Arizona Random Forest Flood Mapping

Travis Zalesky[1]

[1]University of Arizona,

Corresponding author: Travis Zalesky, travisz@arizona.edu

Abstract

Federal Emergency Management Administration 100-year flood risk maps are expanded across the state of Arizona using a random forest, machine learning classification utilizing eight topographic explanatory variables.

Plain Language Summary

Flood mapping across Arizona.

1 Background

A critical component of the Arizona Department of Water Resources (ADWR) Tri-University Recharge Project (TURP, a.k.a ATUR) is a state wide assessment of flooding potential. Initial efforts focused on a traditional suitability analysis approach, using the analytical hierarchy process (AHP) for multi-criterion decision making, largely based on the work by Aloui et al. (2024). These methods saw initial success, and are continuing to be developed and refined. However, it became apparent that there were a number of shortcomings inherent in this analysis which are not easily addressed.

Firstly, the results of such an analysis are intrinsically linked to the data layers used, and the weighting schema determined by the AHP. As additional data sets became available, and alternate weighting schemas were tested we generated multiple versions of mapped flood potential which did not necessarily agree with each other. In the absence of high quality ground-truthed data it was difficult to validate these results and it was not clear to the project team which version was the best. This underscores the need for expert involvement at every stage in these AHP based analysis. While there is a wealth of hydrological expertise within the larger ATUR project, method development and implementation has largely been conducted by a GIS technician with marginal hydrologic knowledge, and it has been difficult to foster sustained buy-in from team members on this portion of the project.

Furthermore, it was extremely difficult to develop a single generalized model that would be effective across the whole state. Because of the wide array of ecological and geologic conditions that are present across the state variables that are important for flood risk in one region may not apply in other regions. Lastly, even if these technical issues could be overcome, there was still gaps in the input data layers, resulting in unclassified regions.

While the traditional suitability analysis methods of assessing flood potential is still valuable to the project, and will be retained and developed further, the reality of these challenges lead us to reevaluate our overall approach and consider alternate methods. Work by Mudashiru et al. (2021) summarized the various methods used by other researchers in this field, which includes AHP based methods as well as physical modeling and machine learning applications. The machine learning methods utilized by Tehrany et al. (2019) appeared to be particularly relevant and applicable. In particular, their use of topographic data **only** was particularly intriguing. These data sets are fully calculable from a digital elevation model (DEM), which are readily available, easily accessible, and have full coverage over the study area. These findings lead to a renewed initiative to apply a machine learning based method towards the objective of a state wide flooding potential map.

2 Data & Methods

2.1 Topography Data

All explanatory variables for the model were derived from the NASA Shuttle Radar Topography Mission (SRTM) 30-m DEM. Slope, aspect, curvature, stream power index (SPI), topographic wetness index (TWI), and sediment transport index (STI) were all calculated in ArcGIS Pro (3.4.3). Slope, aspect and curvature were calculated using the Surface Parameters tool (Spatial Analyst). SPI, TWI, and STI were calculated as per Tehrany et al. (2019) using the Raster Calculator according to Equations 1-3

$$SPI = A_s * tan() \tag{1}$$

$$TWI = ln(A_s/tan()) \tag{2}$$

$$STI = (A_s/22.13)^{0.6} * (sin()/0.0896)^{1.3} \tag{3}$$

where A_s is catchment area (m) and ß is slope (radians).

Similarly, Topographic Roughness Index (TRI) was calculated as per Tehrany et al. (2019) using a custom R (4.4.1) function with the package terra (1.7-78) according to Equation 4

$$TRI = \left[\sum (\chi_{ij} - \chi_{00})^2\right]^{0.5} \tag{4}$$

where $_{ij}$ is the elevation at coordinates (i, j) and $_{00}$ is the elevation at coordinates (0, 0) for a 3x3 focal neighborhood. The code used to calculate TRI is available on GitHub.

2.2 Flooding Data

Flood data used for training the model was obtained from the Federal Emergency Management Administration (FEMA) National Flood Hazards Layer, which provides 100-year flood maps for many areas of the US. The data was manually downloaded for each county in AZ from the FEMA data viewer (accessed 3/15/2025). Data layers were merged in ArcGIS Pro (3.4.3), and the vector data was converted to a raster with a 10 m resolution. Additionally, the FEMA data was reclassified to a binary output, either flooded or not flooded (during a 100-year flood event), eliminating details such as survey methods and flow depth.

2.3 Variable Collinearity

The machine learning model was performed in Google Earth Engine (GEE). The SRTM elevation data was access and clipped to the study area natively through GEE servers, all other data layers, including the study area shapefile, were uploaded as an asset to GEE prior to model implementation.

Prior to modeling, the collinearity of the explanatory variables was explored using a series of pair-wise linear regression plots shown in Figures A1-A36. For collinearity analysis 5,000 points (the maximum number of points which can be plotted in GEE) were randomly sampled across the study area . The collinearity of each pair-wise regression is summarized visually in Figure 1 using the R-squared statistic of each comparison. While some relationships, e.g. slope and TWI, share a complex relationship that is not captured by a linear regression, the R-squared statistic is a simple indicator of collinearity which is readily understood. Although a formal variance inflation factor (VIF) analysis was not performed, efforts were made to limit model complexity by manually testing variable combinations, especially those that showed high degrees of collinearity, and at equivalent model accuracy, simpler models were preferred.

	Flood	Elevation	Slope	Aspect	Curveature	SPI	TWI	TRI
Elevation	3.70E-02							
Slope	2.00E-02	4.40E-02						
Aspect	8.83E-04	6.13E-06	1.47E-03					
Curveature	1.66E-06	3.82E-04	1.10E-02	2.28E-03				
SPI	1.67E-03	4.65E-05	7.41E-03	1.11E-05	3.41E-03			
TWI	1.10E-02	2.43E-03	1.51E-03	3.98E-03	1.82E-01	8.79E-03		
TRI	1.80E-02	3.80E-02	9.55E-01	1.31E-03	1.40E-02	9.78E-03	9.49E-04	
STI	1.91E-06	1.36E-05	1.70E-02	7.13E-05	7.81E-03	8.07E-01	8.07E-03	2.70E-02

Figure 1: Color coded R-squared statistic for each pair-wise linear regression (green = high, red = low), representing the collinearity of each variable used for modeling.

2.4 Initial Model Testing and Development

Many models were iteratively explored using several machine learning algorithms, various combinations of explanatory variables, and many hyperparameterization values. For all initial model trials the study area was reduced to the San Pedro watershed, a well characterized watershed with approx. 66% FEMA flood map coverage (visual estimate). A 70:30::training:testing data structure was adopted, and while a range of sampled points were tested, this ratio was maintained throughout. Model performance was primarily assessed through an overall accuracy score, with confusion matrix analysis performed for highly accurate models.

Tested models included Classification and Regression Tree (CART, a.k.a Decision Tree), Random Forest (RF), and Support Vector Machine (SVM). Generally, CART classification produced very noisy results which tended to overestimate flood waters, and averaged around 79.5% accuracy (data not shown). SVM classification was too computationally demanding, even within the smaller study area of the San Pedro, and with the generous cloud computing resources of GEE. As a consequence SVM classification can not be evaluated, other than to say that it is inefficient and implementation is impractical. RF classification proved to be the most promising method of classification, and the most effort was spent on developing that model.

2.4.1 Random Forest Model Development

Over 400 RF models were tested for the San Pedro watershed. Model optimization parameters tested included the number of trees, the number of sampling points (from 20,000 up to 60,000), and combinations of explanatory variables. Many RF models were tested simultaneously with between 5 to 100 trees using a custom GEE function modified from Nicolau et al. (2023). The referenced accuracy scores for preliminary models refers to the most accurate model, using the fewest number of trees. Tested RF models ranged from 73.9% to 87.4% (Table 1). The most accurate model tested used 35,000 sampling points, consisting of 30,000 dry land points (not flooded) and 5,000 flooded points and with all explanatory variables except for TRI, achieving an 87.4% overall accuracy.

Table 1: Random Forest algorithm optimization and accuracy. All recorded sampling points grouped together before being partitioned into a 70:30::training:testing structure.

Variables	Sampling Points (dry land, flooded)	Trees	Accuracy (%)
All	15000, 5000	50	79.2
All	20000, 5000	85	82

Table 2: The confusion for the most accurate random forest classifier, including overall, producer's and consumer's accuracy.

	Actual			
Predicted	**Dry**	**Flood**		
Dry	8882	1199	*0.881*	*Consumer's*
Flood	126	275		
Producer's	*0.986*		*0.874*	*Overall*

Variables	Sampling Points (dry land, flooded)	Trees	Accuracy (%)
All	20000, 10000	85	73.9
All	25000, 5000	50	84.8
All	30000, 5000	45	87
All	30000, 10000	80	78.9
All	40000, 10000	50	82.1
All	50000, 10000		Error
No TRI	25000, 5000	90	84.9
No slope	25000, 5000	70	84.8
No TRI	30000, 5000	70	87.4
No TRI or STI	30000, 5000	40	87.2
No STI	30000, 5000	??	87.1
No elev	30000, 5000	90	86.3
No slope	30000, 5000	80	87.2
No aspec	30000, 5000	60	87
No curve	30000, 5000	40	87.1
No SPI	30000, 5000	70	87.1
No TWI	30000, 5000	40	87.1

Confusion matrix analysis for this model showed much higher producers accuracy (98.6%) compared to consumers accuracy (88.1%; Table 2). While these results are still satisfactory, they reveal that the model is generally favoring dry land classification over flooded. This can easily be explained by the relative abundance of dry land pixels vs. flooded pixels, both in the sampling points and within the FEMA training image as a whole. Qualitatively, the model appeared to be somewhat overfit to stream channels (high SPI and STI), and while many of the larger flood plains were effectively captured, many features appeared too narrow. Additionally, the results were somewhat noisy, with noticeable speckling in both the dry and flooded regions.

3 Conclusion

References

Aloui, S., Zghibi, A., Mazzoni, A., Elomri, A., & Al-Ansari, T. (2024). *Identifying suitable zones for integrated aquifer recharge and flood control in arid qatar using GIS-based multi-criteria decision-making. 25*, 101137. https://doi.org/10.1016/j.gsd.2024.101137

Mudashiru, R. B., Sabtu, N., Abustan, I., & Balogun, W. (2021). Flood hazard mapping methods: A review. *Journal of Hydrology*, *603*, 126846. https://doi.org/10.1016/j.jhydrol.2021.126846

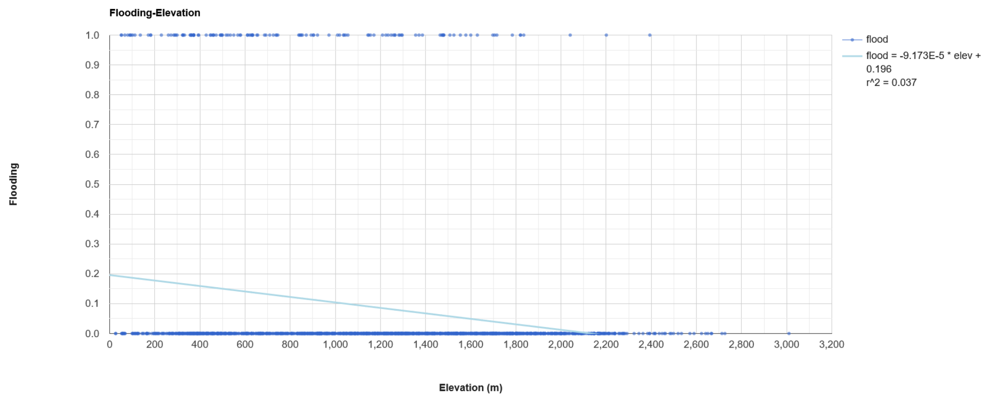
Nicolau, A. P., Dyson, K., Saah, D., & Clinton, N. (2023). Chapter F2.2: Accuracy assessment: Quantifying classification quality. In J. A. Cardille, N. Clinton, M. A. Crowley, & D. Saah (Eds.), *Cloud-based remote sensing with google earth engine: Fundamentals and applications* (1st ed.). Springer Cham. https://docs.google.com/document/d/1UCB900oCdJERca-2WUeDlCu52MjPKJxETJ_jJcLM0bM/edit?tab=t.0

Tehrany, M. S., Jones, S., & Shabani, F. (2019). Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques. *CATENA*, *175*, 174–192. https://doi.org/10.1016/j.catena.2018.12.011

4 Appendix

Code used to render these plots, as well as interactive versions of these plots are available on GEE.

Flooding



A 1: Linear regression analysis of flood risk (binary) and elevation (m) for 5,000 randomly sampled points across the full study area, encompassing Arizona.

Elevation

Slope

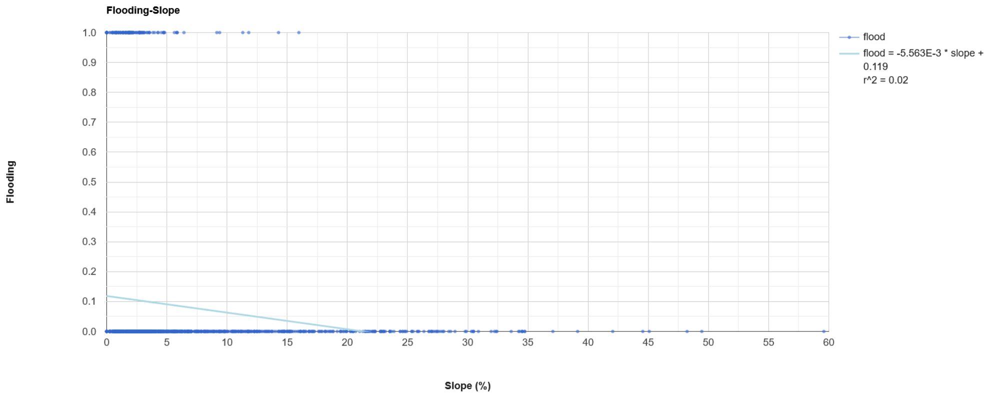
Aspect

Curvature

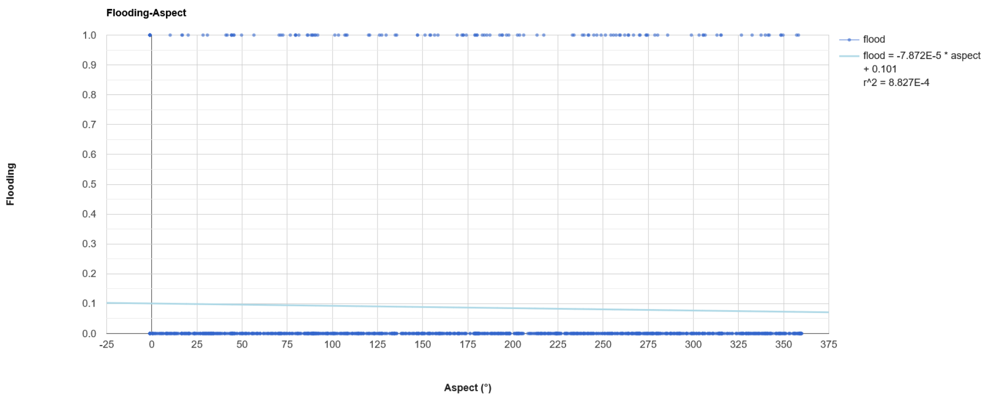
Stream Power Index

Topographic Wetness Index

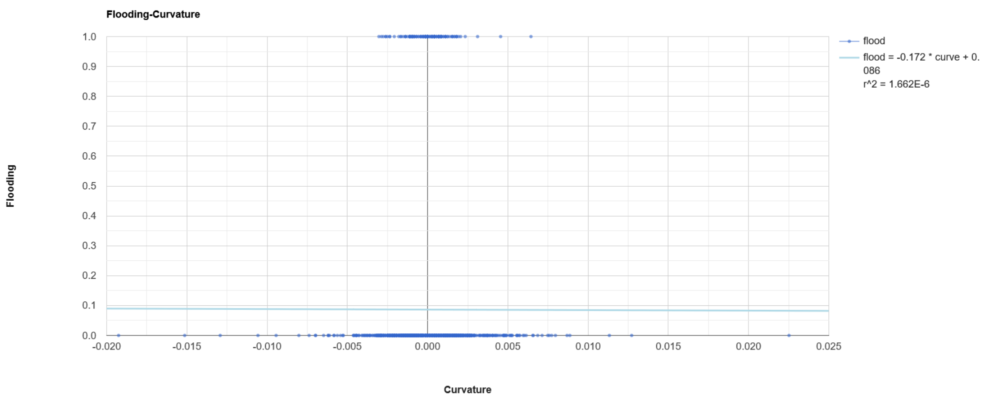
Topographic Roughness Index



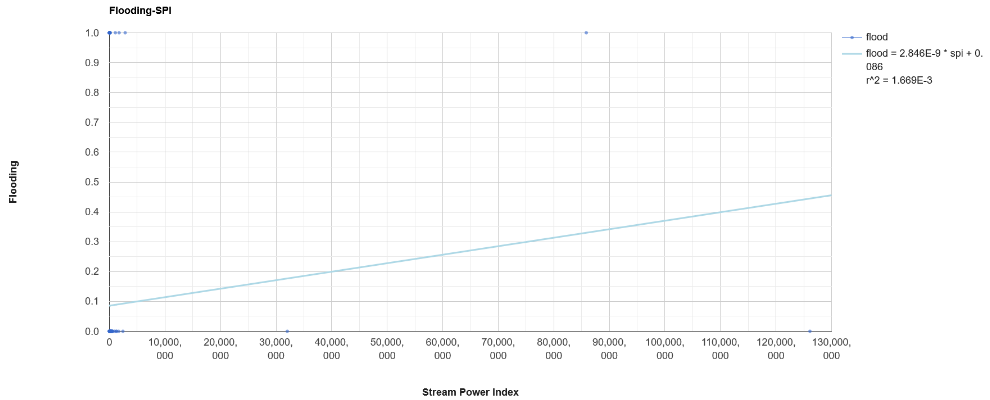
A 2: Linear regression analysis of flood risk (binary) and slope (°) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



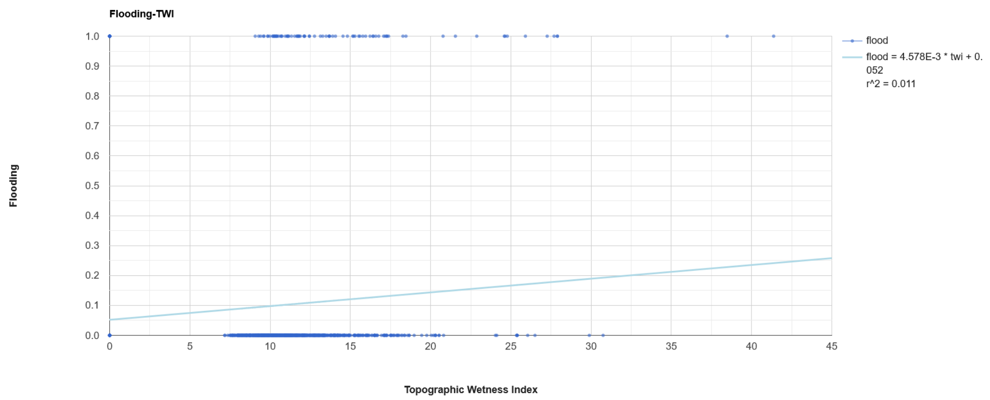
A 3: Linear regression analysis of flood risk (binary) and aspect (°) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



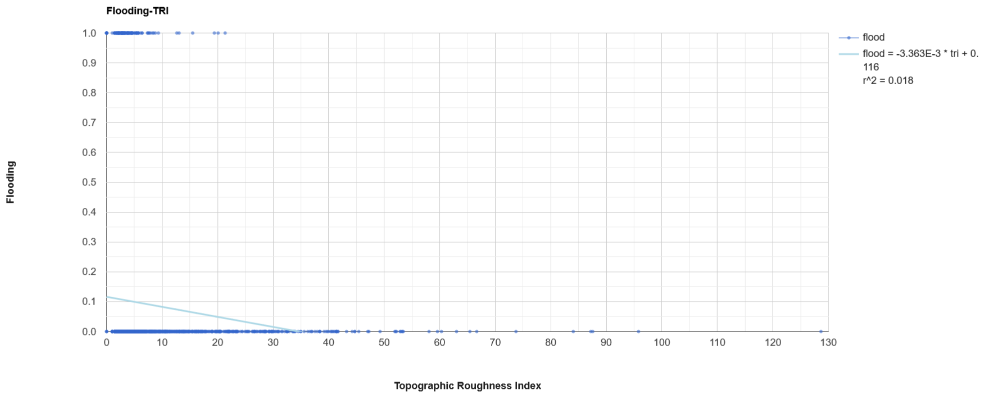
A 4: Linear regression analysis of flood risk (binary) and curvature for 5,000 randomly sampled points across the full study area, encompassing Arizona.



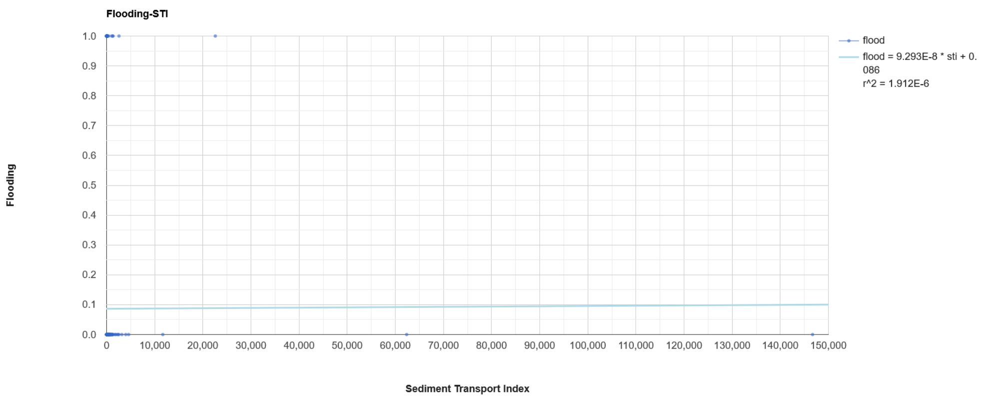
A 5: Linear regression analysis of flood risk (binary) and stream power index (SPI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



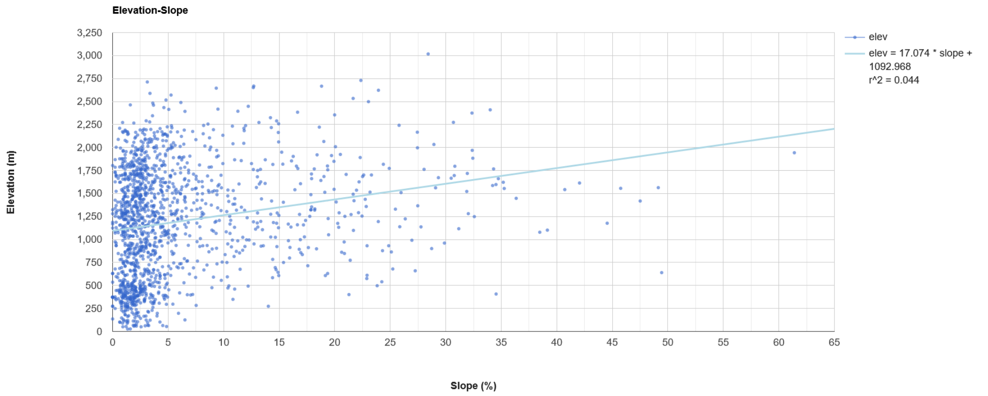
A 6: Linear regression analysis of flood risk (binary) and topographic wetness index (TWI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



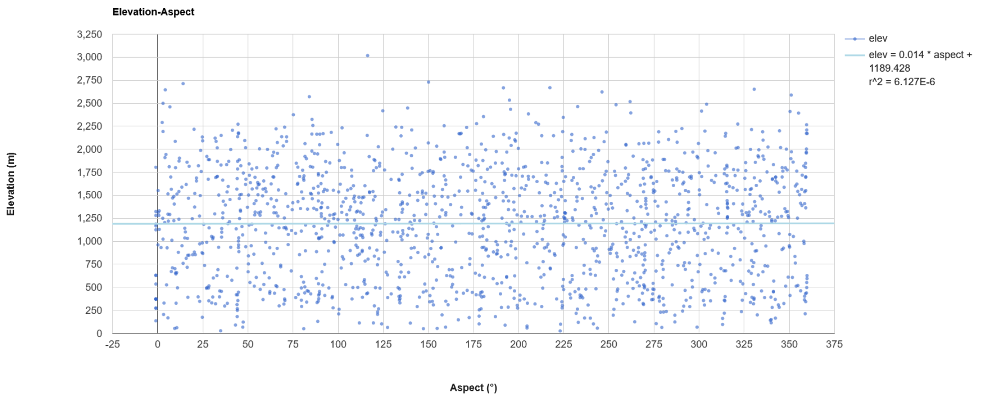
A 7: Linear regression analysis of flood risk (binary) and topographic roughness index (TRI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



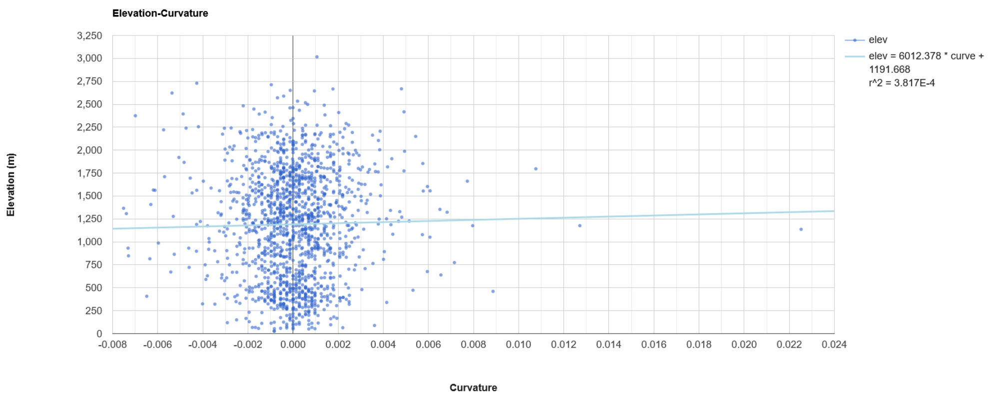
A 8: Linear regression analysis of flood risk (binary) and sediment transport index (STI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



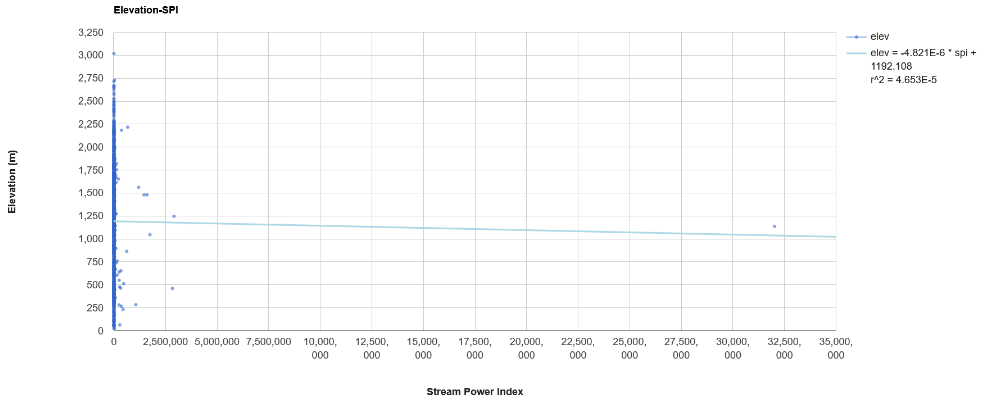
A 9: Linear regression analysis of elevation (m) and slope (°) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



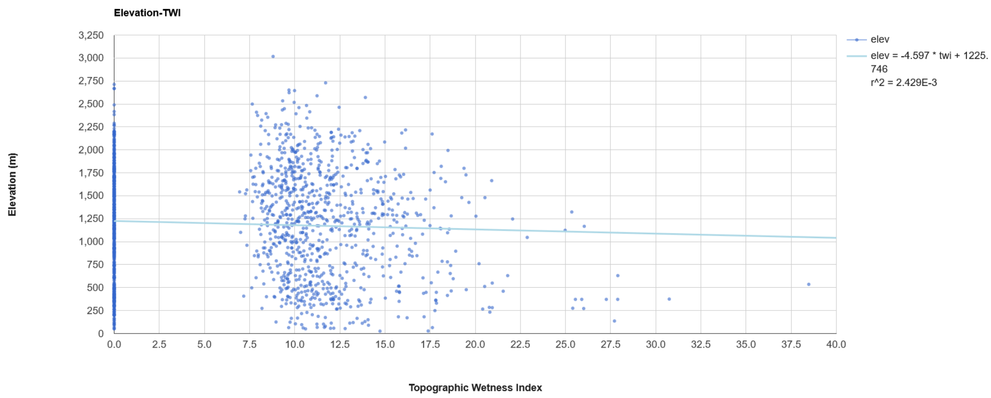
A 10: Linear regression analysis of elevation (m) and aspect (°) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



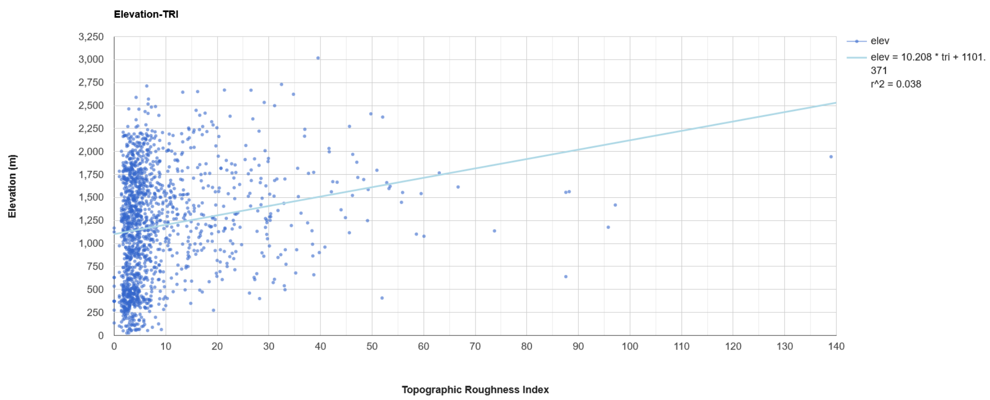
A 11: Linear regression analysis of elevation (m) and curvature for 5,000 randomly sampled points across the full study area, encompassing Arizona.



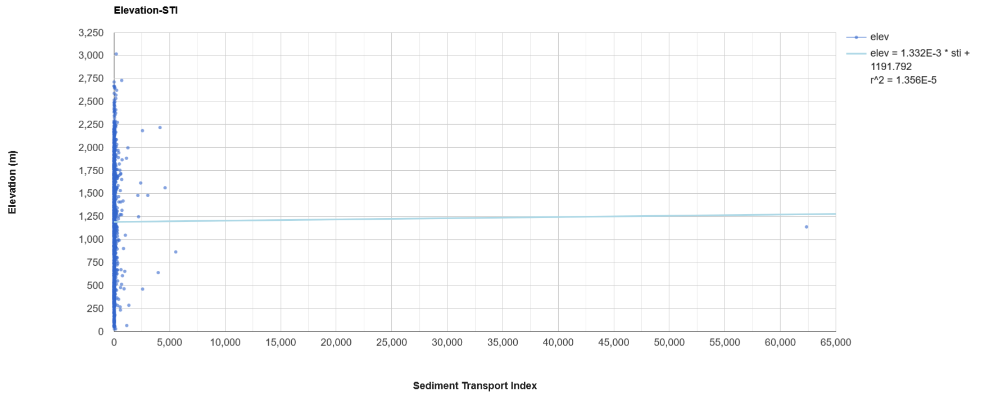
A 12: Linear regression analysis of elevation (m) and stream power index (SPI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



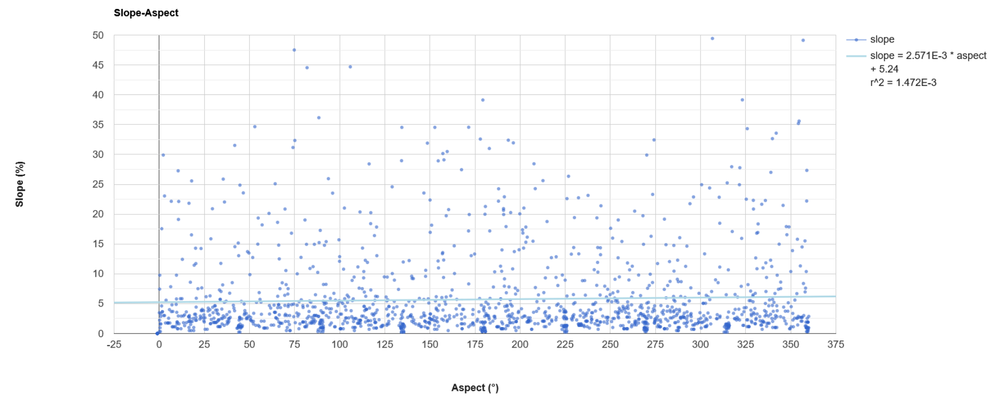
A 13: Linear regression analysis of elevation (m) and topographic wetness index (TWI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



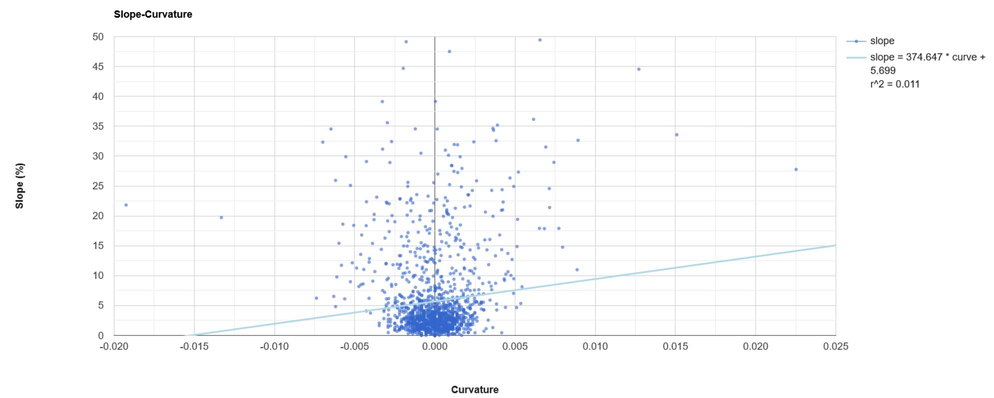
A 14: Linear regression analysis of elevation (m) and topographic roughness index (TRI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



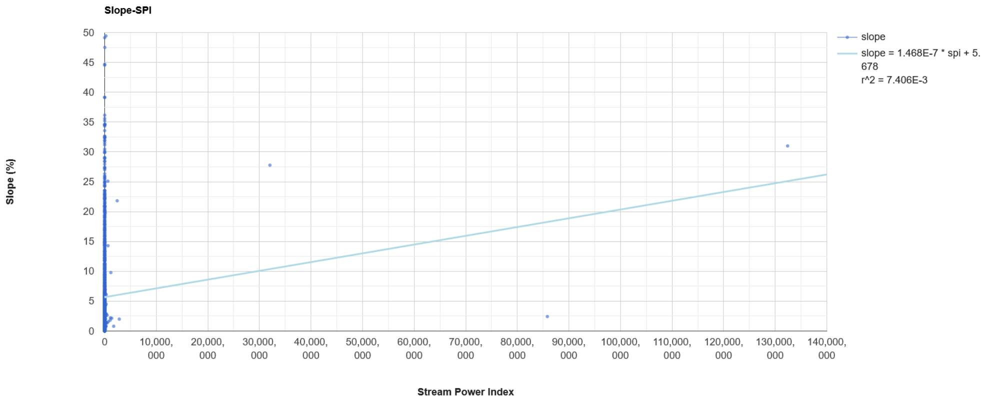
A 15: Linear regression analysis of elevation (m) and sediment transport index (STI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



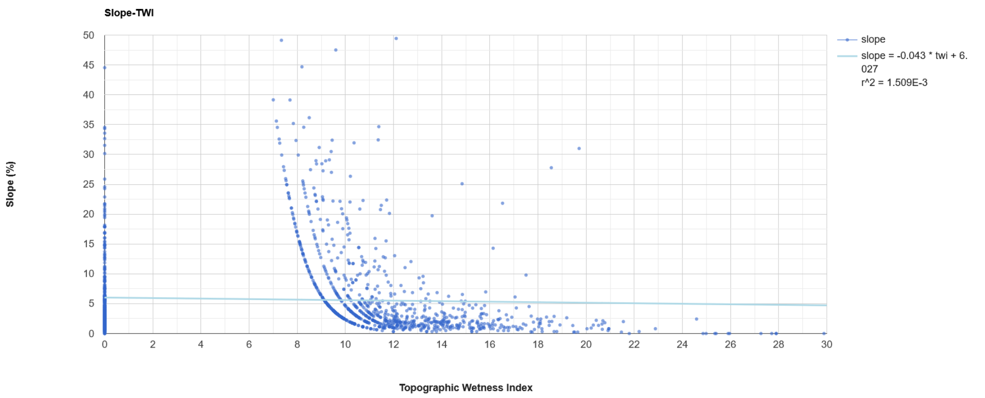
A 16: Linear regression analysis of slope (°) and aspect (°) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



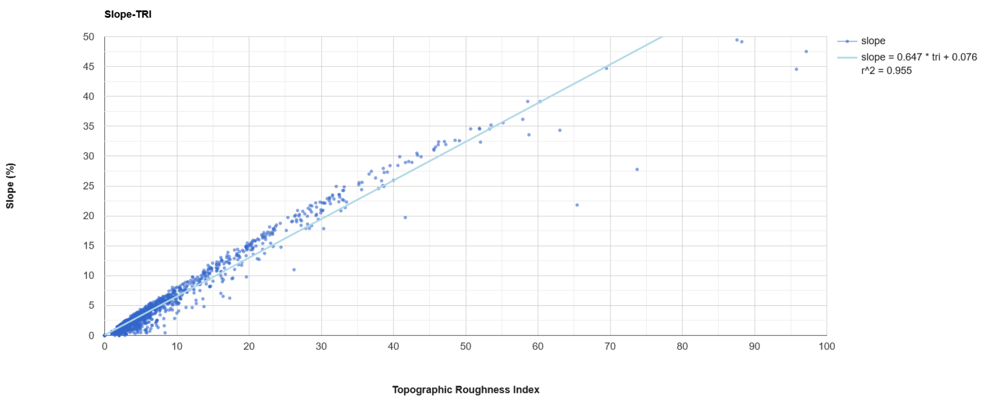
A 17: Linear regression analysis of slope (°) and curvature for 5,000 randomly sampled points across the full study area, encompassing Arizona.



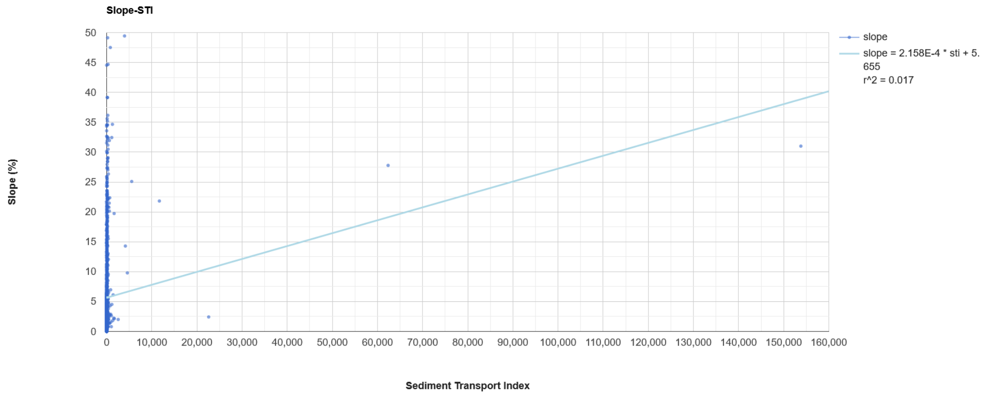
A 18: Linear regression analysis of slope (°) and stream power index (SPI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



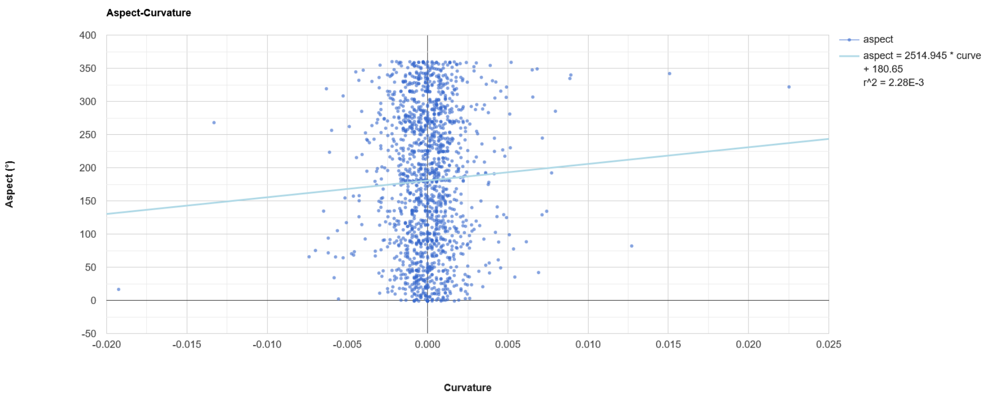
A 19: Linear regression analysis of slope (°) and topographic wetness index (TWI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



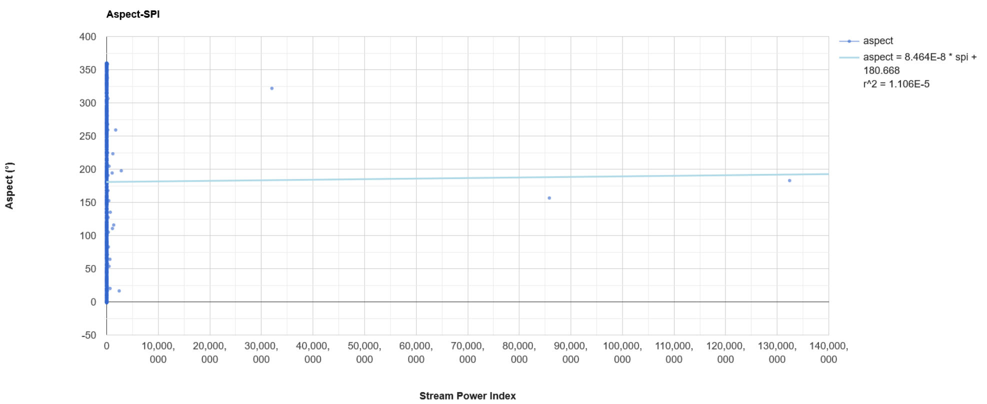
A 20: Linear regression analysis of slope (°) and topographic roughness index (TRI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



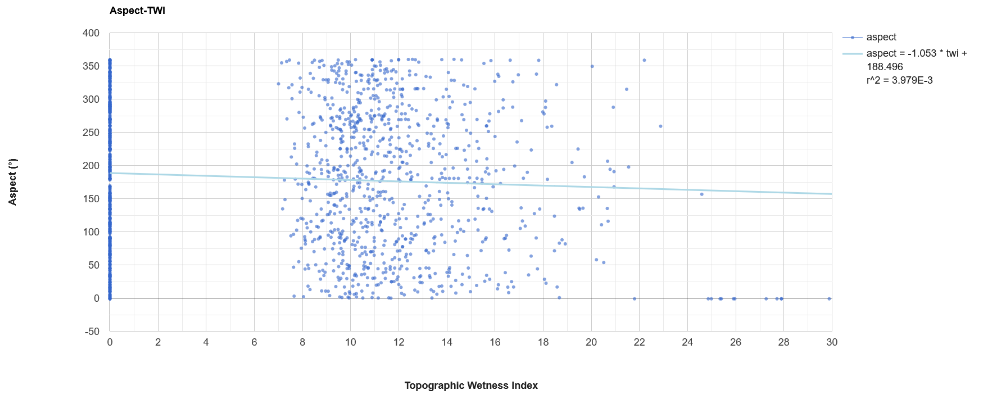
A 21: Linear regression analysis of slope (°) and sediment transport index (STI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



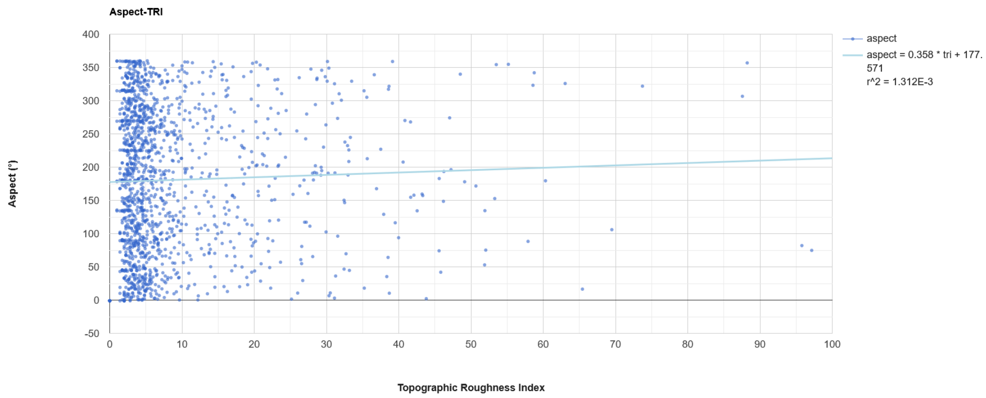
A 22: Linear regression analysis of aspect (°) and curvature for 5,000 randomly sampled points across the full study area, encompassing Arizona.



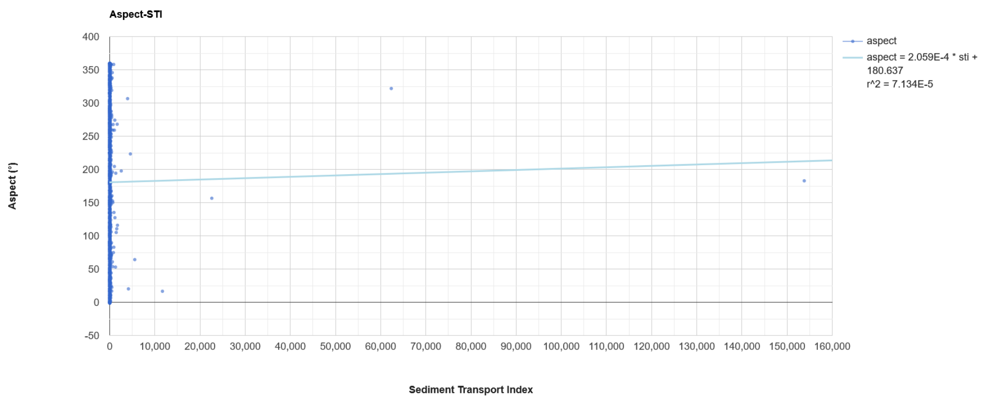
A 23: Linear regression analysis of aspect (°) and stream power index (SPI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



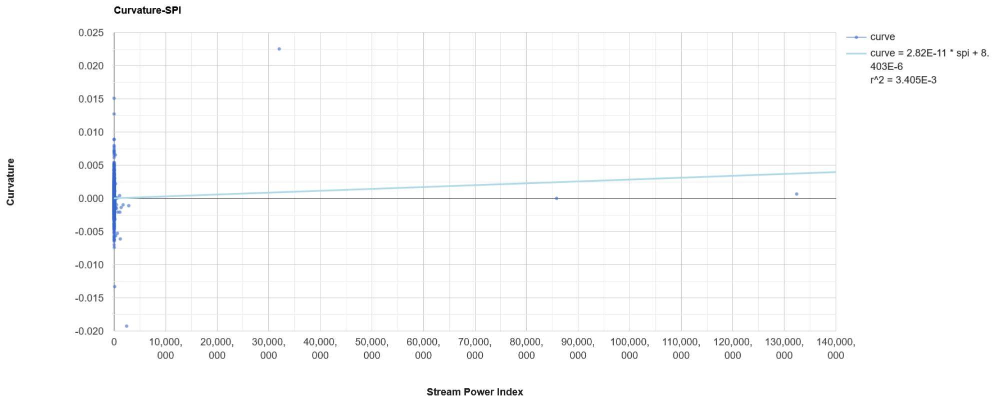
A 24: Linear regression analysis of aspect (°) and topographic wetness index (TWI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



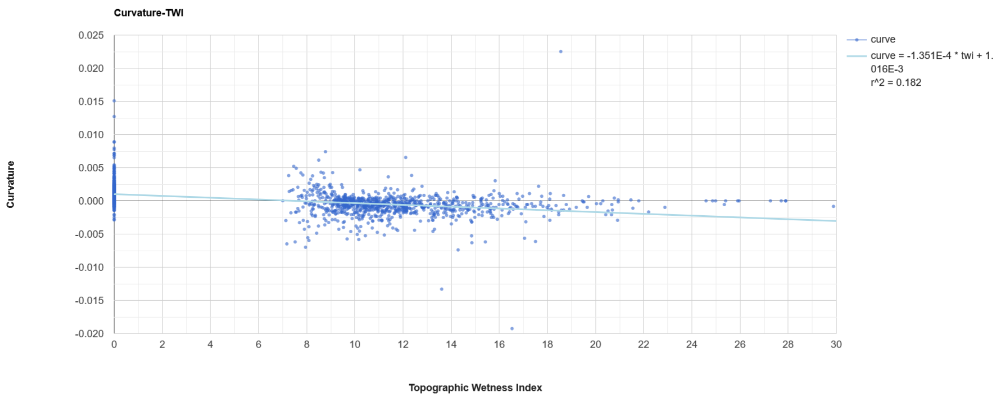
A 25: Linear regression analysis of aspect (°) and topographic roughness index (TRI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



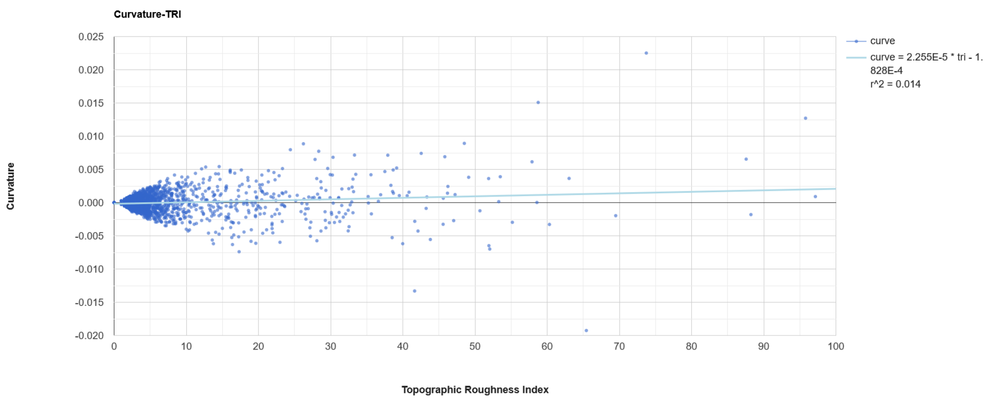
A 26: Linear regression analysis of aspect (°) and sediment transport index (STI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



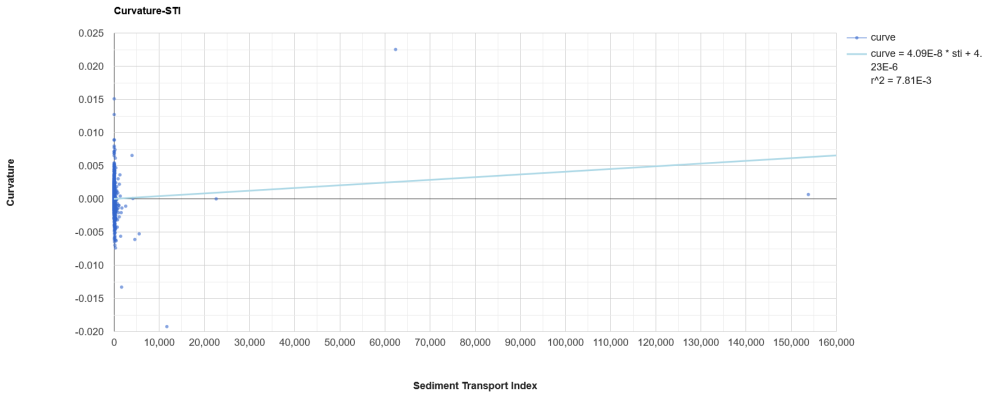
A 27: Linear regression analysis of curvature and stream power index (SPI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



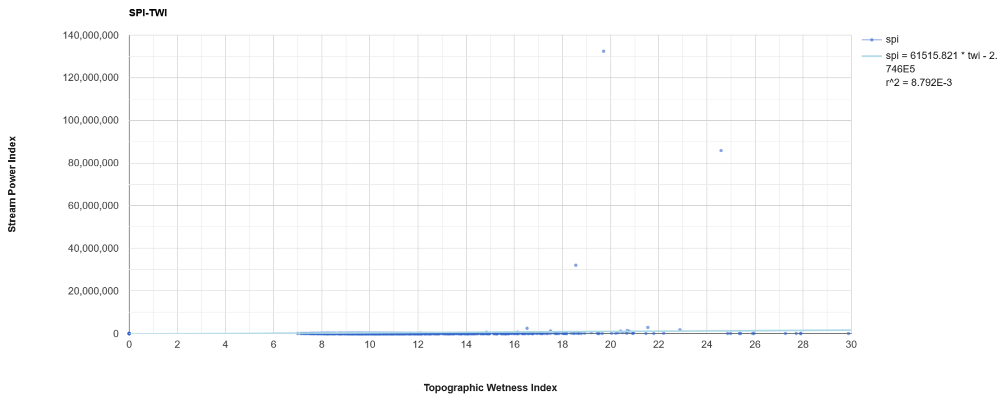
A 28: Linear regression analysis of curvature and topographic wetness index (TWI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



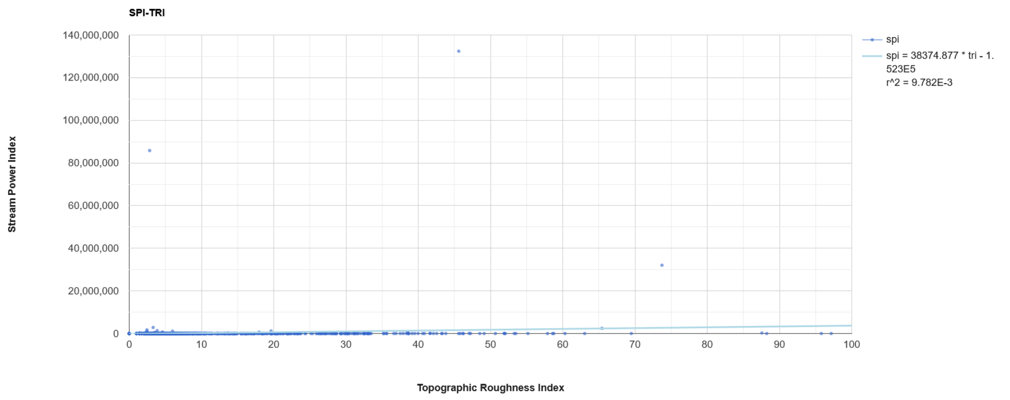
A 29: Linear regression analysis of curvature and topographic roughness index (TRI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



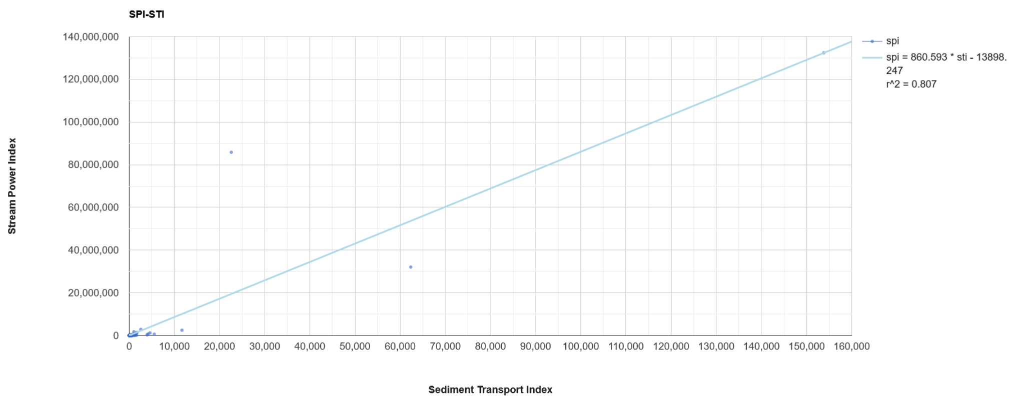
A 30: Linear regression analysis of curvature and sediment transport index (STI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



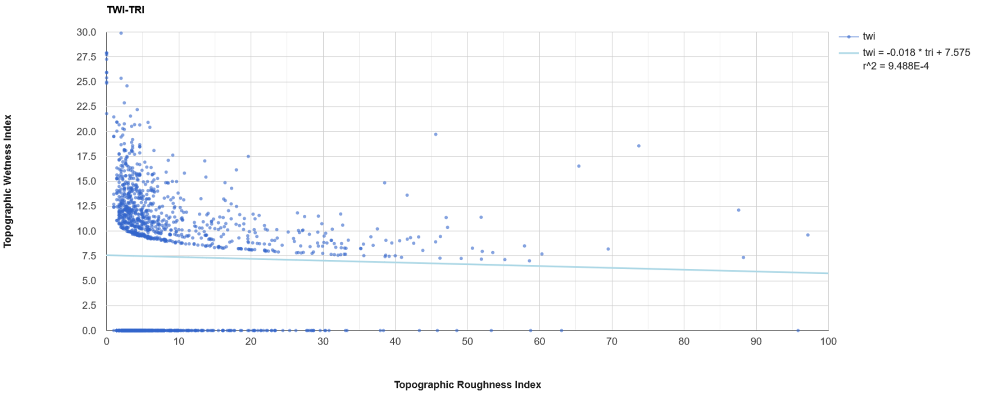
A 31: Linear regression analysis of stream power index (SPI) and topographic wetness index (TWI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



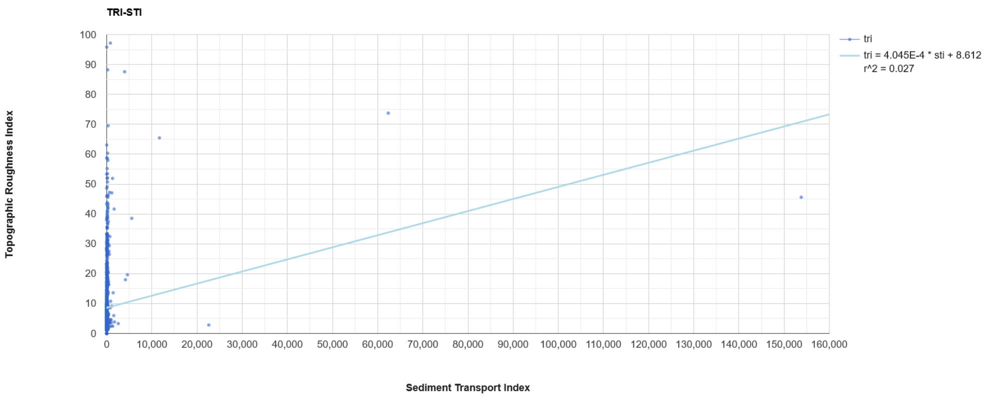
A 32: Linear regression analysis of stream power index (SPI) and topographic roughness index (TRI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



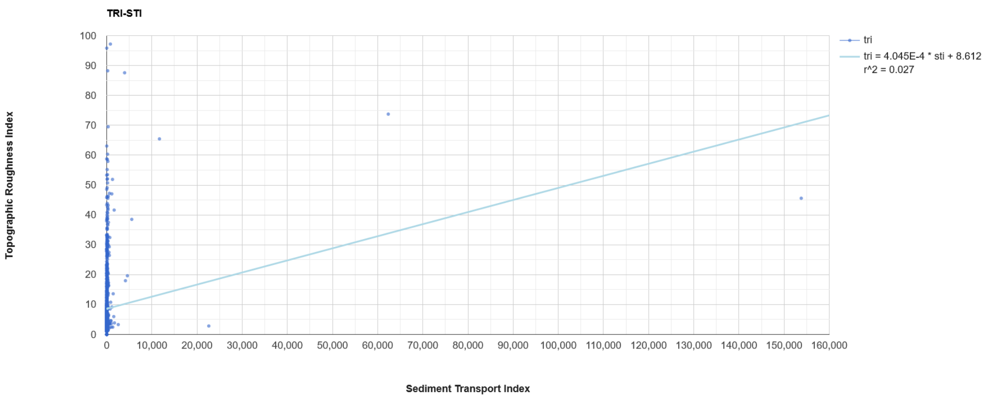
A 33: Linear regression analysis of stream power index (SPI) and sediment transport index (STI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



A 34: Linear regression analysis of topographic wetness index (TWI) and topographic roughness index (TRI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



A 35: Linear regression analysis of topographic wetness index (TRI) and sediment transport index (STI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.



A 36: Linear regression analysis of topographic roughness index (TRI) and sediment transport index (STI) for 5,000 randomly sampled points across the full study area, encompassing Arizona.