

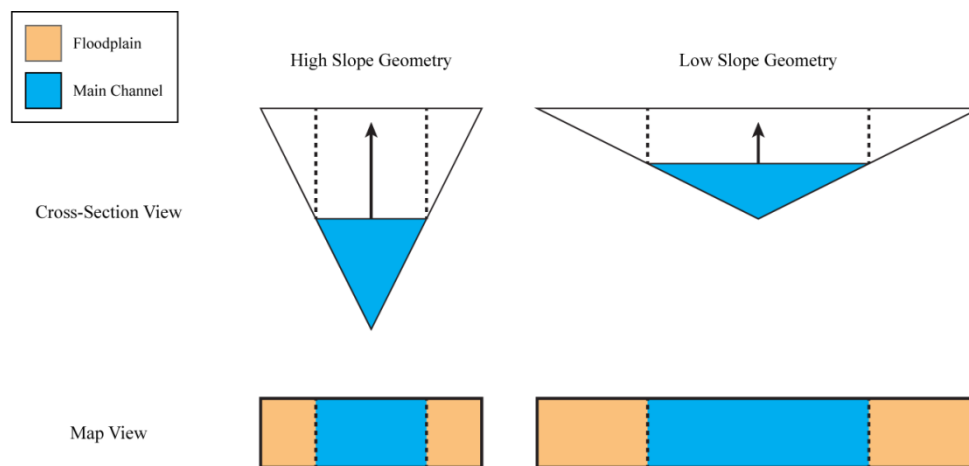
## *Using Topography for the Estimation of Floodplain Area*

### *Purpose*

Our understanding of natural systems and our computational power has improved to the point that the ability to map floodplains and forecast floods in the United States is now a reality. However not all nations have the resources capable of putting together such advanced programs. In fact, a majority of the planet doesn't even have topographic data at resolutions better than ten meters, which is critical to develop such informative flood hazard programs. However, the ability to develop floodplain maps, even at low-resolution scales, from simple topographic metrics could be extremely helpful in informing both the public and emergency personnel of what areas are of most concern during precipitation events. For example, if people knew that 90% of an area was likely located in the floodplain, they would hopefully choose not to construct infrastructure or homes there in the first place.

### *Objective*

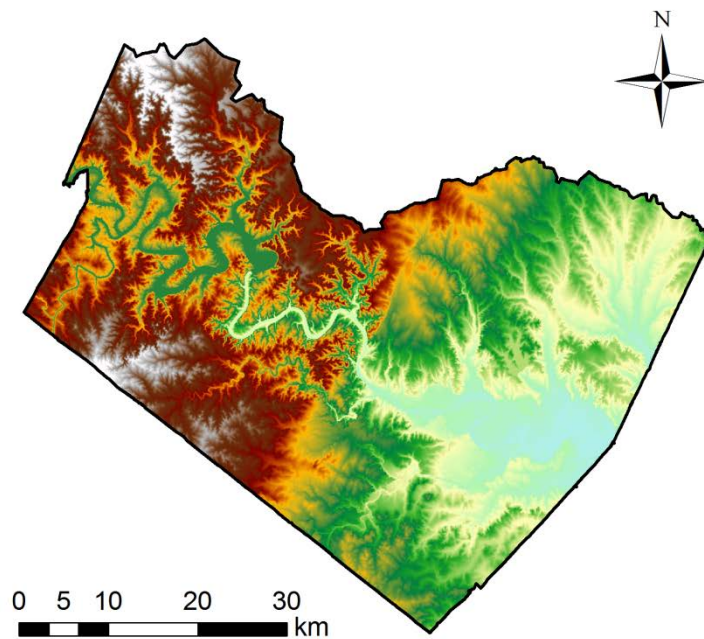
In the field of geomorphology, it has long been recognized that there are relationships between process and topography, and we frequently utilize models that rely upon these relations. The main reason for this is that field and geospatial analysis of topography is much easier to accomplish than direct measurements of geomorphic processes and rates. It seems reasonable that similar metrics could also be capable of predicting the area of floodplains. For instance, a simple observation of channel geometry demonstrates that for an equal increase in flow volume, there are distinct differences in flood stage height and inundation (Figure 1). In areas with steep relief, there is a greater change in stage height, but there is less lateral inundation. However for areas of low relief there is a lower change in stage height, but there is much greater area of lateral inundation. This suggests there should be an inverse scaling relationship between floodplain area and landscape slope.



**Figure 1.** Demonstration of the relationship between slope of adjacent hillslopes and floodplain area. Cross sections of channels have equal area.

Additionally, there are known relationships between drainage area and discharge – specifically discharge generally increases as a power law function of the drainage area. In very small drainage areas the ability to generate discharge from runoff becomes increasingly limited, so overbank flows are increasingly unlikely. Therefore, there should also be a direct scaling relationship between floodplain area and drainage area.

Starting from these basic observations, I hypothesize there is a quantifiable relationship between topography and flood inundation. Furthermore, slope and drainage area are metrics that can easily be calculated from topographic maps or digital elevation models (DEMs). I will test this hypothesis in Travis County in Central Texas (Figure 2). Travis County is an ideal location for a number of reasons: 1) the Balcones Escarpment divides the county into two topographically distinct but hydrologically connected systems (steep, highlands in the west and flat, coastal plain to the east), 2) FEMA flood maps are available for calibrating a topographically-derived model of floodplain areas, 3) flooding is the most impactful natural hazard in the county, and 4) there are hydrologic units of many scales, with the largest being the Colorado River.



**Figure 2.** Digital elevation map of Travis County from the USGS National Elevation Dataset. Of note is the near linear break in elevations running just off vertical through the center of the county. This represents the Balcones Escarpment or Fault Zone. This fault was last active 15 million years ago, but divides the county into the Hill Country to the west and the Coastal plain to the east.

### *Data*

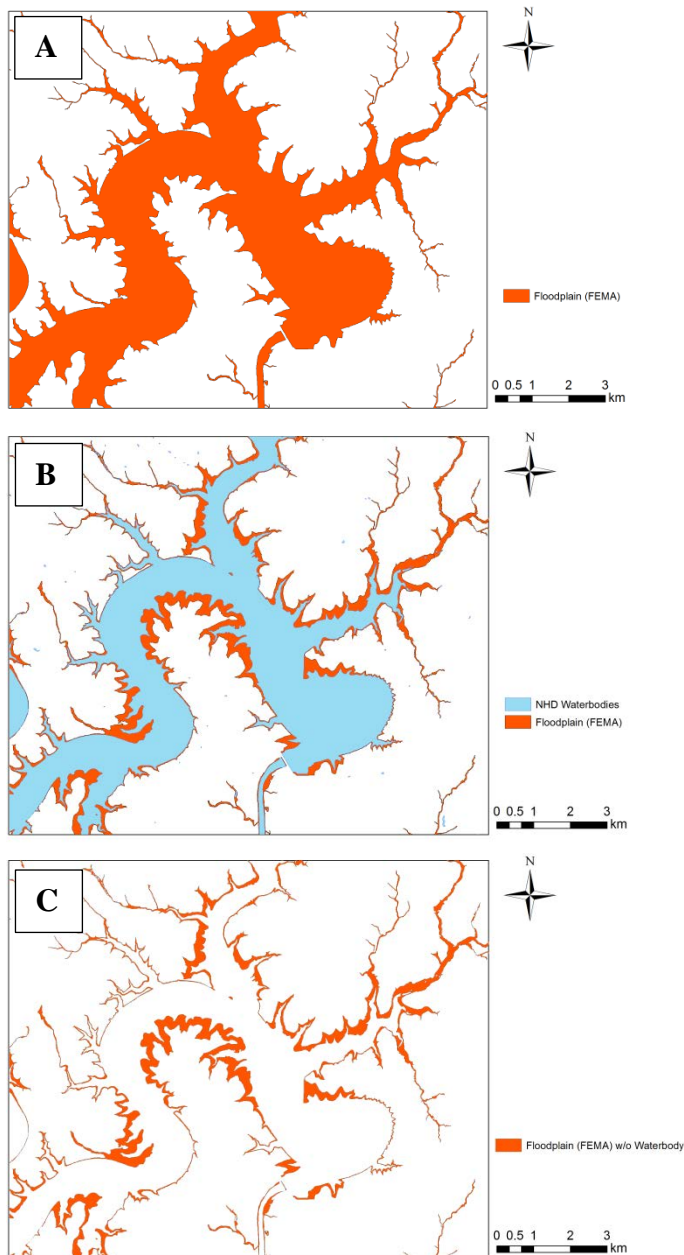
There are only four datasets used in this analysis. The primary dataset is 1/3 arc or approximately 10 meter digital elevation map (DEM) from the USGS National Elevation Dataset (NED). The second dataset is the floodplain polygon for Travis County from Federal Emergency Management Agency (FEMA). The third dataset is the waterbody polygon for Travis County from the National Hydrology Dataset (NHD). And the final dataset is a boundary polygon for the county obtained through ESRI's online map layers.

## Methods

### ArcGIS

#### Floodplain

The FEMA floodplain maps also include the area of waterbodies that are internally located within the floodplain, such as lakes and reservoirs. Since these waterbodies are always present, their inclusion does not accurately represent areas of land that will be inundated during a flood. First, the zone of minimal flood risk was removed from the floodplain polygon using the Query Builder. To select out the minimal risk zone, the query: “FLD\_Zone <> ‘X’ was applied. Exporting this layer then creates a new feature without those zones. The next step was to remove these waterbodies. Using the Erase tool, the waterbody polygon was deleted from the floodplain polygon (Figure 3).



**Figure 3.** A) FEMA floodplain map for the base flood in the Lake Travis area of western Travis County after the minimal risk zone is removed. Orange color designates the region within the base or 100 year floodplain. B) The same FEMA floodplain map with the NHD waterbody layer overlaid on it. This blue waterbody is Lake Travis, a reservoir along the Colorado River. It can be observed that the lake actually consumes a lot of the area actually designated as floodplain in the above map. C) The FEMA base flood map for Lake Travis after the waterbodies have been deleted. This layer provides a much more representative designation of land area that will be inundated during a flood. This is the layer used to calculate fractional area in the floodplain.

Calculating the fractional area of floodplain requires choosing an area to calculate over. The area of Travis County is 2,647 km<sup>2</sup>. In order to limit the number of data points for analysis, I chose to use a grid of 1 km<sup>2</sup> cell size. First, the floodplain was converted from a polygon to a raster. The environment settings used the Fill\_DEM to set the snap raster and the cell size. The value was left as the ObjectID, as this was only temporary. Next, the raster was converted to have binary values, where cells within the floodplain were given an assignment of 1 and outside the floodplain they were given a value of 0. This was done using the raster calculator with the equation: Con(IsNull("Floodplain Raster"),0,1). Averaging this binary raster over a larger grid will then provide a raster of fractional area in the floodplain. This was accomplished using the Aggregate tool, where the mean value was calculated for a 1 km<sup>2</sup> grid.

### *Drainage Area*

In order to calculate the drainage area, I first filled the depressions in the DEM. From this filled DEM, I then ran flow direction and flow accumulation from the hydrology toolbox. Using the raster calculator, the flow accumulation raster was multiplied by the cell area to produce a raster of drainage area.

In order to remove the waterbodies from the drainage area raster, a clipping mask had to be created. First, the Erase tool was used to delete the area of the waterbodies from the county boundary polygon. Then using the Extract by Mask tool, the waterbodies were removed from the drainage area raster.

Then using the Aggregate tool, the maximum value was found for a 1 km<sup>2</sup> grid. The maximum value was chosen, as this would be representative of the largest waterbody within that cell that should be the most significant contributor to flooding. The fractional floodplain raster was used as the snap raster for the aggregation.

### *Slope*

From the filled DEM, the slope tool from the Surface Toolbox was used to calculate local slope. With the same clipping mask used for drainage area, the waterbodies were removed from the slope raster using the Extract by Mask tool. Using the Aggregate tool, the mean value was then found for a 1 km<sup>2</sup> grid. The fractional floodplain raster was used as the snap raster for the aggregation.

### *Extracting Data*

Each of the above datasets (slope, drainage area and floodplain fractional area) were exported from ArcGIS using the Raster to ASCII tool. This created three CSV files that could be read into MATLAB.

## **MATLAB**

The dataset exported from ArcGIS was in matrix form with -9999 used for cells not within county boundaries and a six row header. The header was removed and the individual datasets were transformed to arrays that did not include the non-values and a record of their row and column position within the original matrix was created. Then the entire dataset was log transformed in order to better observe trends. This log transform was necessary because the drainage area spanned orders of magnitude. Linear regressions were run for the log-log plots of

drainage area and slope against fractional area of floodplain. Taking the slope from these two log-log regressions, the two relations were combined and a single fitting coefficient was then applied to try to best-fit the fractional floodplain data:

$$\log(F) = \beta + m \cdot \log(A) - n \cdot \log(S)$$

where  $F$  is the fractional area of floodplain,  $A$  is maximum drainage area,  $S$  is average slope,  $\beta$  is the fitting coefficient and  $m$  and  $n$  are the values of the slopes from the log-log regressions of drainage area and slope respectively. In log space this can be represented as:

$$K = 10^\beta$$

$$F = K \frac{A^m}{S^n}$$

where  $K$  is the fitting coefficient in linear space. The best-fit value of the fitting coefficient ( $\beta$ ) was determined by running an iterative loop that calculated the root mean squared error (RMSE) between the predicted and FEMA-derived floodplain areas for each applied fitting coefficient.

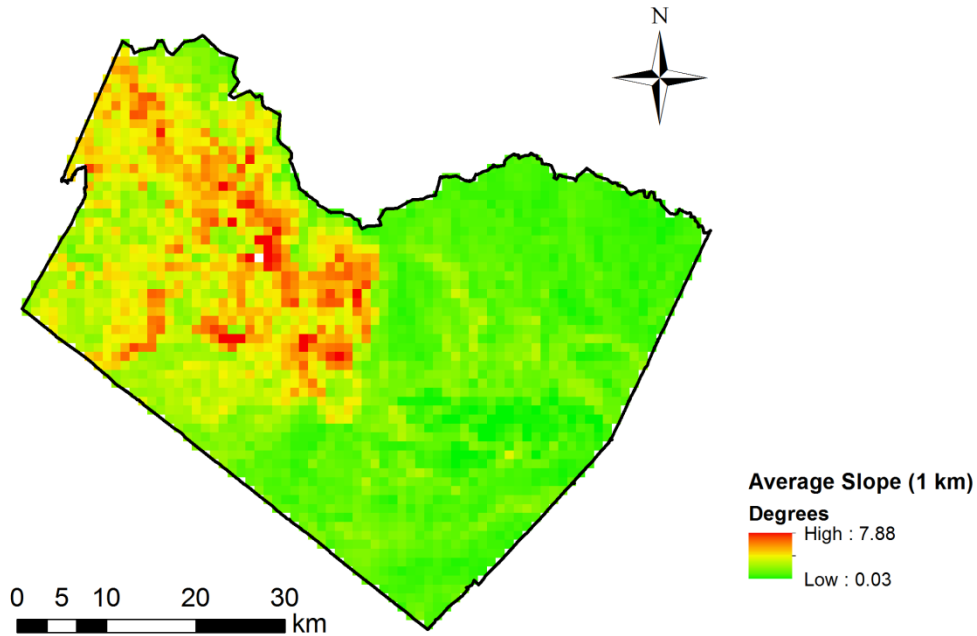
Using the record of original row and column position in the original matrix, a new matrix of the predicted floodplain values ( $F$ ) was created. All of the empty cells were replaced with a value of -9999, and the original header was replaced. This matrix was then written to a CSV file, which was then read back into ArcGIS using the ASCII to Raster tool.

## *Results*

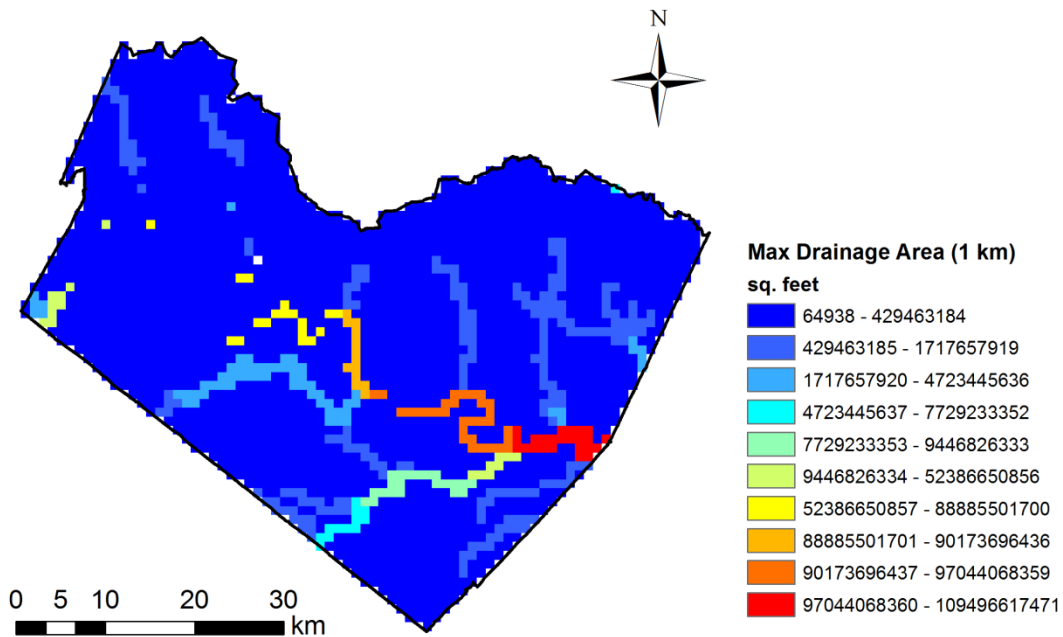
The first products from the analysis were the aggregated maps of slope, drainage area and FEMA-derived fractional floodplain area. As predicted based on the geologic history, the western half of the county is much steeper than the eastern half of the county (Figure 4). This clear delineation is due to the Balcones Escarpment. Uplift in the west has caused rivers to incise into the limestone landscape, creating a rockier, steeper terrain. While in the eastern coastal plain, slopes are very low and uniform.

Maximum drainage areas show predictable patterns (Figure 5). The Colorado River is the largest hydrologic body and flows from west to east, which can clearly be seen in the map. Other large tributaries, such as Onion Creek, can also be seen flowing into the Colorado River. The DEM was limited to Travis County, which clearly influences the result of the drainage area map. A number of tributaries in the south flow into adjacent counties, and so their complete drainage is not represented. Additionally convoluting the result is that the Colorado River flows much farther into West Texas, but is also heavily dammed in this area of Central Texas (e.g. Lake Travis seen earlier). To accurately represent this would require data on the maximum allowed releases of the individual dams.

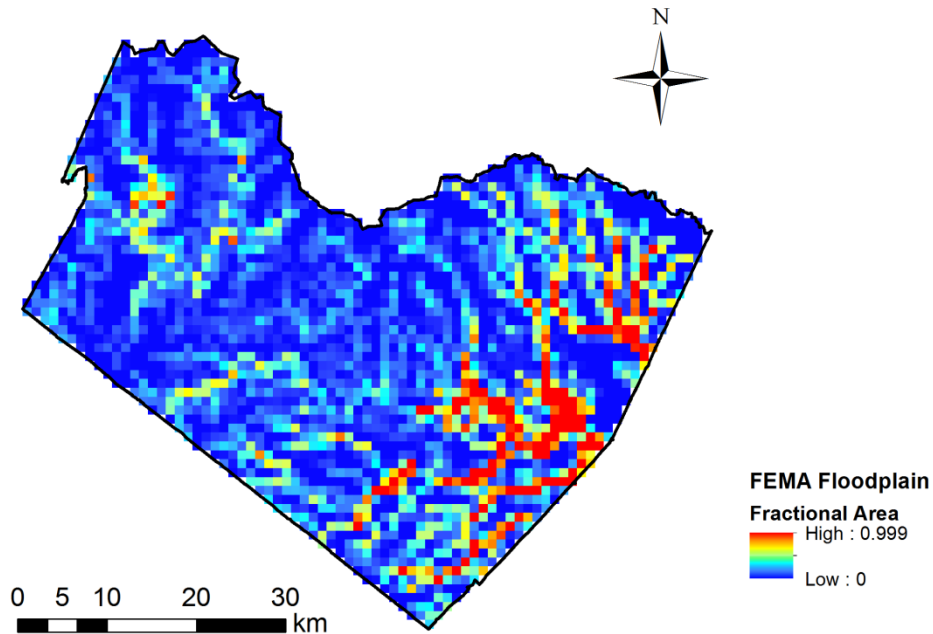
The aggregated FEMA-derived fractional floodplain map demonstrates tributaries in the eastern half of Travis County generally have greater floodplain area than in the west (Figure 6). Qualitatively this map demonstrates that the data is consistent with the hypothesis that floodplain area increases where slopes are lower and drainage areas are higher.



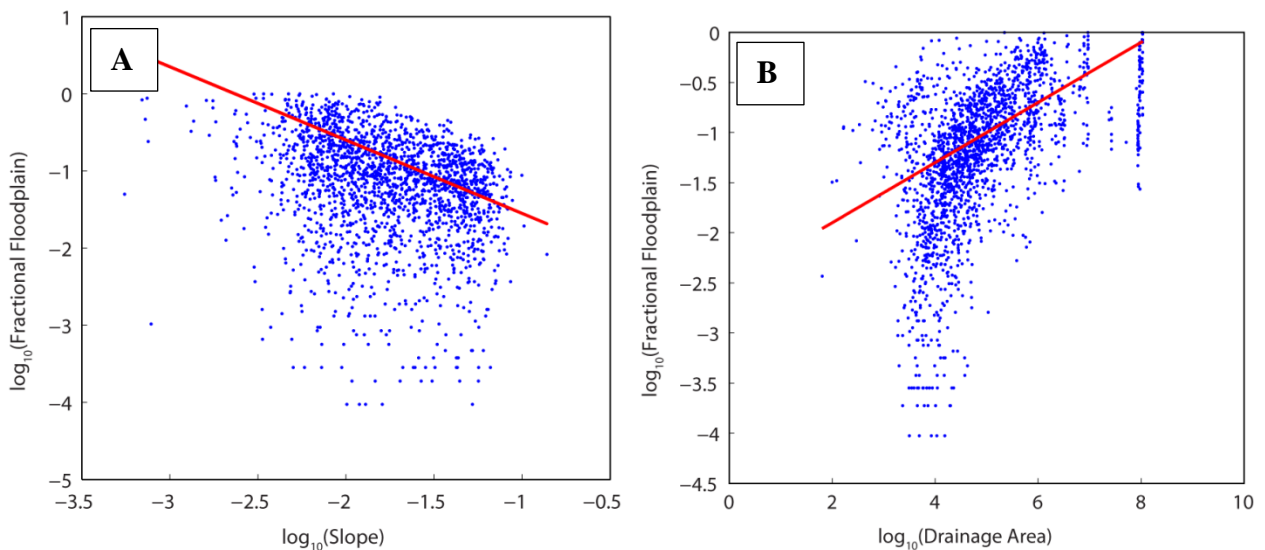
**Figure 4.** Average slope over 1 km grid. There is a distinct difference between the east and western halves of the county, although absolute values only range from 0.03 to 7.88 degrees.



**Figure 5.** Maximum drainage area within 1 km grid cells. The Colorado River, flowing to the east, is the largest hydrologic boundary and thus displays the greatest drainage area. Other large tributaries, such as Onion Creek and Barton Creek can be seen flowing in from the south.



**Figure 6.** FEMA-derived fractional floodplain map for 1 km grid. Areas along the eastern section of the Colorado River display high floodplain area (red) with some cells showing the floodplain consumes as much as 99% of the cell area. Areas in the western half are generally lower (blue), except around the reservoirs of the Colorado River.



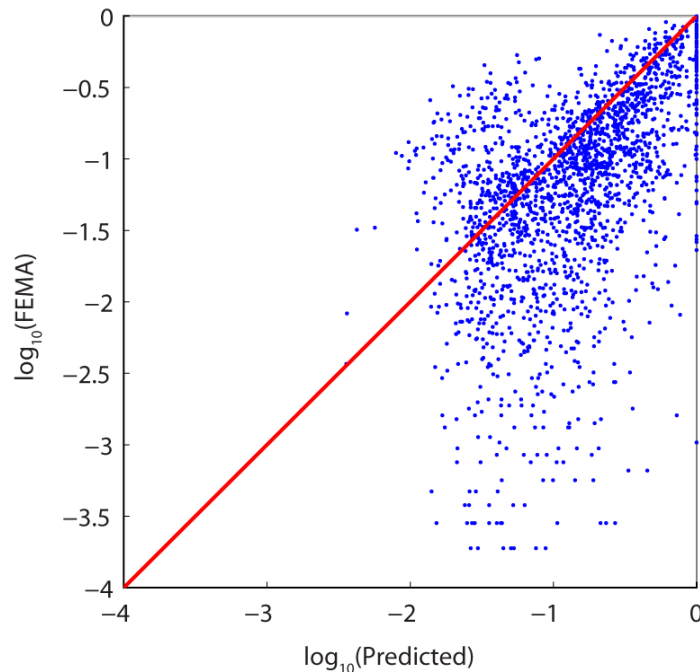
**Figure 7. A)** Log-log plot of average slope with fractional floodplain. There is a lot of scatter, but there is an overall trend of decreasing floodplain with increasing slope. Red line shows regression of data with a slope of -0.95. **B)** Log-log plot of maximum drainage area with fractional floodplain. There is an overall trend of increasing area of floodplain with increasing drainage area. Red line shows regression of data with a slope of 0.30.

There was a lot of spread in the data, but plots of slope and drainage area against fractional area of floodplain showed that the hypotheses were generally correct (Figure 7). There is a definite positive scaling relationship with drainage area and an inverse scaling relationship with slope. Linear regressions of the log-log plots confirmed this observation and suggest that a power-law relationship can reasonably represent both relations.

Combining these two power-law relationships and applying a fitting coefficient derives the equation:

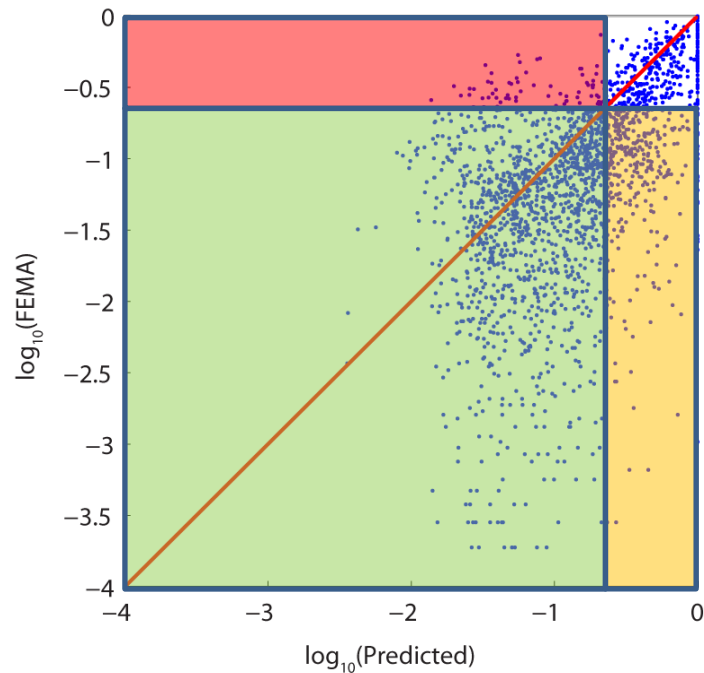
$$F_p = 10^{-4} \cdot \frac{A^{0.3}}{S^{0.95}}$$

where  $F_p$  is the predicted floodplain area. This predictive model broadly collapses the data to a linear line of predicted vs. actual (Figure 8). There is a fair amount of scatter on this plot, but it is also a log-log representation of the data. Most of this scatter actually falls within the area where both predicted and actual floodplain areas are less than 20% of the cell area (Figure 9 – green shaded area). At this low level of predicted and actual floodplain area, the scatter in this area is of minimal concern. The orange area of Figure 9 demonstrates areas where the predicted value is greater than 20% of the cell area while the actual is less than 20%. While this presents more significant error in the model, overestimation presents less of a risk in the interpretation of the result than underestimation. The red area of Figure 9 demonstrates areas where the predicted value is less than 20% of the cell area but the actual area is greater than 20%. Although there are not many points in this region, this is the area of error of most concern as it is considerably underestimating flood risk.



**Figure 8.** Log-log plot of predicted floodplain area vs. the FEMA-derived values. A perfect fit would be represented by the red line. Based on the power-law relations of slope and drainage area, the prediction collapses the data reasonably well along the red-line.

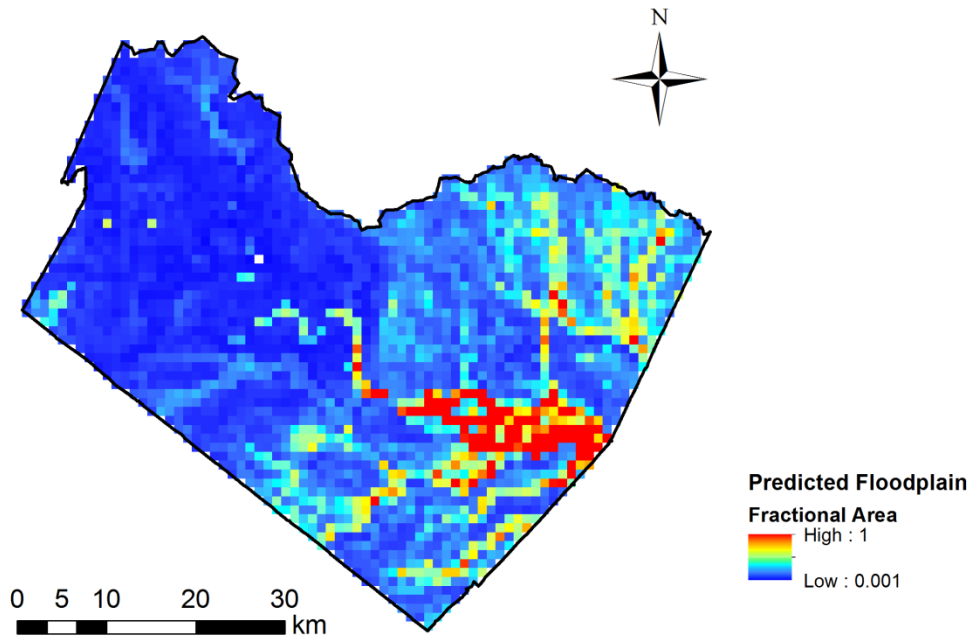




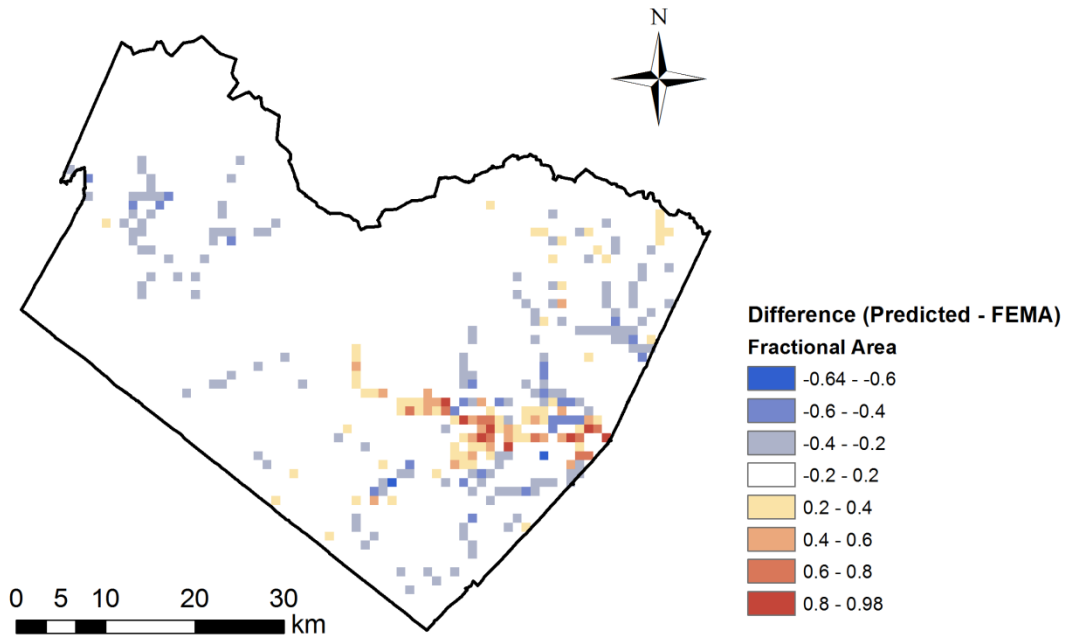
**Figure 9.** Same plot as Figure 8 but with areas designated by type of error and concern over problems with the predictive model. Data points that fall within the green region represent points where the predicted and actual are both  $< 20\%$  of the cell area. Data points that fall within the orange region represent overestimation where the predicted value is  $> 20\%$  of the cell area and actual is  $< 20\%$ . Data points that fall within the red region represent the most concern as underestimated values where the predicted value is  $< 20\%$  of the cell area but actual is  $> 20\%$ .

Interpretations of this predictive model are most useful when the data is redisplayed as a map (Figure 10). Overall, a qualitative, visual comparison of the predicted floodplain area shows that the model does a reasonably good job of replicating spatial patterns. The areas that show the greatest floodplain are accurately within the lower reaches of the Colorado River, along Onion Creek and the Wilbarger Creek watershed. There are some noticeable areas of underestimation, particularly, around Lake Travis and Inks Lake in the western half of the county and along Barton Creek in the southwest. These are notably areas where the watersheds continue outside for the county and therefore drainage area is underestimated. Lady Bird Lake, which runs through downtown Austin, also overestimates flow, but this is a flood-controlled reach of the Colorado River because of its location relative to the city. Drainage area calculations and the impact of engineered waterbodies are the most considerable factors impacting the areas of most considerable over or underestimation.

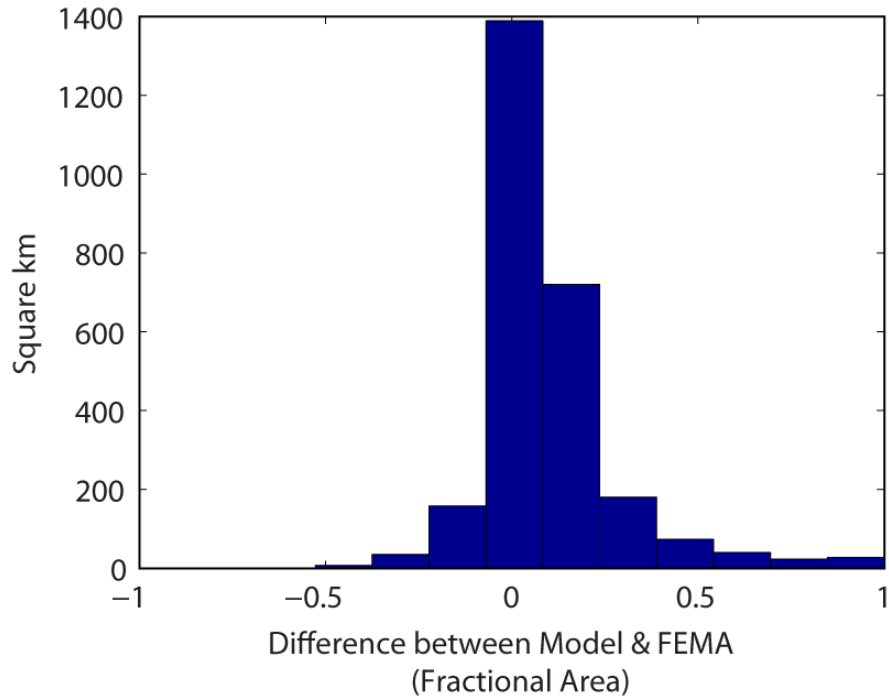
Trying to quantify the error within the predictive model, a difference map between the FEMA and the predicted was created (Figure. 11). This confirms the above observations, but also shows some interesting trends. First the model only appears to largely overestimate floodplain area in cells that are directly adjacent to cells that do have high floodplains down in the lower Colorado River. The spatial distribution of error produced by the model is generally auto-correlated with actual floodplain areas. Higher error is adjacent to high floodplain and lower



**Figure 10.** The predicted model of floodplain area at the 1 km grid resolution. Overall the predicted spatial pattern is consistent with actual values of floodplain area from FEMA maps.



**Figure 11.** Difference between predicted and FEMA derived floodplain area maps. Negative values (blue) represent underestimation and positive values (red) represent overestimation. Areas that are white represent cells where error is  $\pm 20\%$ .



**Figure 12.** Histogram of model error showing total area within each bin. The actual land area of Travis County is 2395 km<sup>2</sup>, showing nearly 60% of the area is within  $\pm 10\%$  of FEMA maps.

error is adjacent to low floodplain. Further analysis of model error, shows that for nearly 60% of the area of Travis County ( $\sim 1400$  km<sup>2</sup>) the predicted floodplain error is only off by  $\pm 10\%$  (Figure 12). Consistent with observations of scatter in Figure 9, there is a slight skew towards overestimation, but the distribution of error is actually quite tight.

### *Conclusions*

The use of publically available topographic data to estimate floodplain area is an approach that works quite well in replicating the spatial patterns of maps produced by FEMA at a 1 km<sup>2</sup> resolution. This result is very promising as most areas of the globe do not have the resources necessary to produce these high resolution maps. While the 1 km<sup>2</sup> resolution is not high enough to evaluate plot by plot where you should and shouldn't build homes, it does provide similar information at a lower resolution scale. This information would also be quite useful for emergency personnel in predicting what areas are most likely to be flooded during large precipitation events. Higher resolutions were not attempted in this study, but given the implications of the results, this model should be tested at higher resolution scales.

Additionally, future work in testing this model would require the use bigger DEMs that are not limited to political boundaries but encompass entire watersheds. It is also clear that this model should do best in areas with limited engineering done to river networks, particularly with respect to the damming of rivers. Where rivers are dammed, the model inputs would have to be adjusted to account for the control of outflows.