.	0.005	0.006	0.006	0.007	0.007	0.010	0.008	0.007	0.009
W2	Mean	0.922	0.946	0.911	0.912	0.891	0.888	0.870	0.927	0.891
	Std. Dev.	0.014	0.011	0.014	0.013	0.018	0.016	0.017	0.012	0.015

Table 2. Cut-off (*T*) dependent statistics calculated in W1 and W2 for the topographic indices and local and transferred statistical models.

		MARS1	MARS2	LR1	LR2	SPI	CTI	TWI	MSPI	MTWI
W1	T	0.952	0.950	0.803	0.794	278.6	52.78	9.696	3245.9	147.8
	κ	0.797	0.761	0.769	0.728	0.766	0.715	0.715	0.795	0.682
	TPR	0.897	0.880	0.894	0.879	0.883	0.817	0.846	0.889	0.817
	TNR	0.900	0.881	0.874	0.849	0.883	0.897	0.868	0.906	0.865
W2	T	0.865	0.889	0.614	0.741	148.2	24.02	9.646	1024.5	80.00
	κ	0.714	0.769	0.672	0.659	0.625	0.675	0.627	0.711	0.633
	TPR	0.850	0.913	0.854	0.835	0.831	0.853	0.783	0.902	0.910
	TNR	0.865	0.857	0.819	0.825	0.794	0.822	0.845	0.809	0.724



Study area
(catchments W1 and W2)
DEM
W1 and W2 flow lines
(dataset including cells with $A_s \geq$ min A_s of event cells)
Gully inventory
of W1 and W2
100 W1 and 100 W2
calibration samples
(25% of event cells +
same number of non-event cells)
Four primary
topographic
attributes
(A_s, S, $PLANC$, CI)
Five topographic
indices
(SPI, CTI, TWI, $MSPI$, $MTWI$)
The value of the topographic indices was used directly as a score to predict the distribution of gully cells
100 W1 and 100 W2
validation samples
(25% of event cells +
same number of non-event cells)
W1 (1)
W2 (1)
W1 (2)
W2 (2)
W1 (3)
W2 (3)
...
...
W1 (100)
W2 (100)
Ensemble LR1 model prepared by averaging the score of the 100 LR1 model runs
Ensemble LR2 model prepared by averaging the score of the 100 LR2 model runs
Ensemble MARS1 model prepared by averaging the score of the 100 MARS1 model runs
Ensemble MARS2 model prepared by averaging the score of the 100 MARS2 model runs
100 LR1 models
(each trained with one of the 100 W1 calibration samples)
100 LR2 models
(each trained with one of the 100 W2 calibration samples)
100 MARS1 models
(each trained with one of the 100 W1 calibration samples)
100 MARS2 models
(each trained with one of the 100 W2 calibration samples)
Accuracy statistics
(AUC, κ, TPR, TNR)
Gully prediction maps

Figure 2 (Color)
Click here to download high resolution image

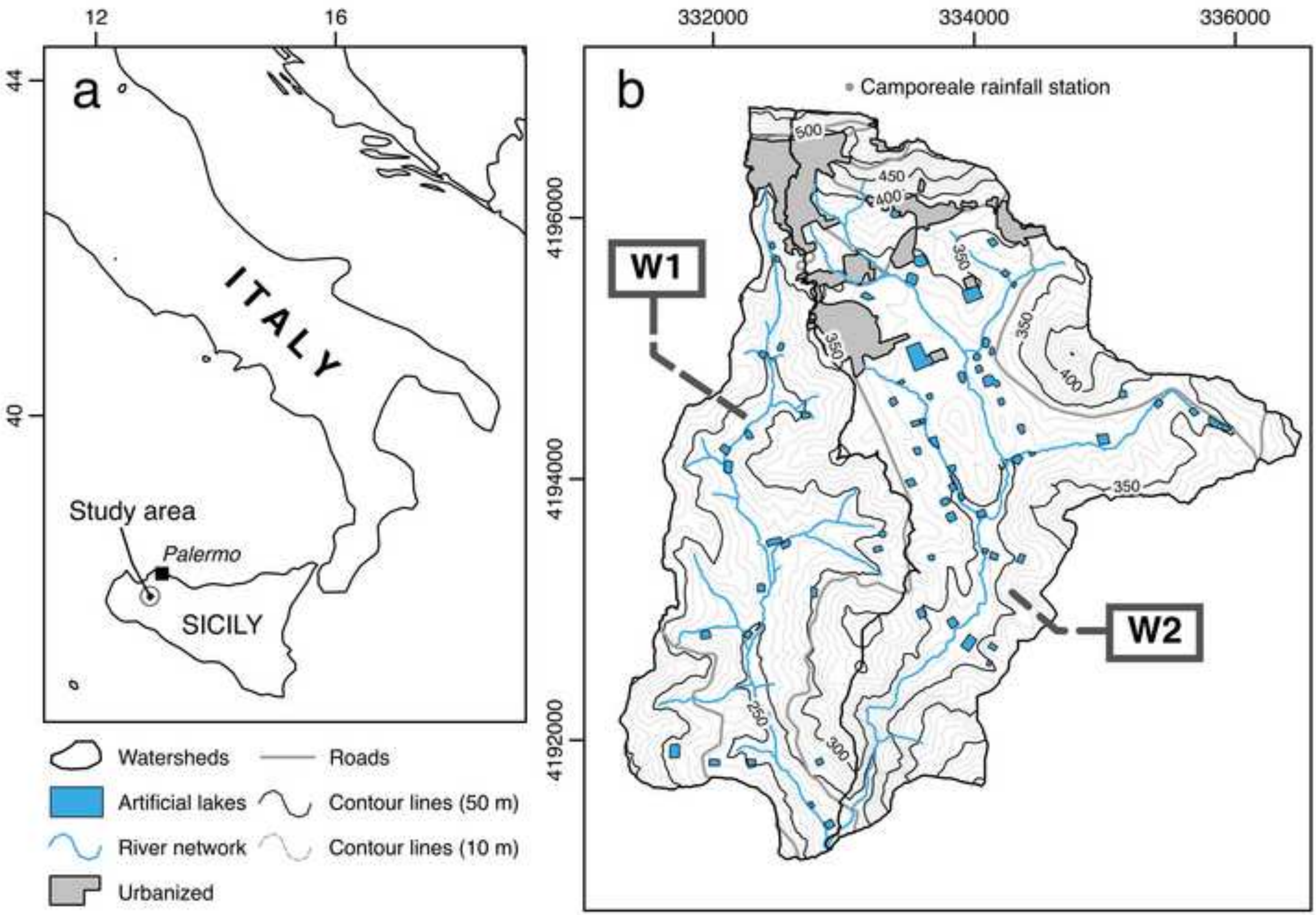


Figure 3 (Color)
Click here to download high resolution image

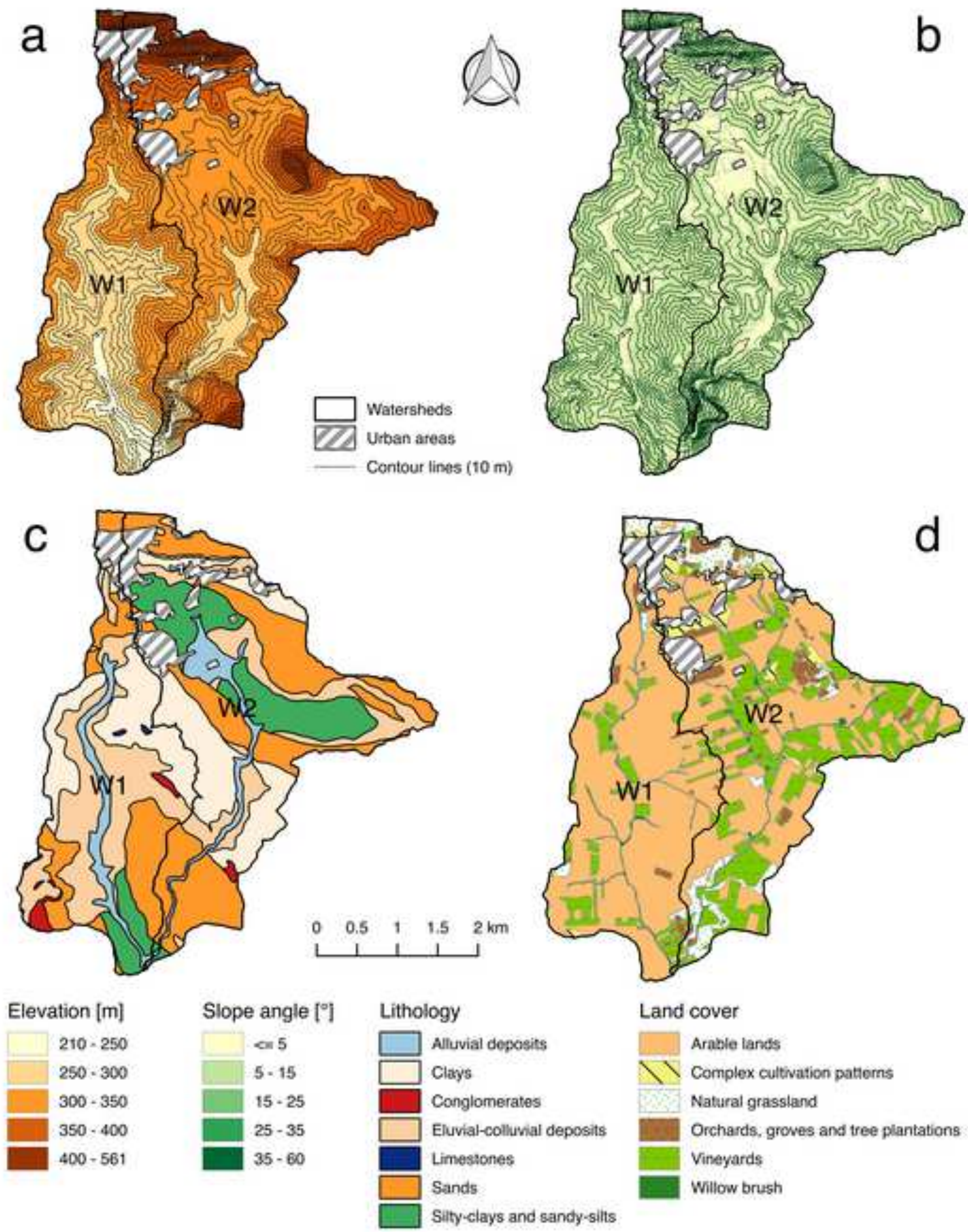


Figure 4 (Color)
Click here to download high resolution image

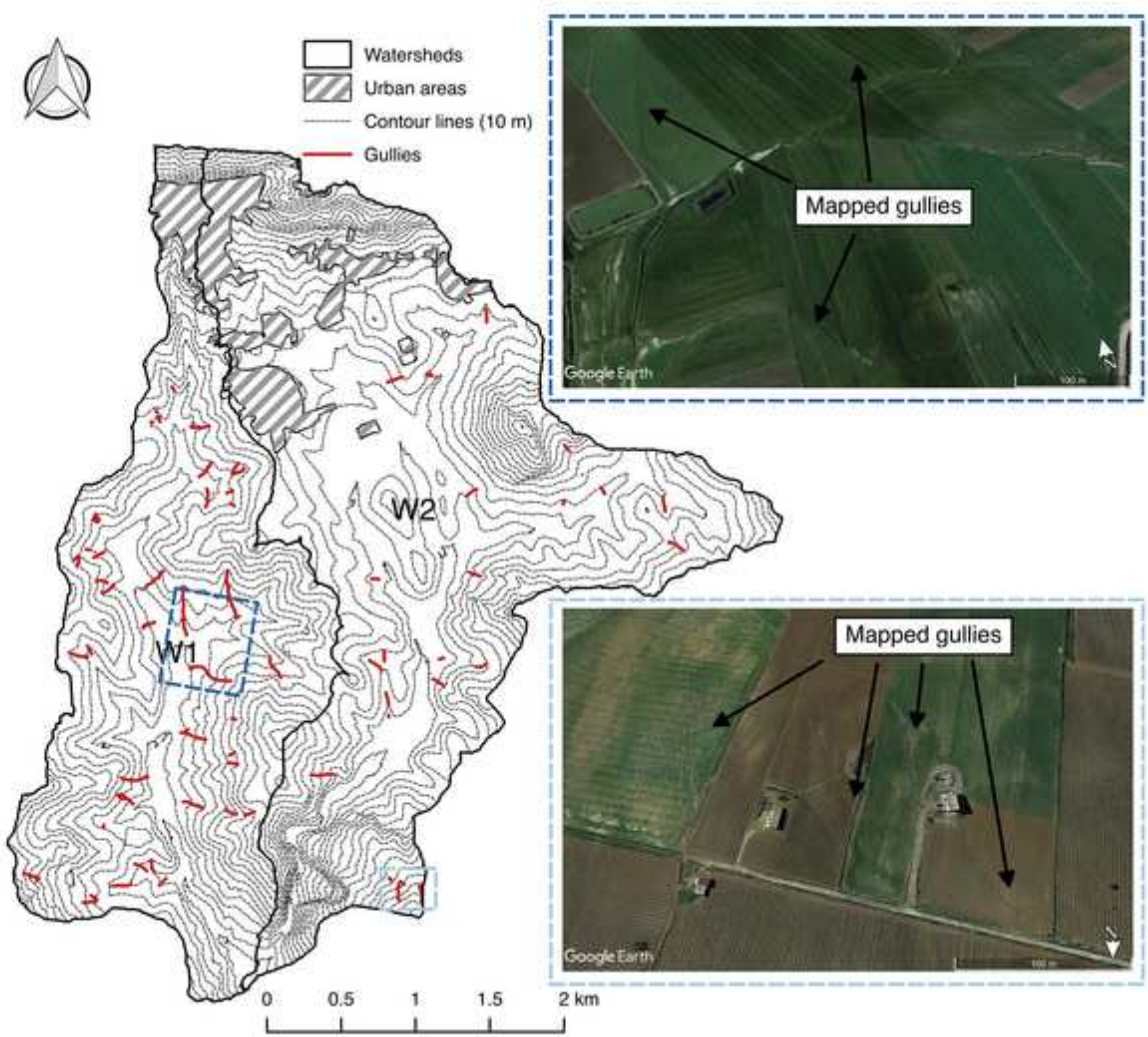


Figure 5 (Color)
Click here to download high resolution image

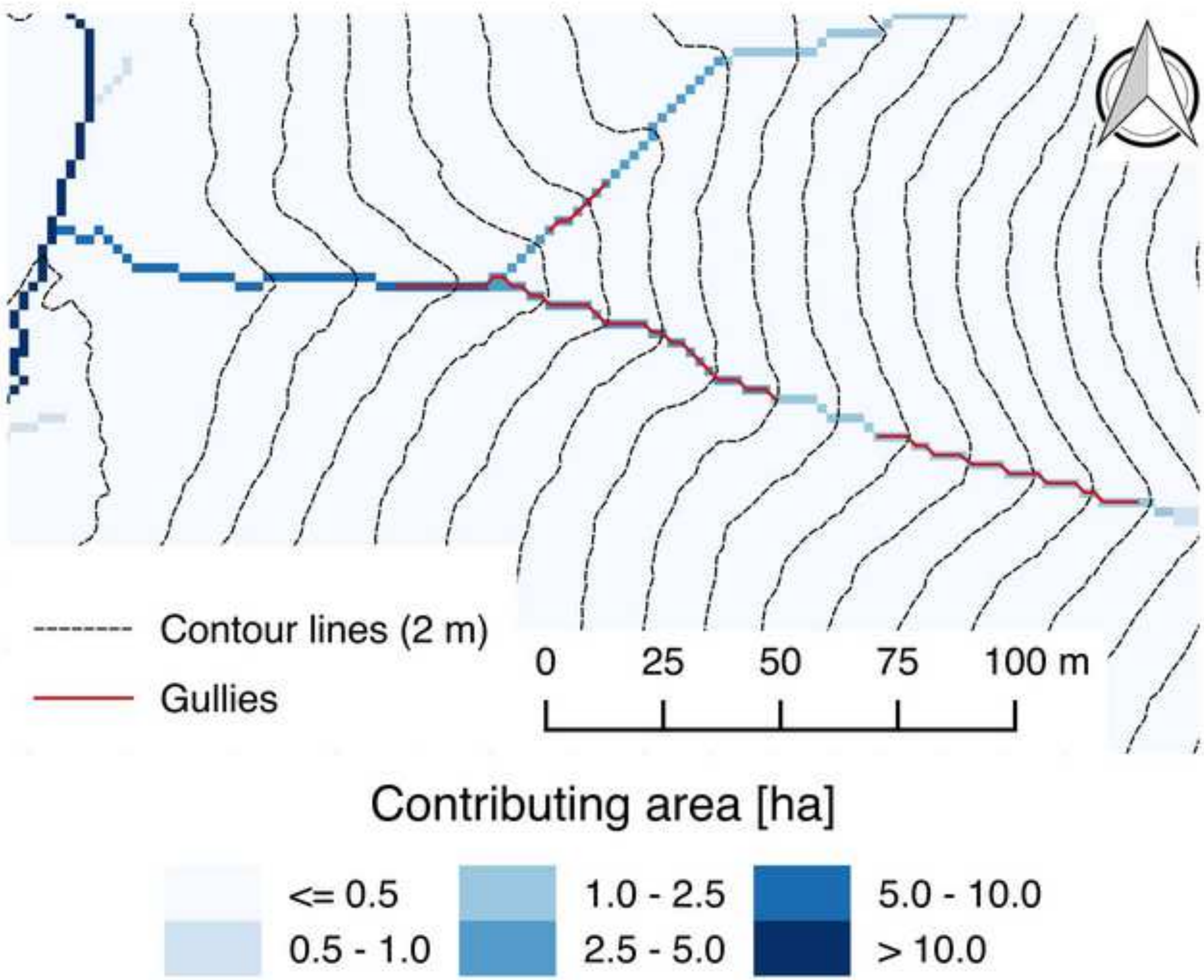


Figure 6 (Color)

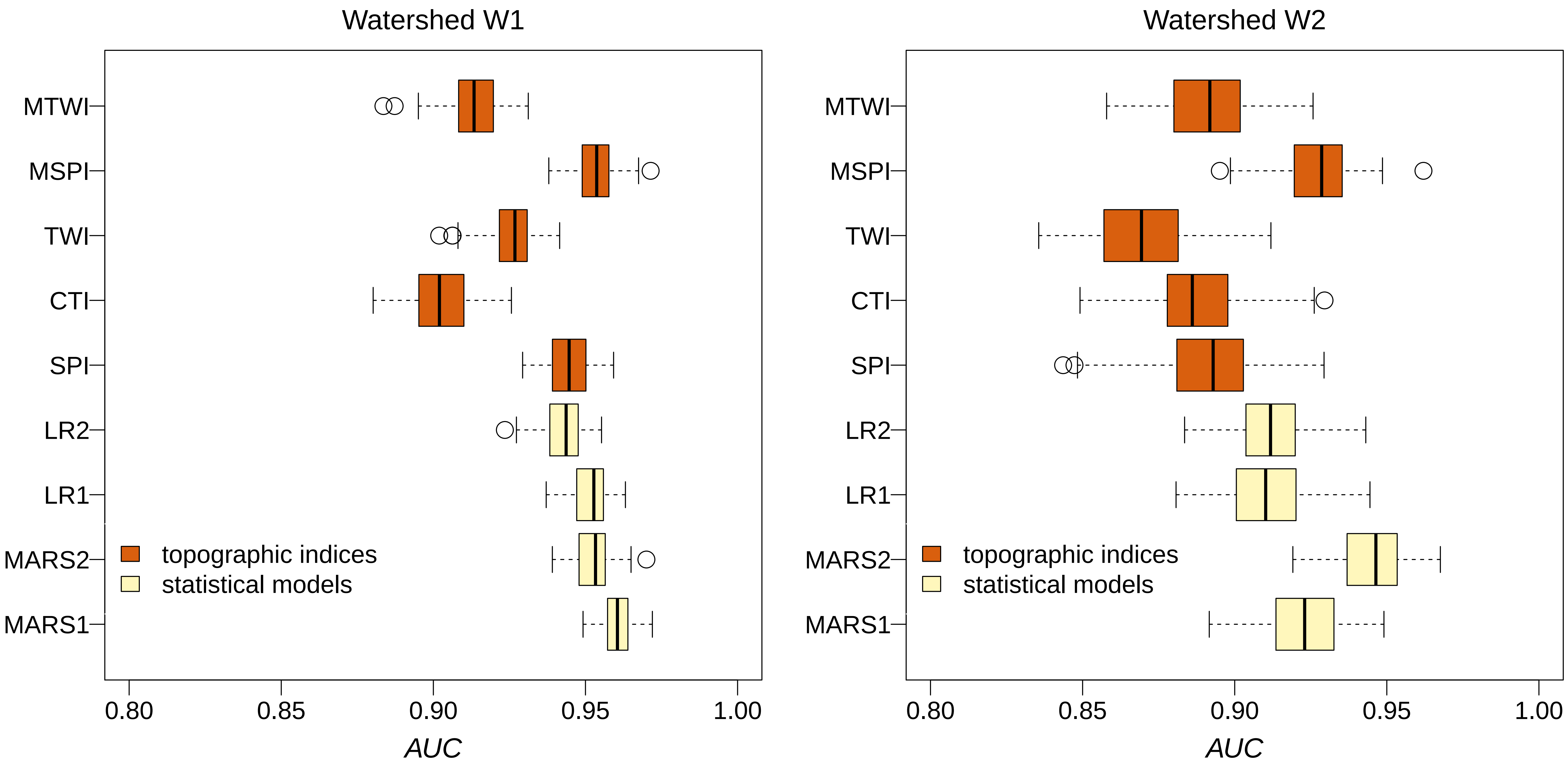


Figure 7 (Color)

Figure 8 (Color)

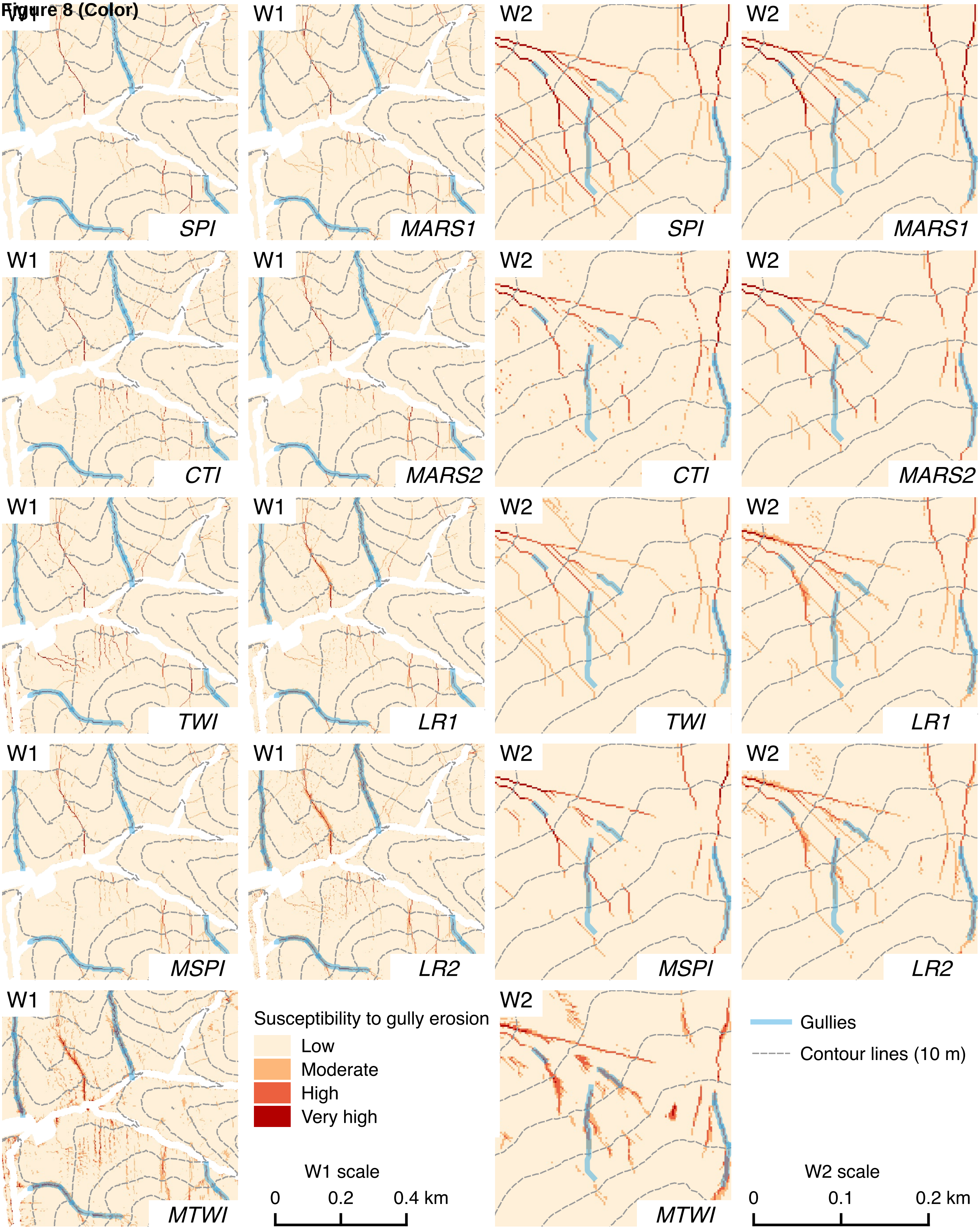
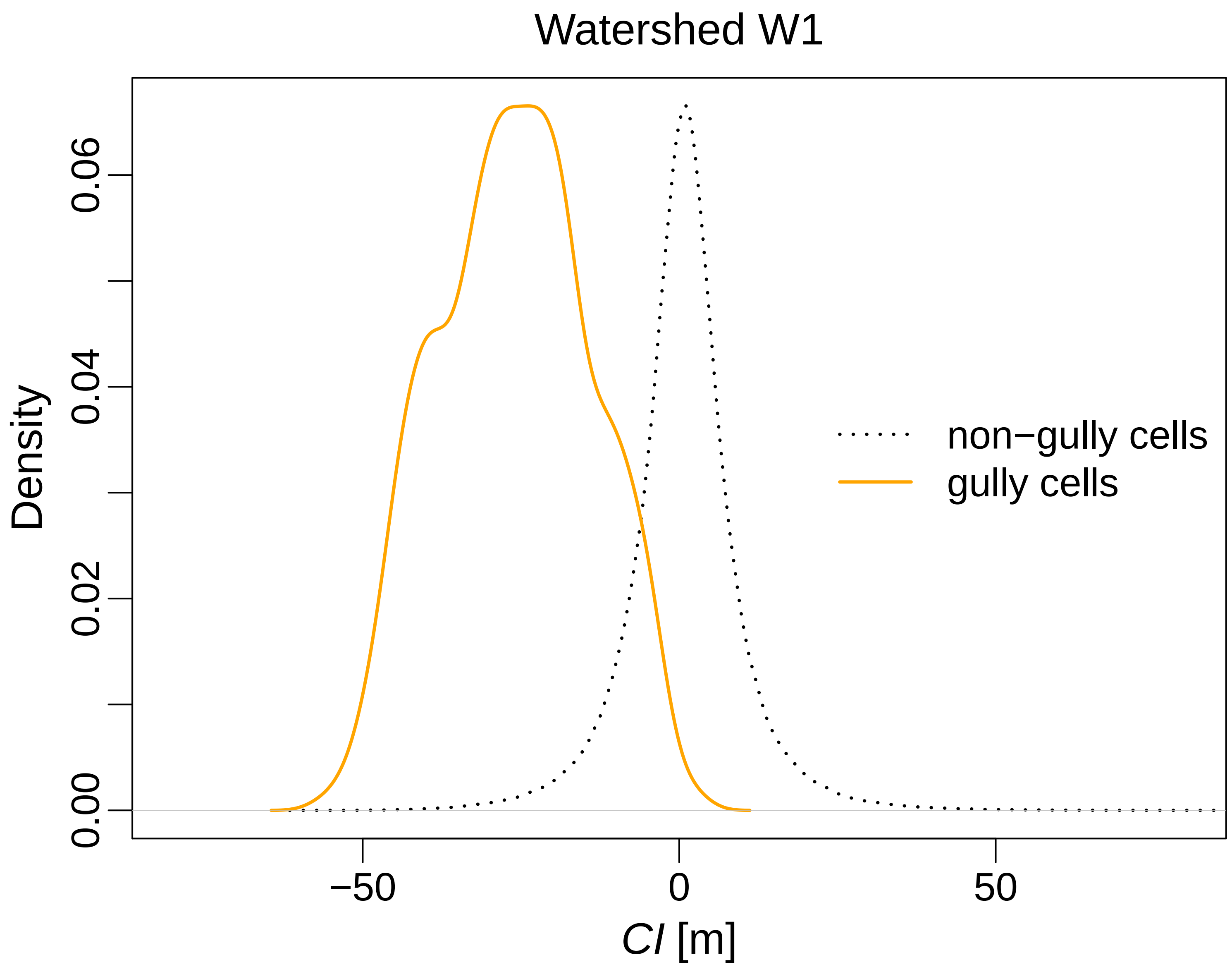
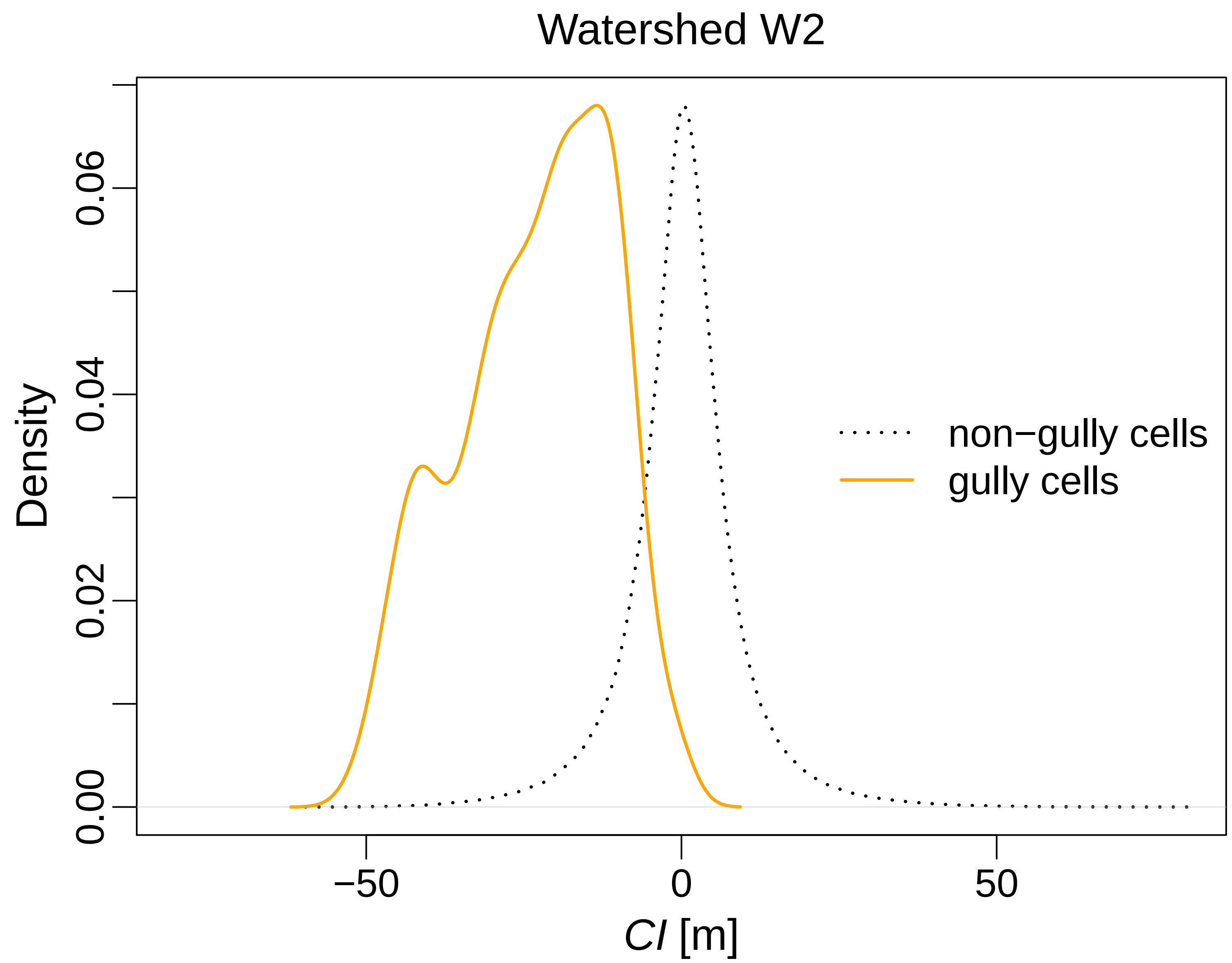
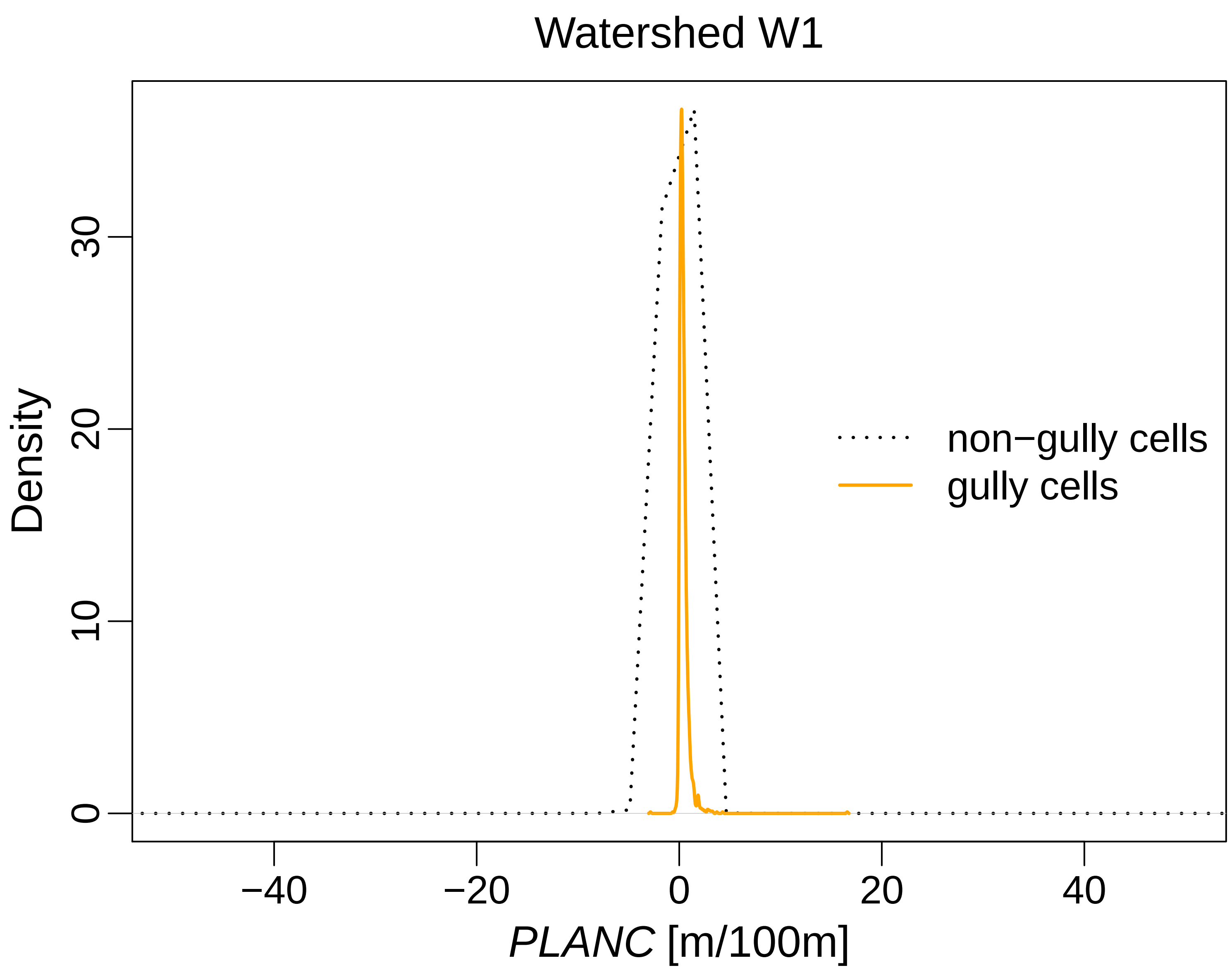


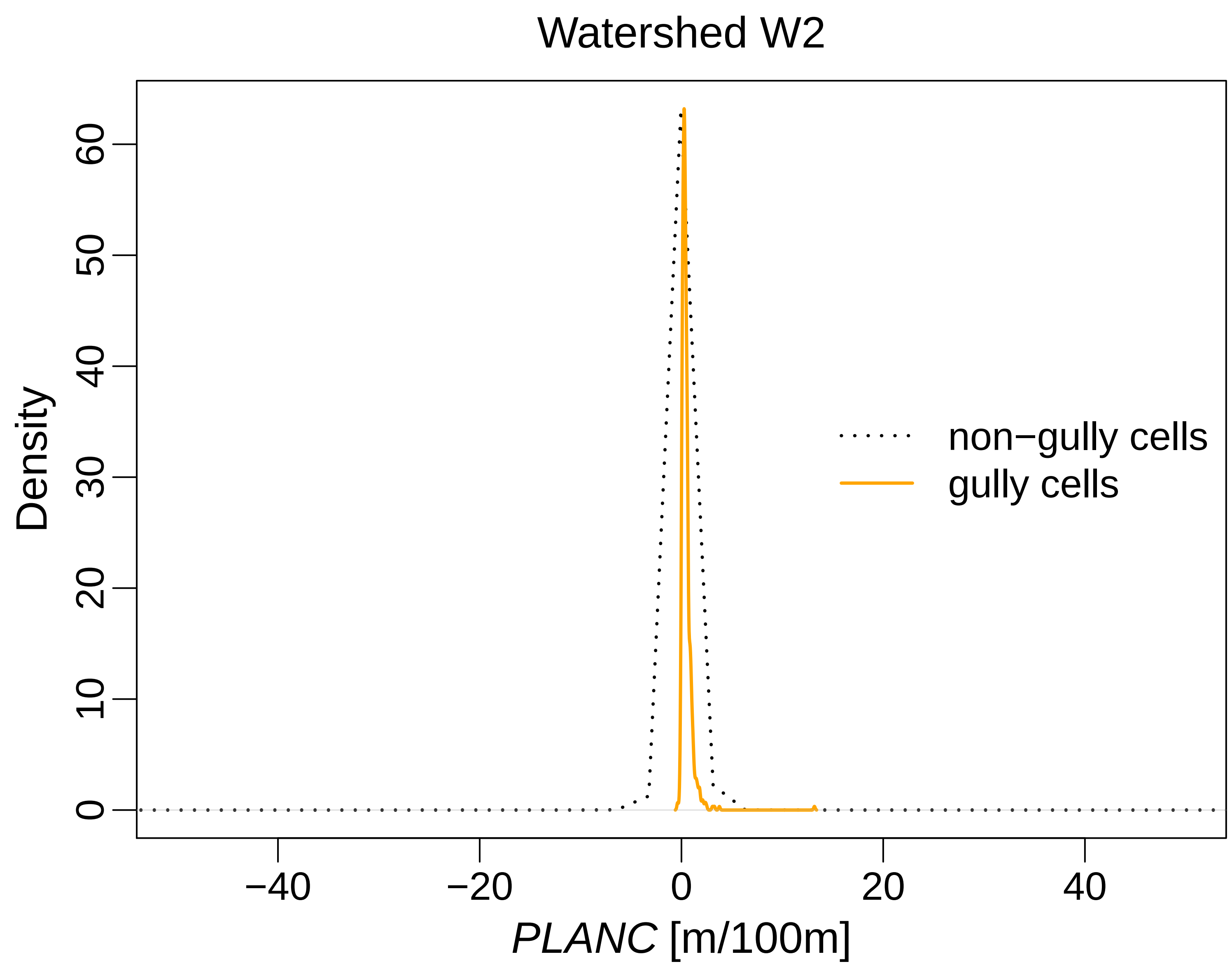
Figure 9 (Color)

Figure 10 (Color)

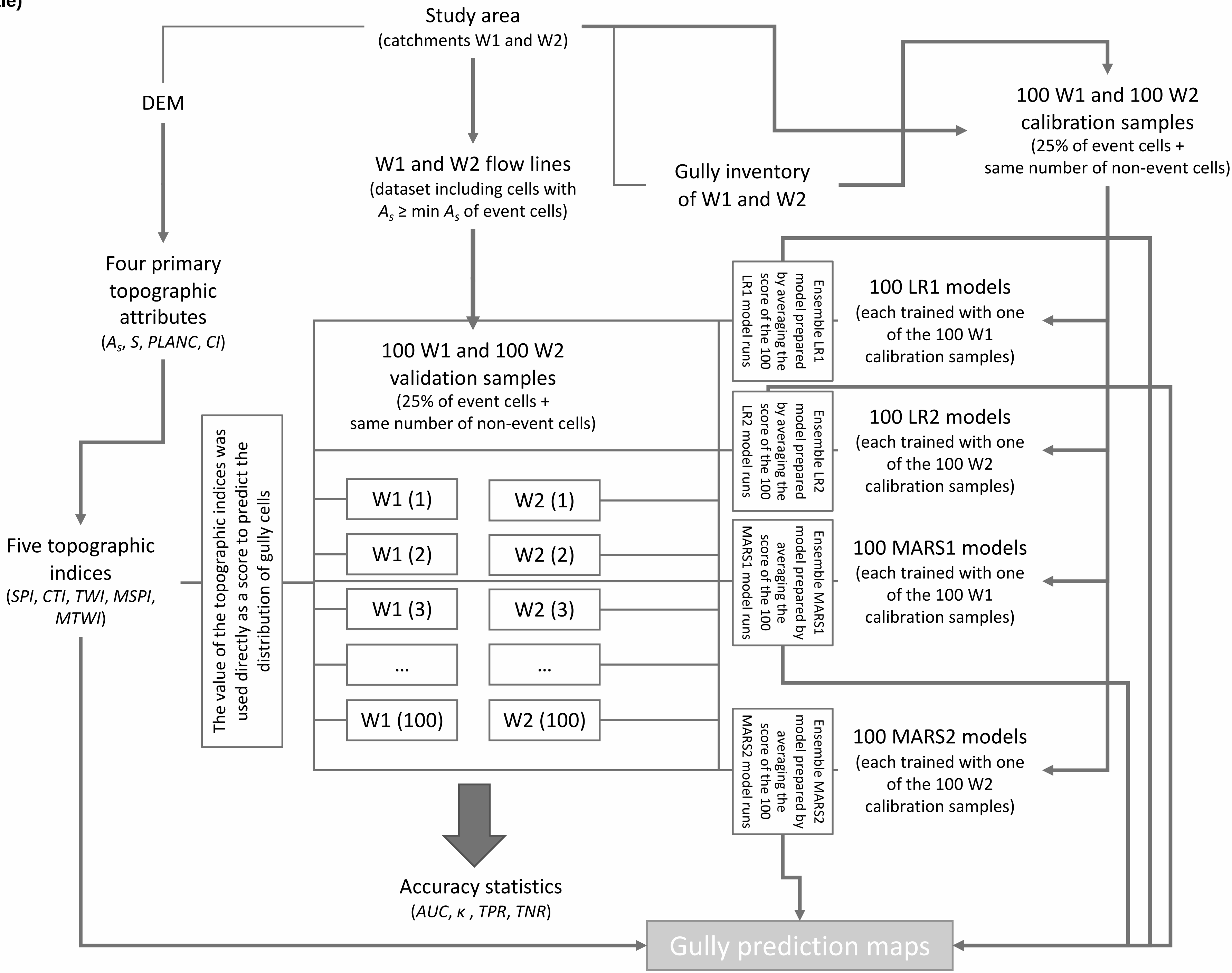








le)



Study area
(catchments W1 and W2)
DEM
W1 and W2 flow lines
(dataset including cells with $A_s \geq$ min A_s of event cells)
Gully inventory of W1 and W2
100 W1 and 100 W2 calibration samples
(25% of event cells + same number of non-event cells)
Four primary topographic attributes
(A_s, S, $PLANC$, CI)
100 W1 and 100 W2 validation samples
(25% of event cells + same number of non-event cells)
Ensemble LR1 model prepared by averaging the score of the 100 LR1 model runs
100 LR1 models
(each trained with one of the 100 W1 calibration samples)
Ensemble LR2 model prepared by averaging the score of the 100 LR2 model runs
100 LR2 models
(each trained with one of the 100 W2 calibration samples)
Ensemble MARS1 model prepared by averaging the score of the 100 MARS1 model runs
100 MARS1 models
(each trained with one of the 100 W1 calibration samples)
Ensemble MARS2 model prepared by averaging the score of the 100 MARS2 model runs
100 MARS2 models
(each trained with one of the 100 W2 calibration samples)
The value of the topographic indices was used directly as a score to predict the distribution of gully cells
Five topographic indices
(SPI, CTI, TWI, $MSPI$, $MTWI$)
W1 (1)
W2 (1)
W1 (2)
W2 (2)
W1 (3)
W2 (3)
...
...
W1 (100)
W2 (100)
Accuracy statistics
(AUC, κ, TPR, TNR)
Gully prediction maps

Figure 2 (Greyscale)
Click here to download high resolution image

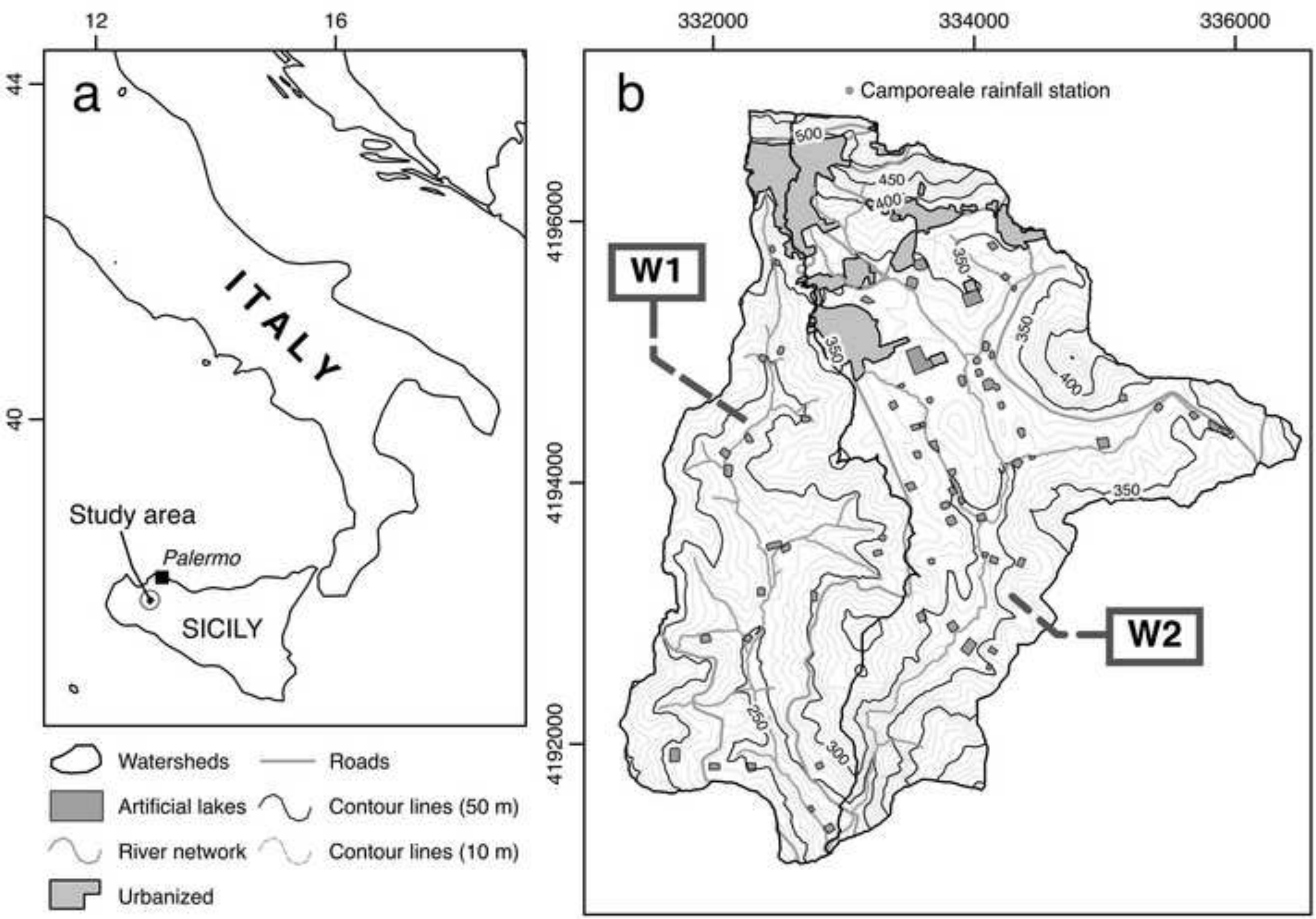


Figure 3 (Greyscale)
Click here to download high resolution image

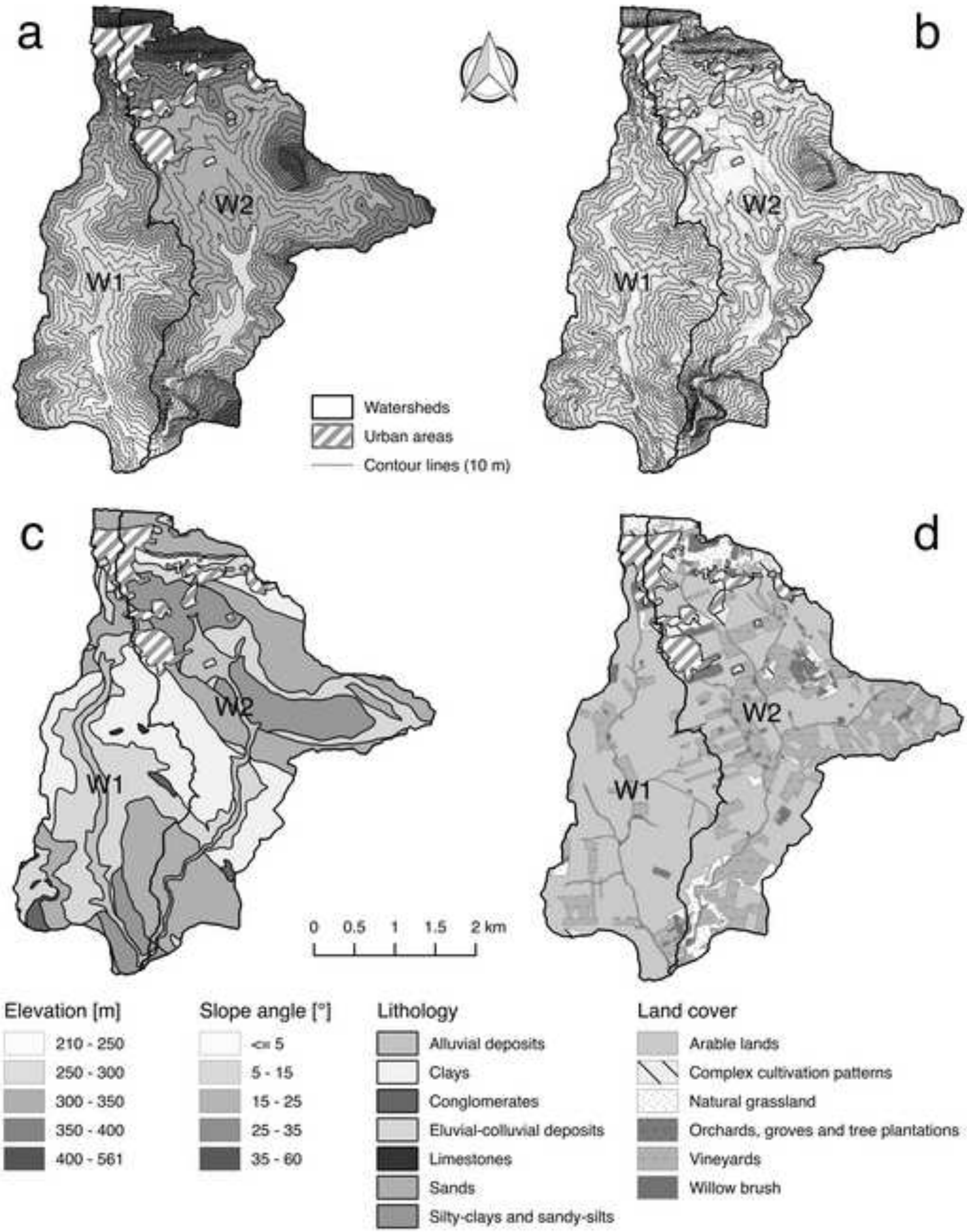


Figure 4 (Greyscale)
Click here to download high resolution image

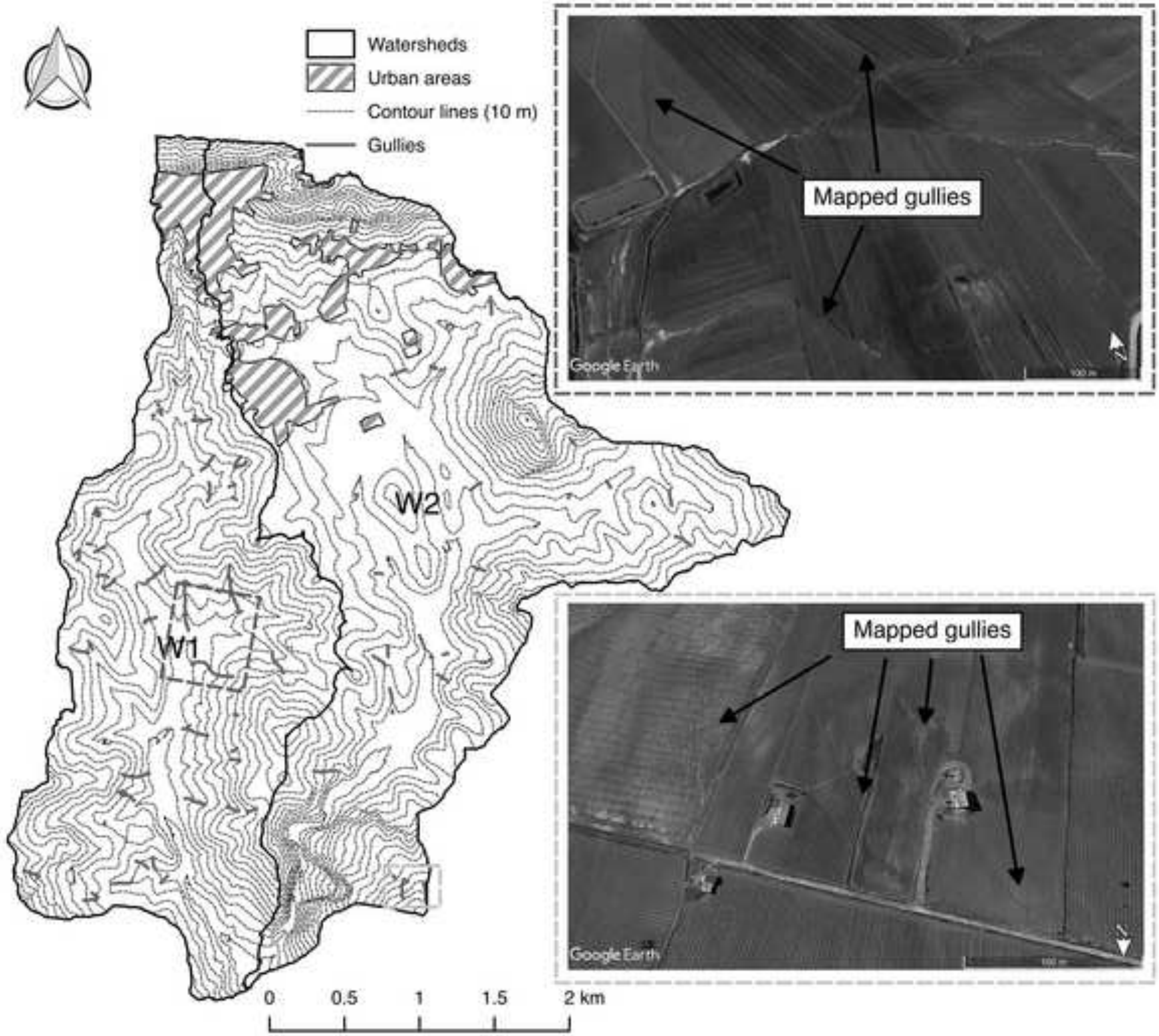


Figure 5 (Greyscale)
Click here to download high resolution image

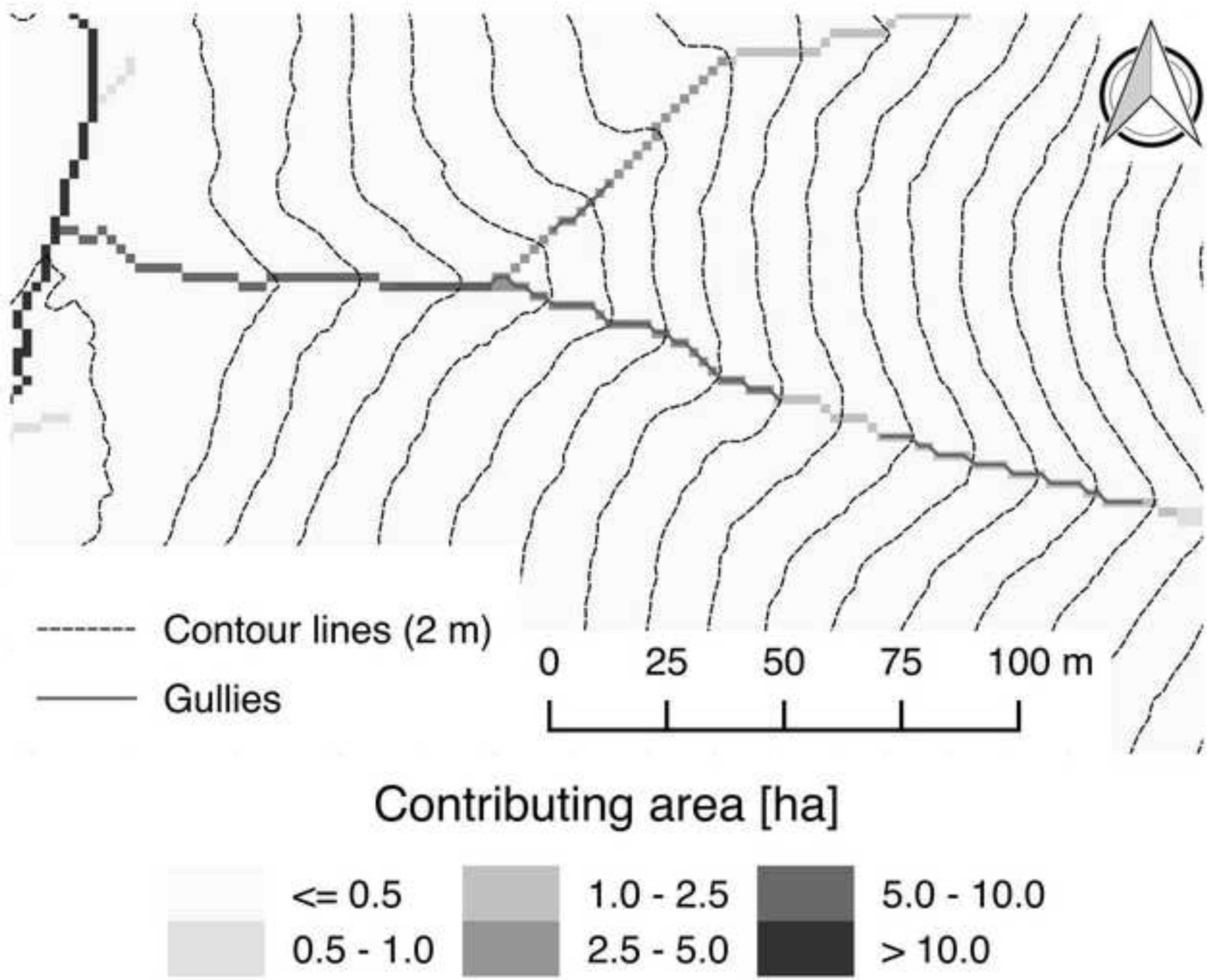


Figure 6 (Greyscale)

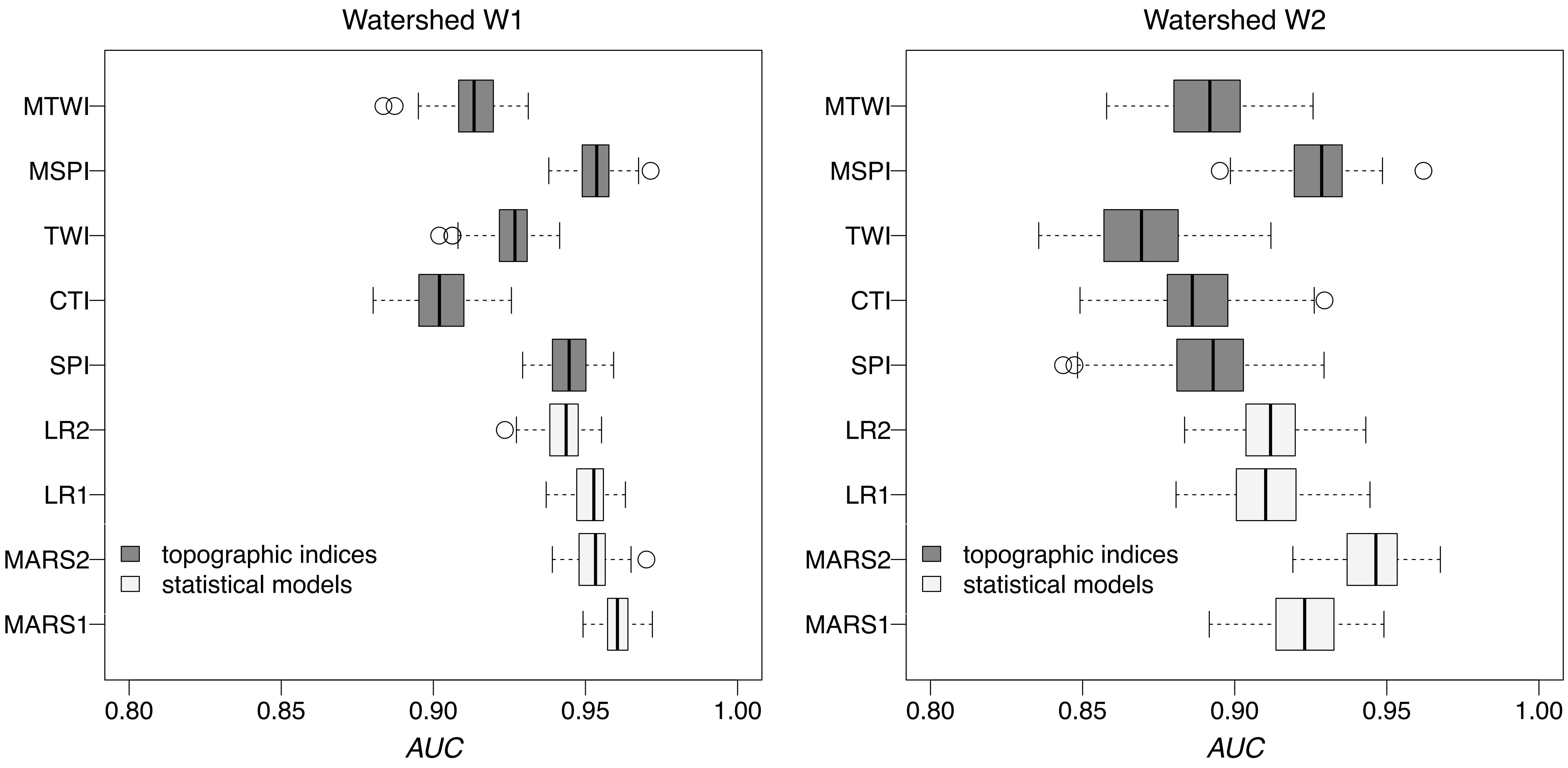
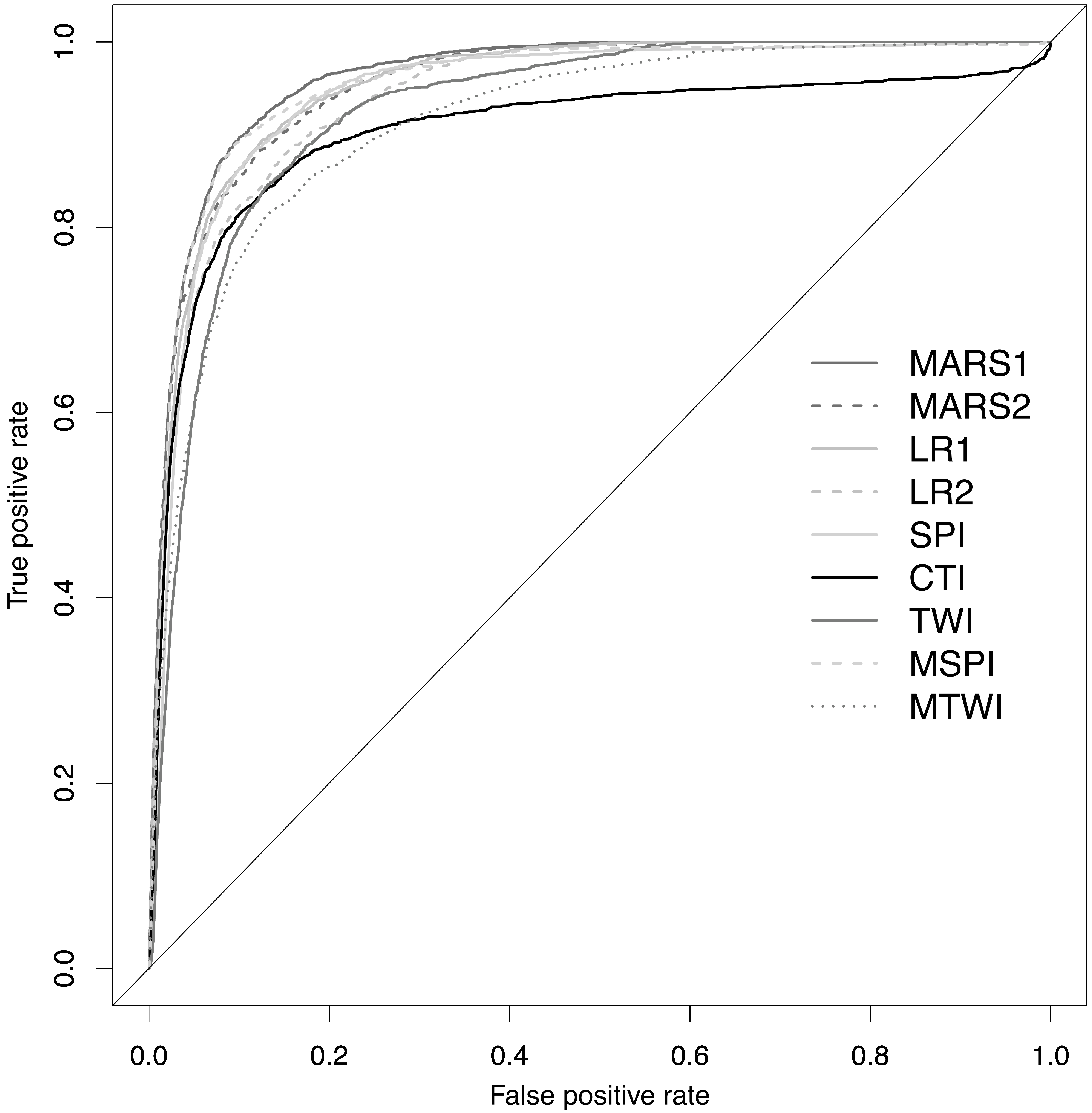


Figure 7 (Greyscale)



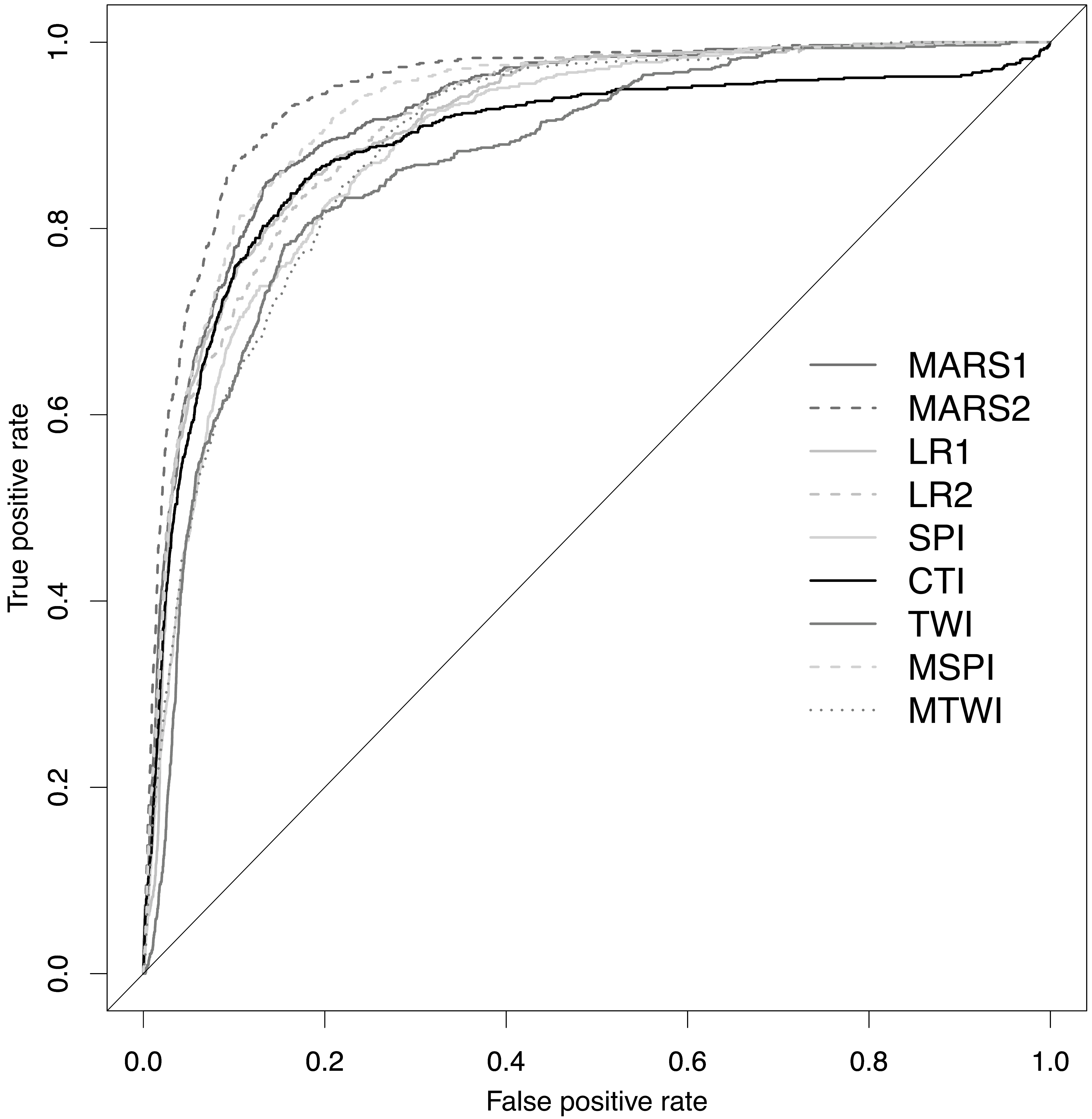
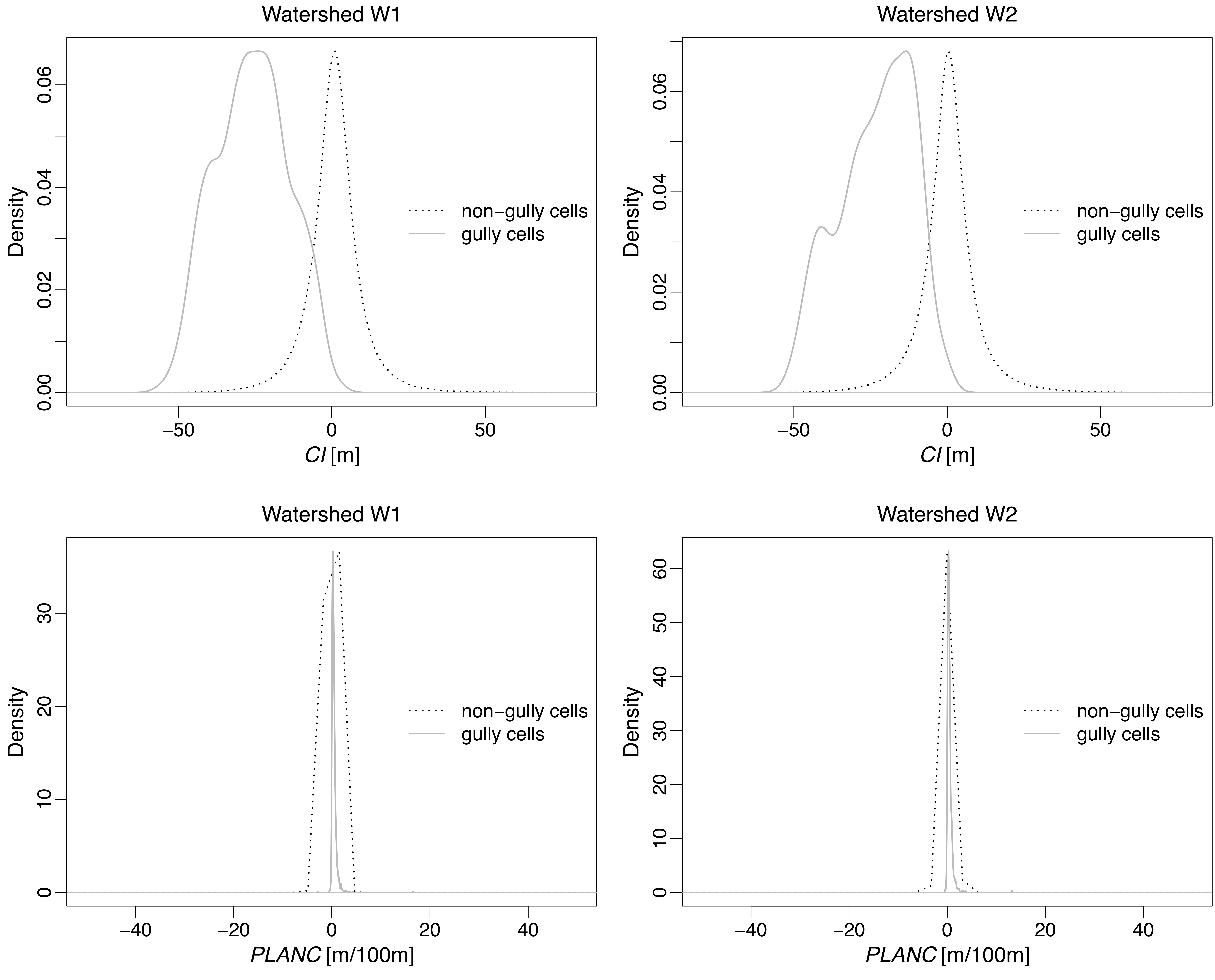


Figure 8 (Greyscale)
Click here to download high resolution image

Figure 9 (Greyscale)

Figure 10 (Greyscale)



*Declaration of Interest Statement

Declaration of interests

☒ 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.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: