Supplement of Nat. Hazards Earth Syst. Sci., 24, 1–12, 2024
https://doi.org/10.5194/nhess-24-1-2024-supplement

Natural Hazards and Earth System Sciences
Open Access

Supplement of

Slope Unit Maker (SUMak): an efficient and parameter-free algorithm for delineating slope units to improve landslide modeling

Jacob B. Woodard et al.

Correspondence to: Jacob B. Woodard (jwoodard@usgs.gov)

Contents of this file

Introduction

Here we provide more details on how SUMak was developed within R (Section S1), a description of the hyperparameters tuned in XGBoost (Section S2), a table providing the number of landslide samples used for developing and evaluating the landslide models (Table S1), an illustration of the slope-unit delineation method (Figure S1), figures showing the existence of landslides within slope units over the Oregon watersheds (Figure S2) and Puerto Rico (Figure S3), a figure showing the slope units over the entire Oregon watersheds (Figure S4), susceptibility maps created using the different machine learning algorithms and sampling methods for the Oregon watersheds (Figures S5-S8) and Puerto Rico (Figures S9-S10), a figure showing the cumulative distribution function of susceptibility probability for the Oregon watersheds (Figure S11) and Puerto Rico (Figure S12), plots showing the percent area as a function of probability for the Oregon watersheds (Figure S13) and Puerto Rico (Figure S14), and a zoomed in portion of the susceptibility maps of Puerto Rico (Figure S15).

Section S1.

We wrote SUMak (Woodard, 2023) as a function in R that builds on tools within TauDEM (Tarboton, 2015) and Geographic Resources Analysis Support System (GRASS) (GRASS Development Team, 2020). The algorithm requires two inputs: (1) a digital elevation model (DEM) and (2) a polygon file outlining the region to be analyzed. While only requiring two inputs, the function has many options for adjusting its performance. By default, the algorithm runs in parallel on all the cores available on the local machine. The algorithm can also be run on a unix cluster. Using the default options, the general processing steps include (Figure S1):

1) creating intermediate scale watersheds (~100 km^2) within the specified area,
2) running TauDEM's 'Dropanalysis' function to determine the optimal flow accumulation threshold for each intermediate watershed and create new watersheds at a scale that captures hillslope processes,
3) dividing the optimal watersheds by their longest flow path to create slope units, and
4) combining the slope units that have unrealistic geometries with the surrounding slope units.

After the initial slope units are delineated, as described in the main text (section 2.1), certain slope units that appear unnaturally long or small can result from the process of delineating watersheds and splitting them with the longest flow path. As such, the algorithm has an option to implement a cleaning technique that eliminates slope units that are less than 3 cells wide in any direction by combining them with adjacent slope units.

Section S2.

Here we briefly describe the effects of the XGBoost hyperparameters used to optimize the model performance. The 'max_depth' and 'min_child_weight' parameters controls how complex the decision tree can be. The ‘subsample’ and ‘colsample_bytree’ parameters determine the proportion of the training data used for growing the model trees. The ‘gamma’ parameter is a threshold loss reduction value that must be exceeded for further tree growth. For a more complete explanation of these parameters and the methods used by XGBoost, see see Chen & Guestrin (2016) and https://xgboost.readthedocs.io/.

Table. S1 Number of landslide samples

Location	**Sampling Method**				
	10m	10m_med	10_m_multi	30m	SU
Umpqua	3499	3499	3707	3090	1237
Calapooia	485	485	4824	484	1983
Puerto Rico	NA	NA	NA	71431	6263

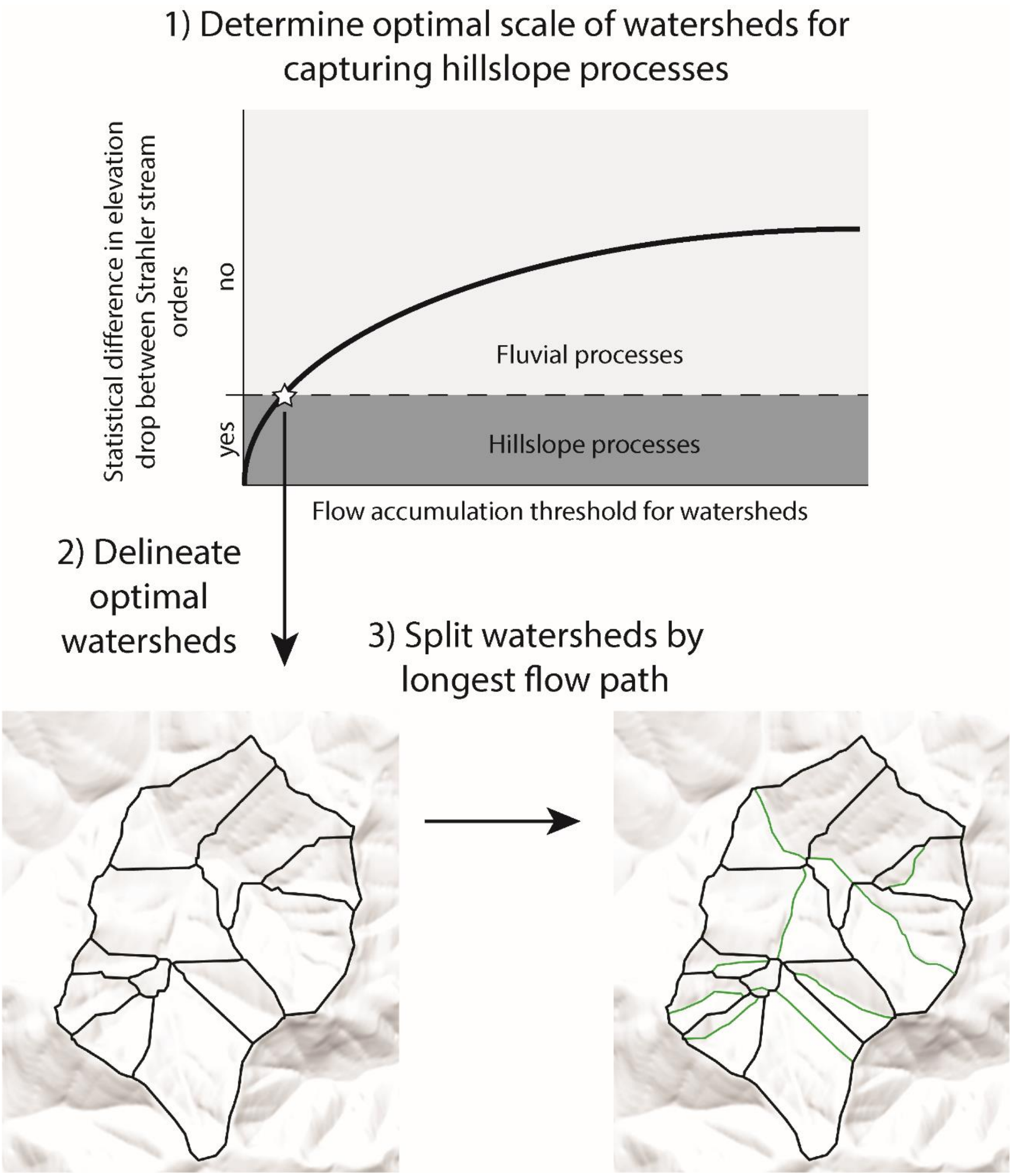


Figure S1. Illustration of the slope unit delineation method. The algorithm first determines the optimal scale (flow accumulation threshold) for capturing hillslope processes using the constant drop law. It then delineates and splits the watersheds by their longest flow paths (green) to create slope units.

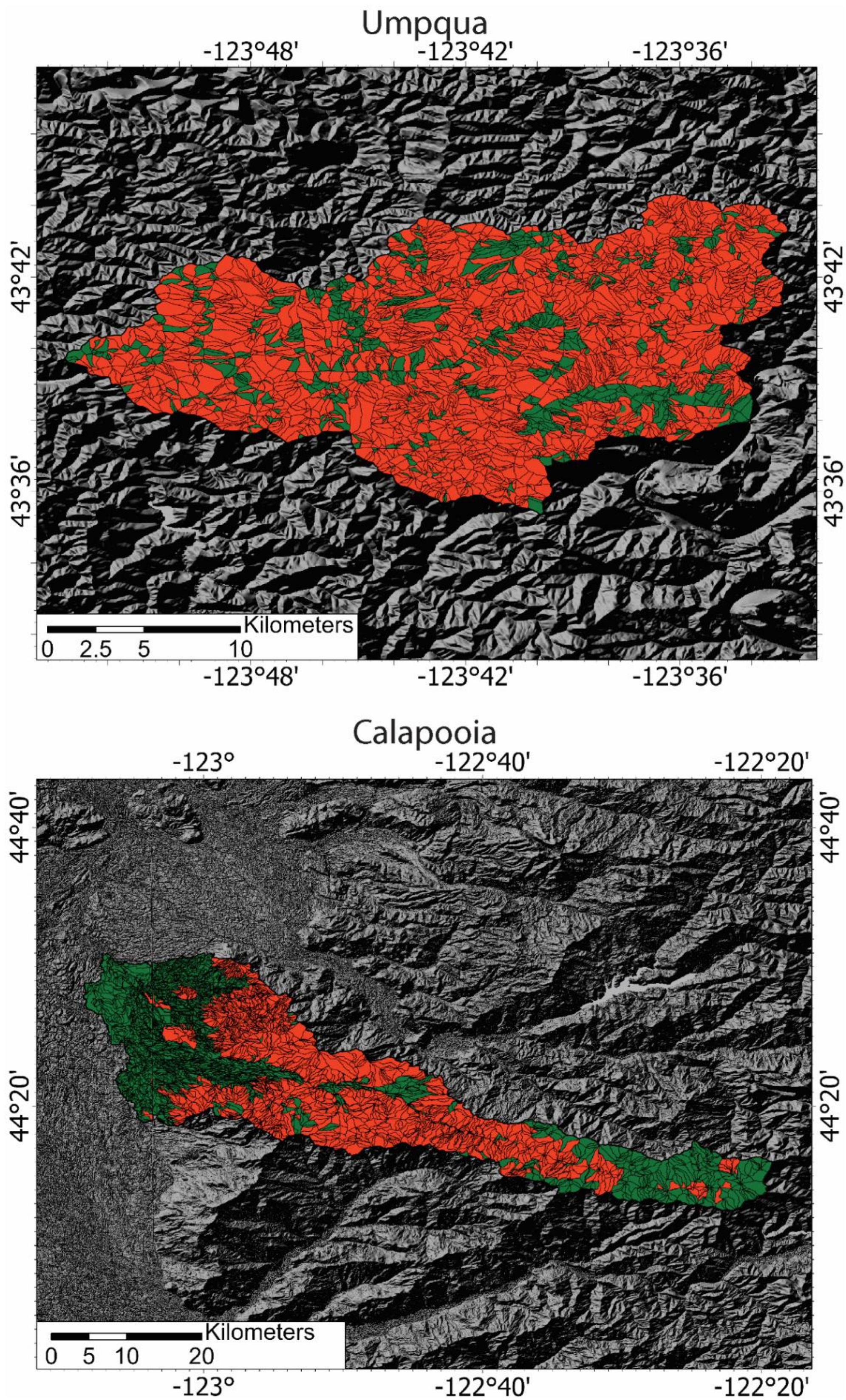


Figure S2: Maps illustrating the existence (red) or non-existence (green) of a landslide within each slope unit over the Umpqua and Calapooia watersheds, Oregon.

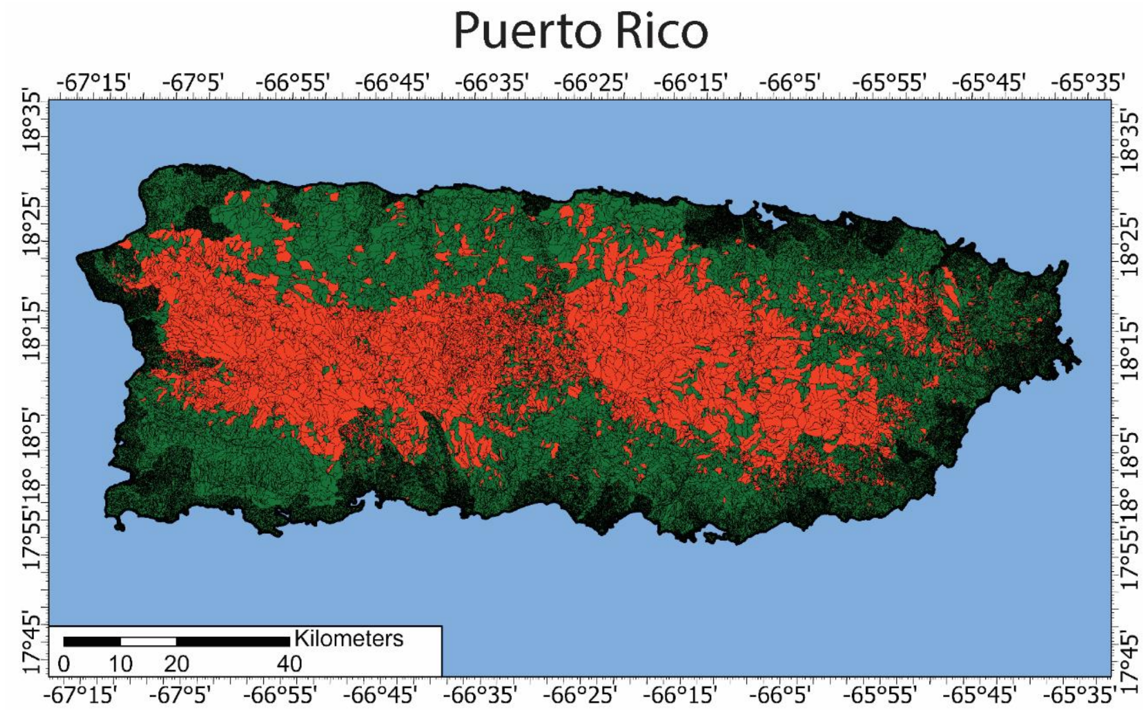


Figure S3: Maps illustrating the existence (red) or non-existence (green) of a landslide within each slope unit over Puerto Rico.

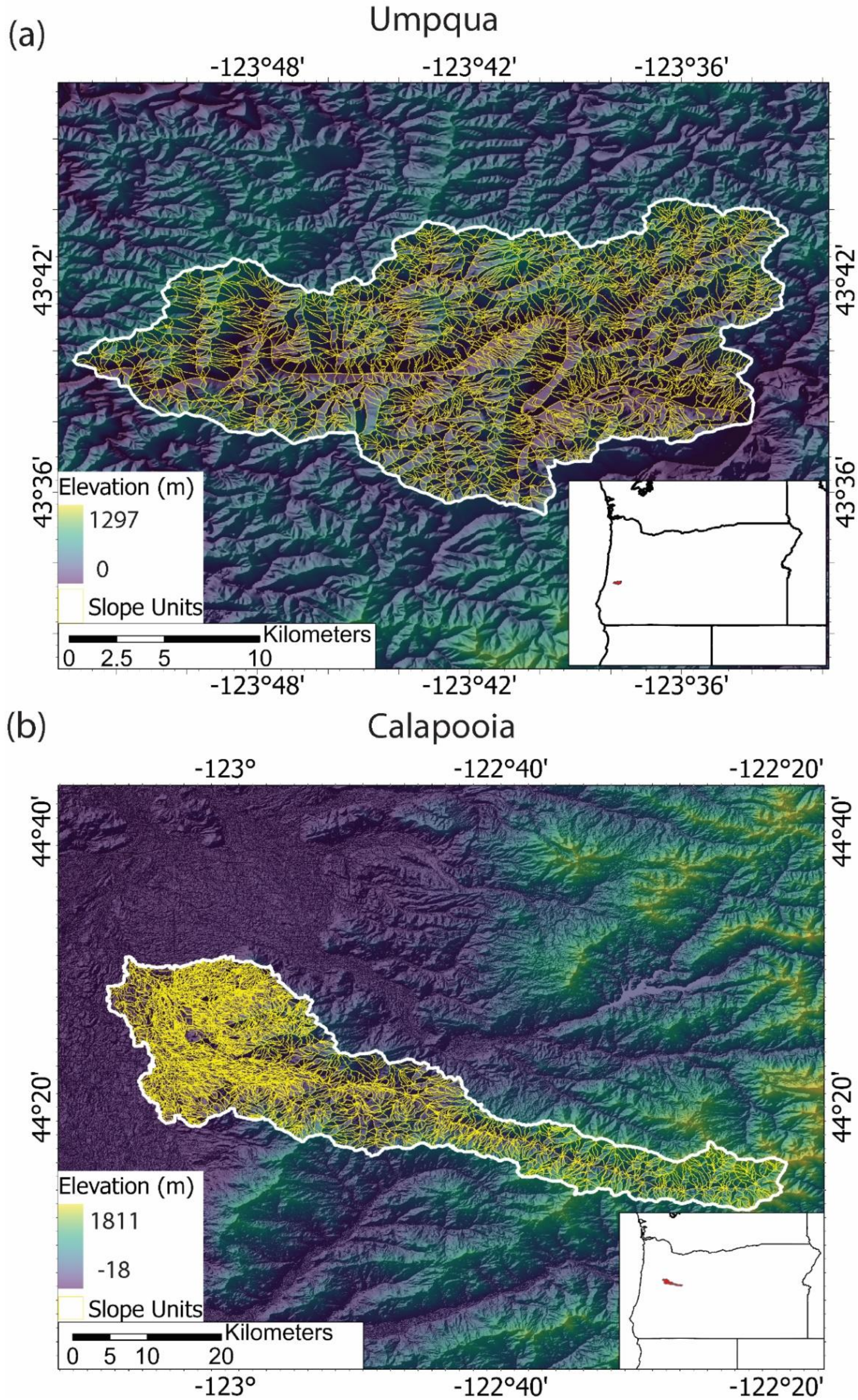


Figure S4. SUMak delineated slope units over the (a) Umpqua and (b) Calapooia watersheds, Oregon.

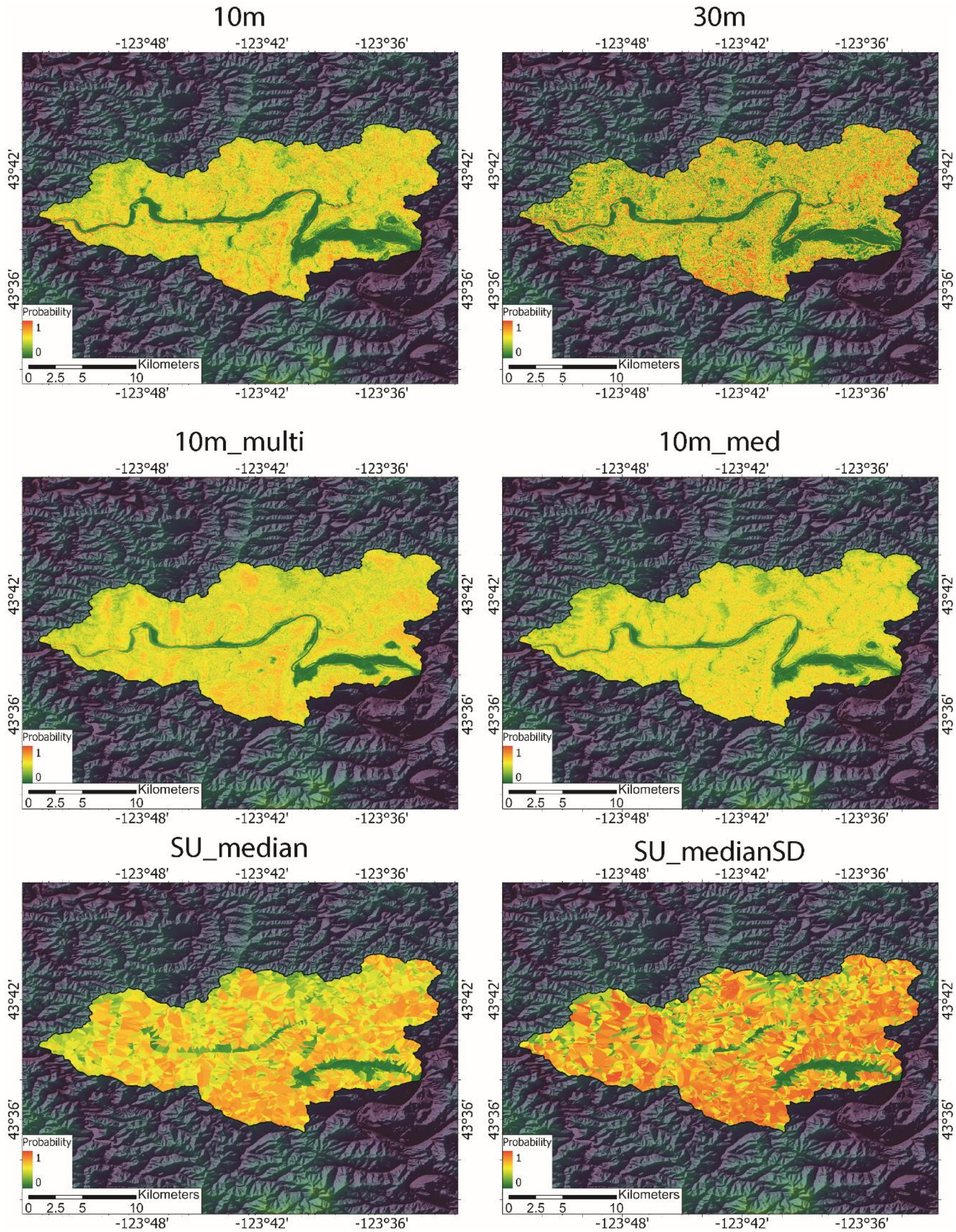


Figure S5. XGBoost susceptibility models over the Umpqua watershed. Maps are for the different sampling methods (10m, 30m, 10m_multi) and the slope unit maps using only the median (SU_median) and the median and standard deviation of the predictor values (SU_medianSD).

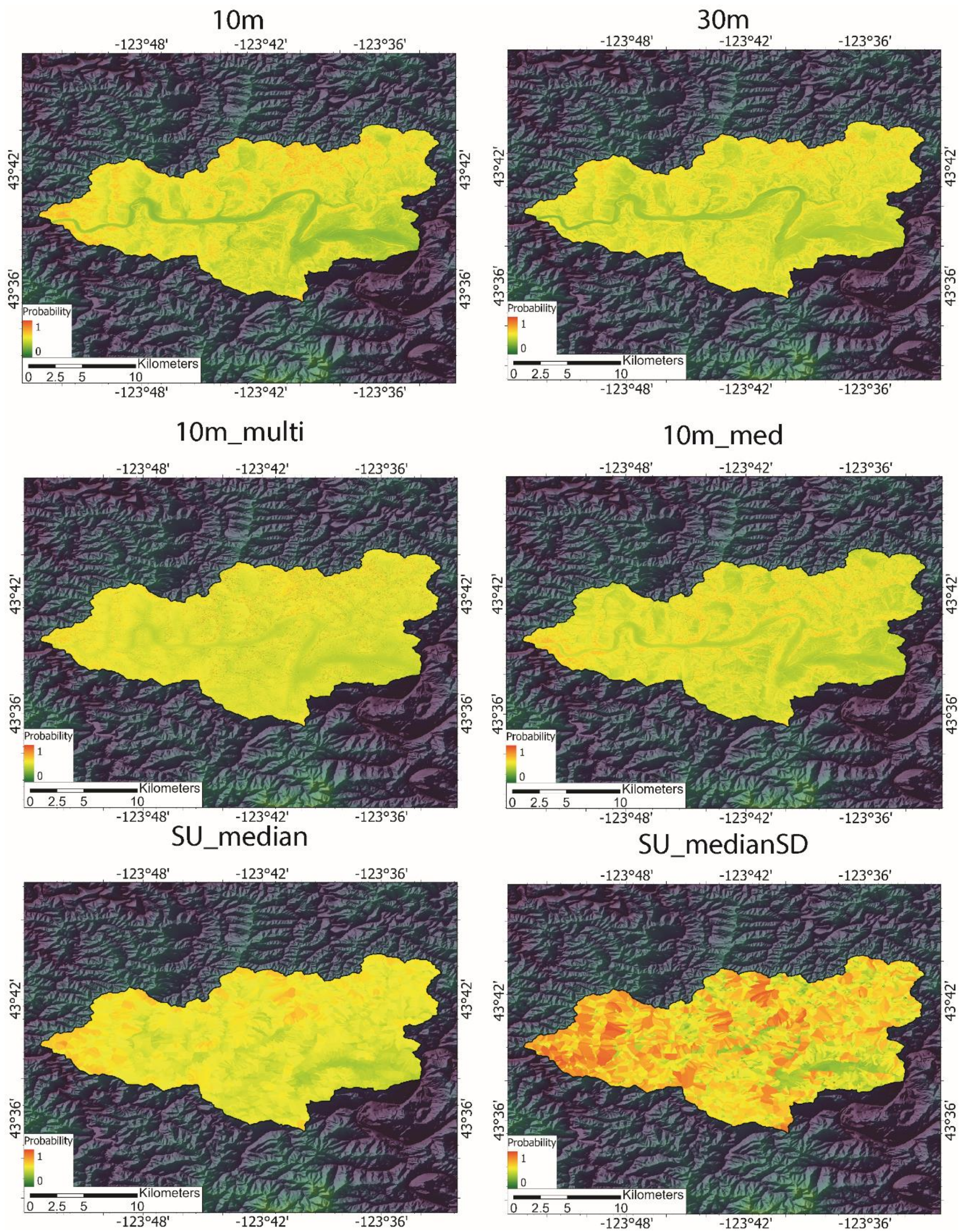


Figure S6. Logistic regression susceptibility models over the Umpqua watershed. Maps are for the different sampling methods (10m, 30m, 10m_multi) and the slope unit maps using only the median (SU_median) and the median and standard deviation of the predictor values (SU_medianSD).

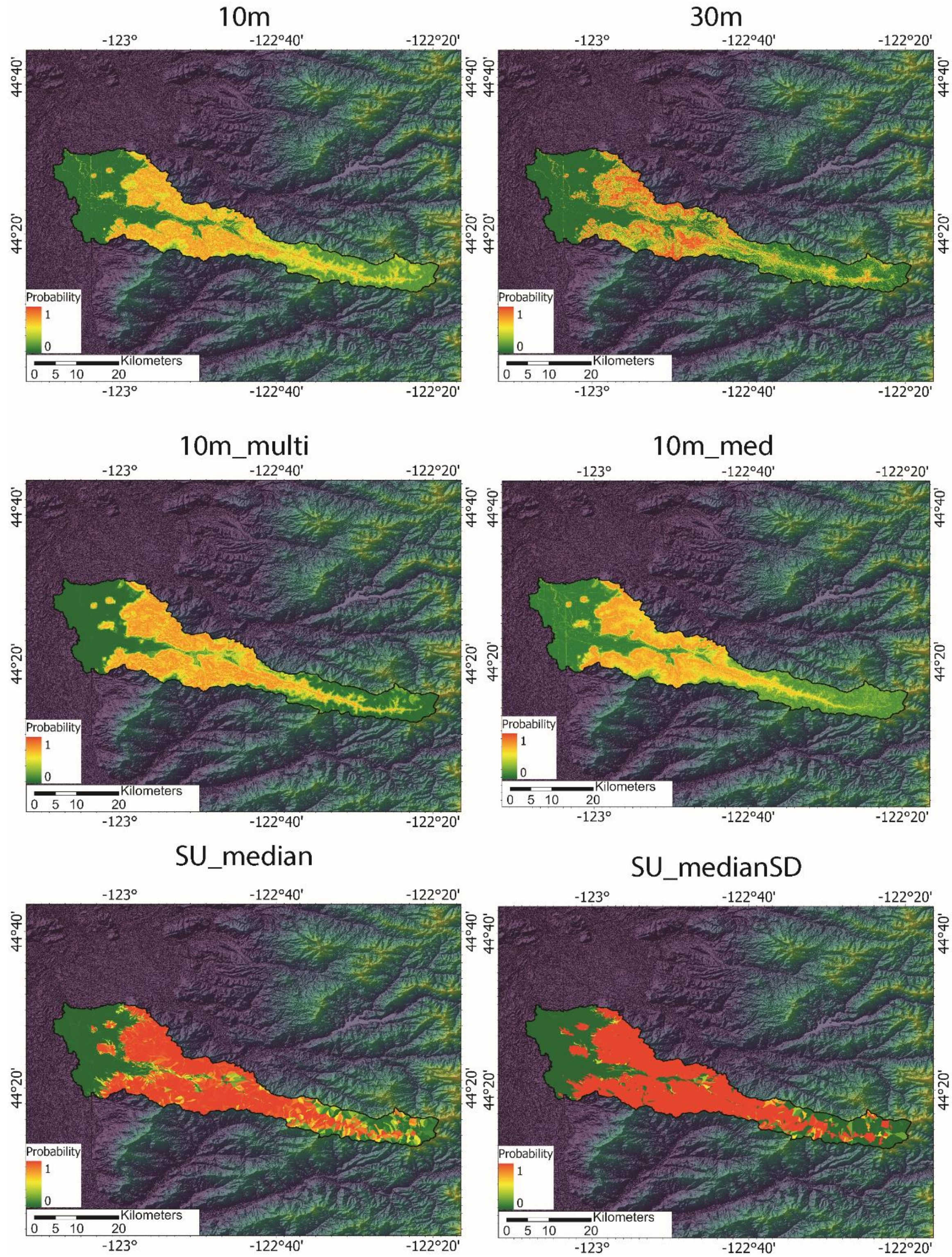


Figure S7. XGBoost susceptibility models over the Calapooia watershed. Maps are for the different sampling methods (10m, 30m, 10m_multi) and the slope unit maps using only the median (SU_median) and the median and standard deviation of the predictor values (SU_medianSD).

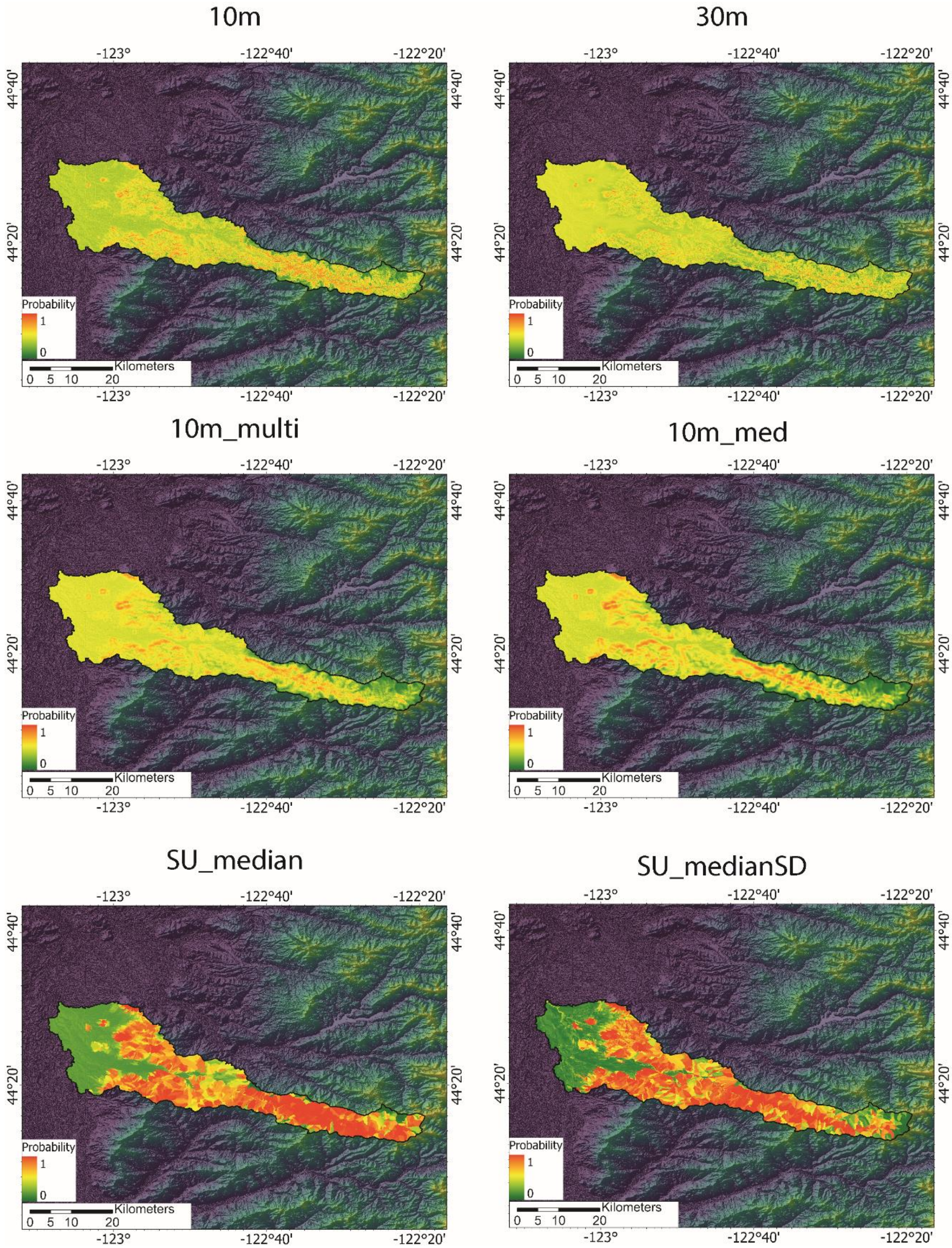
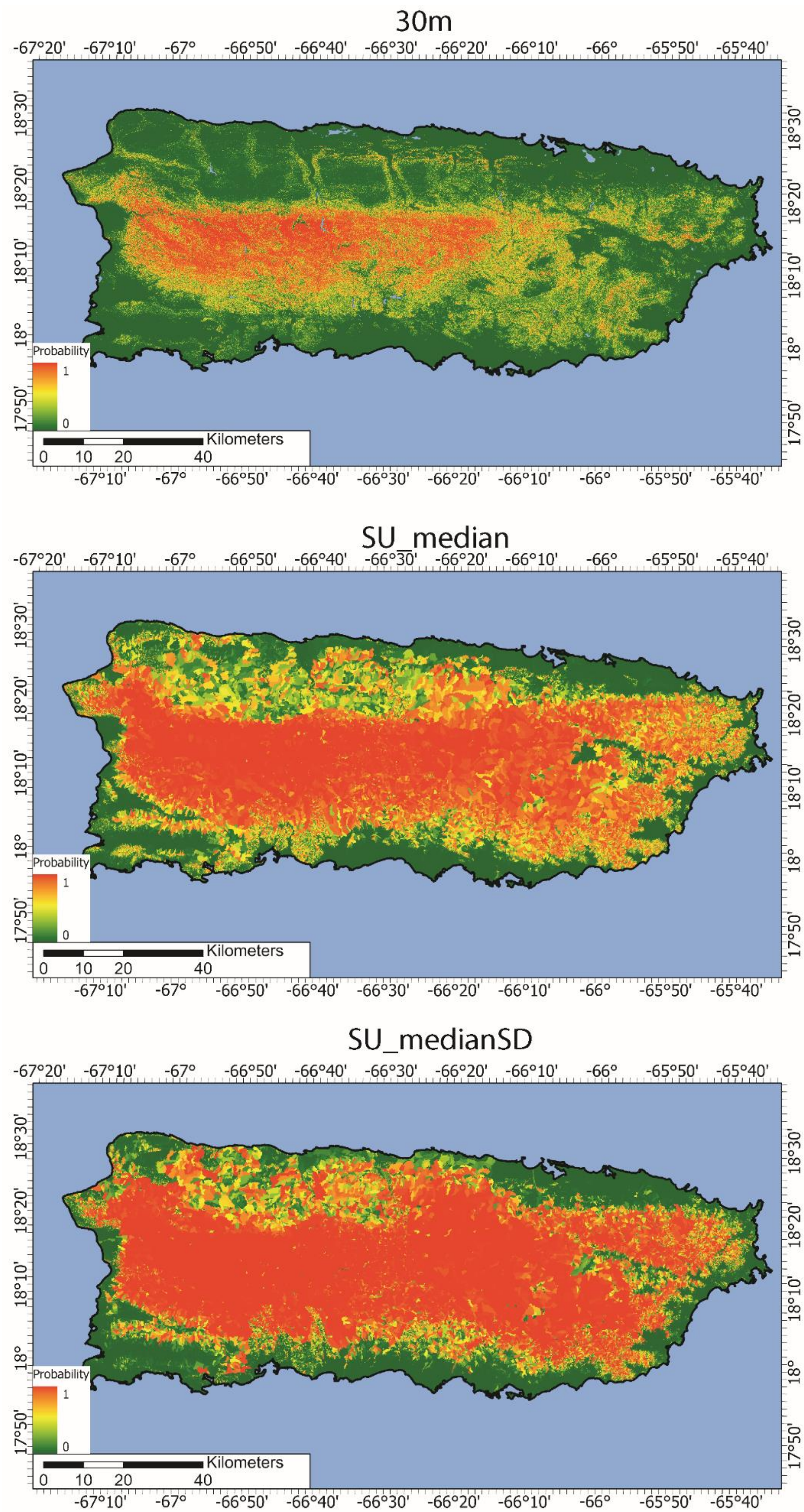


Figure S8. Logistic regression susceptibility models over the Calapooia watershed. Maps are for the different sampling methods (10m, 30m, 10m_multi) and the slope unit maps using only the median (SU_median) and the median and standard deviation of the predictor values (SU_medianSD).


30m
SU_median
SU_medianSD
-67°20' -67°10' -67° -66°50' -66°40' -66°30' -66°20' -66°10' -66° -65°50' -65°40'
18°30' 18°20' 18°10' 18° 17°50'
Probability
1
0
Kilometers
0 10 20 40

95 **Figure S9**. XGBoost susceptibility models over Puerto Rico for the Hurricane Maria landslide dataset. Maps are for the 30m grid-based model and the slope unit maps using only the median (SU_median) and the median and standard deviation of the predictor values (SU_medianSD).

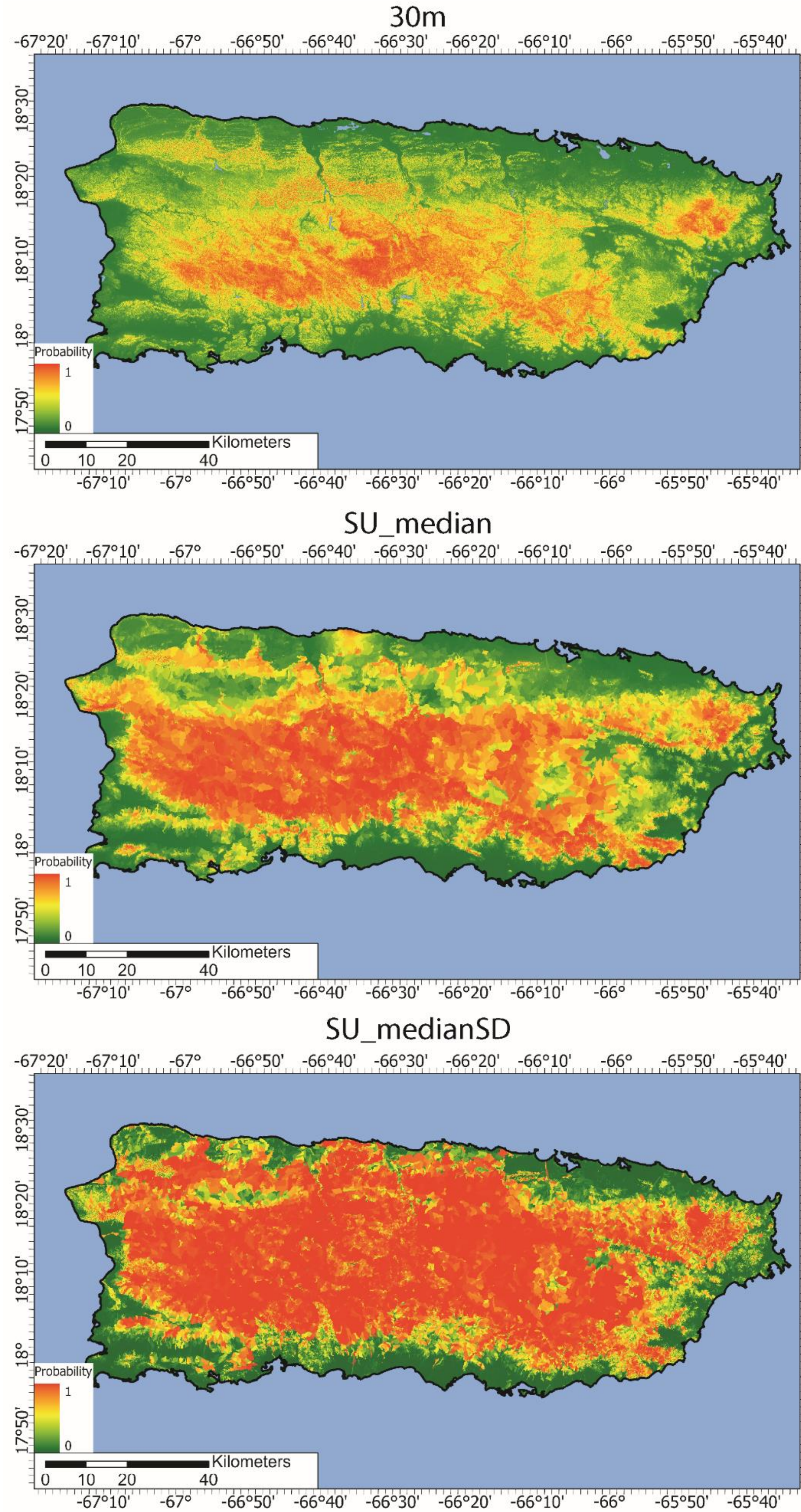

30m
-67°20' -67°10' -67° -66°50' -66°40' -66°30' -66°20' -66°10' -66° -65°50' -65°40'
18°30' 18°20' 18°10' 18° 17°50'
Probability
1
0
Kilometers
0 10 20 40
SU_median
Probability
1
0
Kilometers
0 10 20 40
SU_medianSD
Probability
1
0
Kilometers
0 10 20 40

Figure S10. Logistic regression susceptibility models over Puerto Rico for the Hurricane Maria landslide dataset. Maps are for the 30m grid-based model and the slope unit maps using only the median (SU_median) and the median and standard deviation of the predictor values (SU_medianSD).

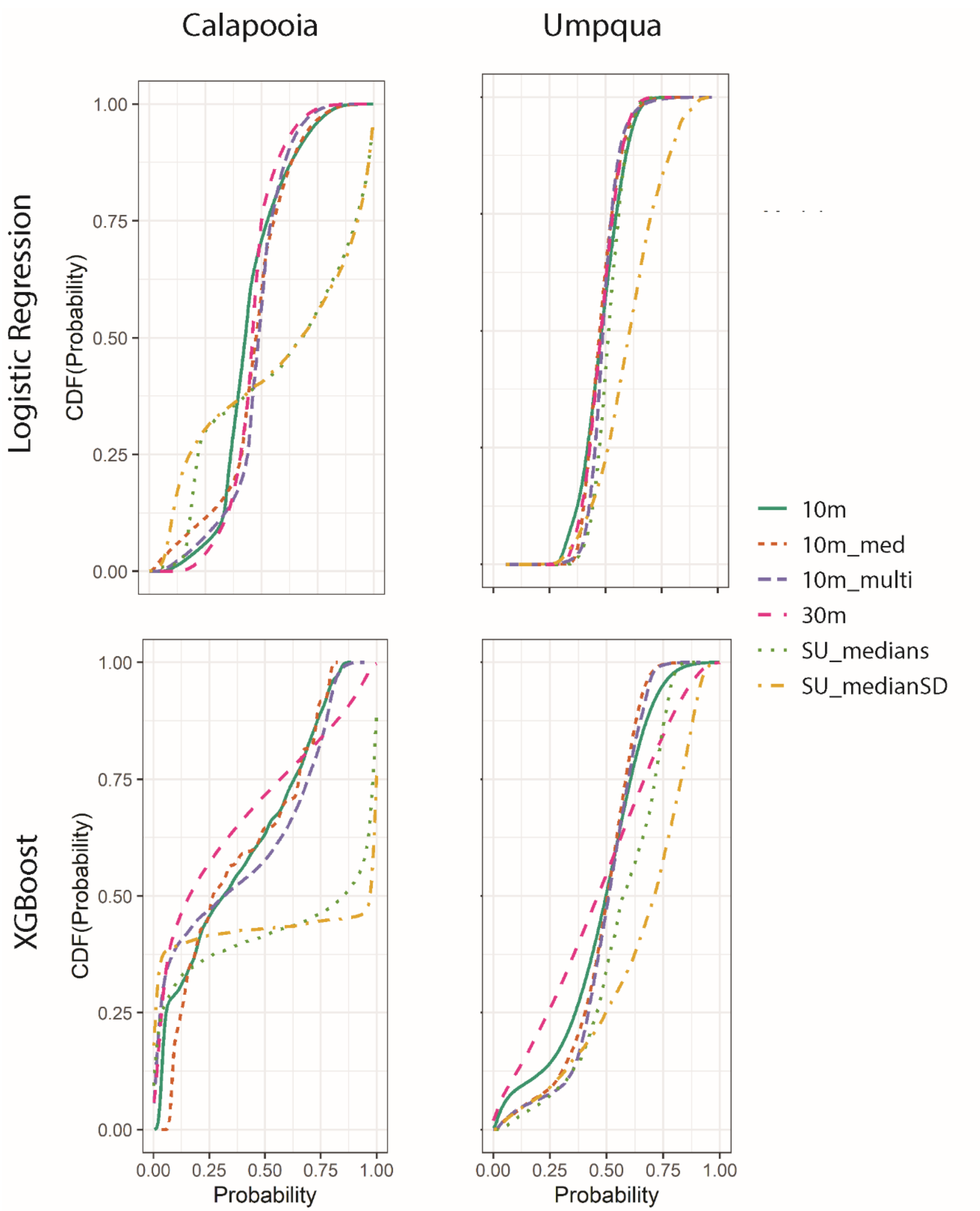
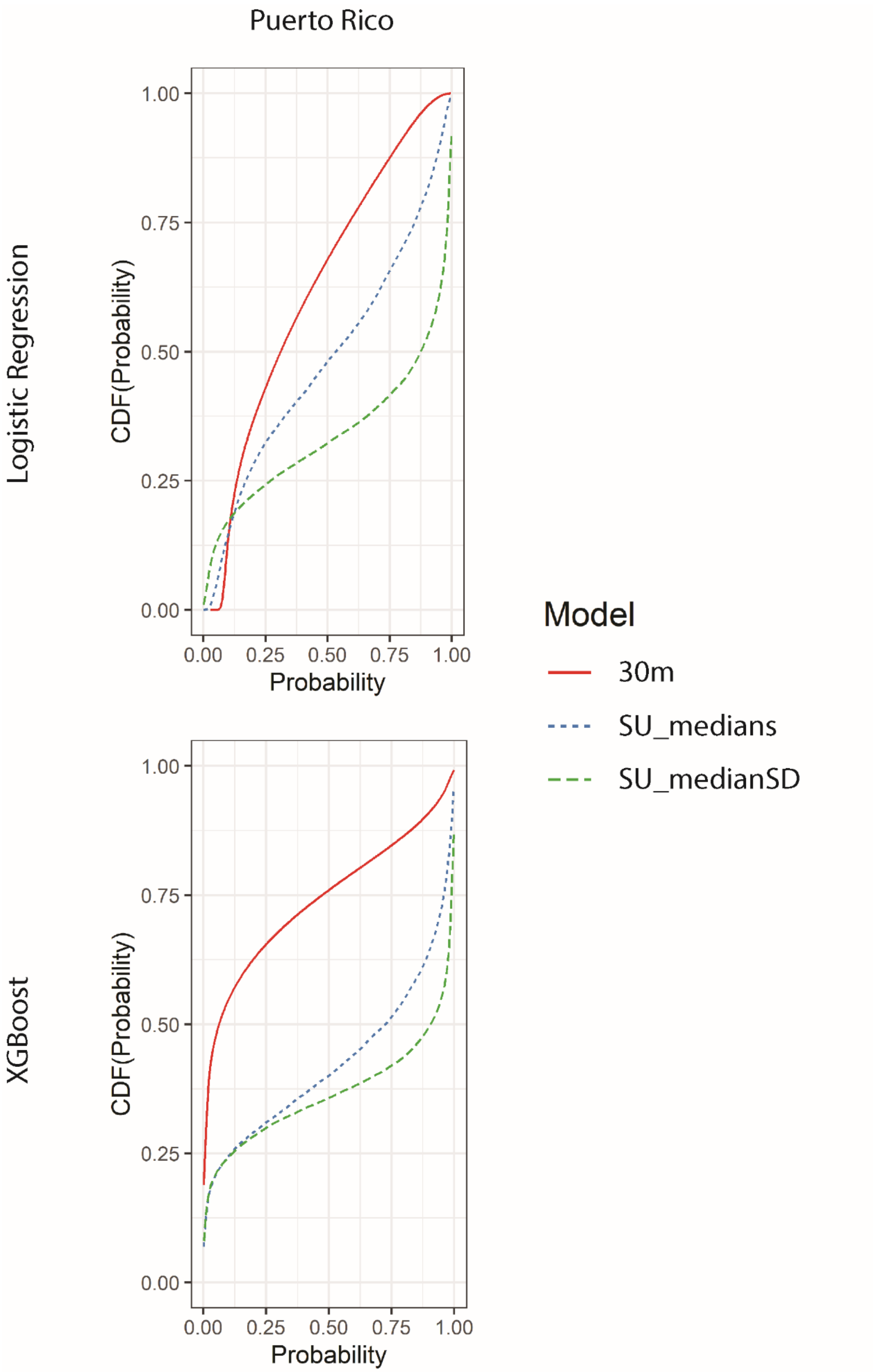


Figure S11. Cumulative distribution functions (CDF) of the Umpqua and Calapooia susceptibility model probabilities. The CDF is the probability that susceptibility model probability distribution function will take a value less than or equal to the value of the x-axis.
110 The different lines represent the different sampling techniques and mapping units used to make the susceptibility maps. For the grid-based maps, "10m" samples were taken at the highest elevation point within each landslide polygon using a 10 m resolution digital elevation model (DEM); "10m_med" samples were taken from the median elevation point within each landslide polygon using a 10 m resolution DEM; "10m_multi" were taken

throughout the landslide polygons with a 200m spacing using a 10 m resolution DEM; “30m” samples ” samples were taken at the highest elevation point within each landslide polygon using a 30 m resolution DEM. For the slope unit-based maps, “SU_medians” used only the median value of each predictor within each slope unit and “SU_medianSD” used the median and standard deviation values within each slope unit.


Puerto Rico
Logistic Regression
CDF(Probability)
1.00
0.75
0.50
0.25
0.00
0.00 0.25 0.50 0.75 1.00
Probability
XGBoost
Model
30m
SU_medians
SU_medianSD

120

Figure S12. Cumulative distribution functions (CDF) of the Puerto Rico susceptibility model probabilities. The CDF is the probability that susceptibility model probability distribution function will take a value less than or equal to the value of the x-axis. The different lines represent the different sampling techniques and mapping units used to make the susceptibility maps. See caption to Figure S11 for details on the sampling techniques.

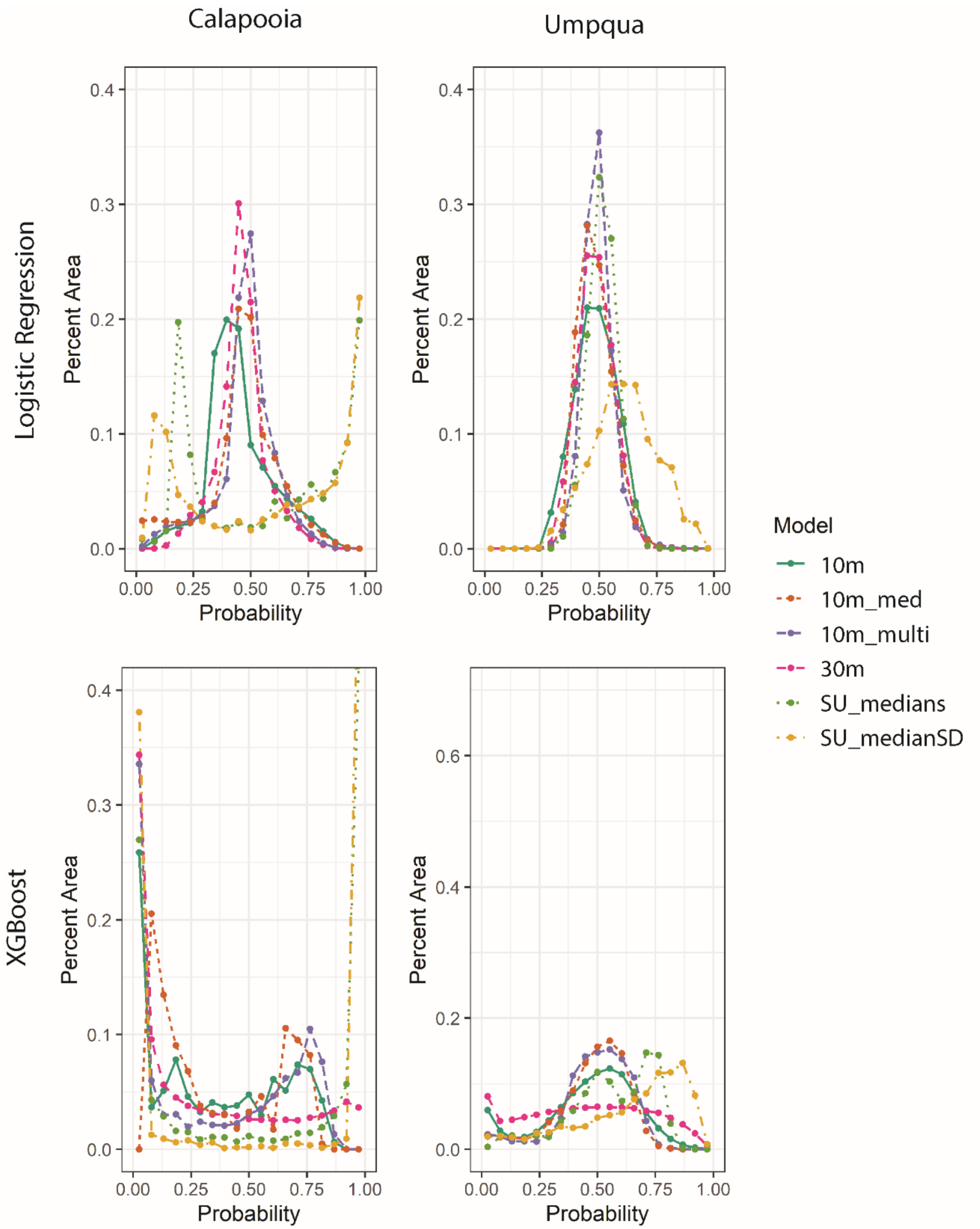


Figure S13. Percent area as a function of probability for the Umpqua and Calappooia susceptibility models. Plots were created using a 20 bin histogram and converting the counts to percent area. Points show the midpoints of each bin. The different lines represent the different sampling techniques and mapping units used to make the susceptibility maps. See caption to Figure S11 for details on the sampling techniques.

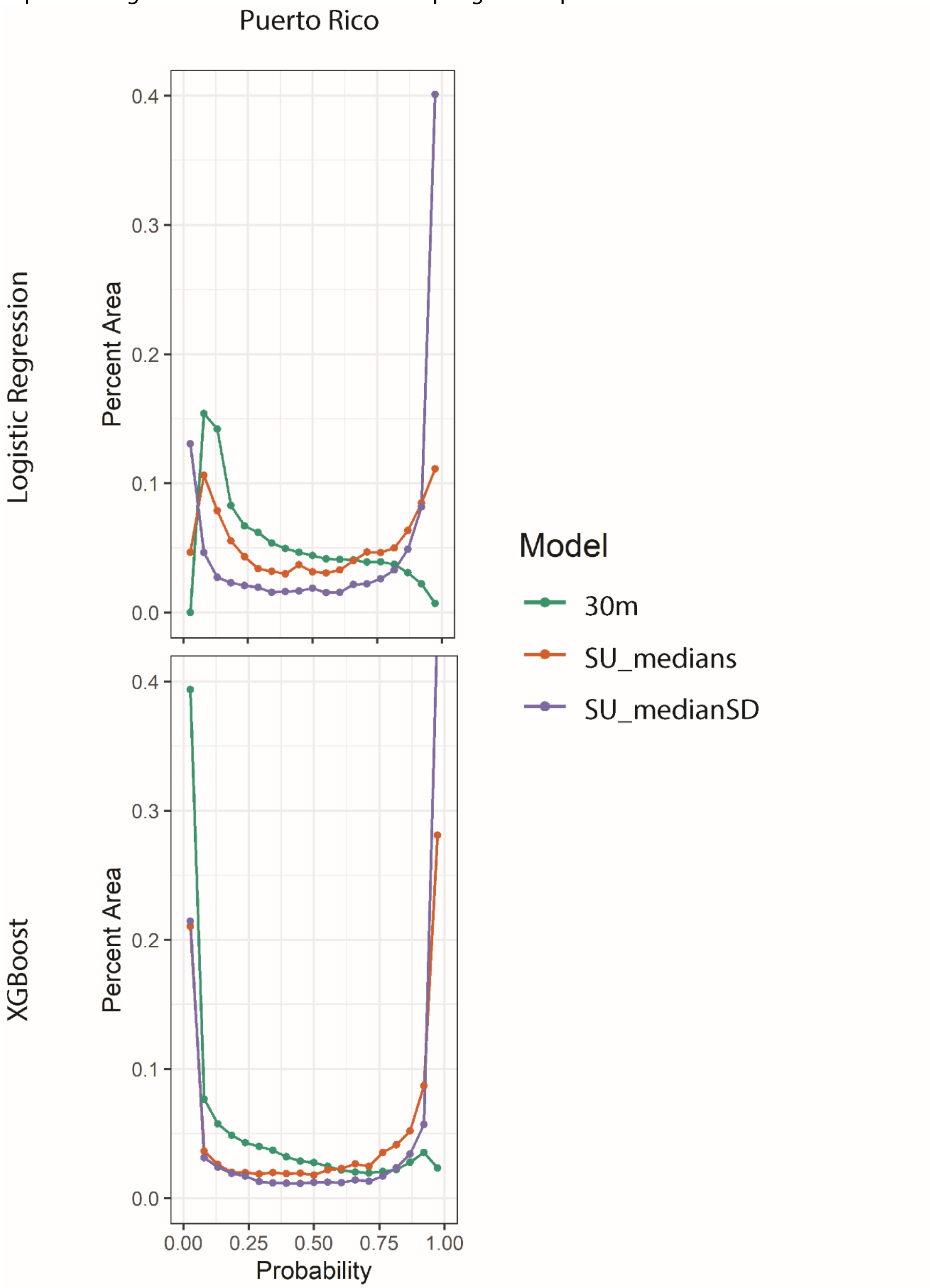


Figure S14. Percent area as a function of probability for the Puerto Rico susceptibility models.
135 Plots were created using a 20 bin histogram and converting the counts to percent area. Points

show the midpoints of each bin. The different lines represent the different sampling techniques and mapping units used to make the susceptibility maps. See caption to Figure S11 for details on the sampling techniques.

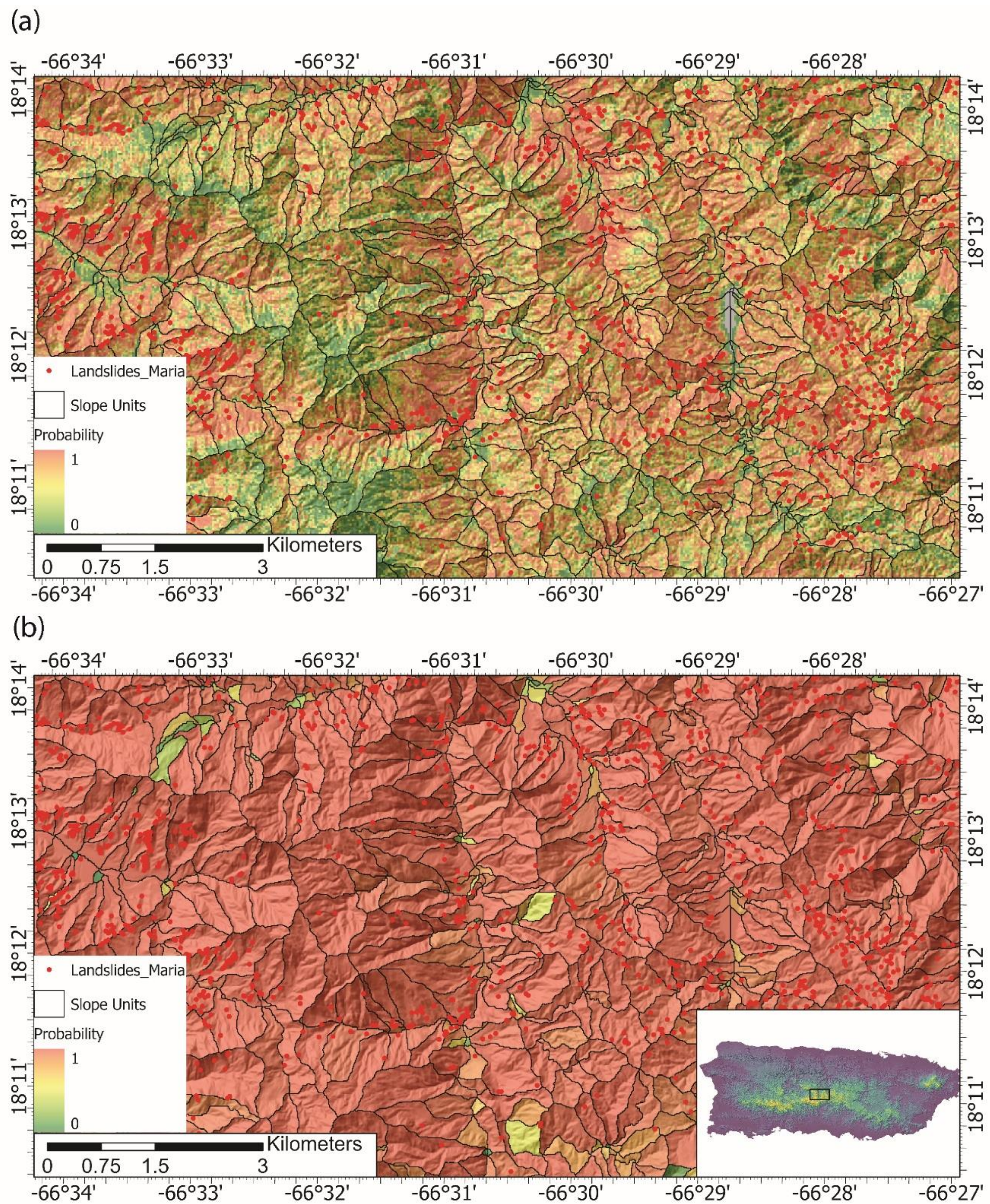


Figure S15. Zoomed in portion of Puerto Rico landslide susceptibility models from the (a) 30m grid-based maps and (b) using slope units with median and standard deviation predictor values with XGBoost. Red dots show the location of mapped landslide scarps. Inset map shows maps extent.

Acknowledgements

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-Augu*, 785–794. https://doi.org/10.1145/2939672.2939785

GRASS Development Team. (2020). Geographic Resources Analysis Support System (GRASS) Software, Version 7.8. Open Source Geospatial Foundation. Retrieved from https://grass.osgeo.org

Tarboton, D. G. (2015). TauDEM. The Free Software Foundation. Retrieved from https://hydrology.usu.edu/taudem/taudem5

Woodard, J. B. (2023). Slope Unit Maker Software. US Geological Survey Software Release. https://doi.org/https://doi.org/10.5066/P98NXFTN