Quantifying Wildfire Susceptibility in Utah Using an Artificial Neural Network

GEOG 6010 - Final Project

By Erik M. Neemann

9 December 2018

Introduction

Wildfire is a common hazard faced by large regions in the western US due to the combination of vegetation/fuels, climate, topography, and ignition sources. In recent decades, with increases in population, communities and residential areas have continually expanded outward from urban centers. In the mountainous west, this expansion typically leads to additional development in the urban-wildland interface, thereby increasing the number of people exposed to the potential wildfire hazard. Identifying areas of wildfire susceptibility is important to ensure at-risk populations have emergency management plans and are able to mitigate fire risk as much as possible. The most effective method of identifying risk areas is producing maps by combining multiple wildfire parameters in a geographic information system (GIS). This study aims to quantify wildfire susceptibility in Utah, identify locations at greatest risk, and analyze the demographics of the at-risk communities.

In addition to the expansion of human populations mentioned previously, the size and occurrence of wildfires has increased since the 1980s (Dennison et al., 2014). Even further, a changing climate may lead to a continued increase in wildfire due to greater fuel aridity in future decades (Abatzoglou et al., 2016). The combination of these factors demonstrates that wildfires will continue to pose a significant threat to communities and populations for years to come, with possible loss of life, major property damage/financial cost, and environmental destruction. In order to manage this threat, a method for quantifying wildfire susceptibility in Utah will be presented in this paper. The susceptibility index will then be extended through a demographic analysis in an effort to improve the understanding of what communities and populations are most likely to encounter wildfires. These communities could then be targeted for outreach, education, and preparedness planning in order to minimize wildfire exposure as much as possible.

Background

Several past studies have looked at quantifying wildfire risk or predicting burned area using a variety of techniques ranging from support vector machines (SVM) to random forests (RF) to artificial neural networks (ANN). Some of these studies have focused primarily on climate factors, while other leverage geographic and vegetation properties in addition to climate factors. Cortez et al. (2007), used a variety of data mining techniques to predict forest fire burned area in northeast Portugal. They found that the best configuration used temperature, relative humidity, wind, and rainfall data as inputs to a SVM. This method produced good results for predicting the

burned area of small fires. However, the method was less effective for large fire area predictions (Cortez et al., 2007). Regardless, the results could potentially be used to inform firefighting decisions and resource allocation. Additionally, Satir et al. (2015) employed a multi-layer perceptron (MLP) ANN to map forest fire probability in southern Turkey near the Mediterranean Sea. Satir's approach fed a variety of climate, anthropogenic (distance to roads, settlements, and farmlands), topographic, and vegetation data sets into the MLP, using the fire history data for training and testing. They found that tree canopy cover, elevation, and temperature were the most effective parameters, while anthropogenic parameters were not significant (Satir et al., 2015). In another study, a fuzzy NN was used to model forest fire potential in the Vietnam central highland (Bui et al., 2017) by applying a Particle Swarm Optimization learning technique. Bui et al. used 10 input parameters relating to terrain, vegetation, climate, and anthropogenic variables and trained the model with historical forest fire incidents. They found that their fuzzy-neural inference system outperformed previous methodologies that used RF and SVM (Bui et al., 2017).

With this previous work in mind, an MLP NN was chosen as the model for this study due to the recent success of other research projects in implementing NNs to predict forest fire susceptibility (Satir et al., 2015; Bui et al., 2017). Specifically, it has been noted that single hidden layer NN architectures with 2-3 times the number of neurons in the hidden layer than the input layer have shown useful results (Satir et al., 2015). Neural Networks are a natural choice for analyzing fire potential because they are good at approximating complex functions, are able to capture both linear and nonlinear relationships that exist between predictors and dependent variables, and adaptively learn from training data (Kantardzic, 2011; Satir et al., 2015).

Data

All data used in this research came from free, publicly-available sources. The original data sets were initially gathered on disparate grids (different cell sizes, projections, dimensions, etc.) and were aligned to a uniform grid via projections in ESRI's ArcMap software and with the use of ArcPy scripts. The data were also clipped down to the same geographic areas of interest (figure 1). This resulted in raster data sets with a Universal Transverse Mercator coordinate system in zone 12 North (UTM 12N). Each grid cell is 30 by 30 m with total dimensions of 2867 by 4251 pixels.

Historical Fires

Historical fire occurrence and burn severity data was collected from the Monitoring Trends in Burn Severity (MTBS) project. MTBS provides annual burn severity data for each state at 30 m grid resolution from 1984-2016. Each raster cells contains a burn severity value based on the scale in Table 1.

Value	Burn Severity
1	Unburned to Low
2	Low
3	Moderate
4	High
5	Increased Greenness
6	Masked/No Data

Table 1: MTBS Burn Severity data categories

For this study, data from 2007-2016 was used to train the NN in order to place an emphasis on the most recent decade of fire data, which should best correspond with recent vegetation and infrastructure data. The ten annual data sets from 2007-2016 where then reclassified to only include burned areas (data value of 1-4), and summed up into a single data set. This resultant MTBS fire data set accounts for a combination of cumulative fire occurrences (multiple fires) and burn severities, called the historical fire index. In this sense, a grid cell that experienced one high-severity burn (value of 4) would receive the same score a grid cell that experienced two low- severity burns (value of 2 multiplied by 2). That value is then divided by the maximum value in the entire region to arrive at an index between 0 and 1. The historical fire index data is shown in figure 2.

Terrain

The Digital Elevation Model (DEM) used in this study was downloaded from the Utah Wildfire Risk Assessment Portal (WRAP). This DEM is distributed on a native UTM 12N grid with 30 m grid cells and was the standard grid that other data sets were aligned to using GIS methods. Slope and aspect data sets were calculated from the DEM and used as inputs into the NN model. These three terrain-related data sets are important for the neural network model because each has an influence on wildfire and fire-related parameters. Elevation influences fire spread, climate variables, and vegetation patterns, all of which influence wildfire behavior (Bui et al., 2017). Slope also impacts fire spread behavior, while aspect influences wind and precipitation patterns and, therefore, secondarily affects vegetation patterns (Bui et al., 2017). The elevation, slope, and aspect data sets are shown in figures 1, 3, and 4, respectively.

Vegetation

Vegetation data sets were also gathered from the Utah WRAP, each on the same UTM 12N, 30 m grid as the DEM data. The vegetation type data set categorizes each grid cell into one of 19 different categories (see table). These categories implicitly included some land use information through the inclusion of “water,” “developed,” and “agriculture” classes (Table 2).

Class	Vegetation Type
1	Agriculture
2	Barren
3	Water
4	Developed
5	Sparse Vegetation
6	Grassland
7	Exotic Herb
8	Riparian
9	Hardwood
10	Mixed Fir Forest
11	Pine Forest
12	Subalpine Forest
13	Pinyon-Juniper
14	Mountain Mahogany
15	Desert Scrub/Steppe
16	Shrubland
17	Gamble Oak
18	Sage Shrub/Steppe
19	Chaparral

Table 2: Vegetation Type classes

The vegetation type information is important for identifying the type of biomass available for burning, where some vegetation types burn more readily than others due to their biophysical properties (Bui et al., 2017). Two additional vegetation data sets, canopy cover (%) and bulk canopy density ($kg/m^3 * 100$) are also included to help quantify the amount of biomass fuel that is available for burning on a grid cell. The vegetation type, canopy cover, and bulk canopy density data sets are show in figures 5-7.

Infrastructure

A statewide Utah roads data set was gathered from the Utah Automated Geographic Reference Center (AGRC). The vector road features were used to calculate a distance-to-roads raster layer using Euclidean distance, meaning the minimum distance from a pixel to a road is calculated "as the crow flies." This data was generated on a 30 m grid cell raster data set with the same UTM 12N projection as the rest of the data used in the study. Distance to roads is employed as a mechanism to measure the anthropogenic influence on wildfires and relates to the potential cause of wildfires (Satir et al., 2015; Bui et al., 2017). It is presumed that locations closer to roads are more exposed to anthropogenic fire ignition sources, while locations far from roads are less influenced by humans and are more likely to be considered wilderness. The distance to roads data is depicted in figure 8.

Climate

Climate data for this study was collected from Oregon State's PRISM Climate Group. Monthly 30-year normal temperature and precipitation data sets were downloaded with 800 m grid cells. Only data from the months of June through October were used, representing the primary fire season in Utah. Precipitation amounts and temperatures from other months of the year may have some role in setting the preconditions for vegetation biomass and aridity, but for the purposes of this study it is assumed that these factors are dominated by weather conditions during the actual fire season. The monthly climate data sets, for both temperature and precipitation, were averaged into a single data set representing the mean conditions for the entire fire season. These data were then projected onto the UTM 12N coordinate reference system and resampled to 30 m grid cells with cubic interpolation. The temperature and precipitation data is shown in figures 9 and 10.

Demographics

Finally, demographic data was acquired from the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates from 2012-2016. Census tract-level data were gathered related to income, educational attainment, and poverty rate. Each of the relevant data tables were joined to their associated geographic area using the ArcMap software.

Methods

Data Preprocessing

In order to prepare the data for input into the neural network model, several preprocessing steps were necessary. First, the data was scaled with a max-min normalization to place each data set between a range of 0 to 1. This was needed to ensure that the scale variations among the different data sets didn't lead one or two variables to overwhelm the training process and contaminate the results. The only exception to data scaling was the vegetation type data, which was the only categorical data set used in the study. Because vegetation type is categorical, the class numbers have no quantitative meaning (e.g., class 8 is not "greater" than class 2) and scaling isn't appropriate. Instead, this data was converted into dummy variables to handle the class differences. With this methodology each vegetation type was broken out into its own variable that contained a binary value (0 for absence, 1 for presence), representing whether or not a pixel belonged to that particular class. The end result was the creation of several new variables, which allow the categorical vegetation type data to be included in the neural network.

Due to the computational challenges of employing a neural network on an area spanning 10,968 km^2 (figure 1) with such a fine grid cell resolution of 30 m (12.2 million total pixels), the data was transformed onto a coarser, 90 m grid. This resulted in a data set that was much more manageable to work with for a neural network (956 by 1417 pixels for 1.35 million total pixels or about 11% of the original size). All data sets were resampled to 90 m grid cells with a nearest neighborhood methodology that maintained the same extent and projection as the original data.

The historical fire severity MTBS data, used for training the NN, was thinned in order to improve the training process. The vast majority of the area of interest is comprised of pixels that

have not been burned (~98%), but it was desirable to have better balance among burned and unburned pixels in the training data set. For this reason, a subset of the entire AOI was used for training, where polygons around the burned areas where chosen to provide a more balanced mix of burned/unburned pixels (black polygons in figure 2). This also improved the efficiency of training the NN, as data set used for training was about 10% of the original size.

Finally, the census data was preprocessed in order to more cleanly categorize income and education levels. Census tract-level income data was used to calculate the percentage of the population in the following income groups: under $30k, $30-60k, $60-90k, $90-120k, and greater than $120k. Similarly, the percent of population at each of the following education levels was calculated: high school graduate, bachelor's degree or higher, and master's degree or higher. In addition to these variables, median income and percentage of families with income below the poverty level were also gathered for each census tract.

Artificial Neural Network

The R statistical language was used to implement the NN using RStudio software, specifically with the 'neuralnet' library. A forward-feed MLP NN was employed with a resilient backpropagation learning algorithm. This varies slightly from a traditional backpropagation algorithm in that it is known to operate more efficiently and requires fewer parameter decisions (McCaffrey, 2015). The network was "fully-connected," meaning that every neuron in a layer was connected to every neuron in adjacent layers. Initial challenges in getting the NN to complete the training process were overcome with a few modifications. The original NN architecture was targeted to have 9 input neurons, a single hidden layer with 25 neurons, and one output neuron representing wildfire susceptibility. By convention, this architecture is denoted as 9:25:1. However, limited computational resources resulted in a very slow training process and led the number of hidden layer neurons to be reduced from 25 to 15. This resulted in a more efficient, and ultimately successful, NN training process when combined with the aforementioned thinning of the input fire training data. The transformation of vegetation types into dummy variables ultimately resulted in a total of 27 input variables, for a final architecture of 27:15:1 (figure 11).

In addition to building a NN model with the full complement of input data sets, a second model was generated without the vegetation type data. This was done to examine the influence of vegetation type and whether its breakout into 18 dummy variables might have affected the results. Other than the removal of the vegetation type categories, there were no other differences between the models. Once vegetation type was removed, only 8 input neurons existed, resulting in a NN architecture of 8:15:1 (figure 12.). This model will hereafter be referred to a "NVT."

Finally, the output of each model was normalized with a min-max normalization method to put the Fire Susceptibility Index on a 0-1 scale. The results from each model were then categorized into 6 fire susceptibility groups with a natural breaks methodology. However, because the categories from the Original and NVT models were very similar, the final categorization for each model was standardized based on table 3. This created a consistent symbology for generating fire susceptibility maps and allowed for a more direct comparison between the two models.

Category	Fire Susceptibility Score Range	Numeric Category
Very Low	0 - 0.4	1
Low	0.4 - 0.5	2
Moderate	0.5 - 0.55	3
High	0.55 - 0.6	4
Very High	0.6 - 0.7	5
Extreme	0.7 - 1	6

Table 3: Fire Susceptibility Index categories

Demographic Analysis

To analyze potential characteristics and trends in demographic data associated with susceptible wildfire areas, a brief demographic analysis was performed at the census tract level. This analysis employed the aforementioned income, education, and poverty variables. For each census tract, a mean fire susceptibility score was calculated based on the Original NN model output. This allowed for correlations between mean fire susceptibility and income, education, or poverty to be examined for each census tract.

Results

Fire Susceptibility

The output from each of the NN models showed results with many similarities and some notable differences. A histogram showing the distribution of Fire Susceptibility Index values from both models is shown in figure 13. Both models had a strong peak centered on 0.55, but the distribution of values around that peak varies. In the Original model, the vast majority of pixels were placed between 0.4 and 0.8, with a narrower distribution and shorter tails. The NVT model, however, had a wider distribution and longer tails, with a generally broader range between 0.2 and 0.85. This indicates that the Original model has many more values scoring in the middle of the range, while the NVT model's larger range has more pixels assessed with "very low" or "extreme" values. Visually, this is depicted in figures 14 and 15, which display a map of the Wasatch Front region with the Fire Susceptibility Index calculated for each pixel. Both models are plotted with the same color bar, allowing for easy comparison between the two. Compared to the Original Model, the NVT model has more values in the "extreme" category for high-threat areas such as Antelope Island in the Great Salt Lake, the northern Oquirrh Mountains, the Lake Mountains, and the mountain range south of Utah Lake. The NVT model also has a larger quantity of "very low" pixels along the high peaks of the Wasatch Front and Oquirrh Mountains.

Another notable difference between the models is the apparent influence of the vegetation type variable in the Original model and its lack of influence in the NVT model (where it was removed). The Original model has a generally noisier appearance, reflective of the noisy nature of the vegetation type variable, whereas the NVT model has smoother, more gradual transitions

from low to high index values. The Original model also highlights the area between Salt Lake International Airport and the Great Salt Lake with "very high" and "extreme" index values, while the NVT model shows "very low" to "moderate" values. Here, the Original model appears to be picking up on increased fire susceptibility from the shrubland and chaparral vegetation types. Additional examples of this difference between the Original and NVT models is displayed in figure 16, which compares the two model outputs to vegetation type between the upper Cottonwood Canyons and Jordanelle Reservoir/Heber city area. The white circle shows the strong influence of "grassland" and "exotic herb" vegetation types, where the Original model clearly highlights "very high" to "extreme" index values that stand out from surrounding values, but the NVT model has "high" to "very high" values that blend into the surrounding area. Figure 16 also shows how aspect appears to have a noticeable influence on NVT model results, in the absence of vegetation type. The red circle indicates sharp contrasts in NVT output, which matches up with aspect changes in Big Cottonwood Canyon where secondary north-south oriented valleys and ridgelines empty into the larger canyon. This influence is not nearly as noticeable in the Original model. Finally, the influence of elevation also appears stronger in the NVT model. The blue circle shows a gradual transition from "very low" values in high-elevation areas to "very high" values at moderate elevations in the NVT model. As one continues southeast toward Heber city, these values begin to decrease again into the lower elevations. The Original model doesn't have nearly as smooth a transition among index values in this area, and appears to reflect changes in vegetation type and other variable more so than elevation.

The influence of vegetation type is further examined by plotting variable importance based on NN connection weights from each model, as seen in figure 17. Indeed, the 10 most important variables in the Original model are vegetation types, which also make up 15 of the 18 most important variables. The top non-vegetation variables include elevation, precipitation, and distance to roads. For the NVT model, the top 3 variables are precipitation, elevation, and distance to roads. This comparison shows the importance of the vegetation type variables in the Original model, perhaps even indicating that it is over-influencing results. However, it is encouraging that both NN models have a similar (but not exact) order of variable importance for non-vegetation variables. Interestingly, the NVT model shows aspect as the least important variable. This is in contrast to the previous discussion about its apparent importance along high ridgelines in the Cottonwood Canyons from figure 16. While the importance of aspect may be low for the NVT model as a whole, it appears that it may still be very important in certain regions where other variables don't have a strong signal in the model.

A final, quantitative comparison of the Original and NVT models is presented in the table in figure 18a. Mean squared error (MSE) and mean absolute error (MAE) metrics are calculated for each model using the subset of validation data from training the NN models. These metrics both indicate that the NVT model performs better as its errors are roughly half of the Original model's error. The area under the curve (AUC) metric for the receiver operating characteristic (ROC) curve was also calculated for each model. This value accounts for a balance between true positive rate and false positive rate when using a broad range of thresholds to classify model results. The AUC metric indicates that the Original model (0.8875) outperformed the NVT model (0.8127) as the Original model had a higher value and was closer to the perfect score of

one. These two sets of metrics (errors and AUC) show conflicting results without a clear winner on which model performs best. It should also be noted that these metrics only consider about 3% of all pixels in the study area, those used for validation of the training data. Another metric, Fire Density is useful to consider because it looks at all pixels in the study area to consider performance, following the method of Pham et al., 2017. For each category in the Fire Susceptibility Index, the Fire Density is calculated by dividing the number of fire pixels present in that category by the total number of pixels in the category and multiplying by 100.

$$Fire\ Density\ (\%) = \frac{\#\ fire\ pixels\ in\ category}{total\ \#\ pixels\ in\ category} * 100$$

The results of the Fire Density calculations are shown in tables 18b-c. In both cases we see Fire Density increase for each category of increasing fire severity. This is an encouraging result, because it indicates that the regions the model assigns high index values to, have a greater percentage of fire pixels within them (from the historical fire data). This is especially true for the Original model, which shows 27% of its "extreme" category are actual fire pixels, compared to 17% for the NVT model. Similar differences are seen for the "high" and "very high" categories. When this is combined with the number of pixels each model categorizes as between "high" and "extreme" (674,119 for the Original model; 1,104,355 for the NVT model), more confidence can be placed in the higher-threat categories of the Original model, than the NVT model.

This section has used a variety of qualitative and quantitative methods to compare the performance of the Original and NVT models. Each model has strengths and weaknesses depending on the metric or region that one focuses on. Overall, both models appear to have generally good performance, but one model cannot clearly be declared as better than the other. Further qualitative and statistical analysis is necessary to make a more definition determination.

Demographic Analysis

Output from the Original model was used to conduct a preliminary demographic analysis to see if any trends or characteristics could be summarized for the populations most at-risk along the Wasatch Front. Scatterplots of the Fire Susceptibility Index and various income-related variables are shown in figure 19. Plots are shown for median income and percentage of population in several income categories, along with a best-fit linear model and Pearson's correlation coefficient. The income data show very weak correlations, broadly indicating that higher income levels may related to increased fire susceptibility and lower income levels may relate to decreased fire susceptibility. This is somewhat intuitive, as one might expect lower income families to be concentrated in cities where wildfire threat is low, and higher income families in the suburbs and closer to the urban-wildland interface where wildfire threat would be greater. However, none of these results is statistically significant and the weak correlations cannot be taken with confidence. The weakness in correlation can also be seen in the wide distribution of points in each scatterplot.

Education levels were also examined for each census tract (figure 20), specifically considering percentage of population with a high school diploma, bachelor's degree or higher, and master's degree or higher. Like the income variables, education levels show weak correlations (~ 0.2)

that might align with a priori hypotheses, where greater educational attainment relates to higher fire threat. However, these correlations are also very weak and do not achieve statistical significance. For this reason, true relationships among education and fire susceptibility cannot be drawn.

Finally, scatterplots of poverty rate and fire susceptibility were generated for another comparison of economic status (figure 21). The result is another weak correlation (-0.18) where poverty rate is negatively correlated with fire susceptibility and high-threat areas tend to have lower poverty rates. Once again, this relationship should be taken with very low confidence due to the large scatter of available data, weak correlation values, and lack of statistical significance.

In summary, the demographic analysis of fire susceptibility and census tract data showed inconclusive results. Weak correlations do exist, but conclusions cannot be made due to a lack of significance. Additional discussion on potential reasons why is included in the following section.

Conclusions and Discussion

Overall, this study demonstrates that the NN methodology produced reasonable results in quantifying wildfire susceptibility along Utah's Wasatch Front. Particularly, it shows that vegetation type may play a prominent role in estimating wildfire susceptibility, as was noted by the model differences between the Original an NVT models. The inclusion of vegetation type often produced noisier results, reflecting the noisy nature of the vegetation type data itself. It also demonstrated local maxima that appeared to be directly tied to the presence of specific vegetation types (e.g., grassland, exotic herb, and chaparral). A comparison of variable importance from the Original and NVT models showed that vegetation type dominated the top 18 variables in the Original model. This reliance on vegetation type may have actually been to the detriment of the model and other methods of employing vegetation information may improve results. The removal of vegetation type in the NVT model allowed other variables to dominate the model results, often varying by geographic location (e.g., elevation, aspect).

Areas with low fire susceptibility were often characterized by the presence of water, location near urban centers, or at very high elevations. Higher fire susceptibility regions tended to exist at moderate elevations and included the presence of specific vegetation types (e.g., grassland, chaparral) in the Original model. The NVT model tended to have a greater range in fire susceptibility values, with a larger number of pixels in the "very low" and "extreme" categories.

The demographic analysis conducted in this study showed inconclusive results. There appeared to be some general, weak trends where higher fire susceptibility tended to be correlated with higher income, higher educational attainment, and lower poverty rates. However, these correlations were quite weak (~0.2 correlation coefficient) and did not have statistical significance. These result may be related to the large size of census tracts. For example, some tracts cover hundreds of square kilometers that encompass the full range of fire susceptibility values. This results in a mean score that falls in the middle of the range, with most census tracts

having a susceptibility score between 3 and 4 ("moderate" and "high"). Ultimately, this washes out any true correlation that may exist and masks potentially meaningful relationships. Finer grain census data, such as block groups or even blocks may prove more useful. On the other hand, they may still suffer from the same limitations where low population density produces large census boundaries that cover a large area. Regardless, further examination and more rigorous statistical evaluation is required to determine whether fire susceptibility is truly linked to income, education, or poverty levels along the Wasatch Front.

Future Work

While some meaningful results have been gathered from this study, additional analysis is needed to gain better insight into the wildfire threat along the Wasatch Front. With this in mind, several recommendations for future work are listed below. First, the neural network architecture could be further optimized and more extensively tested in order to gain the best possible results. This includes experimentation with different numbers of hidden layers and hidden layer neurons. It is noted, however, that additional experimentation would be greatly facilitated by high-performance computing resources, which would allow for more efficient testing of possible architectures. Second, the inclusion of additional geographic and vegetation variables as model inputs is recommended. Specifically, results may be improved by excluding the categorical vegetation type and instead replacing it with quantitative vegetation parameters that might implicitly account for the type of vegetation available to burn (e.g., surface fuel load, rate of fire spread, etc.). The addition of a data set representing the amount of time passed since the most recent fire may also be useful. Third, combining climate variables into a single, fire weather index, in the manner of Cortez et al. (2007), is recommended. This may produce more coherent results, but would require humidity and wind speed data, which were not available for this study. Fourth, a more comprehensive accuracy evaluation would better help determine which models truly perform the best. The dearth of verification data may make this difficult, but employing more advanced statistical techniques would prove helpful. Finally, it is recommended that additional demographic variables and finer-grain census divisions be examined to more thoroughly investigate potential links between wildfire susceptibility and characteristics of at-risk populations.

References

Abatzoglou, J. T. and A. P. Williams, 2016. Impact of anthropogenic climate change on wildfire across western US forests, Proceedings of the National Academy of Sciences, 113(42): 11770-11775.

Automated Geographic Reference Center (AGRC), 2018. Roads and Highway System. https://gis.utah.gov/data/transportation/roads-system/ (accessed October 13, 2018).

Bui, D. T., et al., 2017. A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area, Agricultural and Forest Meteorology, 233: 32–44.

Cortez, P. and A. Morais, 2007. A data mining approach to predict forest fires using meteorological data, New trends in artificial intelligence: proceedings of the 13th Portuguese Conference on Artificial Intelligence (EPIA 2007), Guimarães, Portugal, 2007.

Dennison, P. E., et al., 2014. Large wildfire trends in the western United States, 1984–2011, Geophysical Research Letters, 41: 2928–2933.

Eidenshink, J., et al., 2007. A Project for monitoring trends in burn severity, Fire Ecology Special Issue, 3(1): 3-21.

Kantardzic, M., 2011. Data Mining: Concepts, Models, Methods, and Algorithms, 2nd edition. Hoboken, NJ: A John Wiley & Sons, Inc.

McCaffrey, J., 2015. Neural Network Lab: How To Use Resilient Back Propagation To Train Neural Networks, Visual Studio Magazine. https://visualstudiomagazine.com/articles/2015/03/01/resilient-back-propagation.aspx (accessed December 8, 2018).

Pham, B. T., et al., 2017. Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS, Catena, 149: 52-63.

PRISM Climate Group, Oregon State University, 2018. 30-Year Climate Normals. http://prism.oregonstate.edu/normals (accessed October 8, 2018).

Satir, O., et al., 2015. Mapping regional forest fire probability using artificial neural network model in a Mediterranean forest ecosystem, Geomatics, Natural Hazards and Risk, 7(5): 1645-1658.

U.S. Census Bureau, 2018. TIGER/Line with Selected Demographic and Economic Data: American Community Survey 5-Year Estimates. https://www.census.gov/geo/maps-data/data/tiger-data.html (accessed 15 October, 2018).

U.S. Census Bureau, 2018. American Fact Finder: Guided Search. https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_17_5YR_DP03&prodType=table (accessed 8 December, 2018).

Utah Department of Natural Resources, 2018. Wildfire Risk Assessment Portal. https://www.utahwildfirerisk.com/ (accessed 13 October, 2018).

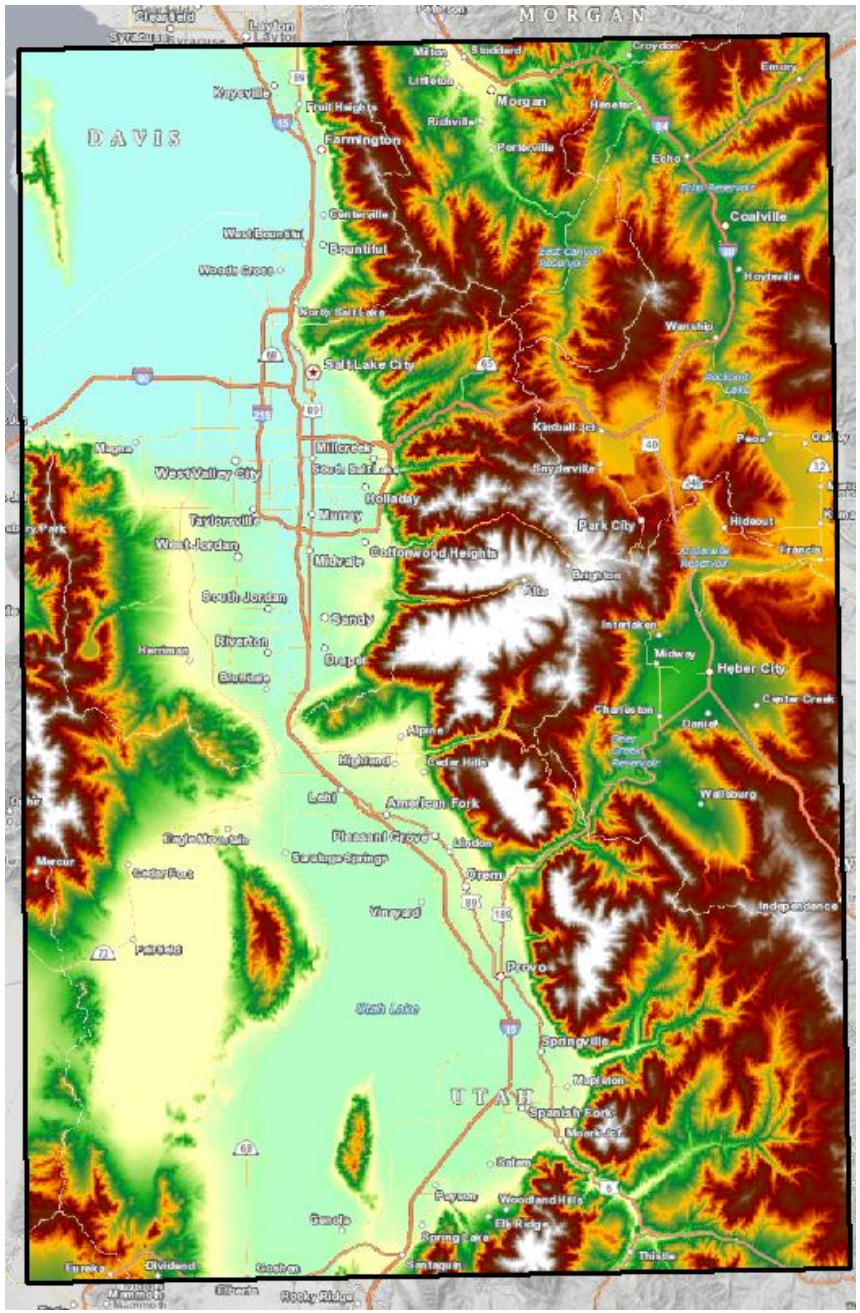


Figure 1: Study area of interest covering much of Utah's Wasatch Front. Elevation (in meters) is colored according the legend on the right.

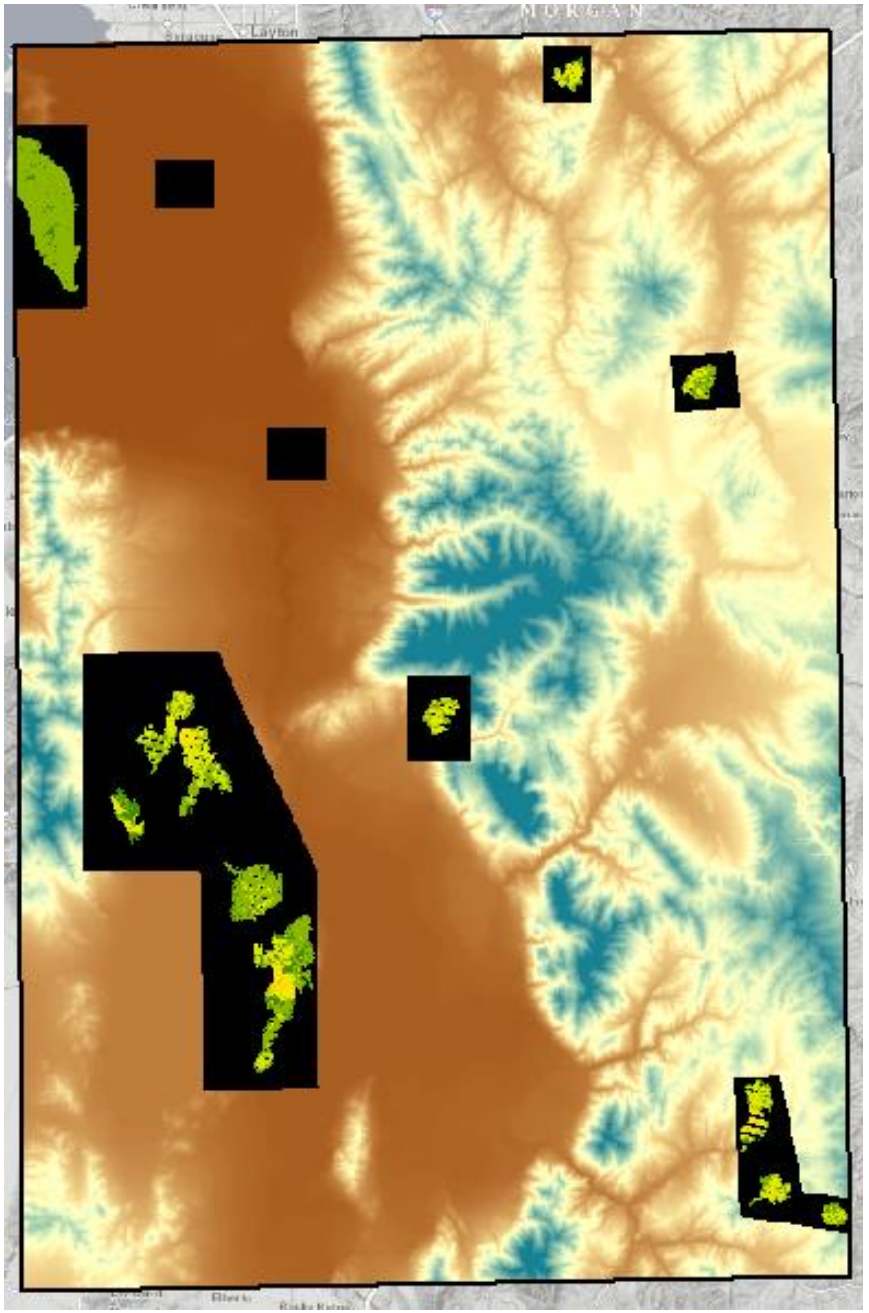


Figure 2: Historical Fire occurrence Index (2007-2016) with elevation in the background.

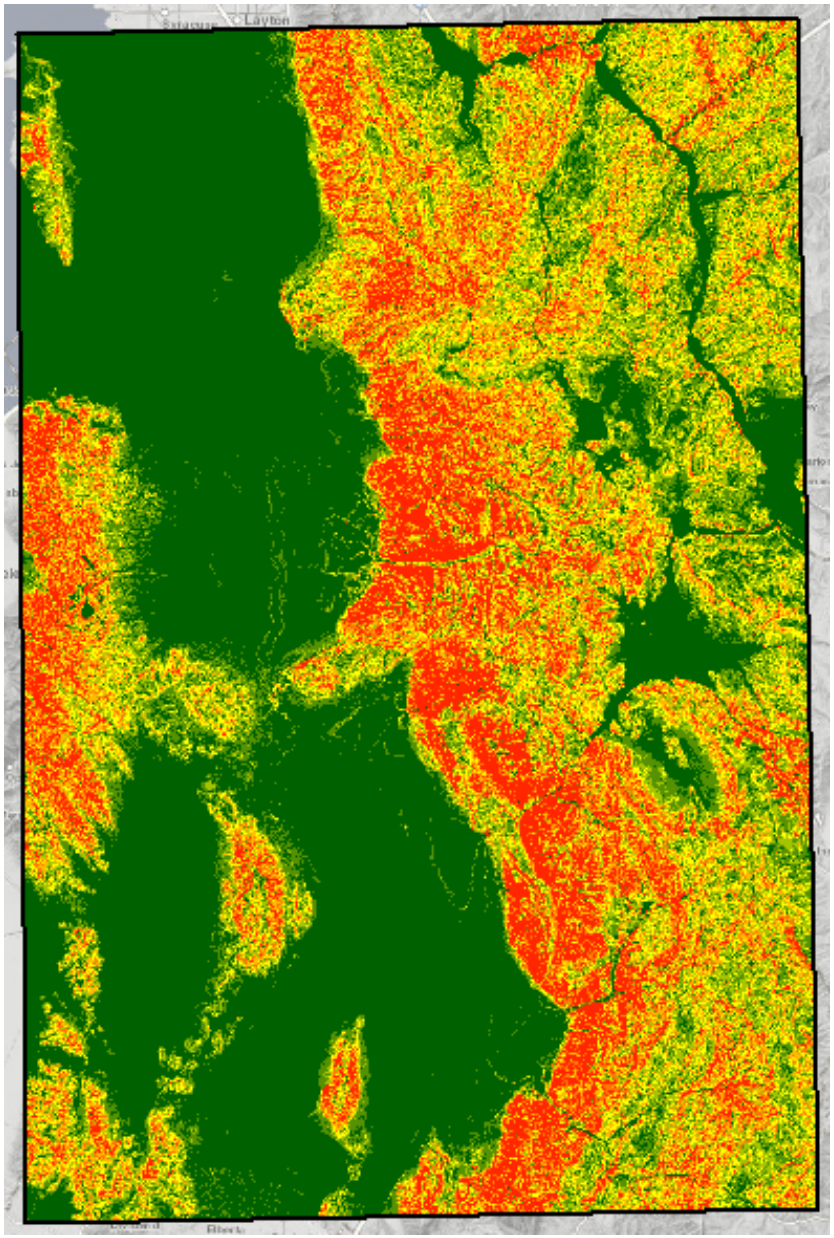


Figure 3: Slope derived from digital elevation model.

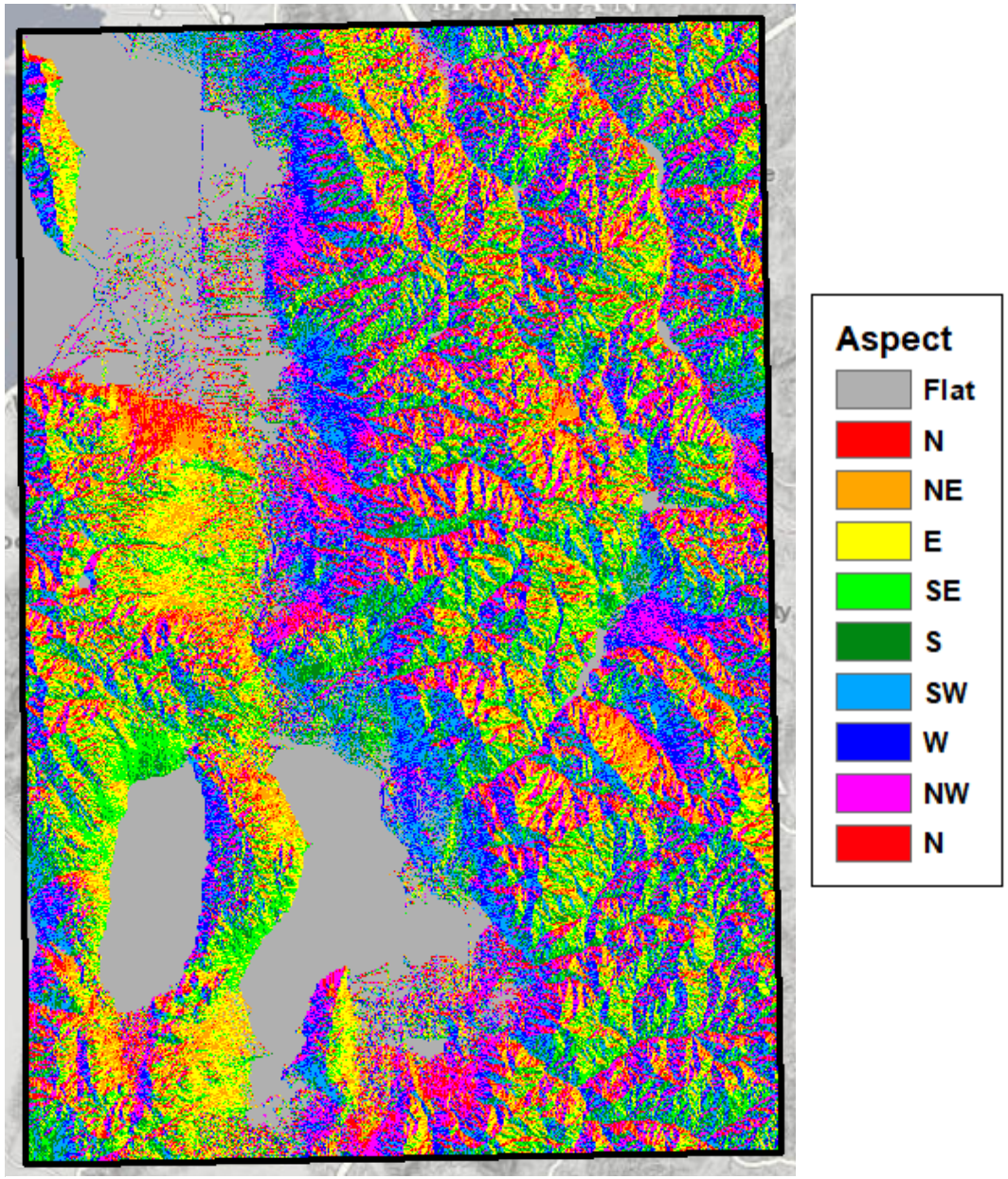


Figure 4: Aspect derived from digital elevation model.

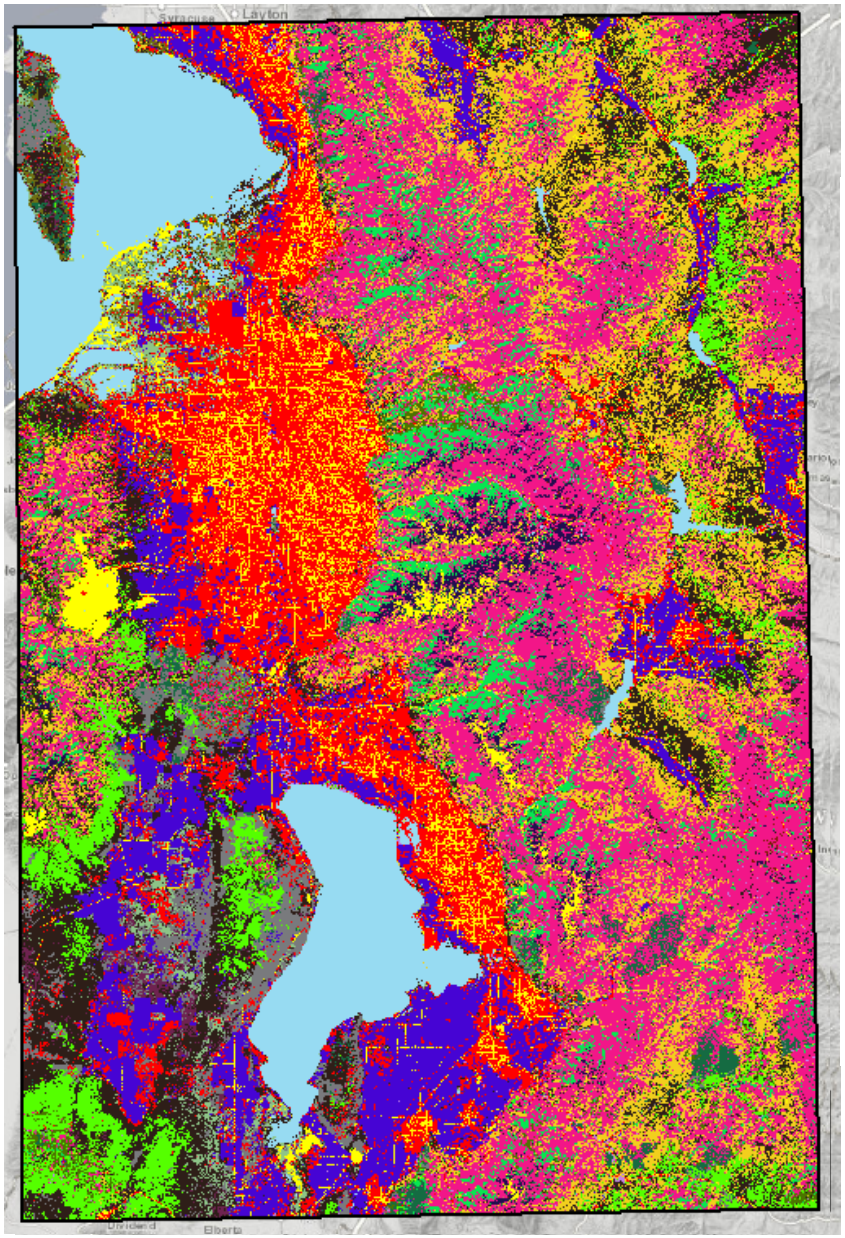


Figure 5: Vegetation Type.

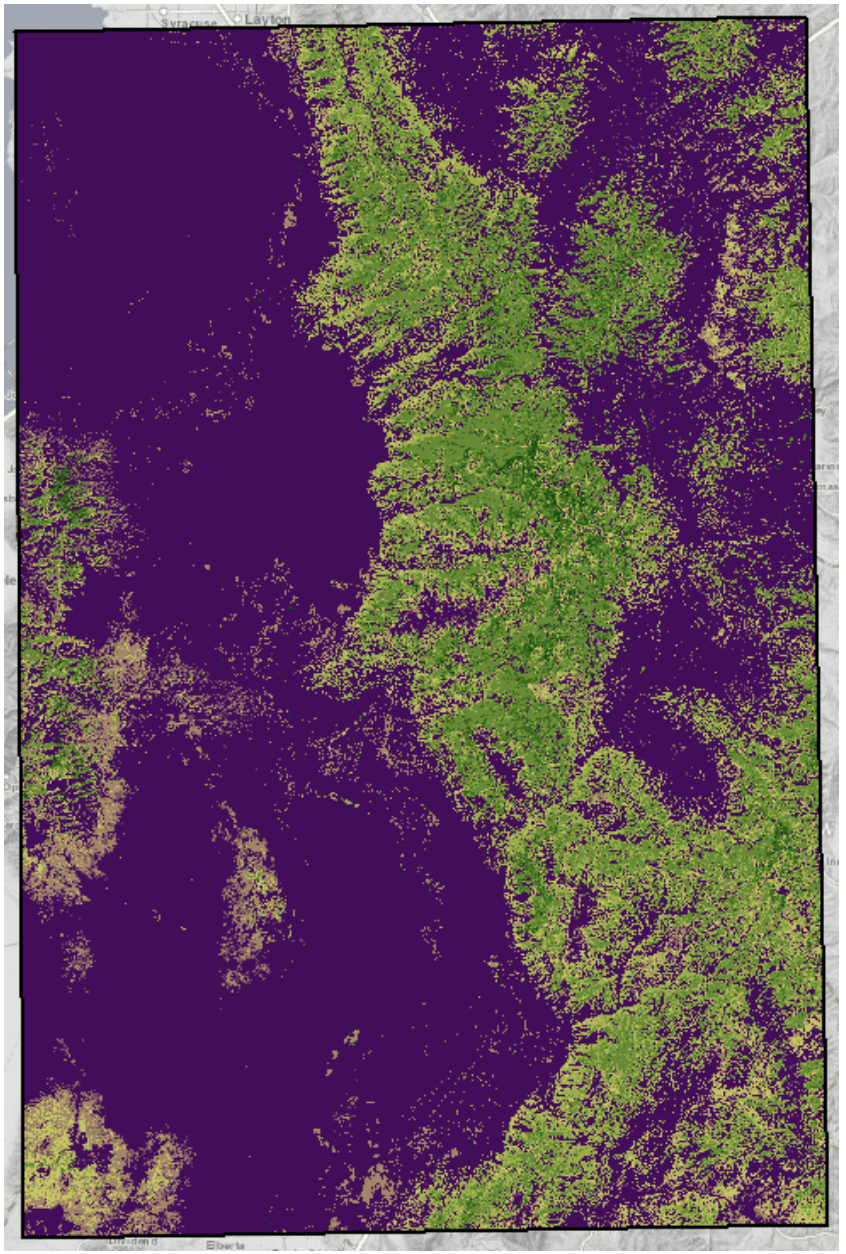


Figure 6: Vegetation Canopy Cover as a percentage.

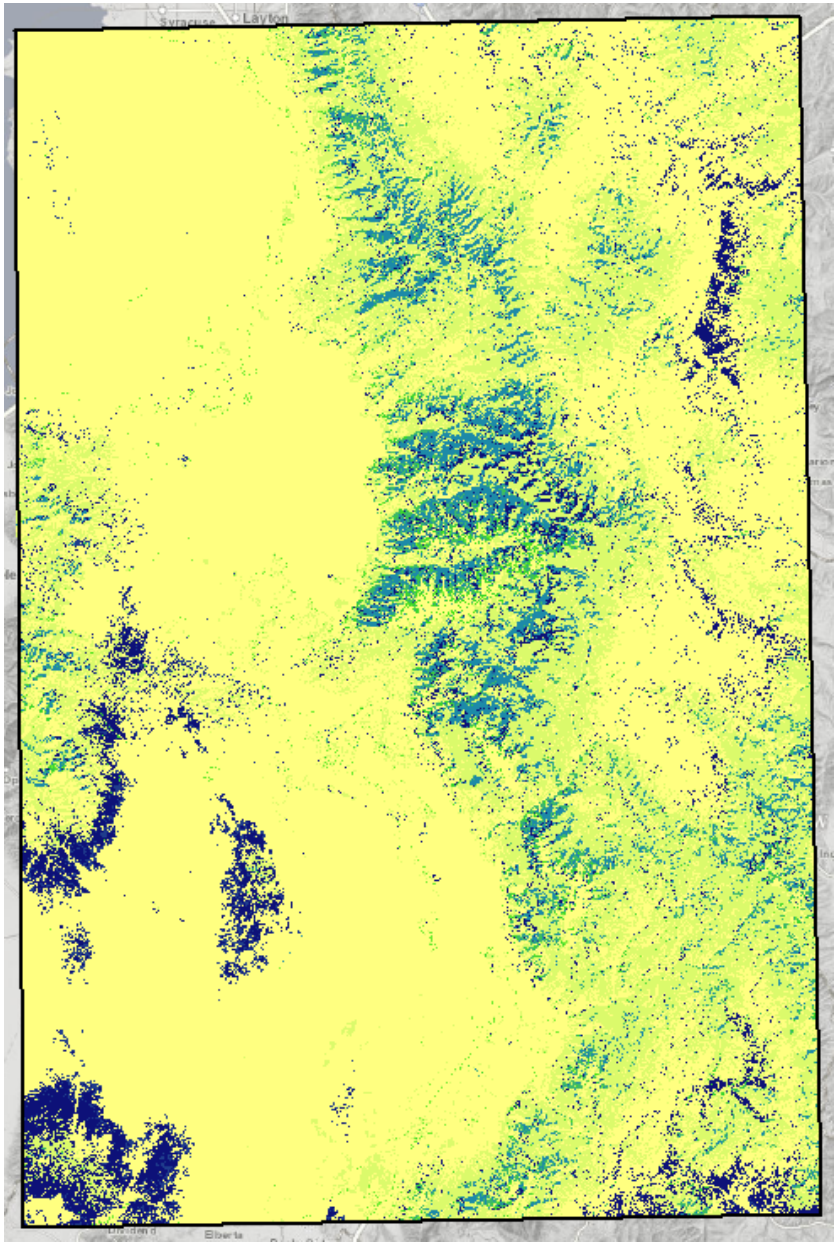


Figure 7: Vegetation Bulk Canopy Density (BCD) in kg/m^3 * 100.

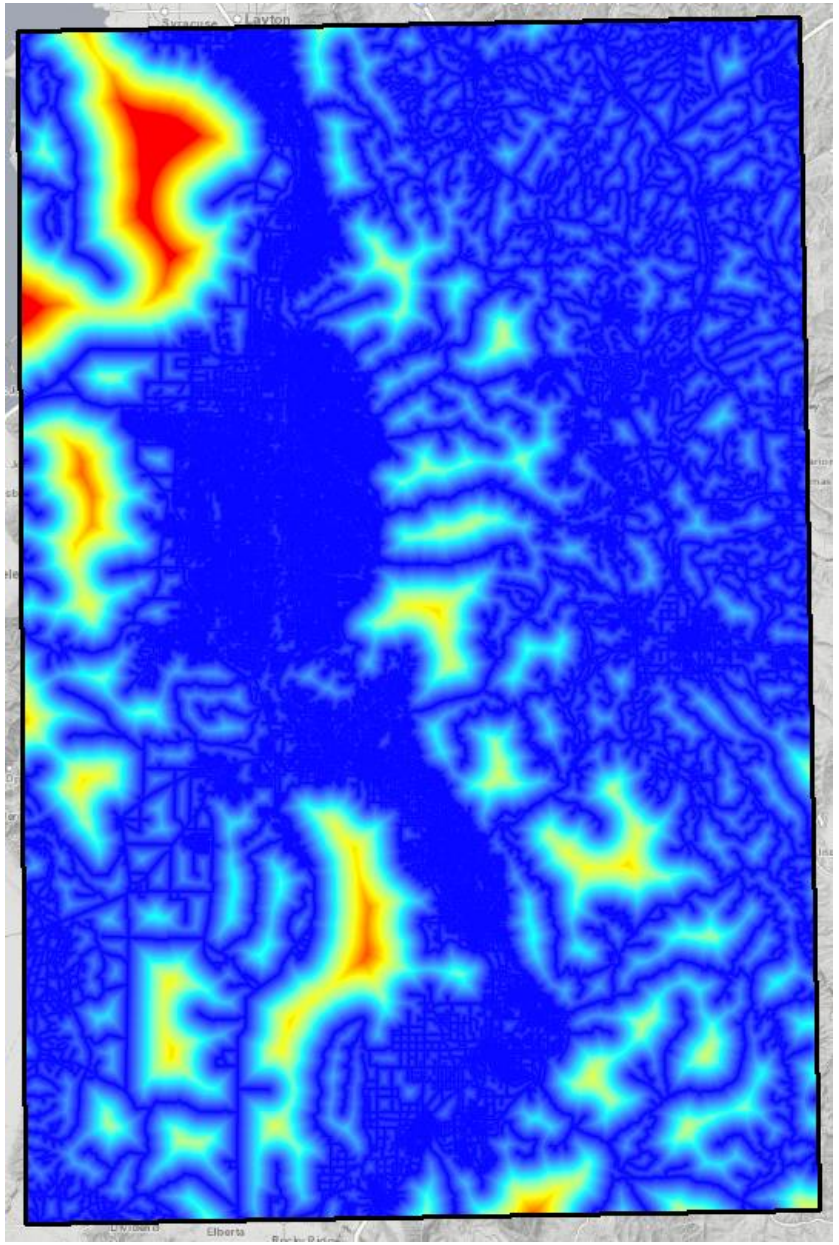


Figure 8: Euclidean distance from roads in meters.

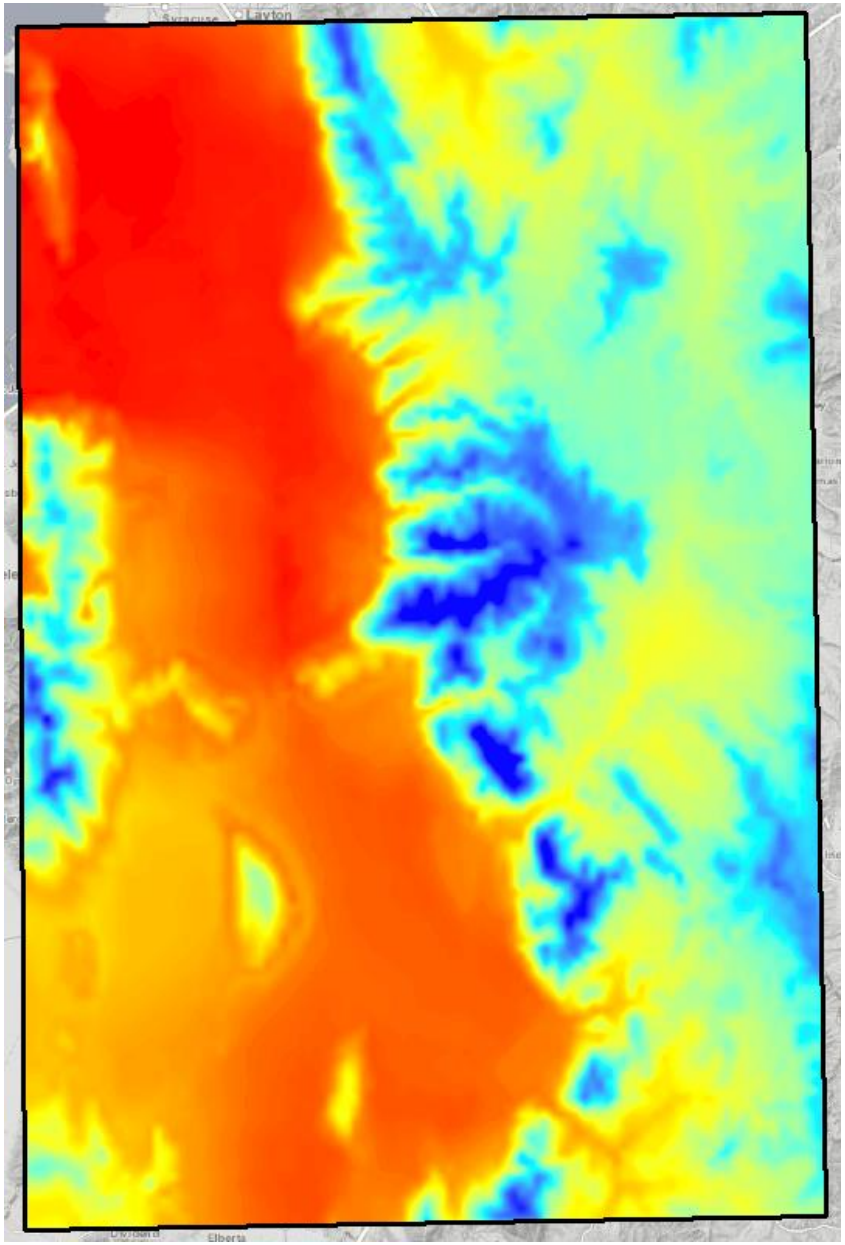


Figure 9: Monthly mean temperature during the wildfire season (June - October), in degrees C.

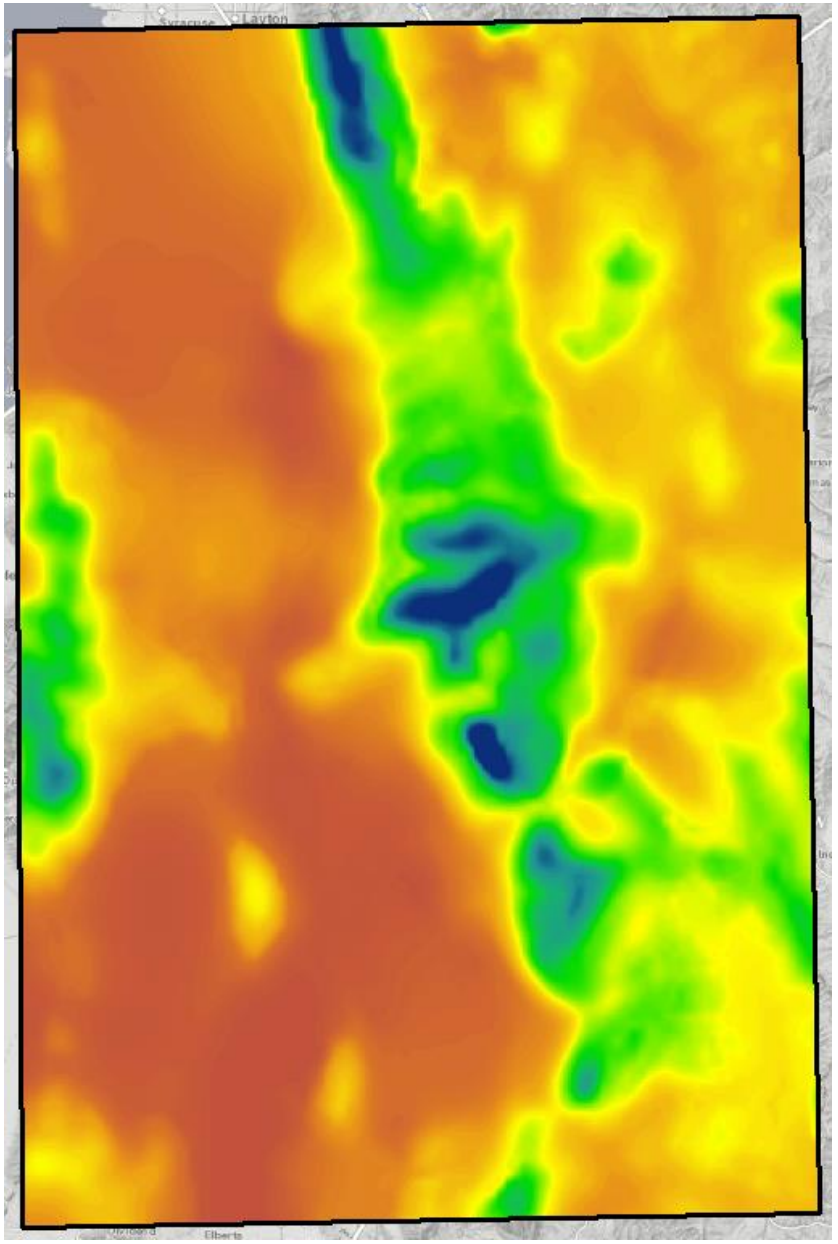


Figure 10: Monthly mean precipitation during the wildfire season (June - October), in mm.

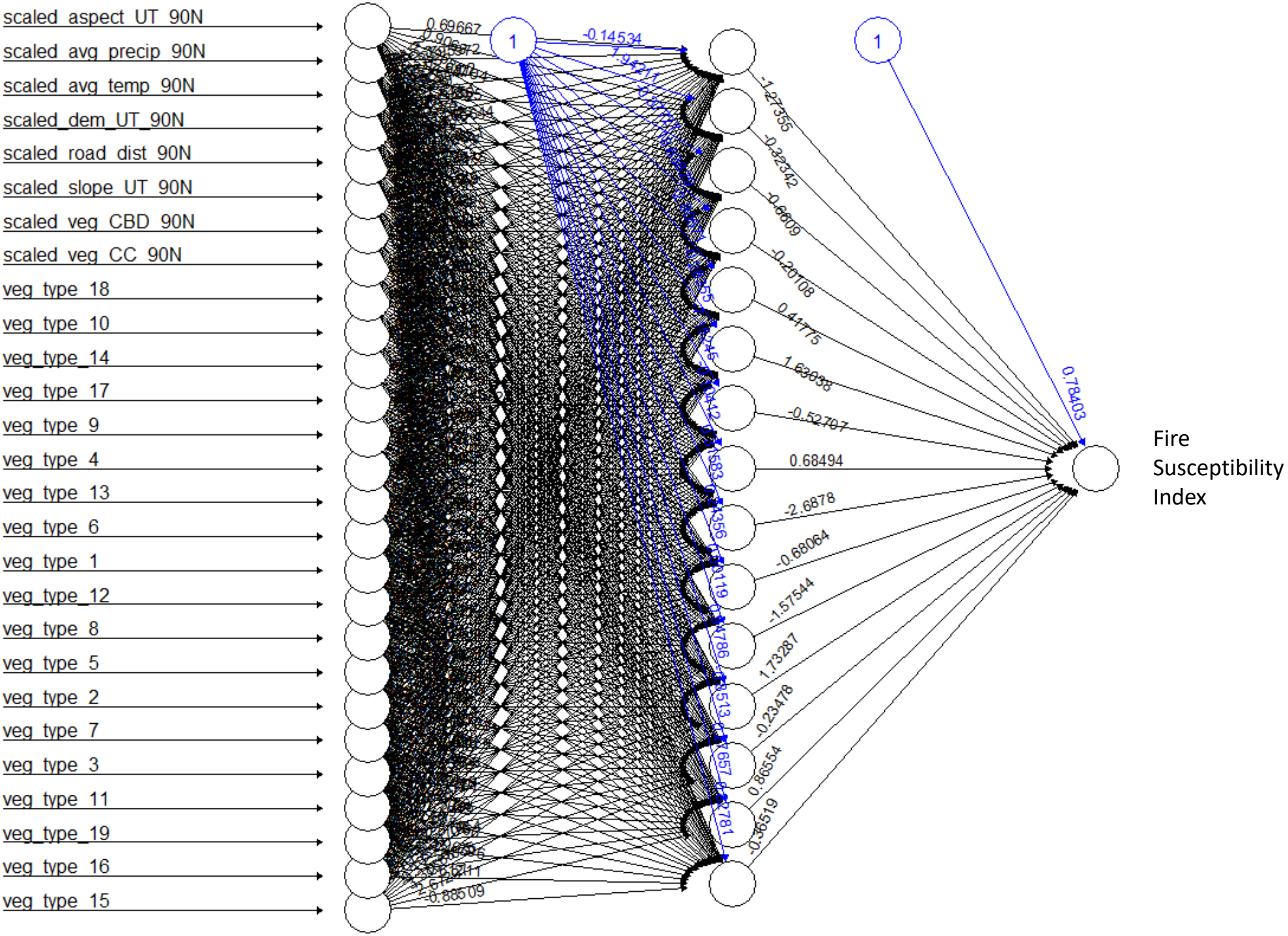


Figure 11: Final neural network architecture and weights for the Original model (27:15:1).

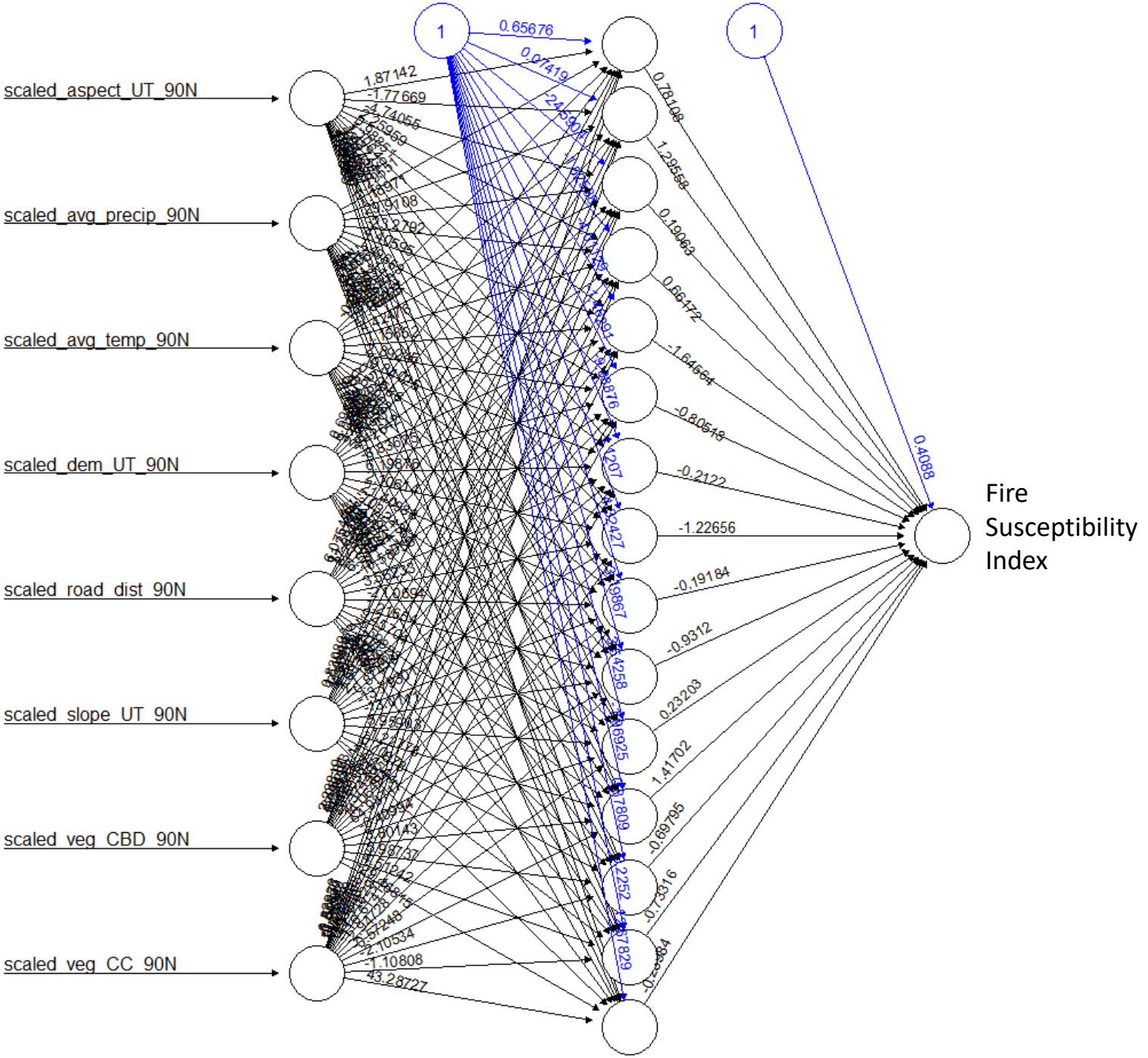


Figure 12: Final neural network architecture and weights for the No Vegetation Type (NVT) model (8:15:1).

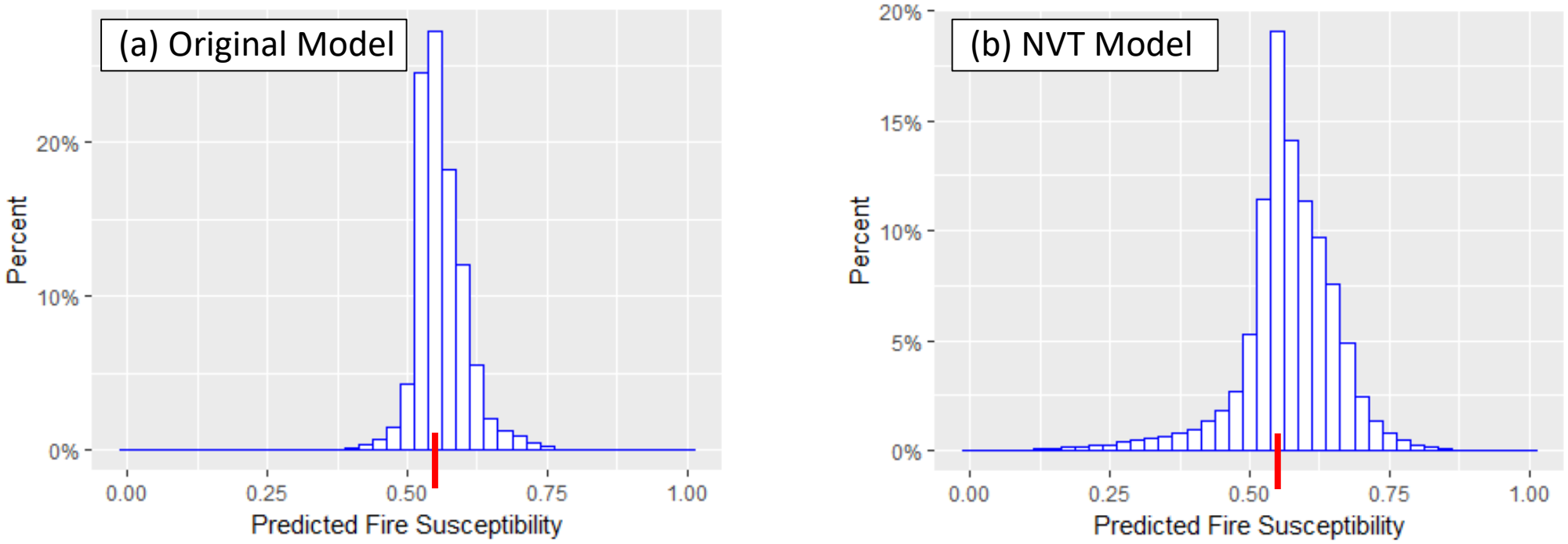


Figure 13: Histogram showing frequency distribution (%) of Fire Susceptibility Index values from (a) Original model and (b) NVT model

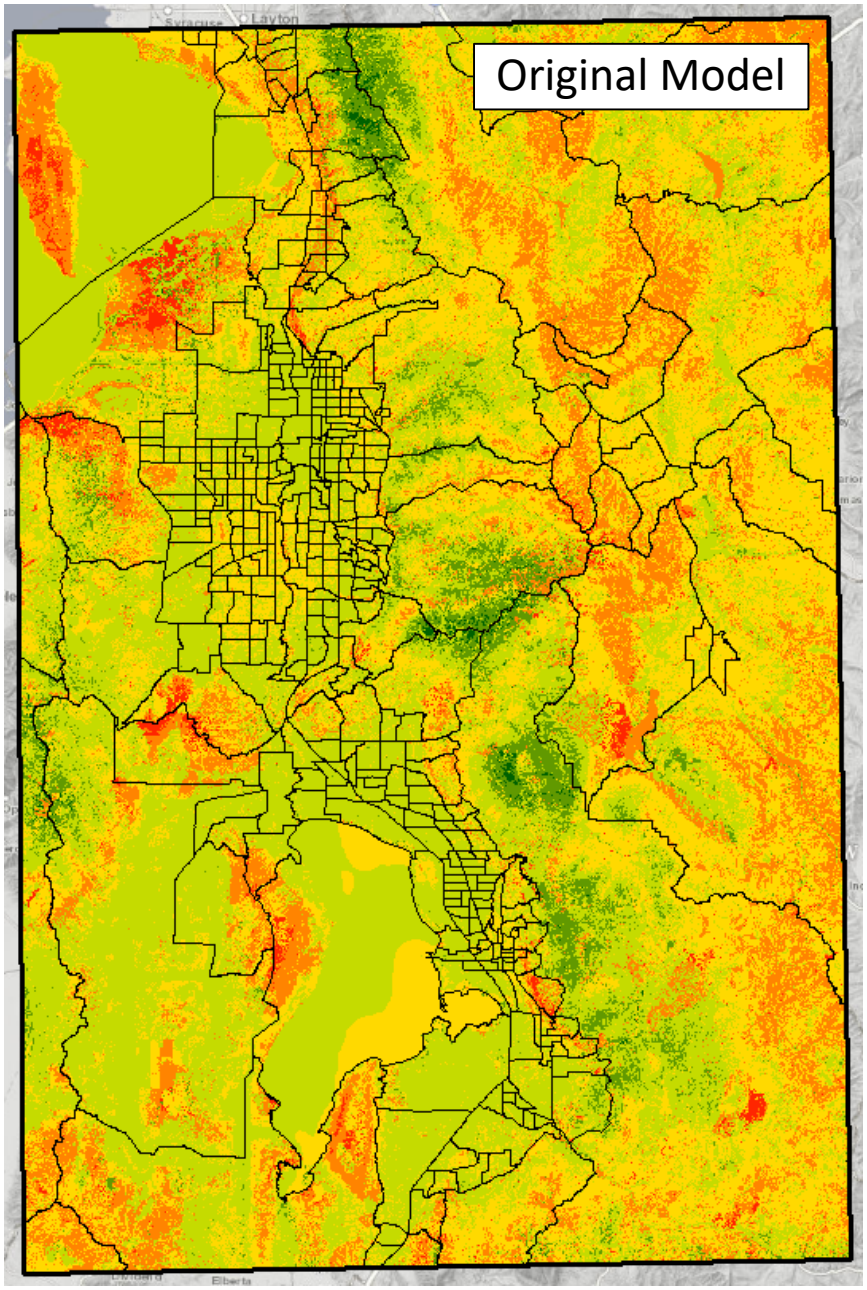
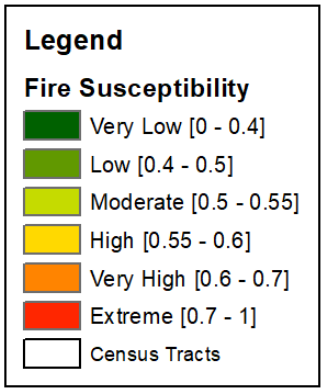


Figure 14: Fire Susceptibility Index from the Original model with census tracts outlined in black.



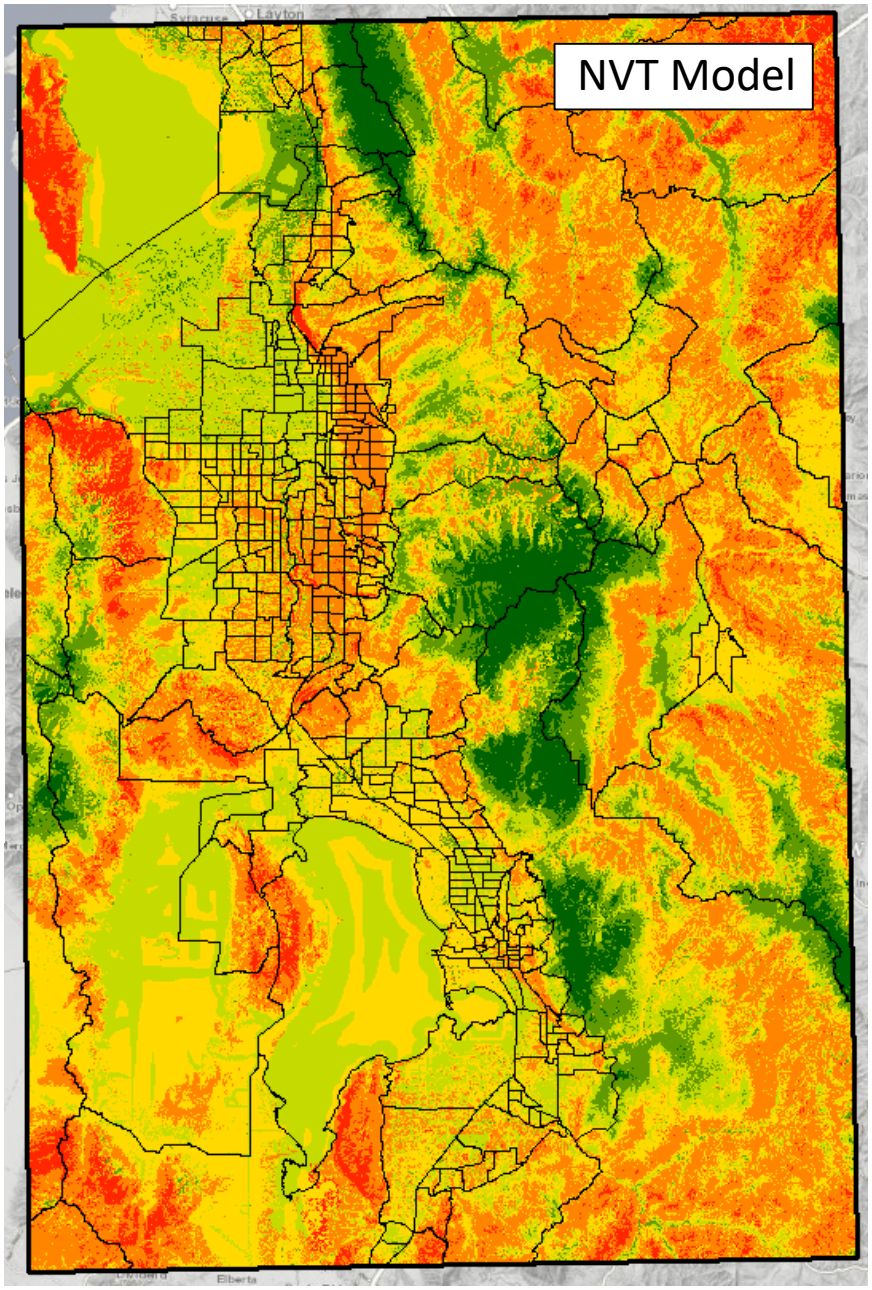


Figure 15: Fire Susceptibility Index from the No Vegetation Type Model (NVT) model with census tracts outlined in black.

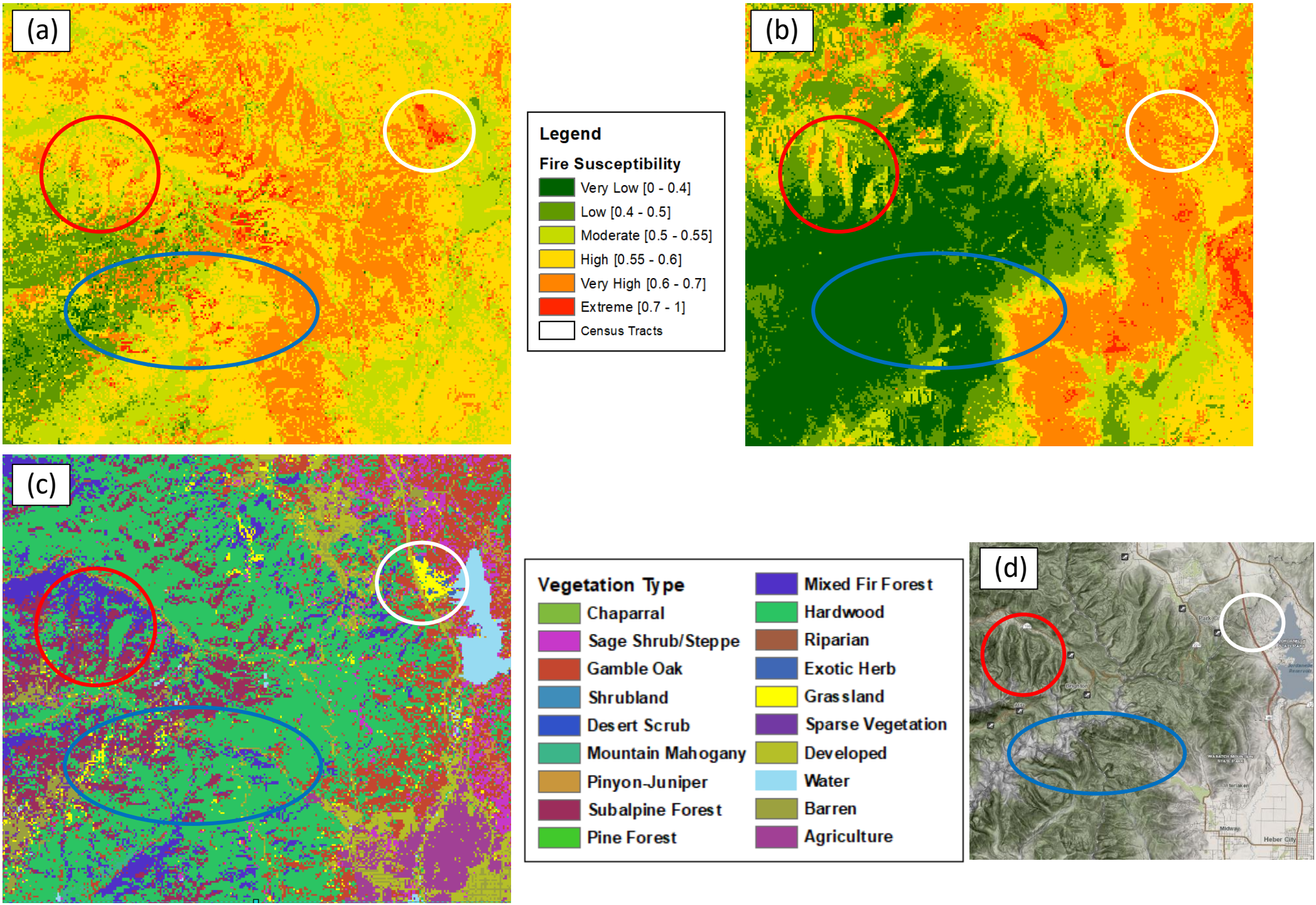


Figure 16: Comparison of Fire Susceptibility Index results with Vegetation Type, where (a) is Original model output, (b) is NVT model output, and (c) is Vegetation Type. (d) shows the geographic location of the graphics, focused on the Cottonwood Canyons (left half) with the Jordanelle Reservoir on the right and the town of Heber in the bottom-right.

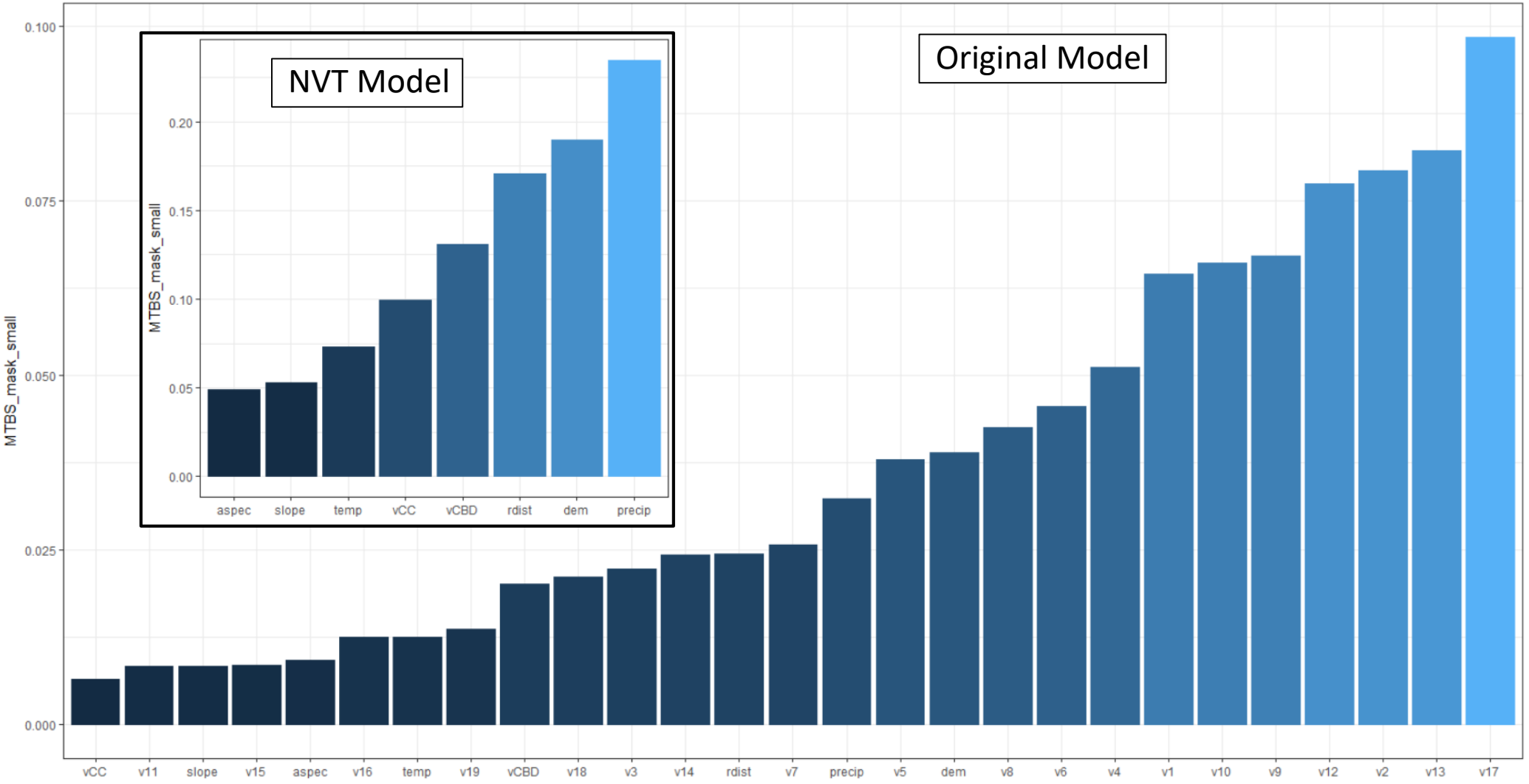


Figure 17: Variable Importance plots built from NN model weights. The larger plot is the Original model, the smaller, inset plot is from the NVT model.

(a)

	Original Model	NVT Model
Mean Squared Error (MSE)	2.1398	0.8981
Mean Absolute Error (MAE)	4.4087	1.8097
Area Under the Curve (AUC)	0.8875	0.8127

(b) Original Model

Category	Total Pixels	Fire Density (%)
Very Low	3250	0.00%
Low	52494	0.02%
Moderate	535672	0.15%
High	470312	1.47%
Very High	188226	7.96%
Extreme	15581	27.01%

(c) NVT Model

Category	Total Pixels	Fire Density (%)
Very Low	56313	0.00%
Low	104867	0.02%
Moderate	311982	0.21%
High	372354	0.94%
Very High	367234	3.71%
Extreme	52785	17.27%

Figure 18: Quantitative results comparing Original and NVT models. (a) Mean Squared Error, Mean Absolute Error, and Receiver Operating Characteristic Area Under the Curve metrics, (b, c) Fire Density metrics for Original and NVT Models, respectively.

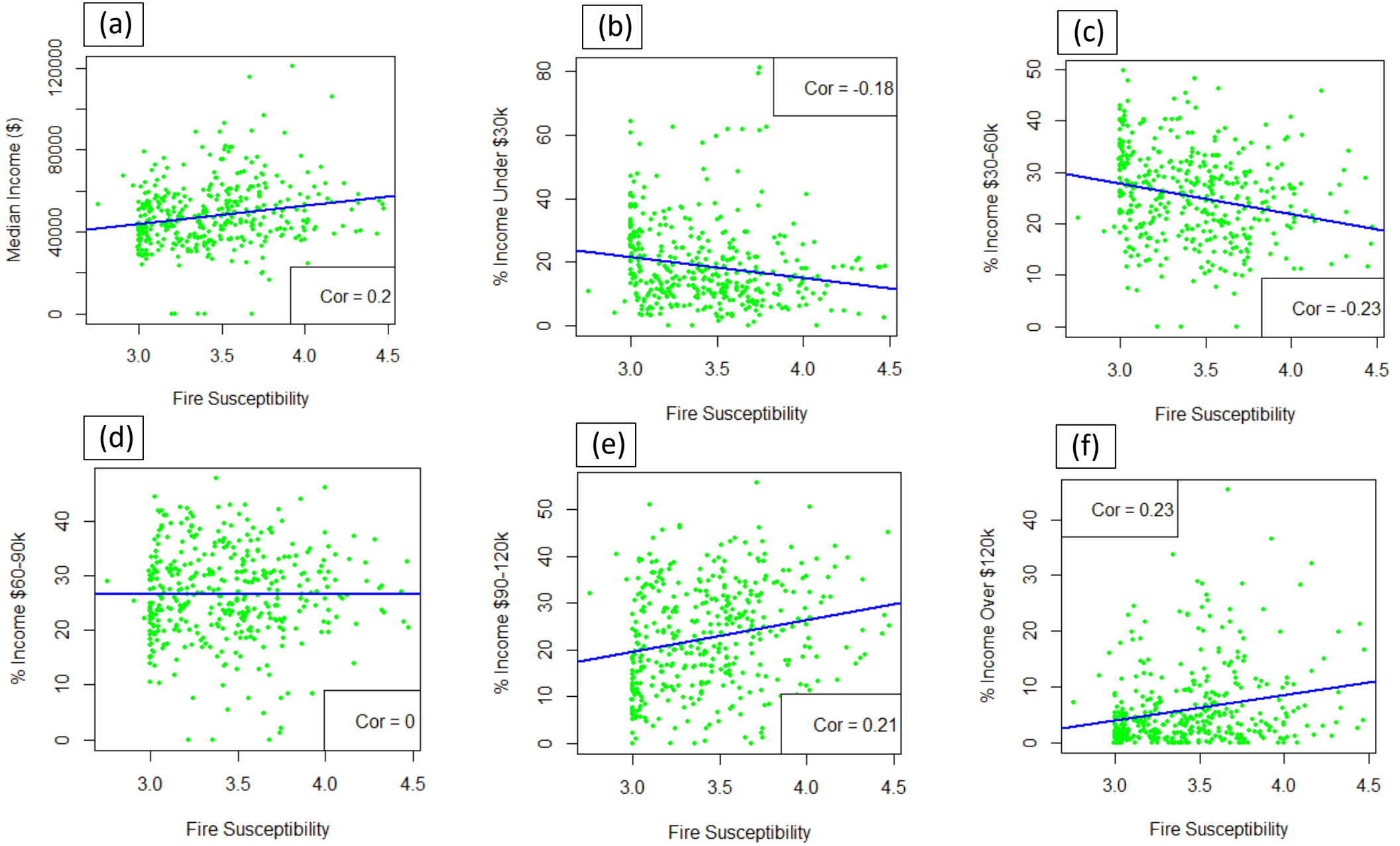


Figure 19: Demographic analysis: scatterplot of mean Fire Susceptibility Index (converted to 1-6 scale) and Income variables for Wasatch Front census tracts. Linear model trend line and correlation coefficient are also displayed. (a) Median income, (b) percent population under $30k, (c) percent population $30-60k, (d) percent population $60-90k, (e) percent population $90-120k, (f) percent population over $120k.

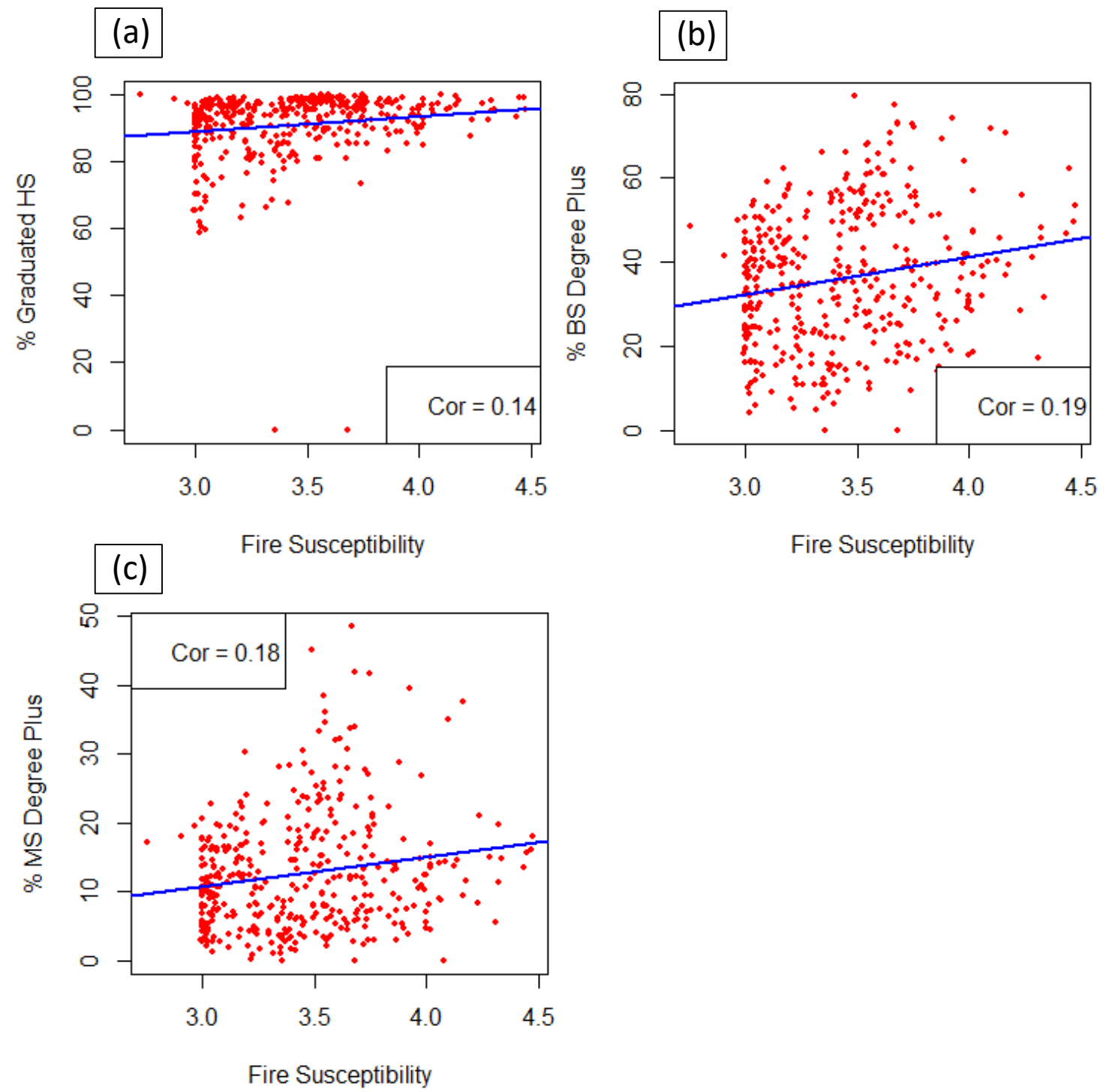


Figure 20: Demographic analysis: scatterplot of mean Fire Susceptibility Index (converted to 1-6 scale) and Education variables for Wasatch Front census tracts. Linear model trend line and correlation coefficient are also displayed. (a) percent population graduated high school, (b) percent population with bachelor's degree or higher, (c) percent population with master's degree or higher.

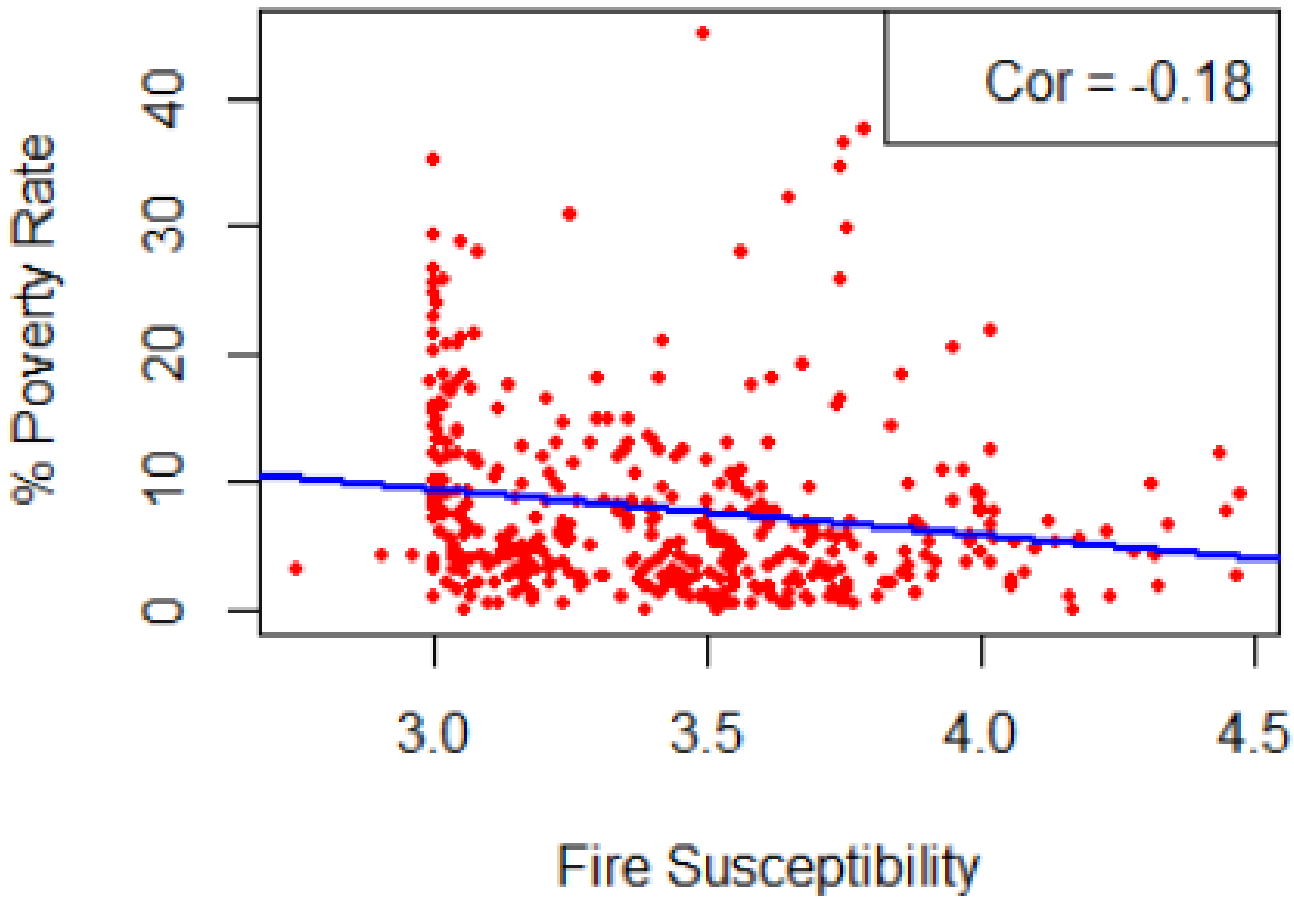


Figure 21: Demographic analysis: scatterplot of mean Fire Susceptibility Index (converted to 1-6 scale) and Poverty Rate for Wasatch Front census tracts. Linear model trend line and correlation coefficient are also displayed. Percentage of census tract families that are below the poverty threshold.