Supporting Information for

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

J. B. Woodard[1], B. B. Mirus[1], N. J. Wood[2], K. E. Allstadt[1], B. A. Leshchinsky[3], M. M. Crawford[4]
[1] U.S. Geological Survey, Geologic Hazards Science Center, Golden, CO, USA
[2] U.S. Geological Survey, Western Geographic Science Center, Portland, OR, USA
[3] Department of Forest Engineering, Resources and Management, Oregon State University, Corvallis, OR, USA
[4] Kentucky Geological Survey, University of Kentucky, Lexington, KY, USA

Contents of this file

Introduction

Here we provide more details on how SUMak was developed within R (Text S1), an illustration of the slope-unit delineation method (Figure S1), a figure showing the cumulative distribution function of susceptibility probability for the Oregon watersheds (Figure S2) and Puerto Rico (Figure S3), plots showing the percent area as a function of probability for the Oregon watersheds (Figure S4) and Puerto Rico (Figure S5), susceptibility maps created using the different machine learning algorithms and sampling methods for the Oregon watersheds (Figures S6-S9) and Puerto Rico (Figures S10-S11), and a zoomed in portion of the susceptibility maps of Puerto Rico (Figure S12).

Text 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.

Creating intermediate watersheds allows the algorithm to adapt the scaling of the slope units according to the characteristics of the local topography. If the intermediate watershed has significant variation in topography, TauDEM may choose a threshold that doesn't adequately characterize every area within the watershed. Thus, intermediate watersheds must be small enough to limit the variation in topography but large enough to avoid significantly reducing computational efficiency. While experimenting with different watershed dimensions on the topographically diverse regions of Sicily, Puerto Rico, and the Umpqua and Calapooia watersheds, we found an accumulation threshold of ~100 km^2 to adequately strike this balance. This threshold can be adjusted to meet the user's needs, or SUMak has an option to input predetermined intermediate watersheds. After appropriate intermediate watersheds are created, the algorithm runs the rest of the processing steps individually for each intermediate watershed in parallel.

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.

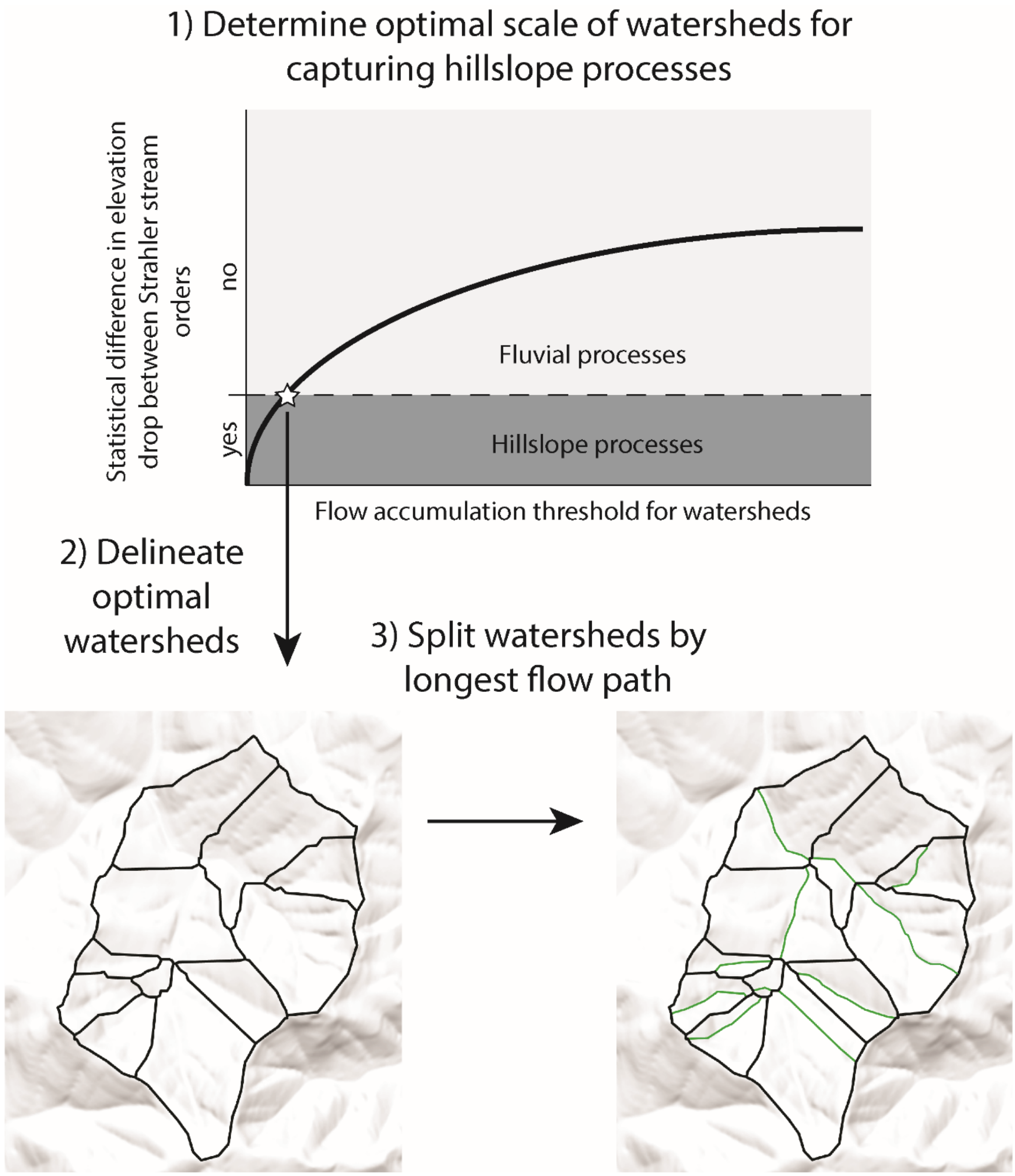


Figure S1. Illustration of the slope unit delineation method. The algorithm first determines the optimal scale (flow accumulation threshold) for capturing hillslope processes using the constant drop law. It then delineates and splits the watersheds by their longest flow paths (green) to create slope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

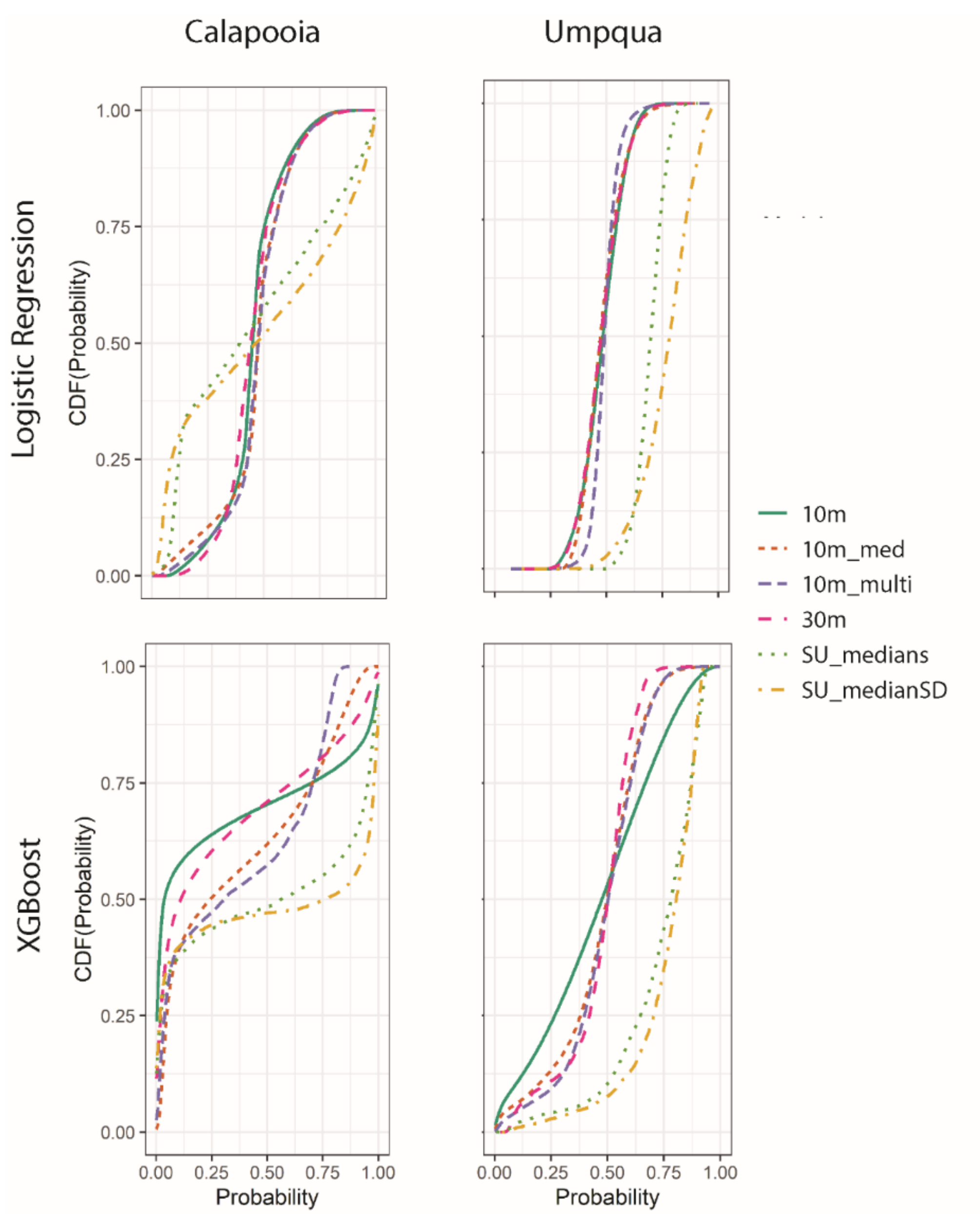


Figure S2. Cumulative distribution functions (CDF) of the Umpqua and Calapooia susceptibility model probabilities. The CDF is the probability that susceptibility model probability distribution function will take a value less than or equal to the value of the x-axis. The different lines represent the 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.

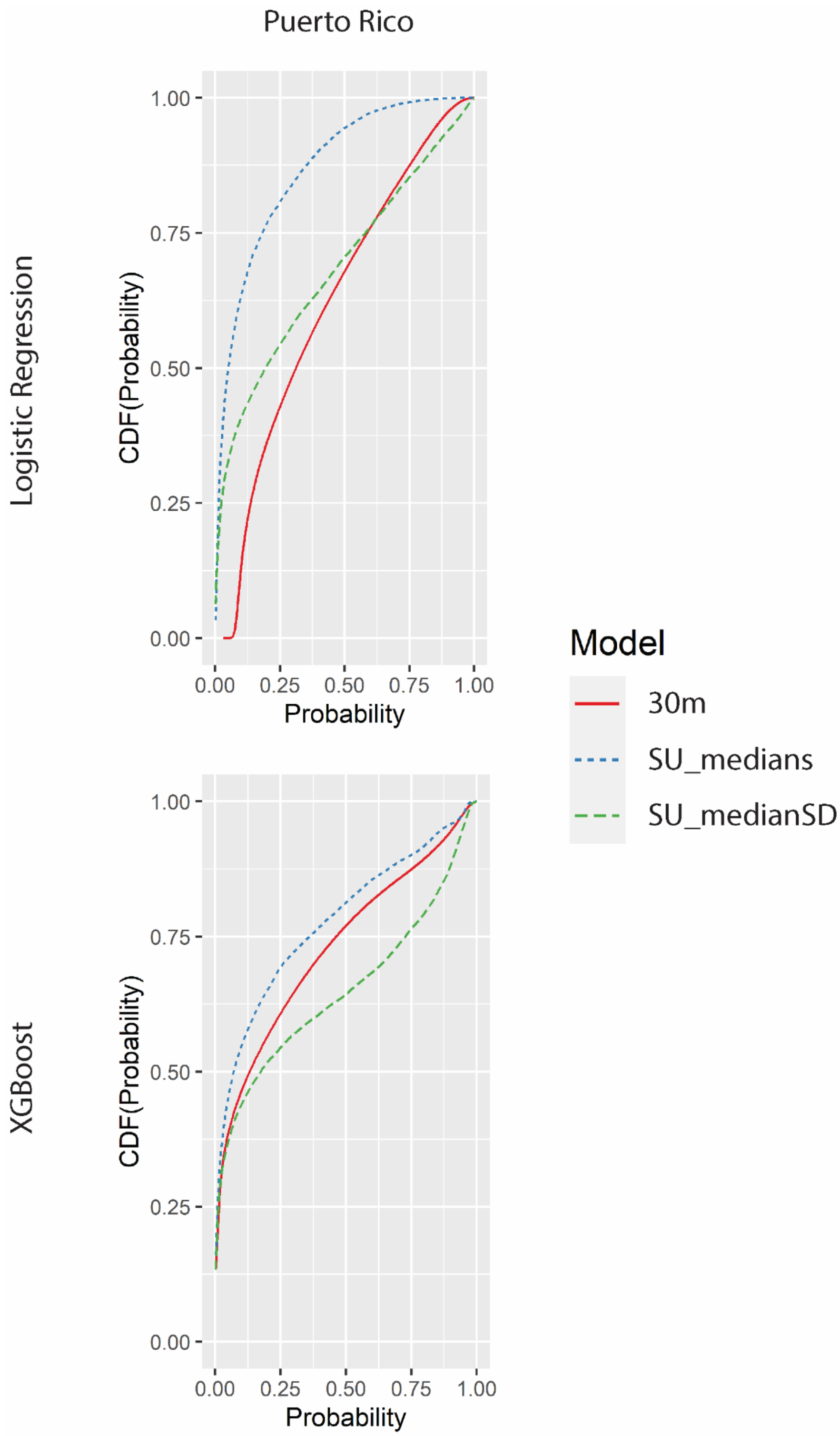


Figure S3. Cumulative distribution functions (CDF) of the Puerto Rico susceptibility model
85 probabilities. The CDF is the probability that susceptibility model probability distribution function will take a value less than or equal to the value of the x-axis. The different lines represent the different sampling techniques and mapping units used to make the susceptibility maps. See caption to Figure S2 for details on the sampling techniques.

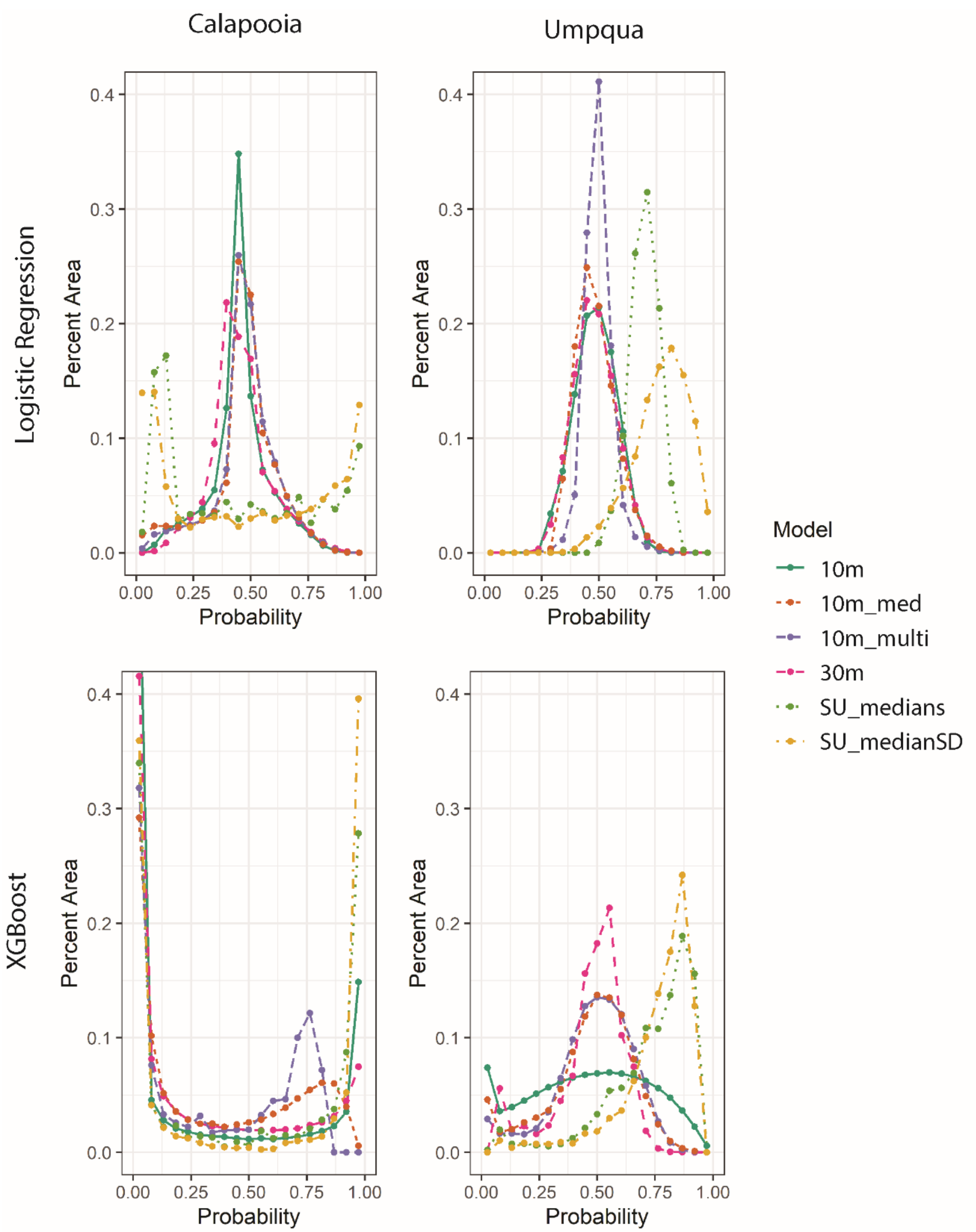


Figure S4. Percent area as a function of probability for the Umpqua and Calappooia susceptibility models. Plots were created using a 20 bin histogram and converting the counts to percent area. Points show the midpoints of each bin. The different lines represent the different sampling techniques and mapping units used to make the susceptibility maps. See caption to Figure S2 for details on the sampling techniques.

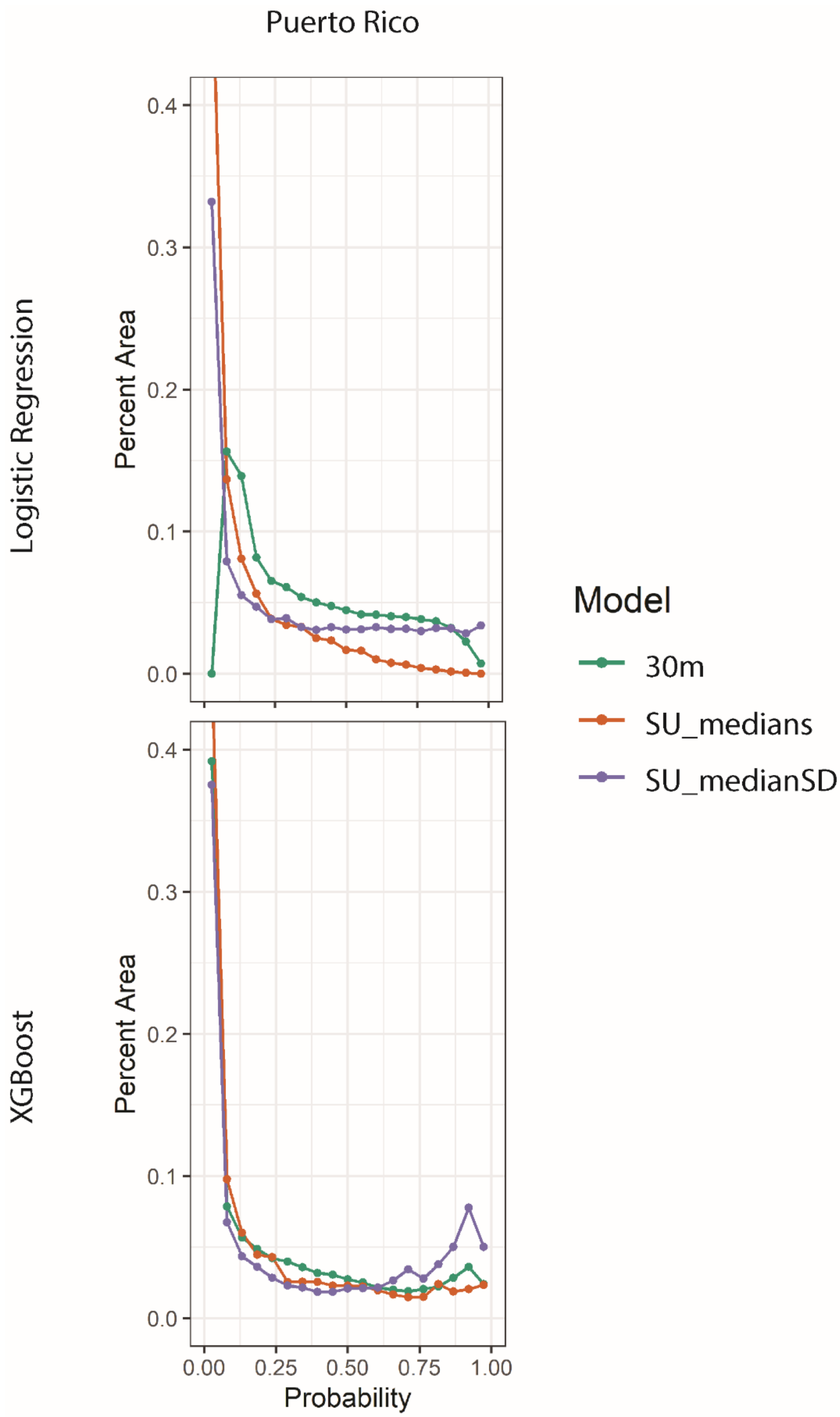


Figure S5. Percent area as a function of probability for the Puerto Rico susceptibility models. Plots were created using a 20 bin histogram and converting the counts to percent area. Points show the midpoints of each bin. The different lines represent the different sampling techniques and mapping units used to make the susceptibility maps. See caption to Figure S2 for details on the sampling techniques.

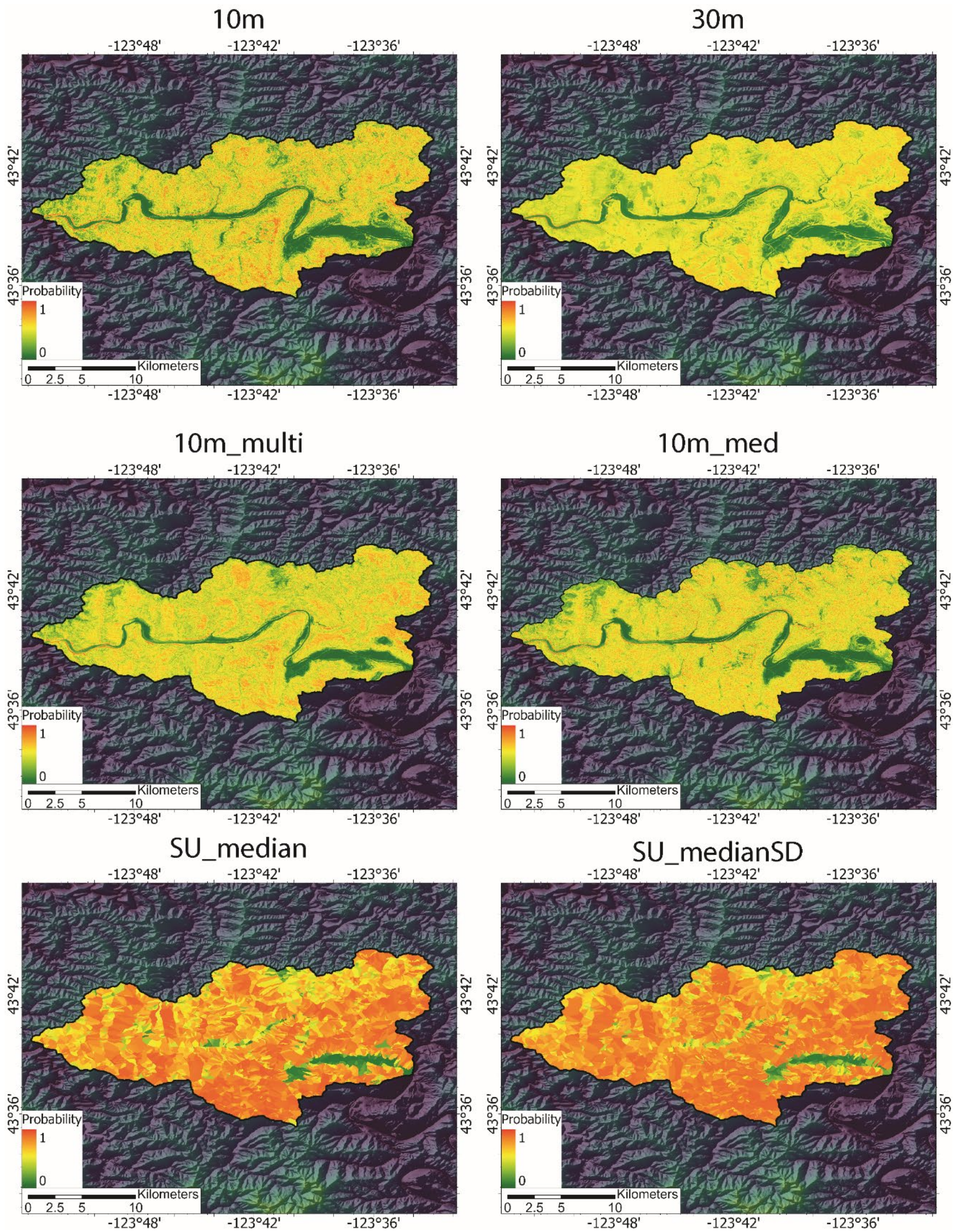
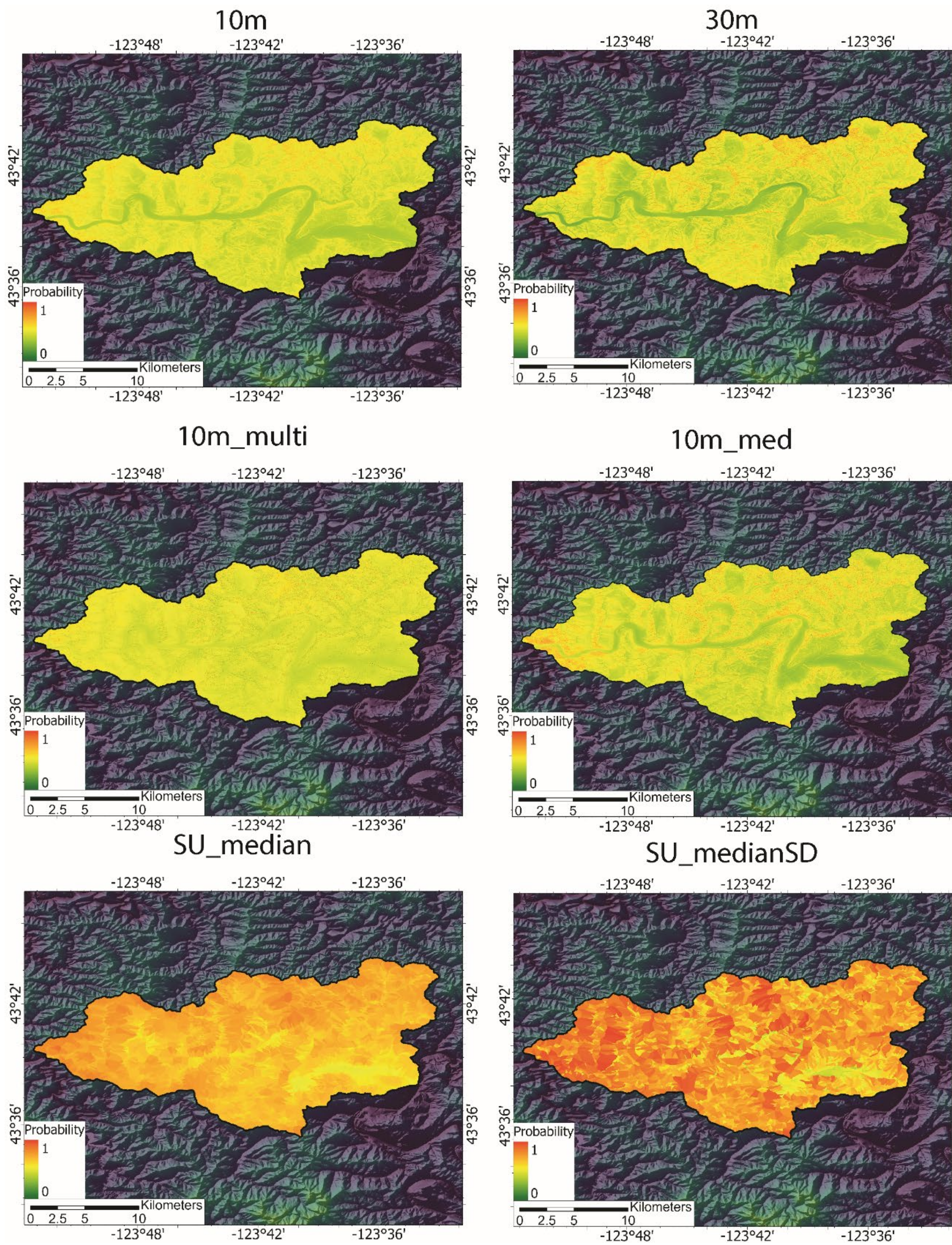


Figure S6. XGBoost susceptibility models over the Umpqua watershed. Maps are for the different sampling methods (10m, 30m, 10m_multi) and the slope unit maps using only the median (SU_median) and the median and standard deviation of the predictor values (SU_medianSD).



110 **Figure S7**. 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).

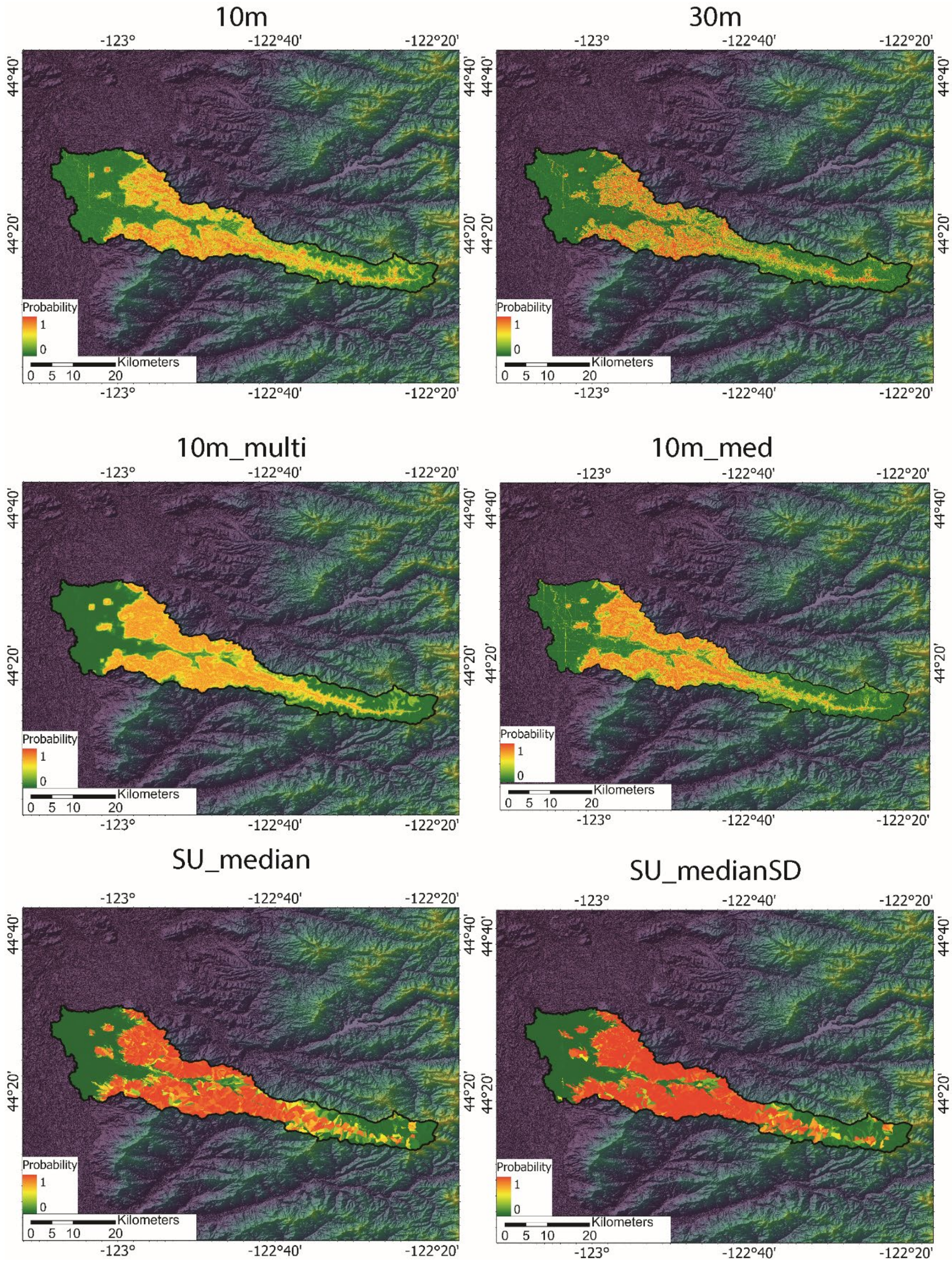


Figure S8. XGBoost susceptibility models over the Calapooia watershed. Maps are for the different sampling methods (10m, 30m, 10m_multi) and the slope unit maps using only the median (SU_median) and the median and standard deviation of the predictor values (SU_medianSD).

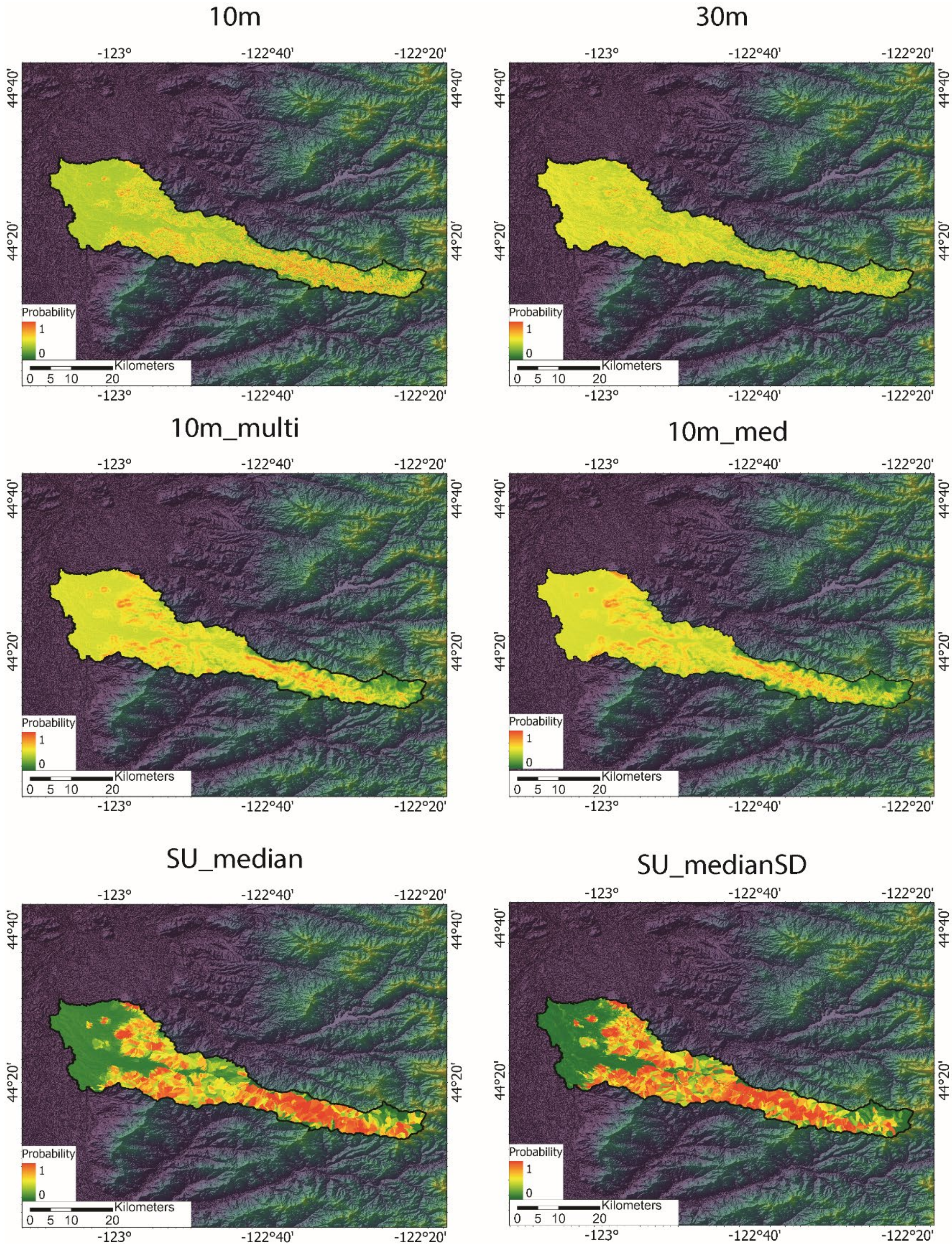


Figure S9. Logistic regression susceptibility models over the Calapooia watershed. Maps are for the different sampling methods (10m, 30m, 10m_multi) and the slope unit maps using only the median (SU_median) and the median and standard deviation of the predictor values (SU_medianSD).

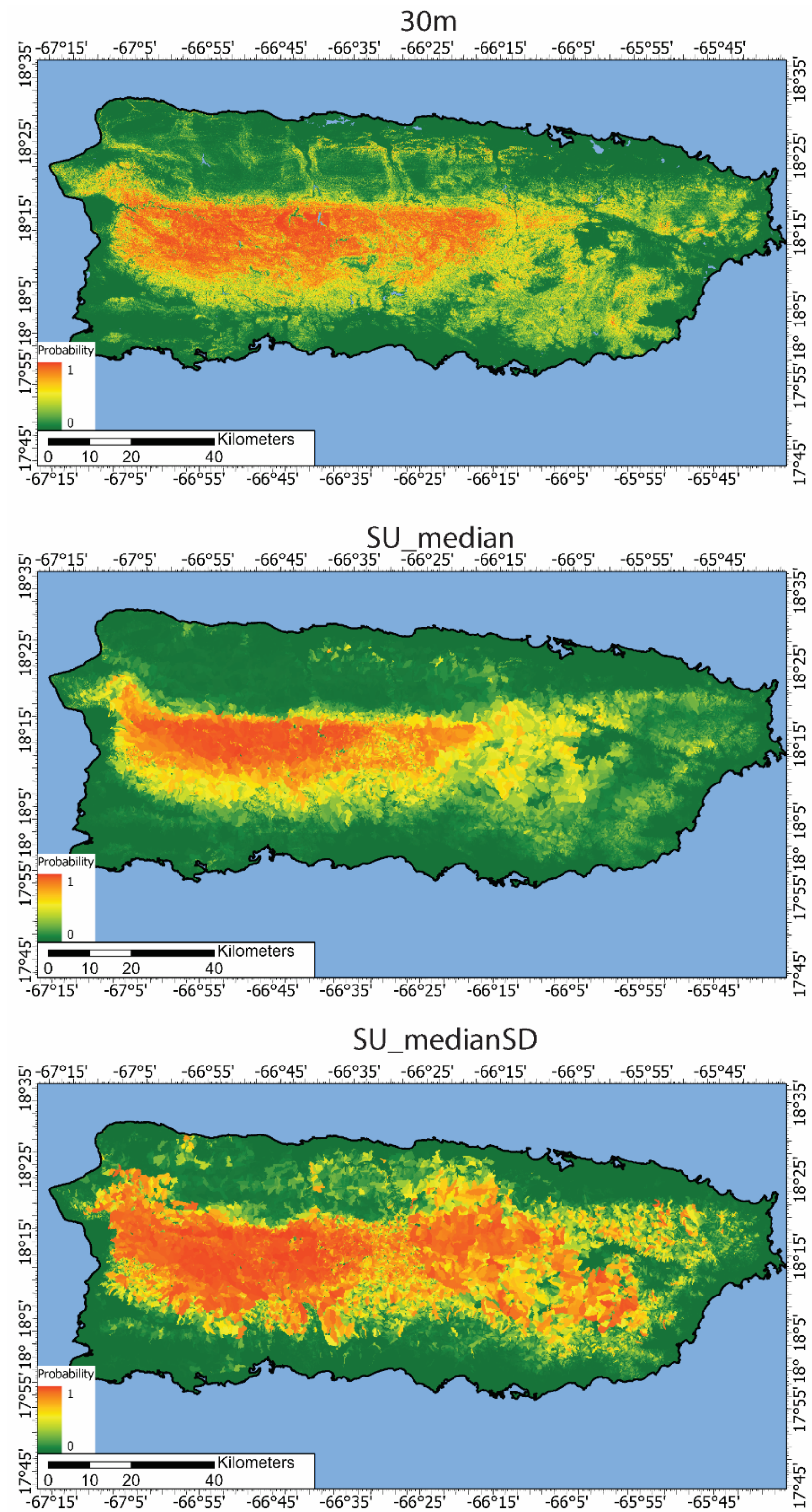


Figure S10. XGBoost susceptibility models over Puerto Rico for the Hurricane Maria landslide dataset. Maps are for the 30m grid-based model and the slope unit maps using only the median (SU_median) and the median and standard deviation of the predictor values (SU_medianSD).

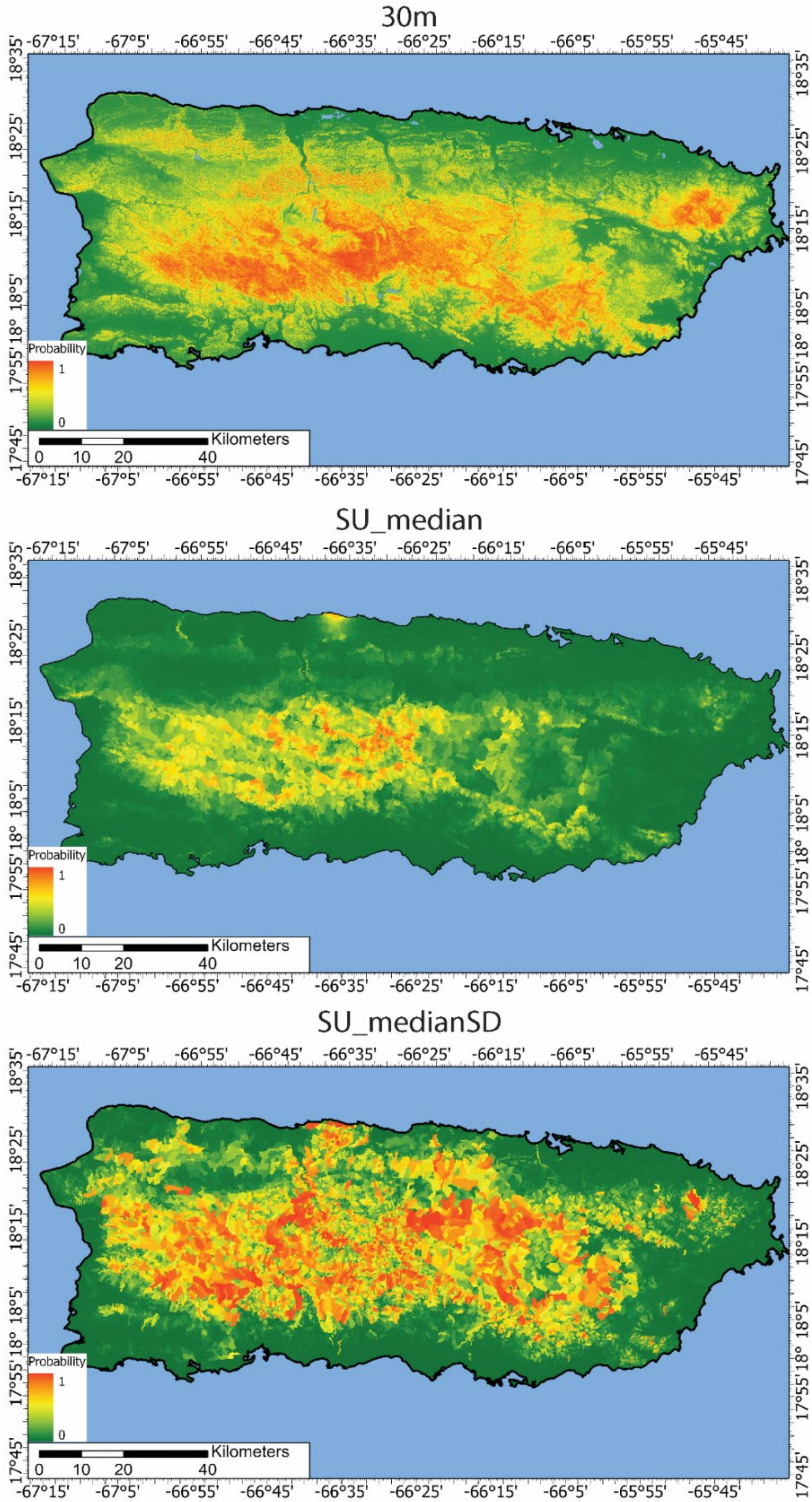


Figure S11. Logistic regression susceptibility models over Puerto Rico for the Hurricane Maria landslide dataset. Maps are for the 30m grid-based model and the slope unit maps using only the median (SU_median) and the median and standard deviation of the predictor values (SU_medianSD).

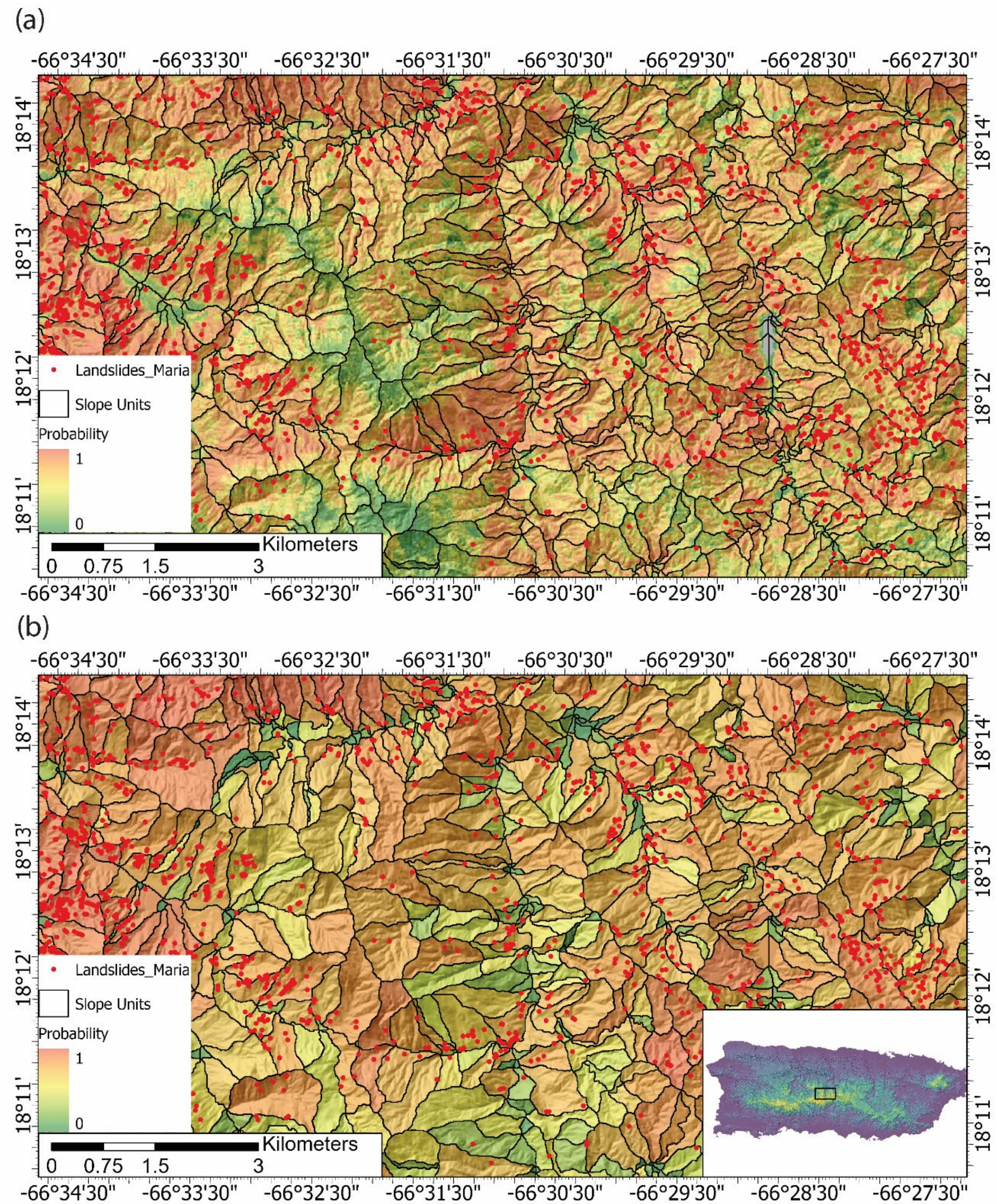


Figure S12. Zoomed in portion of Puerto Rico landslide susceptibility models from the (a) 30m grid-based maps and (b) using slope units with median and standard deviation predictor values with XGBoost. Red dots show the location of mapped landslide scarps. Inset map shows maps extent.

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

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