Combining Risk Factors: Detailed Calculations and Extended Results

Calvin A. Whealton, Jared D. Smith

September 30, 2015

Contents

This document outlines detailed calculations and extended results for the individual risk factor maps and the combined risk factor (common risk segment, CRS) maps. Individual risk factors are addressed first. For each individual risk factor the values used in converting the risk factor to the common play fairway scale (thresholds, minimum, and maximum), the frequency distribution of the risk factor, and the play fairway maps (3- and 5-color versions) of the variable are shown. Some risk factors have additional challenges or assumptions that were used in processing and these are also discussed.

The second section outlines the methods used to combine the individual risk factor maps into a single play fairway map (CRS map). In this section results are presented for combining the 3- and 5-color maps separately, including the distribution of the combined play fairway metric for the whole map. The combinations are completed separately for all four risk factors, the three geology variables (thermal, seismic, and reservoirs), and for a no reservoir version (utilization, thermal, and seismic).

The next section discusses the uncertainty in the maps. Although uncertainty maps are presented near their related maps, this section gives the details of those calculations. The two main issues discussed are the methods used to derive uncertainty of individual risk factor play fairway maps and the uncertainty of a combined map.

The last two sections show the robustness of the methods of combining the risk factors, and show some results for specific project locations. The robustness is illustrated by plotting the results of different combined play fairway metrics against each other. Project-specific results show the individual risk factors for a set of locations thought to be of interest for Phase 2.

Some common terms used throughout this report are defined below.
Play Fairway Metric (PFM): A formula used to generate an aggregate measure of how favorable a place (site, raster cell) is based on input risk factors. *Scaled Risk Factor (SRF)*: A value for an individual risk factor (RF) that has been converted to the play fairway scale of [0,3] or [0,5], depending on the map. *Thresholds*: Values used to assign an input risk factor to the range [0,3] or [0,5], depending on the map.

1 Individual Risk Factors

Individual risk factors are presented in the following sections. If there were special considerations when calculating the risk factor, these are outlined as well. Generally, the input risk factor layer was converted into the play fairway metric by linearly scaling the continuous risk factor value between the provided thresholds. Also, if a value was greater than the maximum or less than the minimum it was assigned the minimum (0) or maximum (3 or 5) on the play fairway scale for that risk factor. In the map figures, areas in white are areas with no data for the calculation because the risk factor could not be evaluated at that location, or because the area was outside the region. The histograms show the location of the thresholds with respect to the original input risk factor map.

1.1 Thermal

The depth to 80 °C is used as the map for the thermal risk factor as per the Statement of Project Objectives (SOPO) requirements. The thresholds are based on the memo titled "Assignment of Thresholds for Depth-to-Temperature and Temperature-at-Depth Maps" (July 15, 2015). Table 1 provides the scaling values and thresholds, Figure 1 shows the histogram of the input risk factor map, and Figures 2 and 3 show the 3- and 5-color maps for the scaled values.

There were no special calculation considerations for this risk factor. All conversions used a linear scale.

The uncertainty was estimated using the mean and standard error of prediction. The distribution was assumed to be normal, with mean and standard deviation defined from the mean and standard error of prediction. After this, the methods discussed in the uncertainty section were used to derive the uncertainty maps shown in Figures 4 and5.

Table 1: Table of minimum, maximum, and thresholds used in scaling the thermal map (depth to 80 °C [m]). The scale is reversed so that high values are unfavorable because shallow depths mean reduced cost.

Min	8750
Max	500
3-Color Thresholds	{3000, 2000}
5-Color Thresholds	{4000, 3000, 2300, 1500}

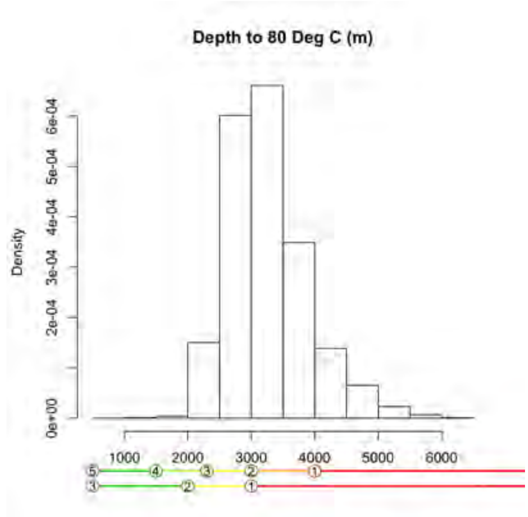


Figure 1: Histogram of the thermal risk factor with the 3- and 5-color thresholds noted. Points beyond the minimum and maximum of the scale are assigned the minimum and maximum of the scale, respectively (e.g. 0, and 3 or 5). Density is proportional to the frequency or count in the bins, but the values have been rescaled so the total area of the histogram is 1.

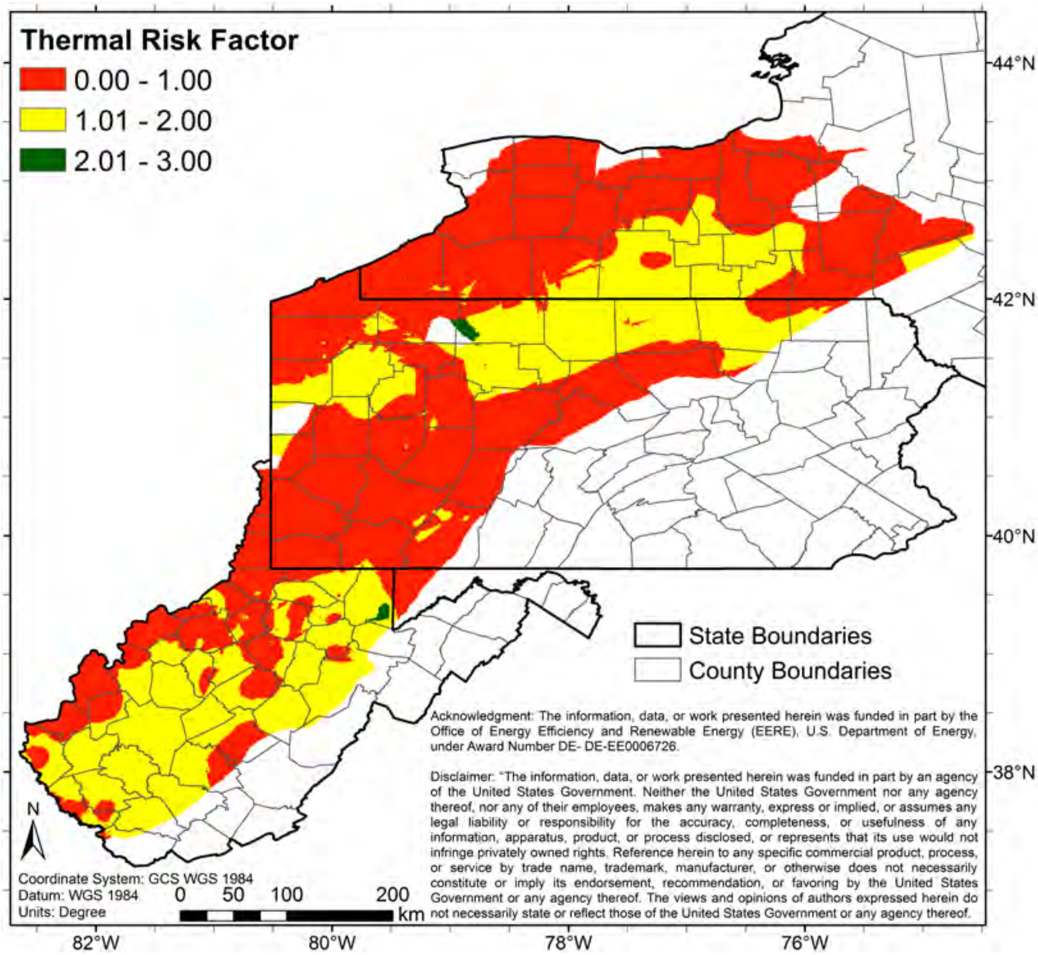


Figure 2: Map of the thermal risk factor with a 3-color scheme. Red areas are unfavorable and green areas are favorable values of the scaled risk factor.

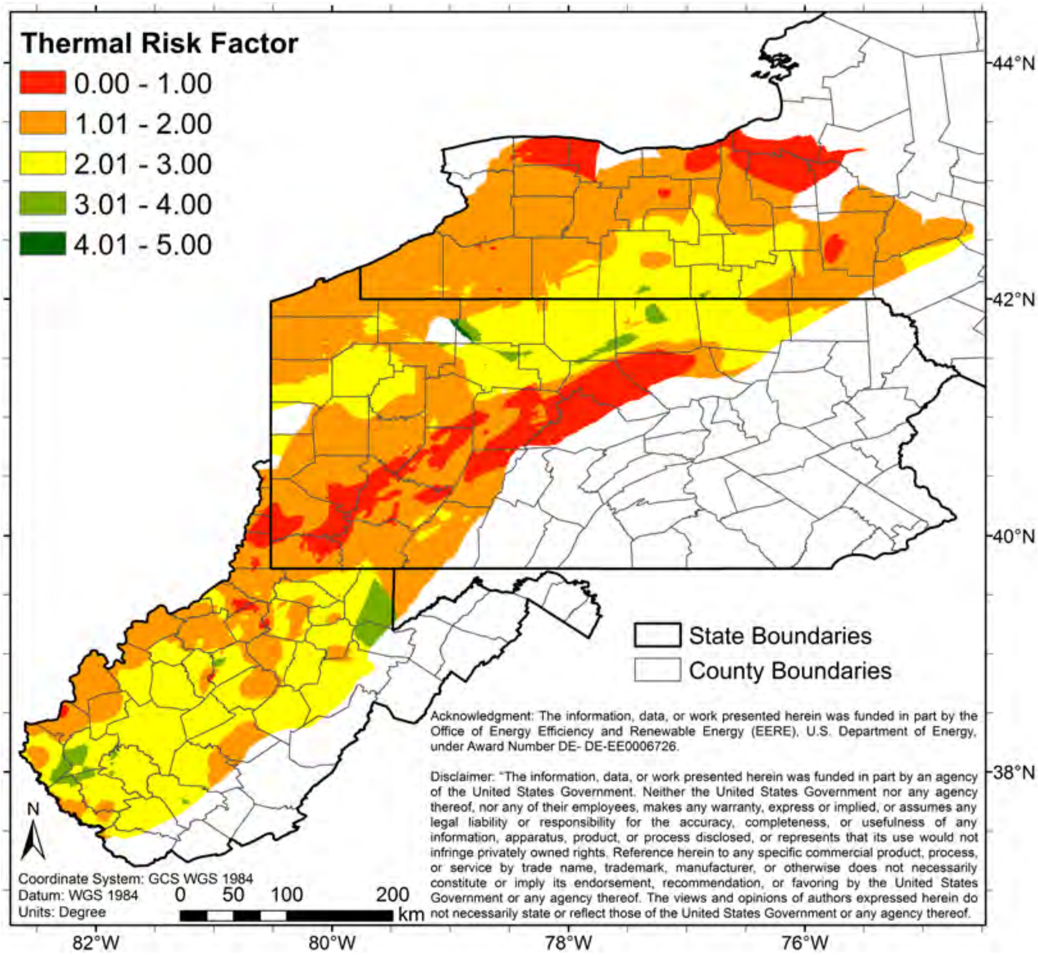


Figure 3: Map of the thermal risk factor with a 5-color scheme. Red areas are unfavorable and green areas are favorable values of the scaled risk factor.

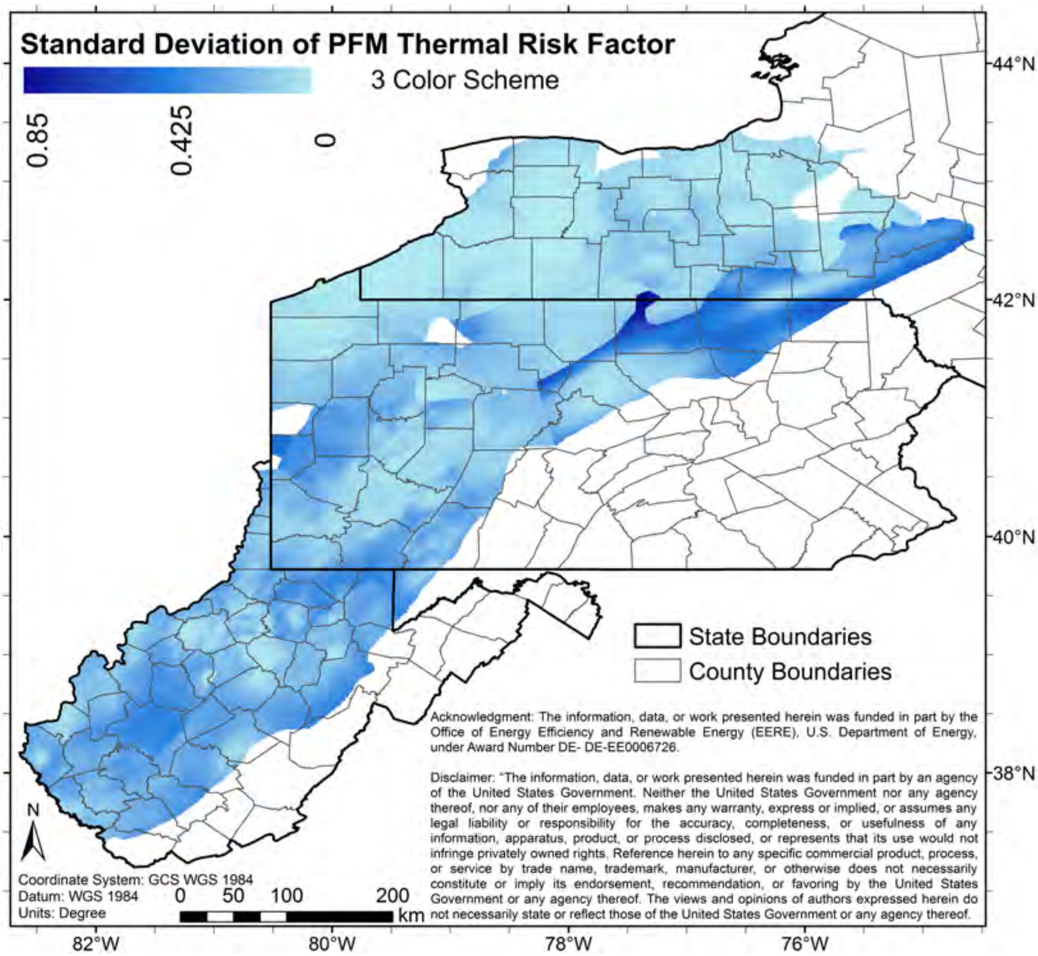


Figure 4: Map of the standard deviation of the scaled risk factor for thermal for a 3-color scheme. Dark colors are less certain and light colors are more certain.

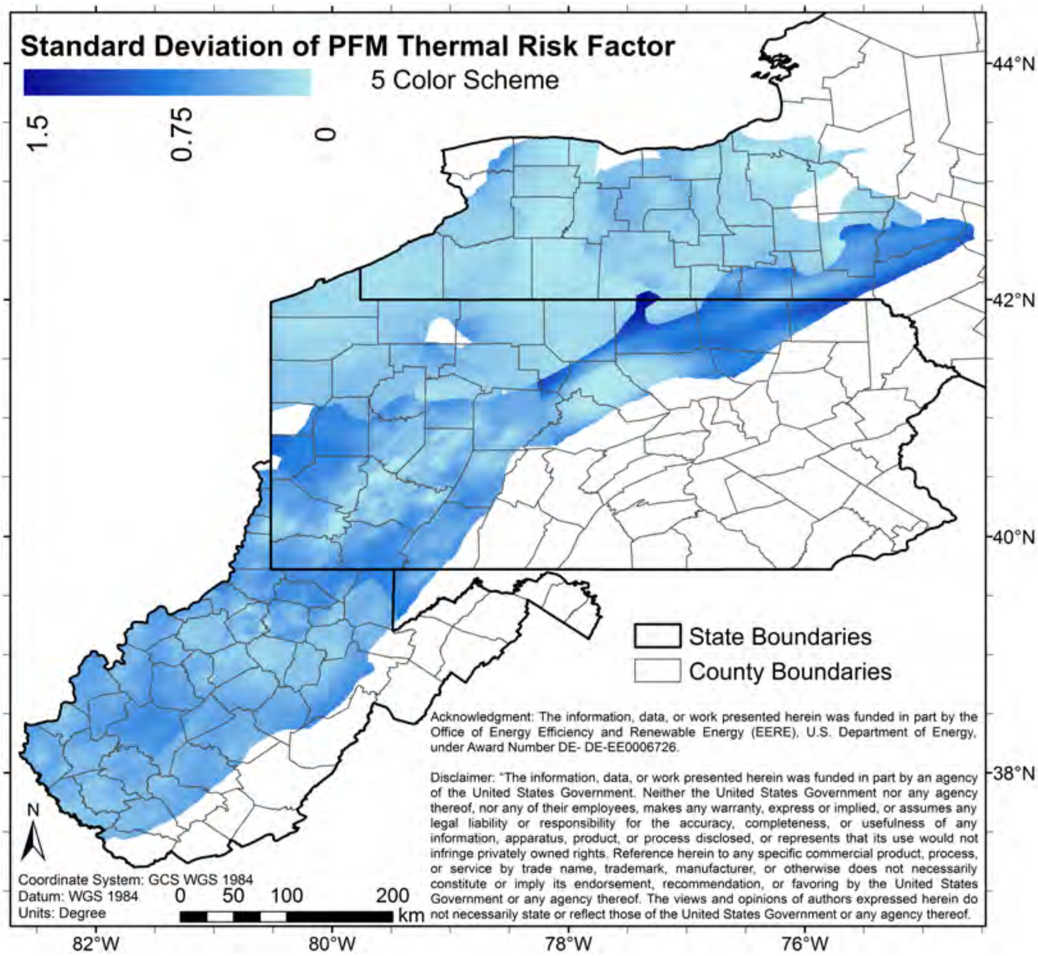


Figure 5: Map of the standard deviation of the scaled risk factor for thermal for a 5-color scheme. Dark colors are less certain and light colors are more certain.

1.2 Reservoirs

The reservoir play fairway risk is in terms of reservoir productivity index (L/MPa-s). The reservoir productivity index is the calculated expression of the reservoir's favorability for geothermal applications. More information is available in the memo "Natural Reservoirs Task Methodology" (September 15, 2015). The maps are divided into different depth slices, which were combined into a single risk factor map by taking the maximum (high values are good) value at each spatial location. Reservoirs shallower than 1000 m were ignored in the combination because they will likely not meet the minimum temperature requirements for our projects. The remainder of the depth slices are for 500 m intervals with the shallowest slice being 1000-1500 m and the deepest being 3500-4000 m. Table 2 gives the thresholds, Figure 6 gives the histogram of the reservoir productivity index (transformed to base-10 log), and Figures 7 and 8 plot the 3- and 5-color maps of the play fairway conversion.

The thresholds are defined mainly for orders of magnitude, so the conversion was linear on a logarithmic scale. The steps in this analysis were first to convert the raster of the maximums by taking the base-10 log of the values. Similarly, the thresholds were converted to base-10 logs. Once the raster and the thresholds were converted, the conversion was linear.

The uncertainty of the reservoir map was calculated by assuming that the mean and coefficient of variation provided defined a log-normal distribution. The parameters were specified in real-space, not log-space, so they had to be converted. The log-space variance, σ^2, can be solved for from the coefficient of variation, CV, as shown in Equation 1. The log-space mean, μ, can be solved using the real-space mean, m, and the log-space variance, σ^2, as shown in Equation 2. The uncertainty of the individual maps was then calculated based the methods discussed in the uncertainty section of this document and the results are shown in Figures 9 and10.

$$\sigma^2 = \ln(1 + \mathrm{CV}^2) \tag{1}$$

$$\mu = \ln(m) - \frac{\sigma^2}{2} \tag{2}$$

Table 2: Table of minimum, maximum, and thresholds used in scaling the reservoir map. The measure of risk is the reservoir productivity index (L/MPa-s).

Min	3×10^{-5}
Max	301
3-Color Thresholds	{0.1,1}
5-Color Thresholds	{0.01,0.1,1.0,10}

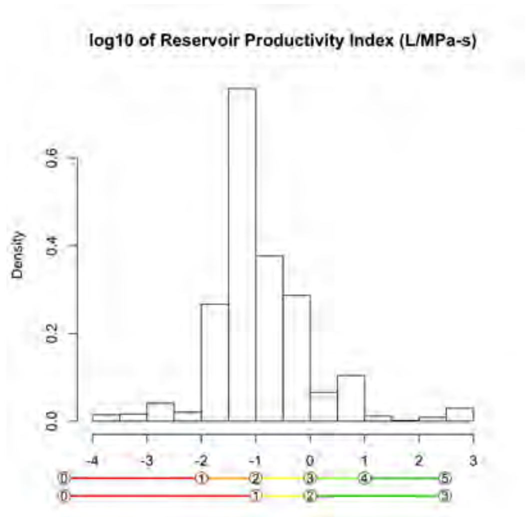


Figure 6: Histogram of the reservoir risk factor with the 3- and 5-color thresholds noted. Points beyond the minimum and maximum of the scale are assigned the minimum and maximum of the scale, respectively (e.g. 0, and 3 or 5). Density is proportional to the frequency or count in the bins, but the values have been rescaled so the total area of the histogram is 1.

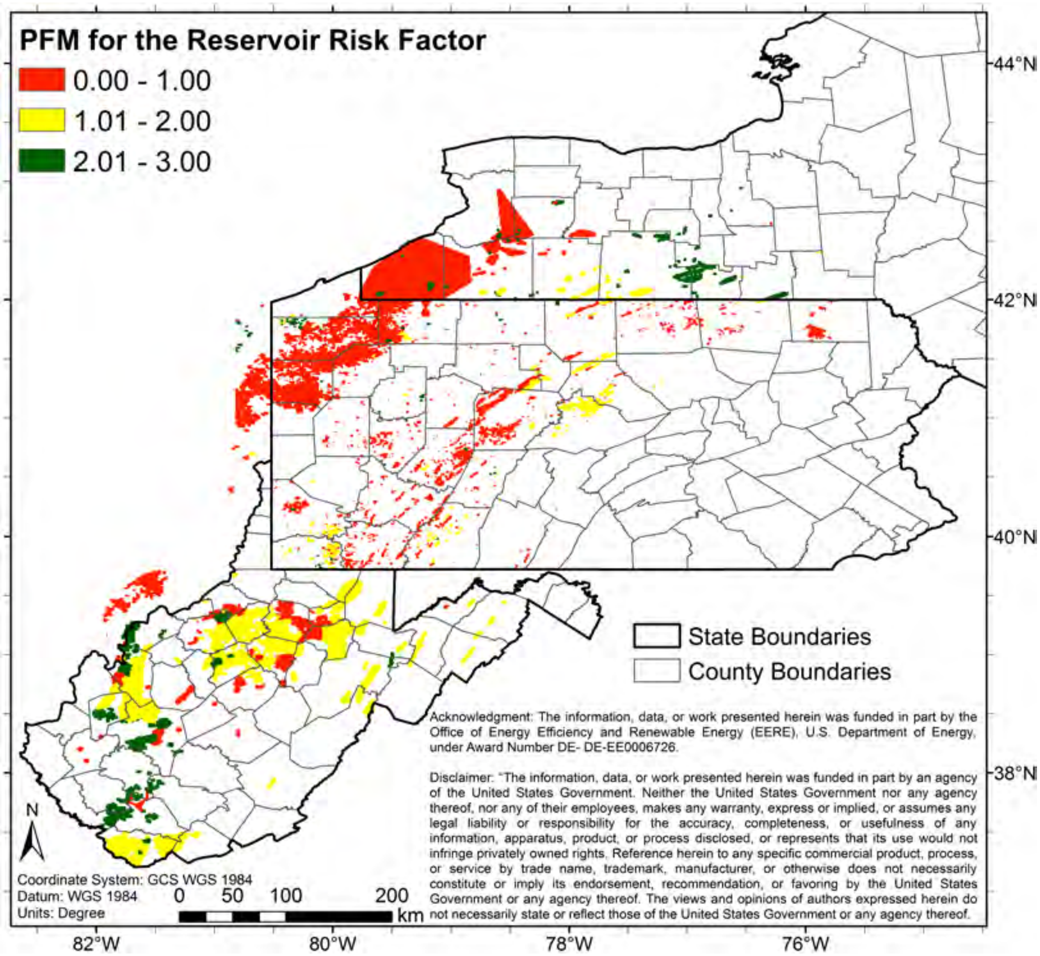


Figure 7: Map of the reservoir risk factor with a 3-color scheme. Red areas are unfavorable and green areas are favorable values of the scaled risk factor.

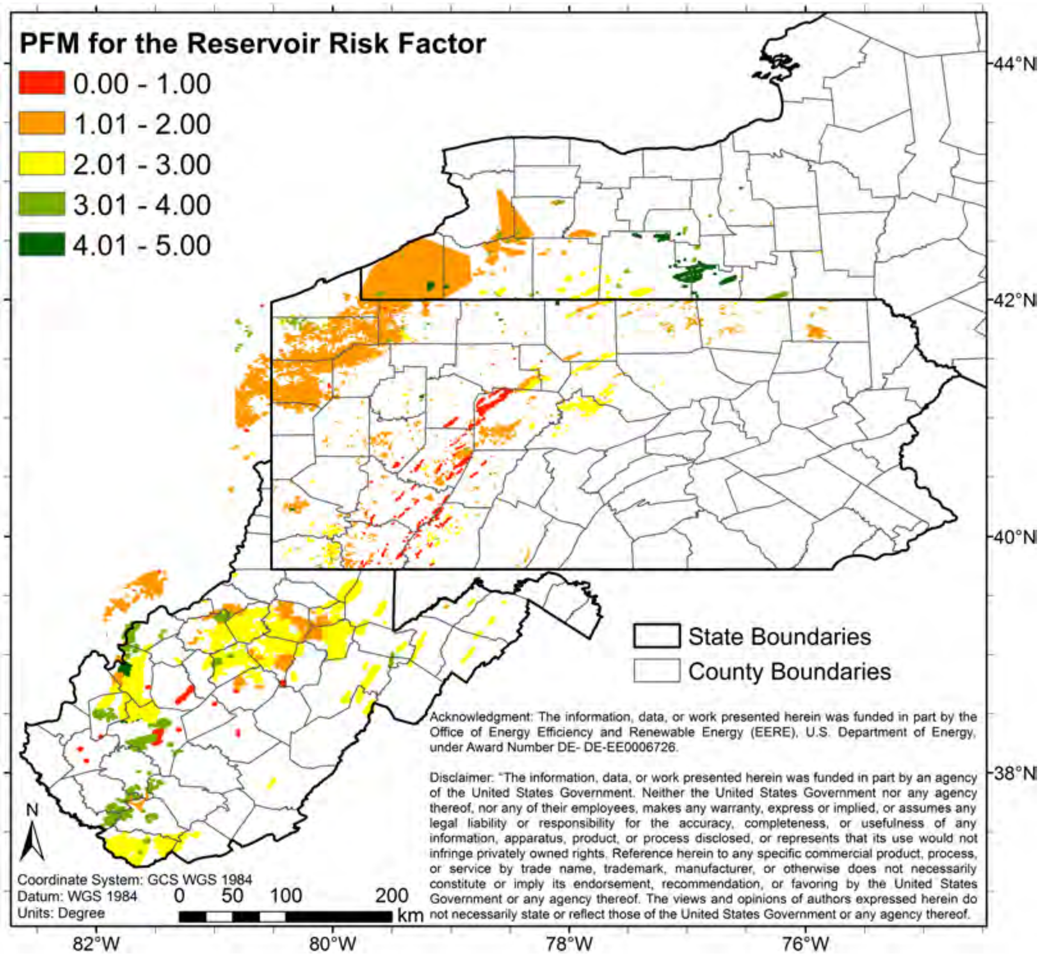


Figure 8: Map of the reservoir risk factor with a 5-color scheme. Red areas are unfavorable and green areas are favorable values of the scaled risk factor.

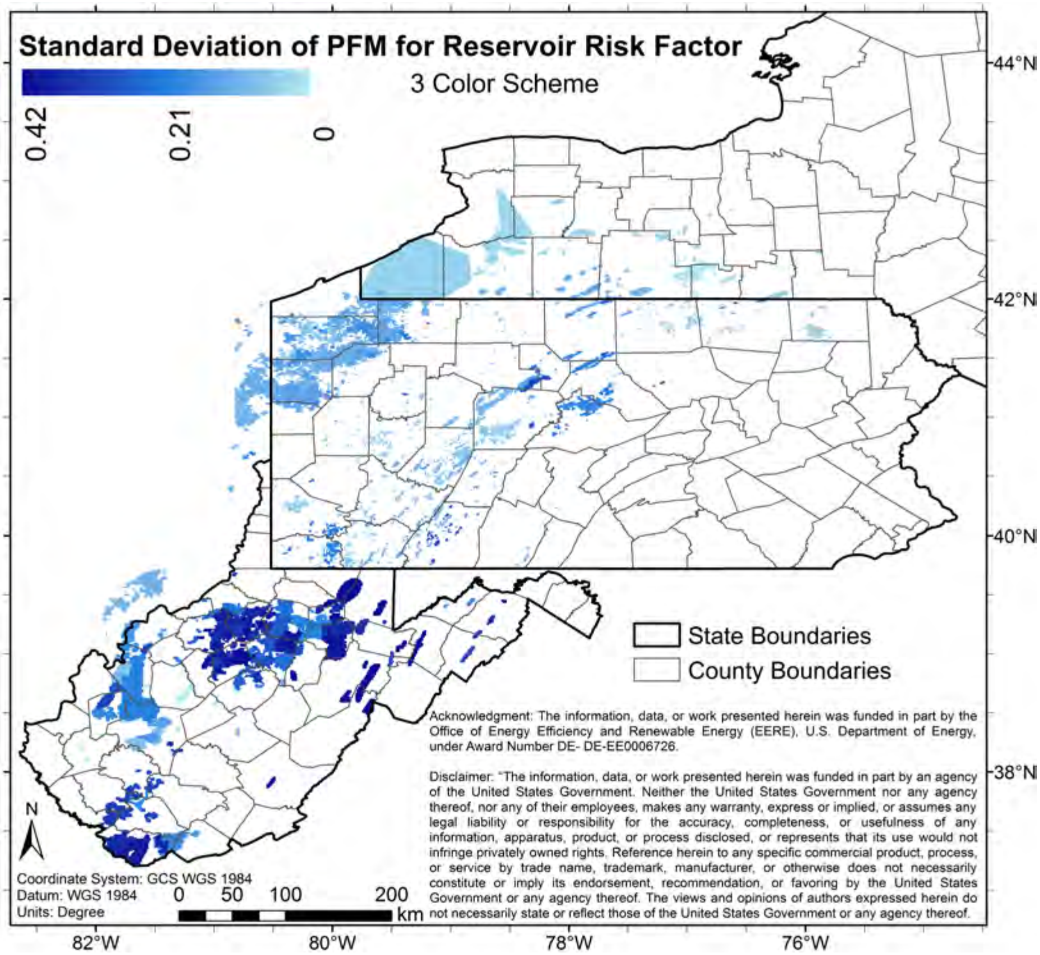


Figure 9: Map of the standard deviation of the scaled risk factor for reservoirs for a 3-color scheme. Dark colors are less certain and light colors are more certain.

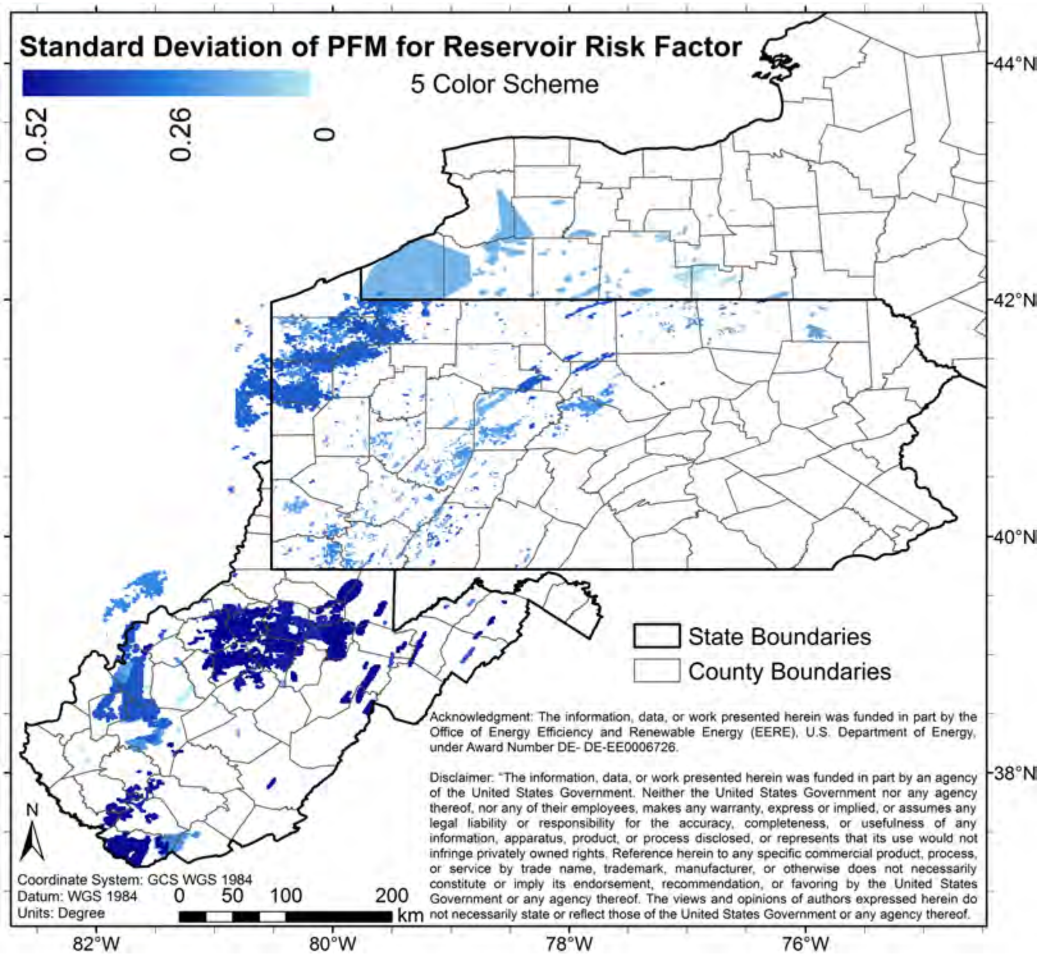


Figure 10: Map of the standard deviation of the scaled risk factor for reservoirs for a 5-color scheme. Dark colors are less certain and light colors are more certain.

1.3 Seismic

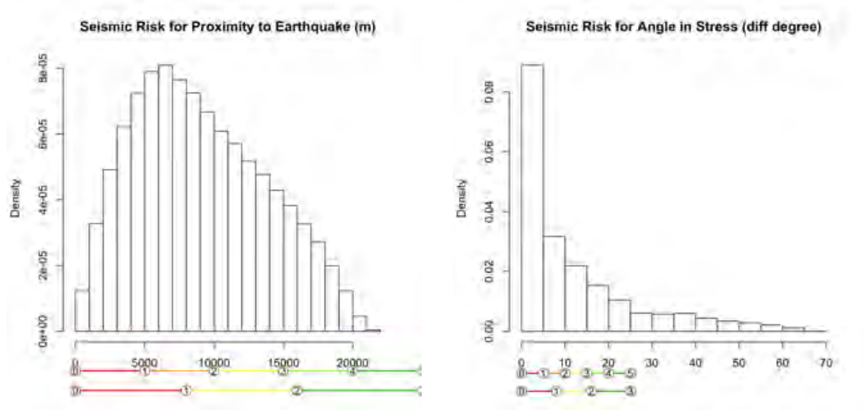
The seismic risk factor had two separate measures of risk: one for proximity to earthquakes and one for orientation of "worms" to the stress field. These were considered separate estimates of the risk, so they were averaged to create a seismic risk map. Both were converted into the play fairway scale before averaging. The stress map has values calculated as the minimum angle between the "worm" and the stress field failure angle. The earthquake map uses proximity to earthquakes and has units of meters.

Table 3 presents the thresholds used in creating the maps and Figure 11 plots the histograms of the earthquake- and stress-based risks. Figures 12 and 13 are the earthquake based risk maps for 3- and 5-color schemes, Figures 16 and 17 are the stress based risk maps for 3- and 5-color schemes, and Figures 20 and 21 are the averaged seismic risk map for the 3- and 5-color schemes. The same scale is used in the averaged map as in the input maps. The averaged seismic maps show that the areas with earthquakes are still high risk, but many of the "worms" with favorable orientation but no earthquakes nearby have been discounted in the average.

The uncertainty for the earthquake-based map was calculated by assuming a normal distribution with mean and standard deviation given as the prediction and the error, respectively. Because being -1 km from a worm should be equivalent to being 1 km from a worm, the absolute value was taken. The methods described in the uncertainty section were used to develop uncertainty maps, which are shown in Figures 14 and 15.

The uncertainty for the stress-based map was calculated by assuming that the mean and standard deviation defined a normal distribution. However, in this case the maximum distance from an angle of failure is approximately 65.2 degrees. Therefore, the absolute values of the normal values was taken. Next, the scaled risk factor was calculated for values between 0 and 65.2 degrees. Next, values between 65.2 were changed to 130.4 minus the value, because being 130.4 degrees from one angle of failure implies being close to another angle of failure. All values greater than 130.4 degrees away were reduced by 130.4 and the process was repeated. This was repeated many times, and then the methods described in the uncertainty section were used to develop uncertainty maps for this metric, which are shown in Figures 18 and 19.

When the two stress maps were averaged (multiplied by 0.5 each), the variance of the average could be calculated by summing the two individual variance uncertainty maps but multiplying the total by a value of 0.25. The results are given in Figures 22 and 23. Generally, the earthquake-based maps high much higher certainty than the stress angle-based maps (more details in



(a) Earthquake-based (m). (b) Stress-based (degrees).

Figure 11: Histograms of the seismic risk factor for earthquakes and stress with the 3- and 5-color thresholds noted. Points beyond the minimum and maximum of the scale are assigned the minimum and maximum of the scale, respectively (e.g. 0, and 3 or 5). Density is proportional to the frequency or count in the bins, but the values have been rescaled so the total area of the histogram is 1.

the memos "Identifying Potentially Activatable Faults for the Appalachian Basin Play Fairway Analysis" and "Conversion of Seismic Risk Data to Risk Maps").

Table 3: Table of minimum, maximum, and thresholds used in scaling the seismic maps. Earthquake units are in meters and stress units are in degrees. Minimum values were not set to zero to avoid possible numerical problems.

Scaling	Earthquake	Stress Angle
Min	0.1	0.001
Max	25000	25
3-Color Thresholds	{8000, 16000}	{8, 16}
5-Color Thresholds	{5000, 10000, 15000, 20000}	{5, 10, 15, 20}

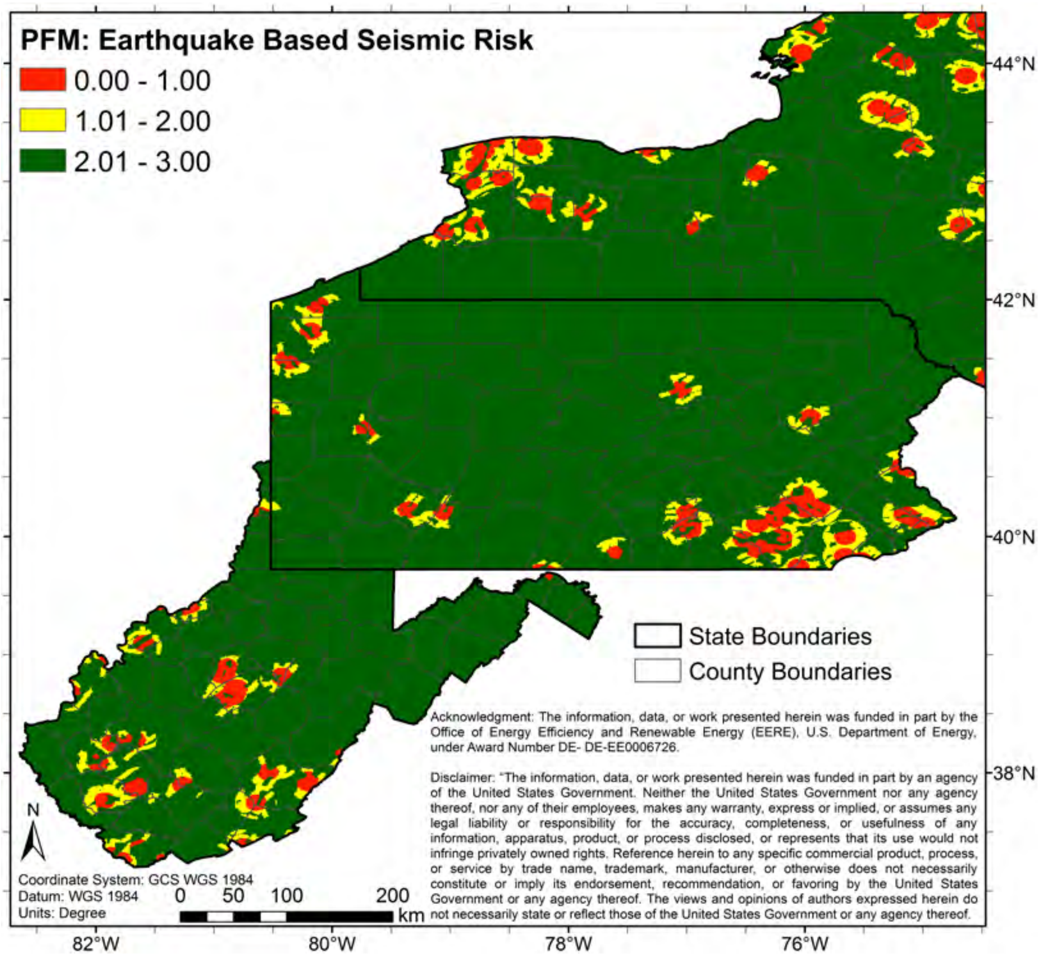


Figure 12: Map of the seismic risk factor for earthquakes with a 3-color scheme. Red areas are unfavorable and green areas are favorable values of the scaled risk factor.

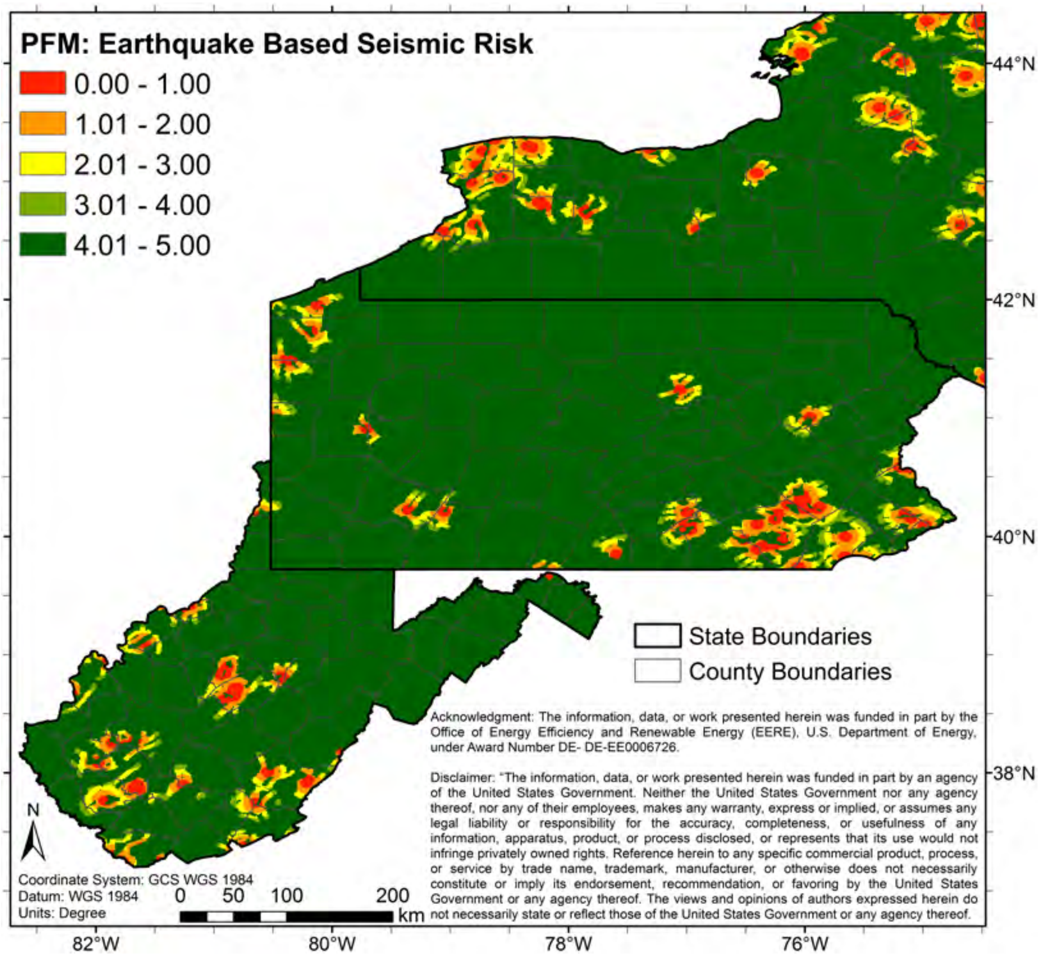


Figure 13: Map of the seismic risk factor for earthquakes with a 5-color scheme. Red areas are unfavorable and green areas are favorable values of the scaled risk factor.

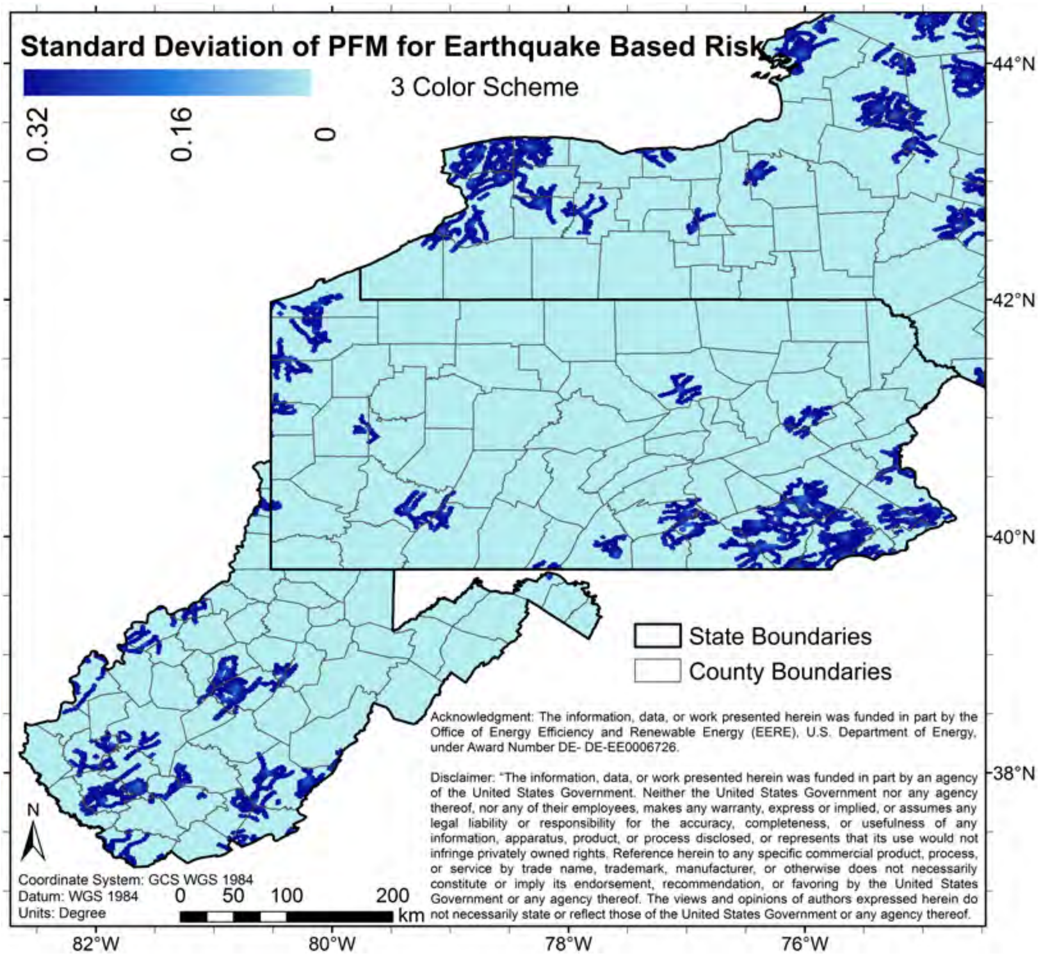


Figure 14: Map of the standard deviation of the scaled risk factor for earthquake-based seismic risk factor for a 3-color scheme. Dark colors are less certain and light colors are more certain.

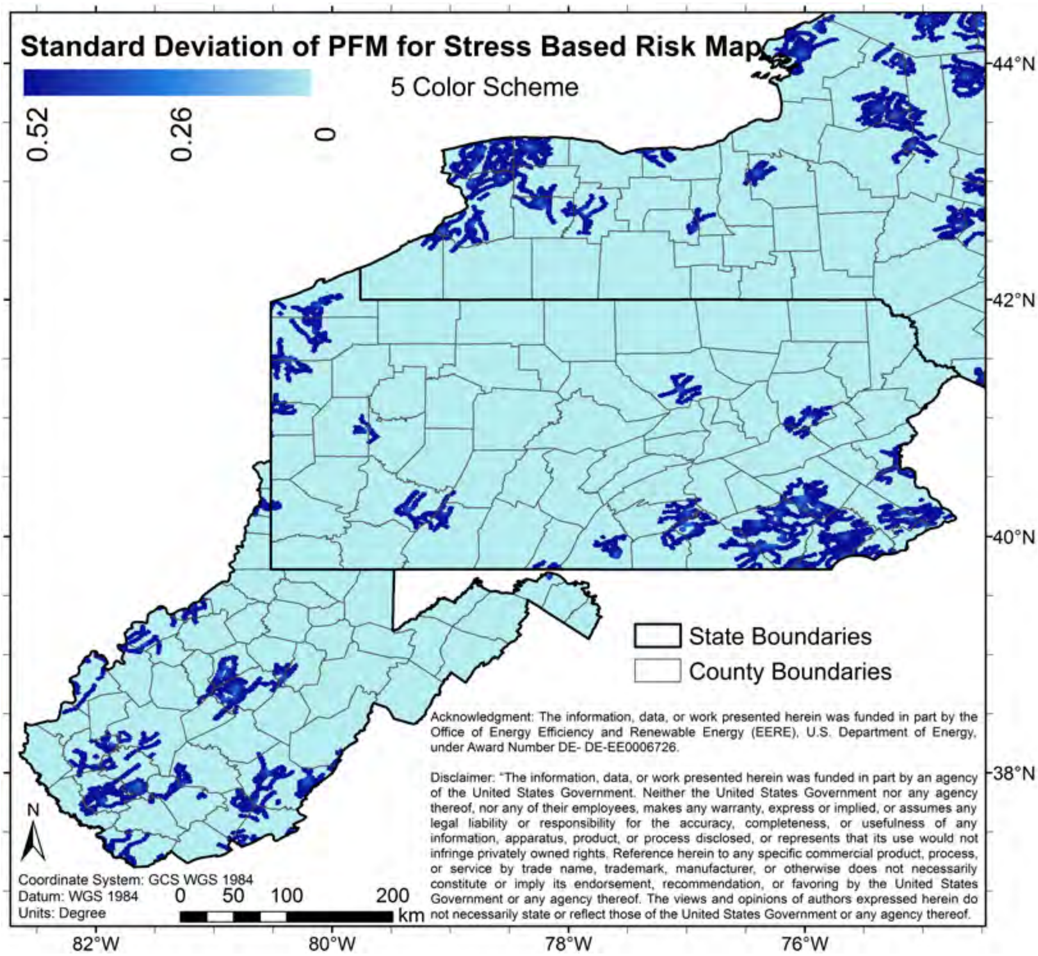


Figure 15: Map of the standard deviation of the scaled risk factor for earthquake-based seismic risk factor for a 5-color scheme. Dark colors are less certain and light colors are more certain.

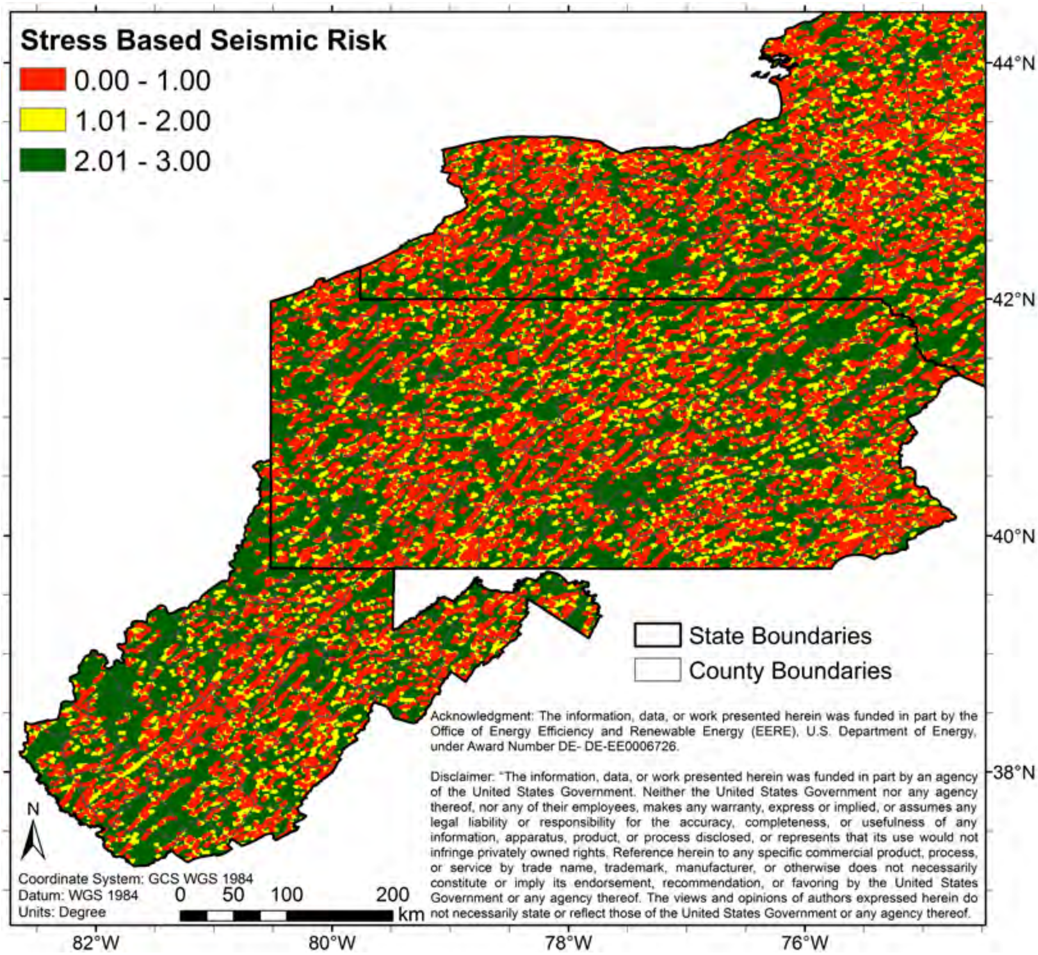


Figure 16: Map of the seismic risk factor for stress with a 3-color scheme. Red areas are unfavorable and green areas are favorable values of the scaled risk factor.

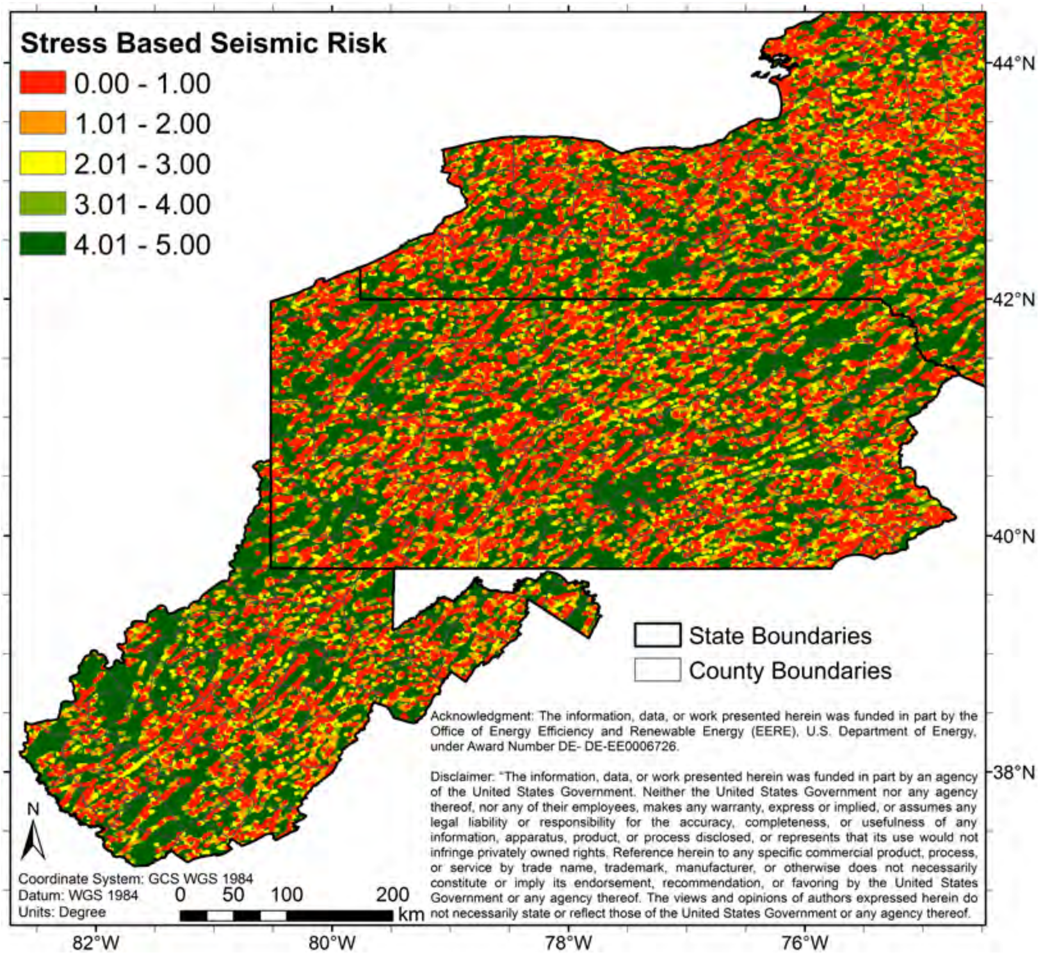


Figure 17: Map of the seismic risk factor for stress with a 5-color scheme. Red areas are unfavorable and green areas are favorable values of the scaled risk factor.

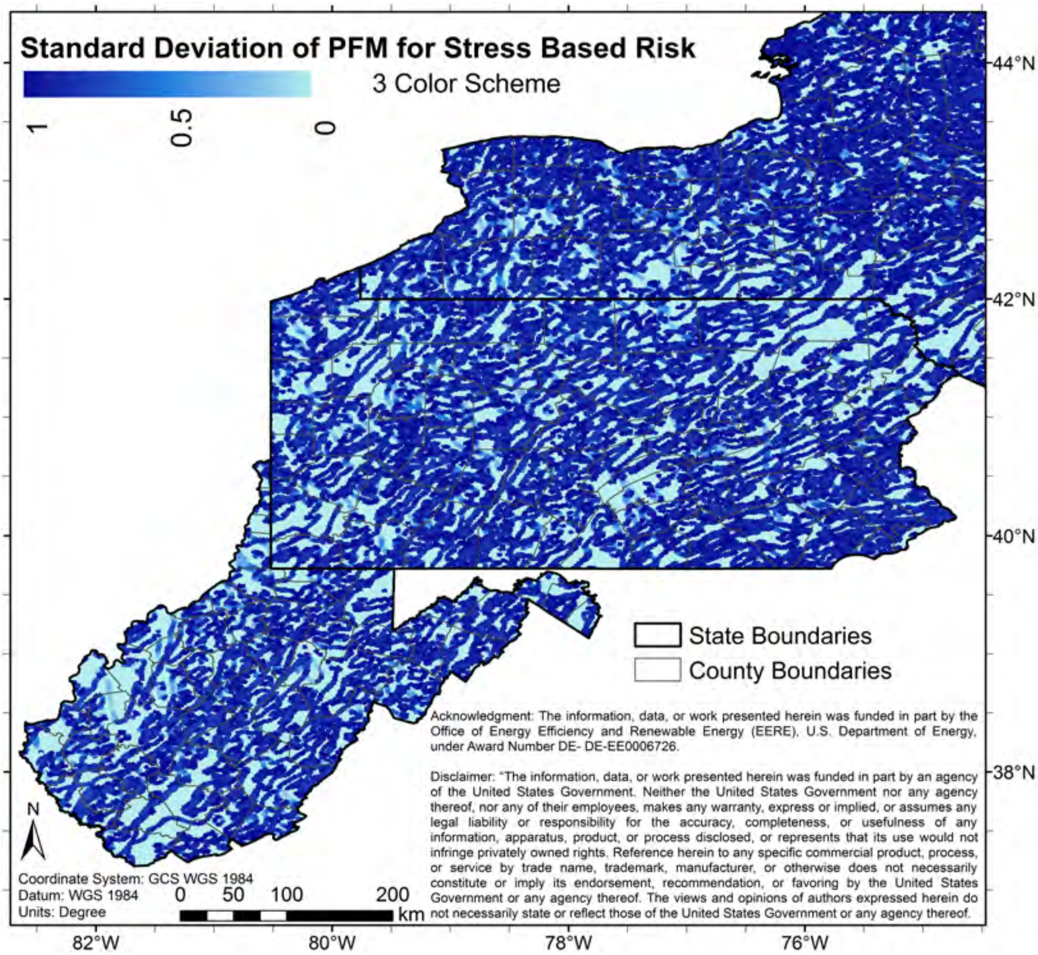


Figure 18: Map of the standard deviation of the scaled risk factor for stress-based seismic risk factor for a 3-color scheme. Dark colors are less certain and light colors are more certain.

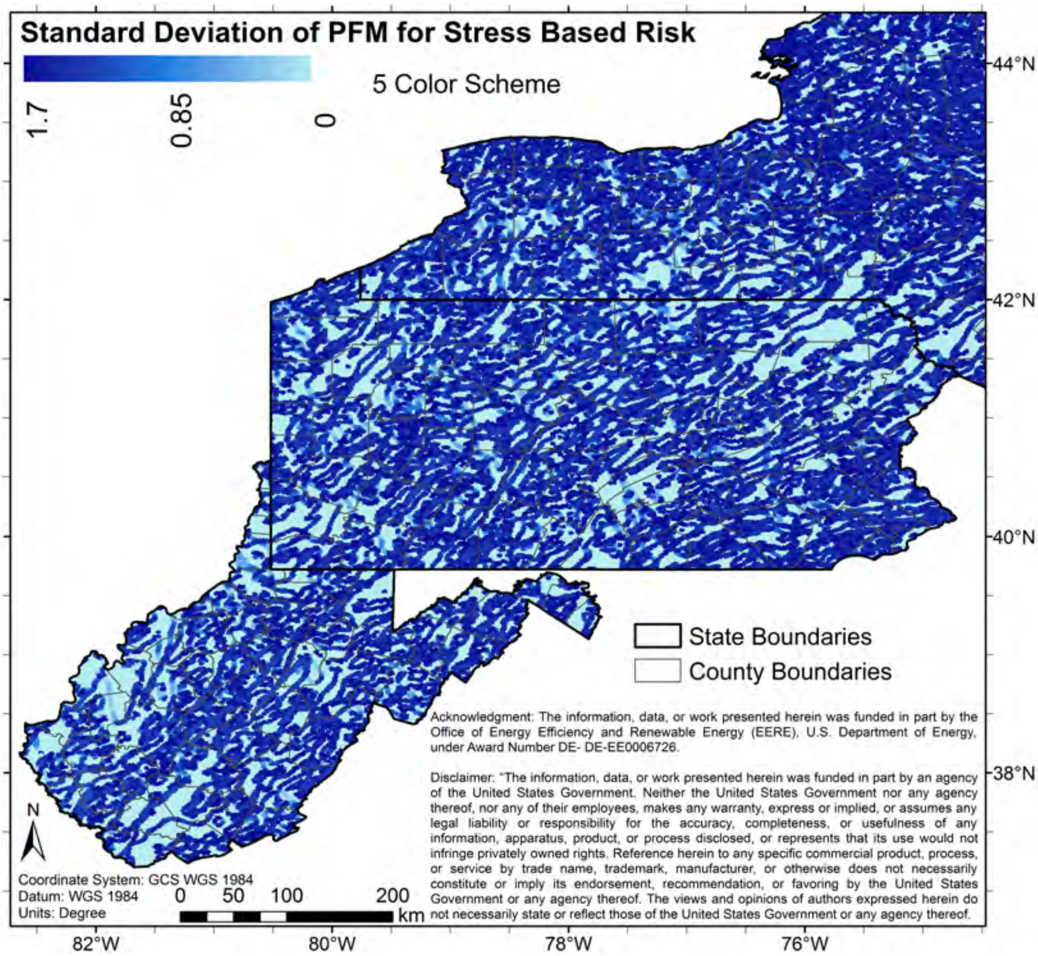


Figure 19: Map of the standard deviation of the scaled risk factor for stress-based seismic risk factor for a 5-color scheme. Dark colors are less certain and light colors are more certain.

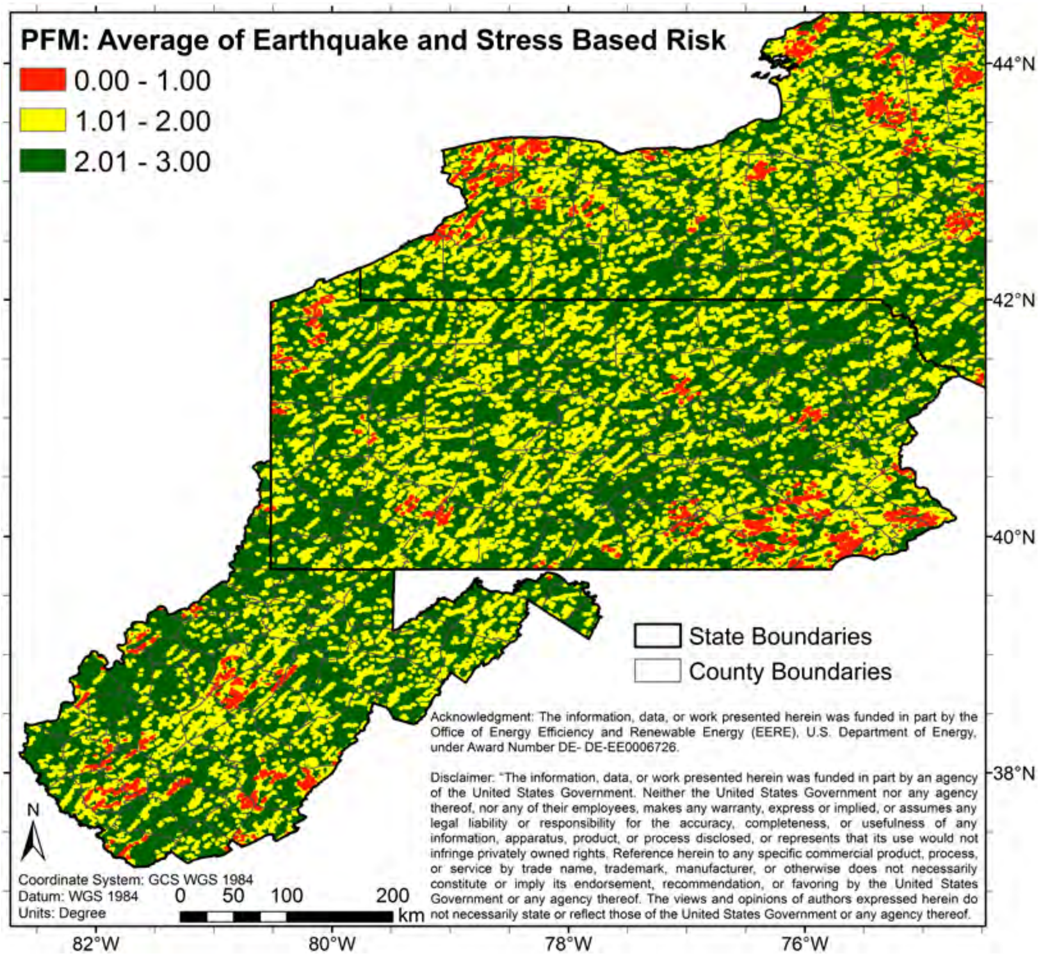


Figure 20: Map of the seismic risk factor (averaged stress and earthquake) with a 3-color scheme. Red areas are unfavorable and green areas are favorable values of the scaled risk factor.

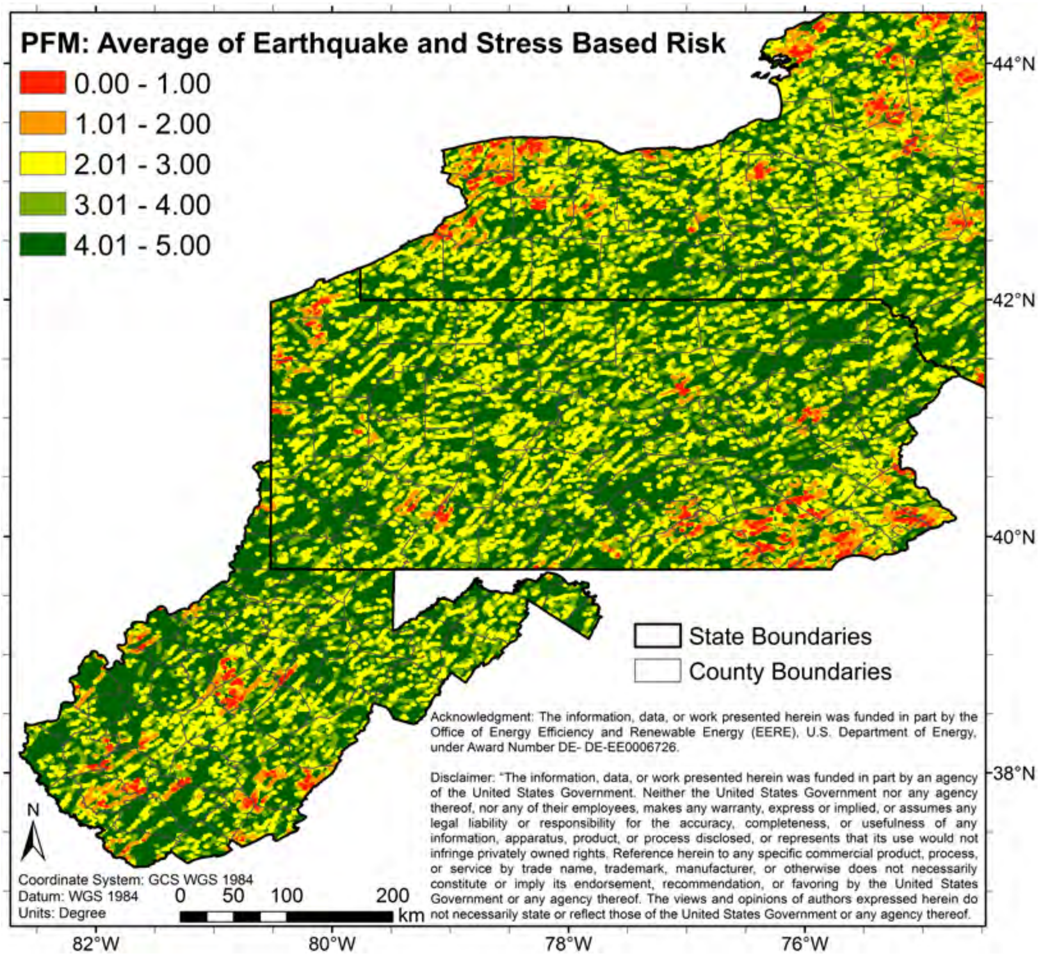


Figure 21: Map of the seismic risk factor (averaged stress and earthquake) with a 5-color scheme. Red areas are unfavorable and green areas are favorable values of the scaled risk factor.

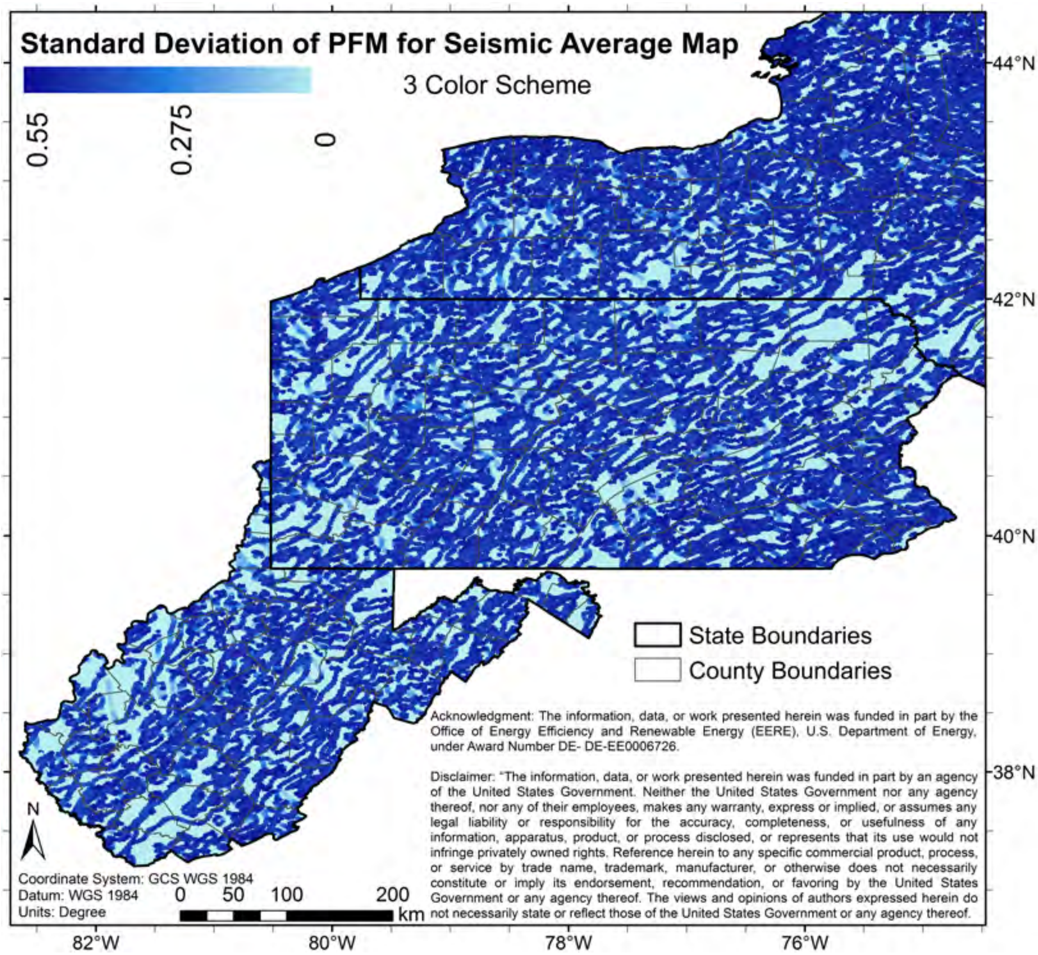


Figure 22: Map of the standard deviation of the scaled risk factor for averaged seismic risk factor for a 3-color scheme. Dark colors are less certain and light colors are more certain.

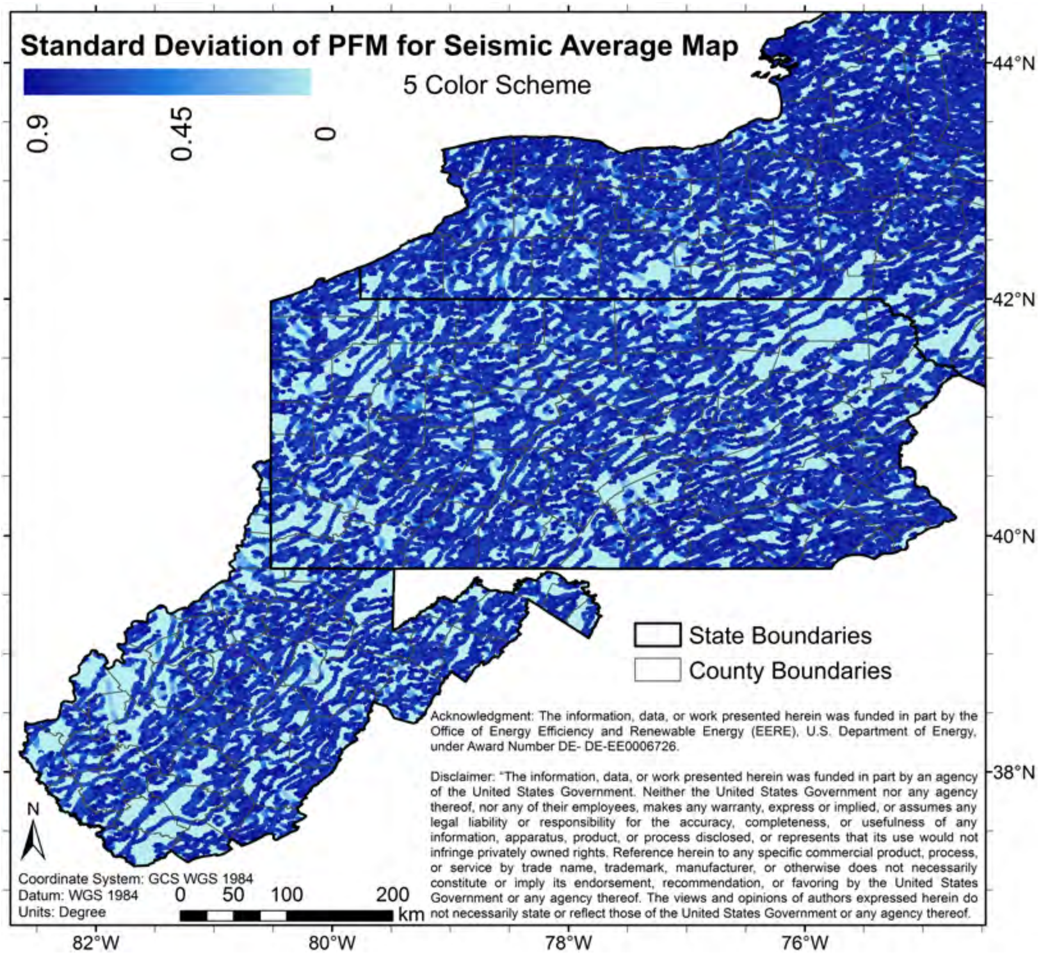


Figure 23: Map of the standard deviation of the scaled risk factor for averaged seismic risk factor for a 5-color scheme. Dark colors are less certain and light colors are more certain.

1.4 Utilization

The utilization risk factor is based on the surface levelized cost of heat (SLCOH, \$/MMBTU), which was calculated using methods described in the utilization methodology section. The calculations are based on district heating only, not other potential direct-use applications. There were several special considerations. Because this analysis focused on geothermal district heating systems, the communities must be of sufficient size to afford the up-front capital cost associated with the project. Therefore, places with fewer than 4,000 people had the SLCOH set to an arbitrary value, 100 \$/MMBTU, to represent that a district heating system of sufficient size to be of interest for this would be infeasible for this population.

The 3-color thresholds are based on recommendations from the utilization team, including K. Welcker, Z. Frone, and M. Richards. Based on these values, the 5-color thresholds were assigned. The minimum and maximum values were assigned as well.

Our project requires co-location of the utilization location with the reservoirs, but it is reasonable to assume that the utilization location can be a small distance from the location of the reservoir. Therefore, the input utilization map was buffered. All pixels (raster cells) whose center was within 5 km of the center of the middle cell were included in the buffer. The best utilization value (lowest SLCOH) among these was assigned to only the center value. Figure 24 shows an example of the cells considered in calculation. Another reason to use the buffer is that many small census places are near a large census place, but the method of calculating the SLCOH did not use information on the proximity of the smaller places to the larger ones. Potentially, a small census place could be part of the neighboring larger place's system, so using the buffer means that the neighboring small place will be assigned the more attractive and often more reasonable SLCOH from the larger place.

Table 4 summarizes the thresholds, Figure 25 gives the histogram of the utilization of the original map (only values less than \$100/MMBTU were plotted), Figures 26 and 27 give the original map converted into the 3- and 5-color play fairway scheme, and Figures 28 and 29 give the utilization map converted into the play fairway scheme with buffering to account for the utilization distance.

No uncertainty was assigned by utilization to their calculations or model prior to these calculations being made.

Table 4: Table of minimum, maximum, and thresholds used in scaling the utilization map. The measure is leveilized cost of heat ($/MMBTU). The scale is reversed because high LCOH is unfavorable.

Min	25
Max	5
3-Color Thresholds	{13.5,16}
5-Color Thresholds	{12,13.5, 16, 20}

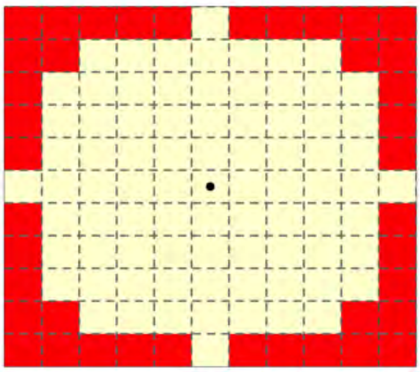


Figure 24: Weighting matrix used when evaluating the best utility for the pixel (raster cell). The maximum of the neighboring pixels within a certain distance is taken (best of cells in yellow). Cells whose center is farther than 5 km from the center (black dot) are not considered and are marked as red.

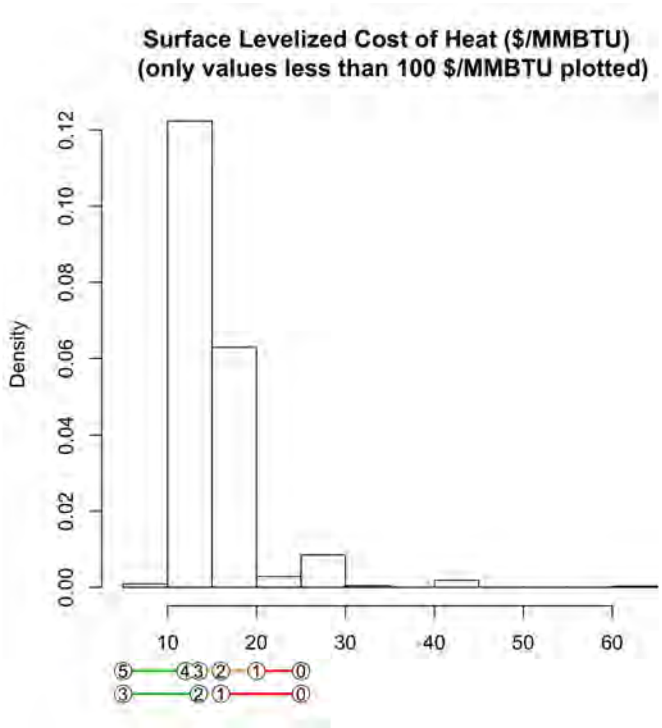


Figure 25: Histogram of the utilization risk factor with the 3- and 5-color thresholds noted. Points beyond the minimum and maximum of the scale are assigned the minimum and maximum of the scale, respectively (e.g. 0, and 3 or 5). Only values less than 100 are plotted to avoid large distortions in the scale. Density is proportional to the frequency or count in the bins, but the values have been rescaled so the total area of the histogram is 1.

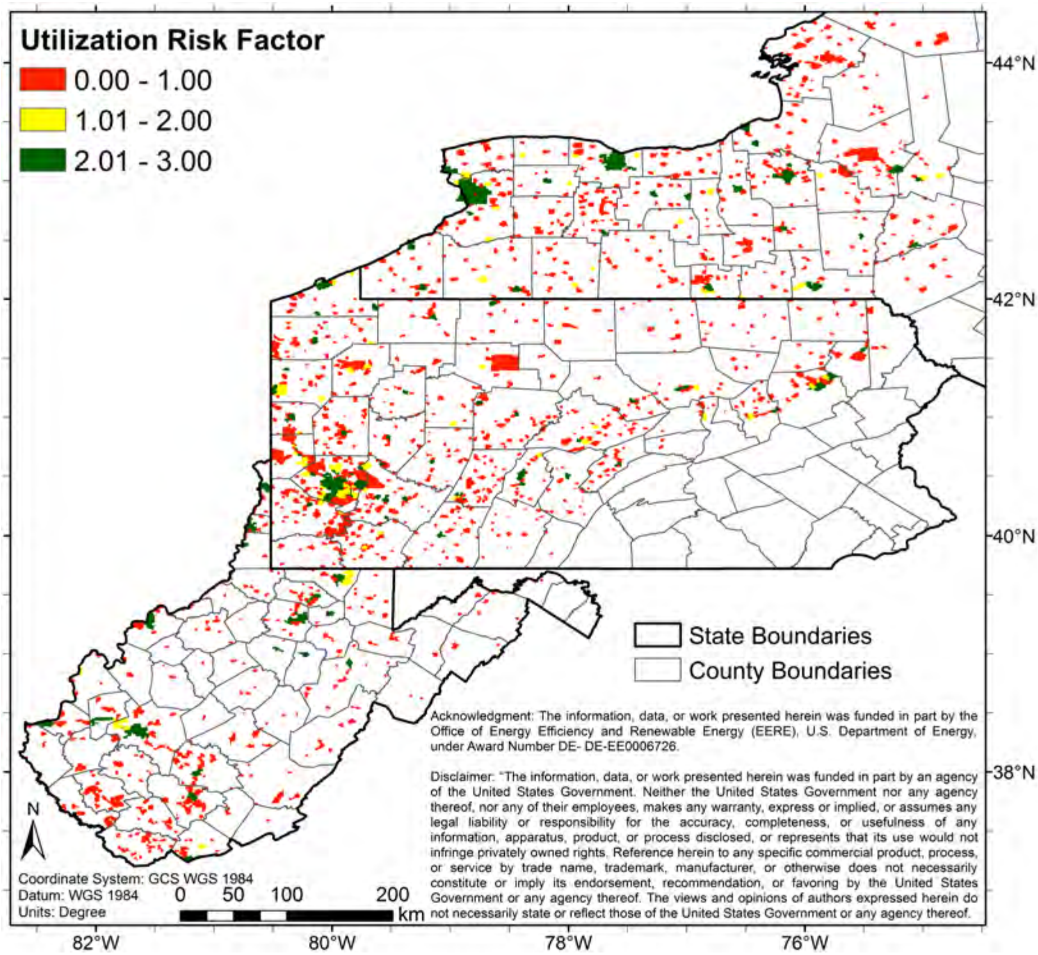


Figure 26: Map of the utilization risk factor with a 3-color scheme with no buffer. Red areas are unfavorable and green areas are favorable values of the scaled risk factor.

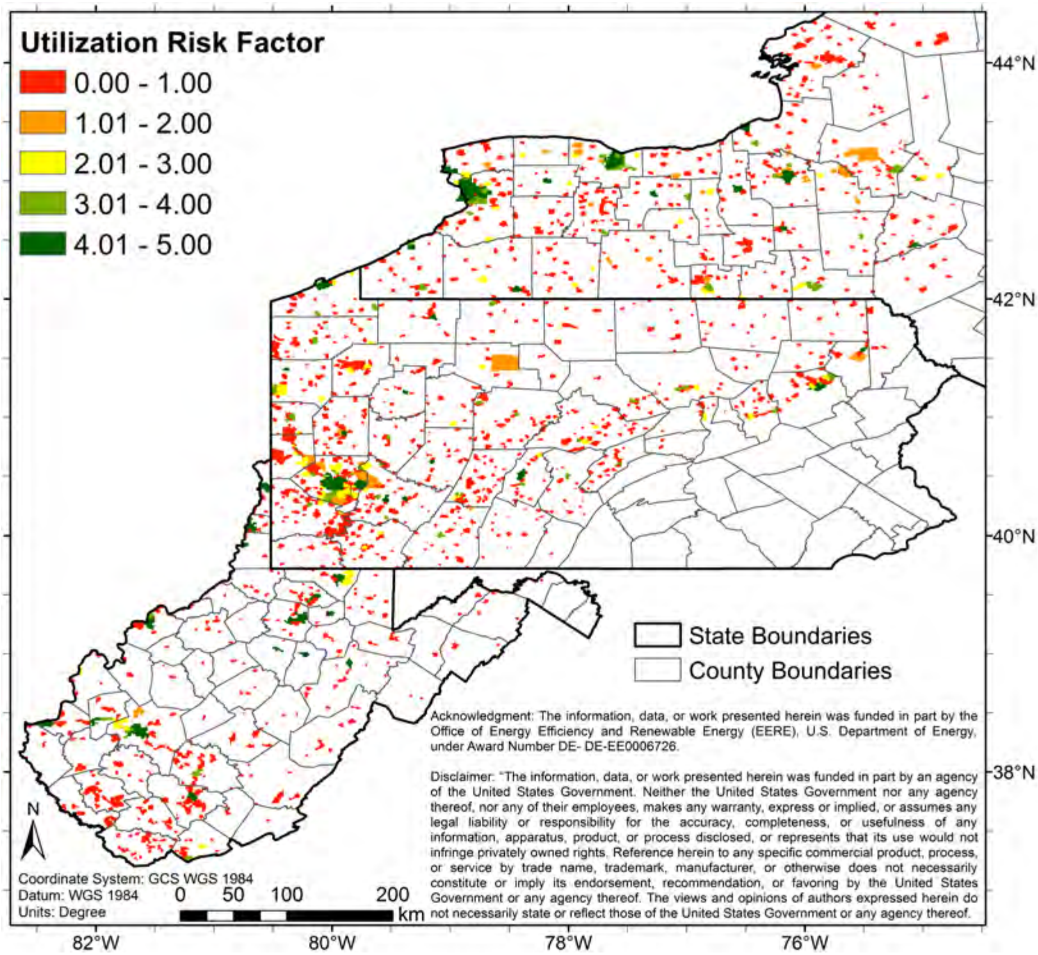


Figure 27: Map of the utilization risk factor with a 5-color scheme with no buffer. Red areas are unfavorable and green areas are favorable values of the scaled risk factor.

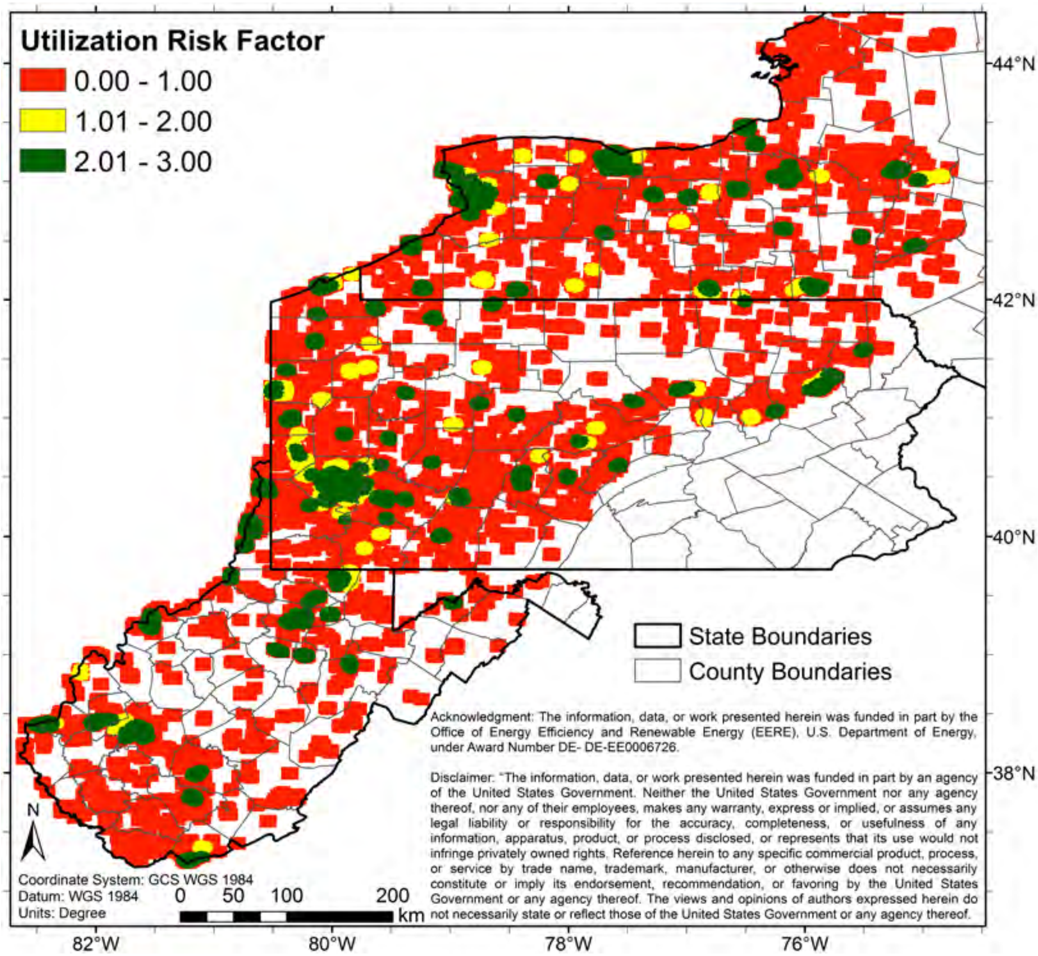


Figure 28: Map of the utilization risk factor with a 3-color scheme with a radius of 5 km buffer. Red areas are unfavorable and green areas are favorable values of the scaled risk factor.

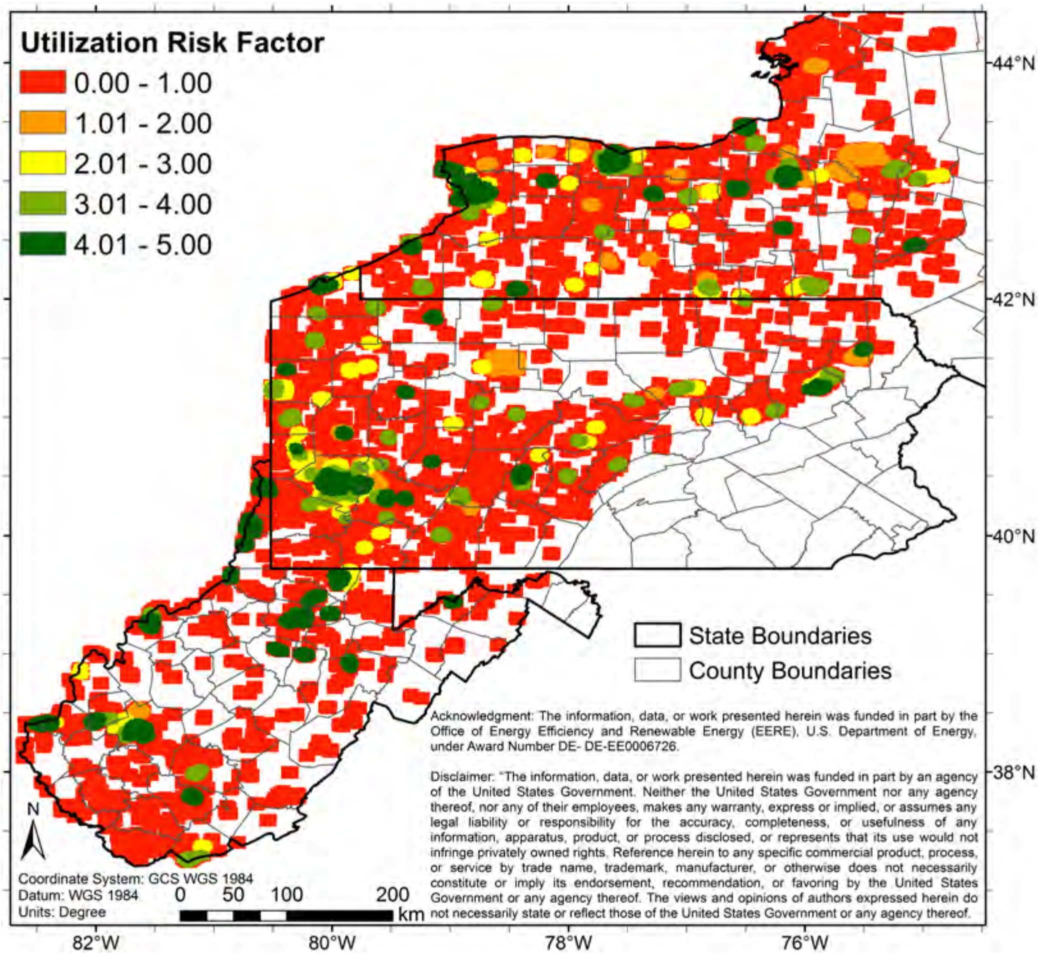


Figure 29: Map of the utilization risk factor with a 5-color scheme with a radius of 5 km buffer. Red areas are unfavorable and green areas are favorable values of the scaled risk factor.

2 Combined Risk Factors

The following sections outline extended results for combining risk factors. The results for all risk factors are presented first, followed by geology variables only (no utilization), and then the results without reservoirs (potentially enhanced geothermal systems [EGS]).

The risk factors can be combined in several ways including by taking the sum, product, or minimum value. Equations 3 to 5 show the formulas used, where PFM is the combined play fairway metric, SRF_i is the risk factor scaled into the play fairway system (e.g. 0 to 3 or 0 to 5), and (j, k) is the raster cell location. Areas without a risk factor defined appear as white in the final map, indicating insufficient data to evaluate the PFM at that location.

The following subsections present the results for 3- and 5-color maps for each method of combining the maps. The thresholds for the combined map are based on preserving the scale of the combined map. For instance, the red-yellow cutoff on the 3-color scale is 1, so in the combined map using a sum that is 4 because there are four risk factors. In the product map, the yellow-green cutoff is 2^4 because in an individual risk factor it is at 2 and now there are four risk factors.

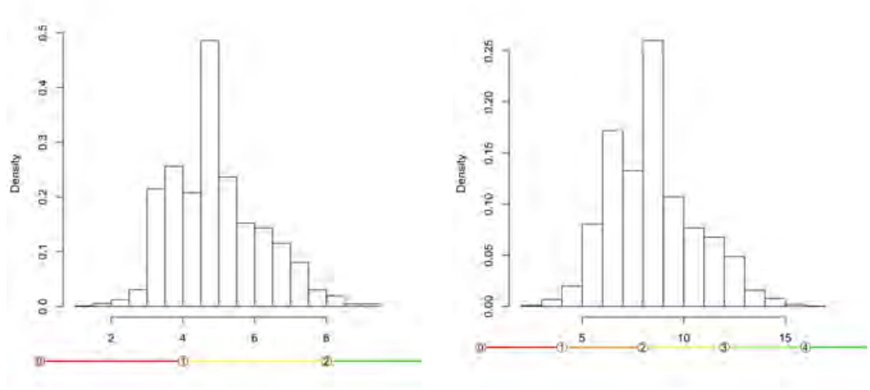
Table 5: Table of thresholds for the different methods of combining risk factors. Thresholds are given generally as a function of n, where n is the number if risk factors. The thresholds are non-dimensional.

Combining	3-Color	5-Color
Sum	$\{0, 1, 2, 3\} \times n$	$\{0, 1, 2, 3, 4, 5\} \times n$
Product	$\{0, 1, 2, 3\}^n$	$\{0, 1, 2, 3, 4, 5\}^n$
Minimum	$\{0, 1, 2, 3\}$	$\{0, 1, 2, 3, 4, 5\}$

$$\text{PFM}_{\text{sum}}(j, k) = \sum_{i=1}^{n} \text{SRF}_i(j, k) \tag{3}$$

$$\text{PFM}_{\text{product}}(j, k) = \prod_{i=1}^{n} \text{SRF}_i(j, k) \tag{4}$$

$$\text{PFM}_{\text{minimum}}(j, k) = \min\{\text{SRF}_1(j, k), \text{SRF}_2(j, k), ..., \text{SRF}_n(j, k)\} \tag{5}$$



(a) Three colors. (b) Five colors.

Figure 30: Histograms of the combined risk metric using a sum with 3- and 5-color schemes. Density is proportional to the frequency or count in the bins, but the values have been rescaled so the total area of the histogram is 1.

2.1 All Risk Factors

This subsection gives the results for combining all risk factors. The results are given for sum, product, and minimum. For each map, histograms of the distribution are also provided for the 3- and 5-color scheme.

2.1.1 Sum

The following results are for summing the risk factors. The thresholds are given in Table 5 with $n = 4$. Figure 30 shows the histograms for the 3- and 5-color maps, respectively, of the resulting metric and Figures 31 and 32 show the 3- and 5-color maps, respectively.

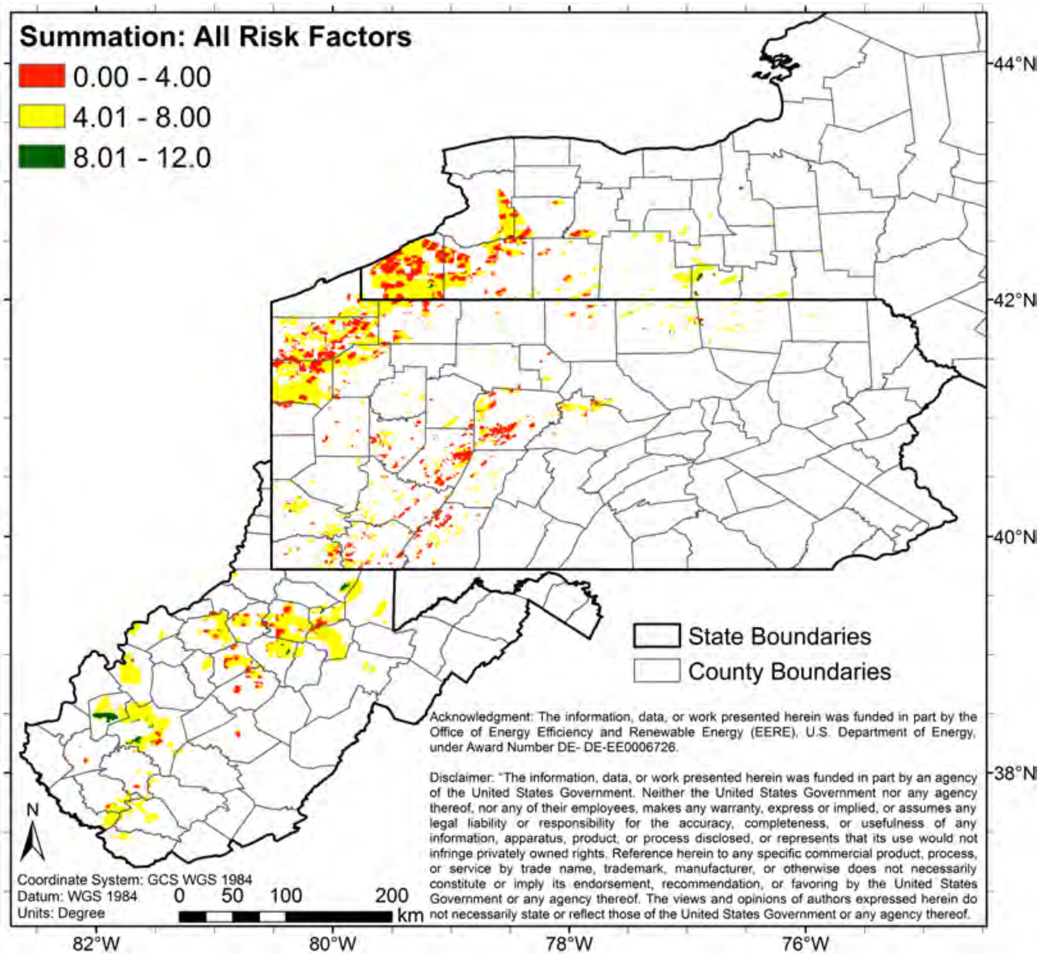


Figure 31: Map of the combined risk map with a 3-color scheme using sums. Red areas are unfavorable and green areas are favorable values of the play fairway metric.

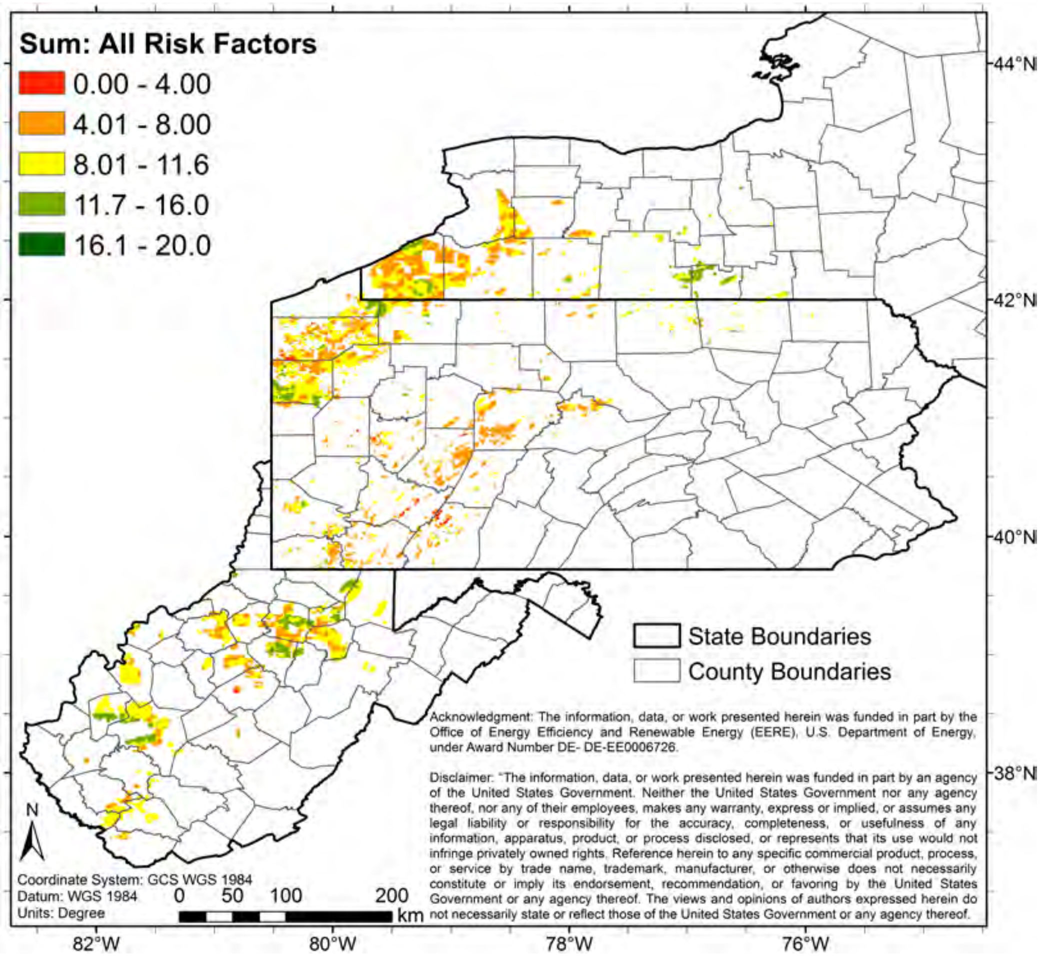
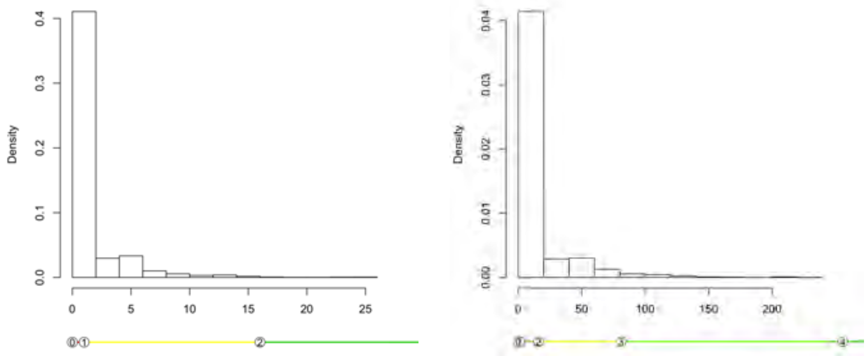


Figure 32: Map of the combined risk map with a 5-color scheme using sums. Red areas are unfavorable and green areas are favorable values of the play fairway metric.



(a) Three colors.

(b) Five colors.

Figure 33: Histograms of the combined risk metric using a product with 3- and 5-color schemes. Density is proportional to the frequency or count in the bins, but the values have been rescaled so the total area of the histogram is 1.

2.1.2 Product

The following results are for taking the product of the risk factors. The thresholds are given in Table 5 with $n = 4$. The histograms and maps are given in Figure 33 and the 3- and 5-color maps are given in Figures 34 and 35, respectively. Note that there is a large portion of the data at zero mainly because the utilization values gave many zeros, which set the product to zero.

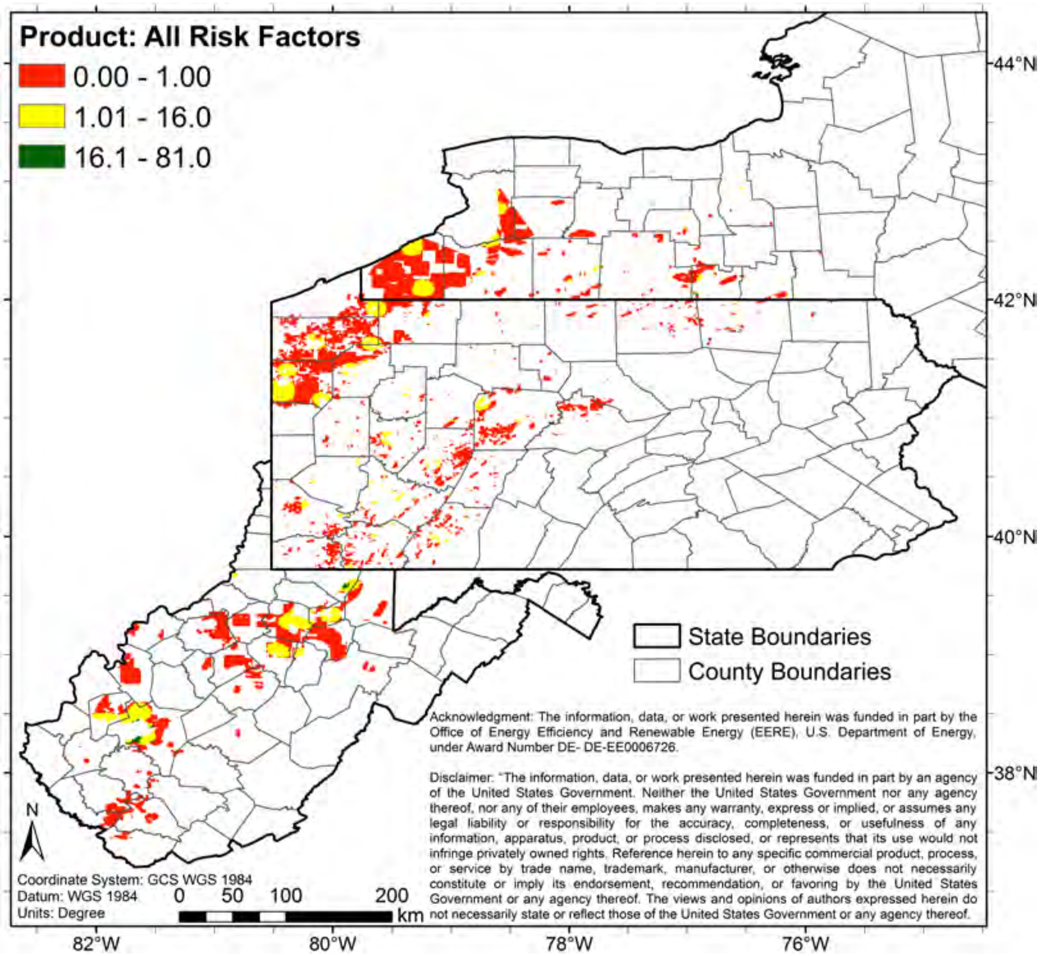


Figure 34: Map of the combined risk map with a 3-color scheme using products. Red areas are unfavorable and green areas are favorable values of the play fairway metric.

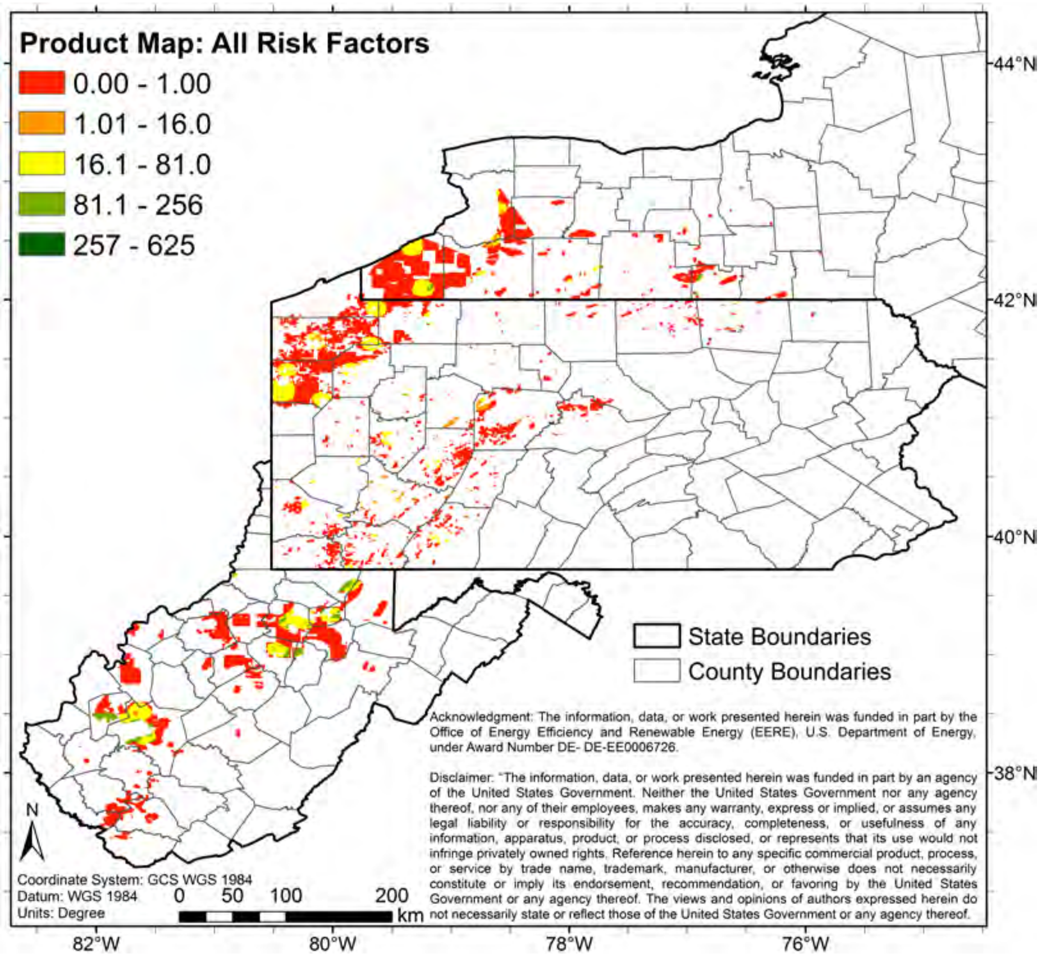
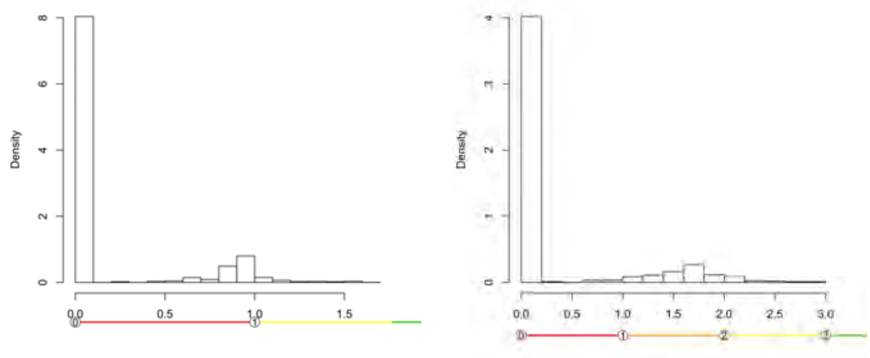


Figure 35: Map of the combined risk map with a 5-color scheme using products. Red areas are unfavorable and green areas are favorable values of the play fairway metric.



(a) Three colors. (b) Five colors.

Figure 36: Histograms of the combined risk metric using a minimum with 3- and 5-color schemes. Density is proportional to the frequency or count in the bins, but the values have been rescaled so the total area of the histogram is 1.

2.1.3 Minimum

The following results are for taking the minimum of the risk factors. The thresholds are given in Table 5. The thresholds are based on the original scale of the colors, so green would mean that the lowest risk metric is green at that location (very good). Because of the choice of scale, the area is dominated by reds. Figure 36 plots the histograms for the 3- and 5-color maps and Figures 37 and 38 plot the 3- and 5-color maps, respectively. The large number of zeros is because many of the utilization locations did not meet the requirement of 4,000 people, so they had the high SLCOH assigned, which in turn means that their scaled risk factor value was zero.

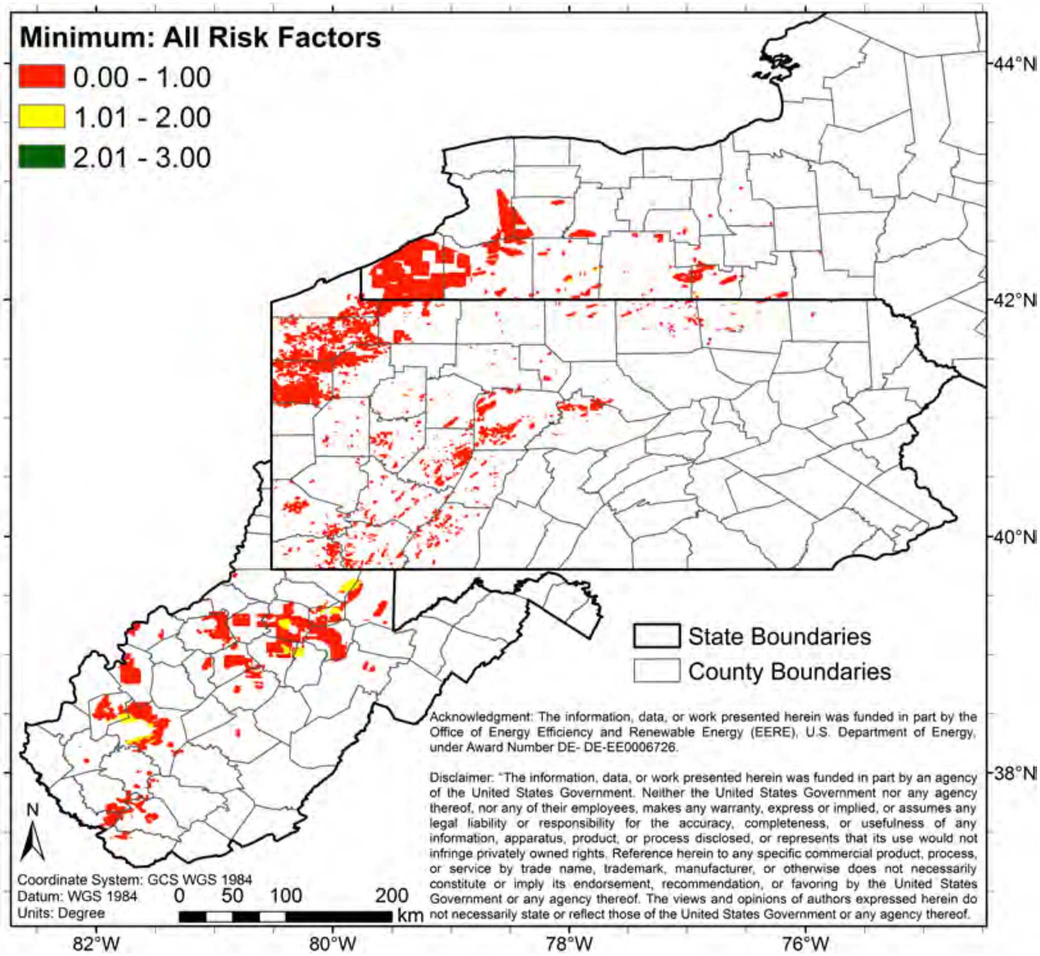


Figure 37: Map of the combined risk map with a 3-color scheme using minimums. Red areas are unfavorable and green areas are favorable values of the play fairway metric.

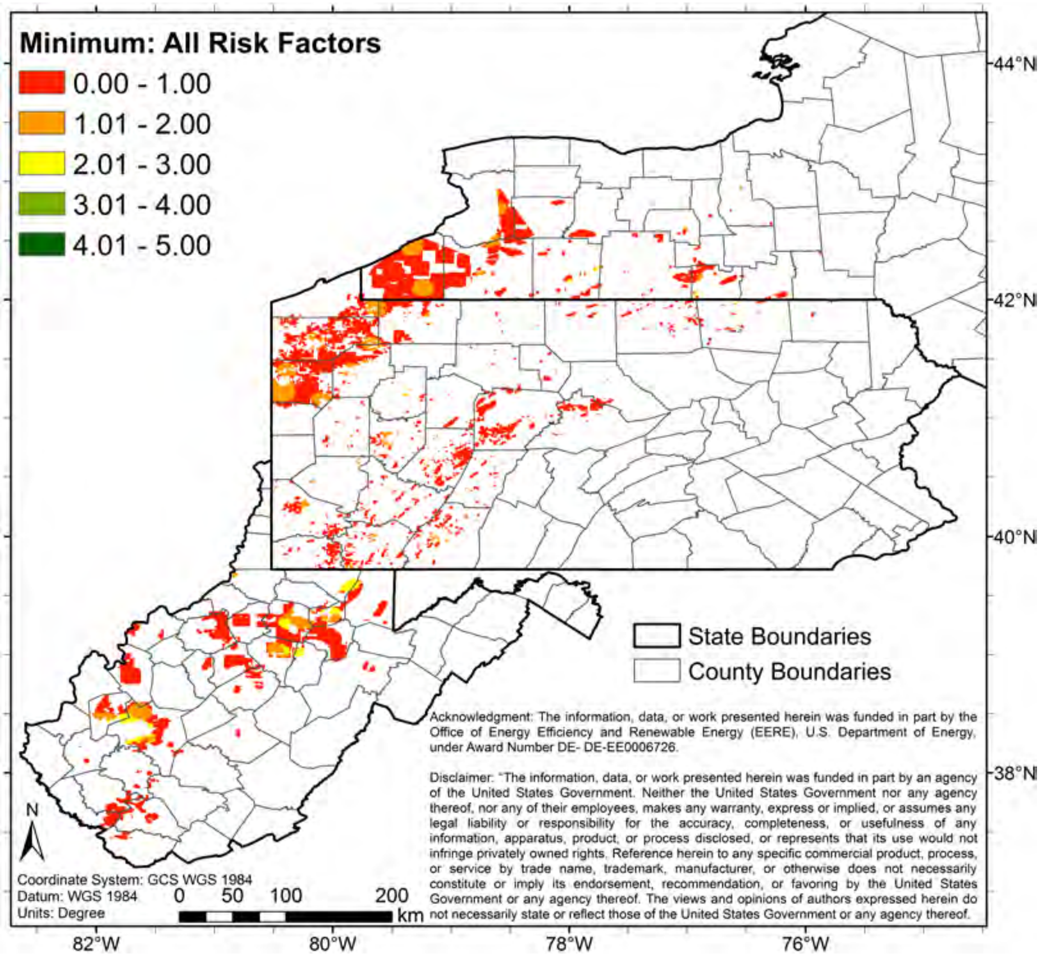
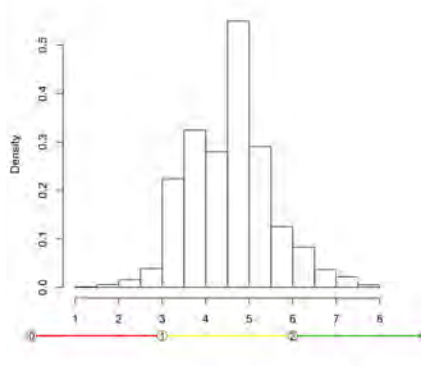
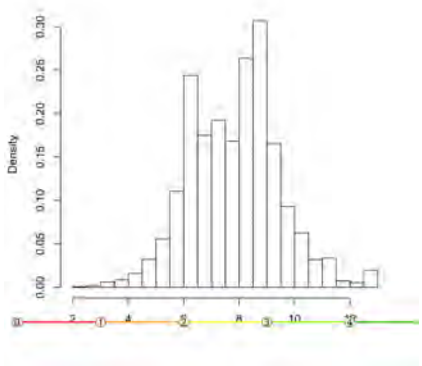


Figure 38: Map of the combined risk map with a 5-color scheme using minimums. Red areas are unfavorable and green areas are favorable values of the play fairway metric.



(a) Three colors.



(b) Five colors.

Figure 39: Histograms of the combined risk metric using a sum with 3- and 5-color schemes for geology only. Density is proportional to the frequency or count in the bins, but the values have been rescaled so the total area of the histogram is 1.

2.2 Geology Only

This section presents the results for the variables relating to geology only (no utilization). This would represent the best areas to develop the resources independent of the current location of population centers.

2.2.1 Sum

The following results are for summing the risk factors. The thresholds are given in Table 5 with $n = 3$. Figure 39 shows the histograms for the 3- and 5-color maps, respectively, of the resulting metric and Figures 40 and 41 show the 3- and 5-color maps, respectively.

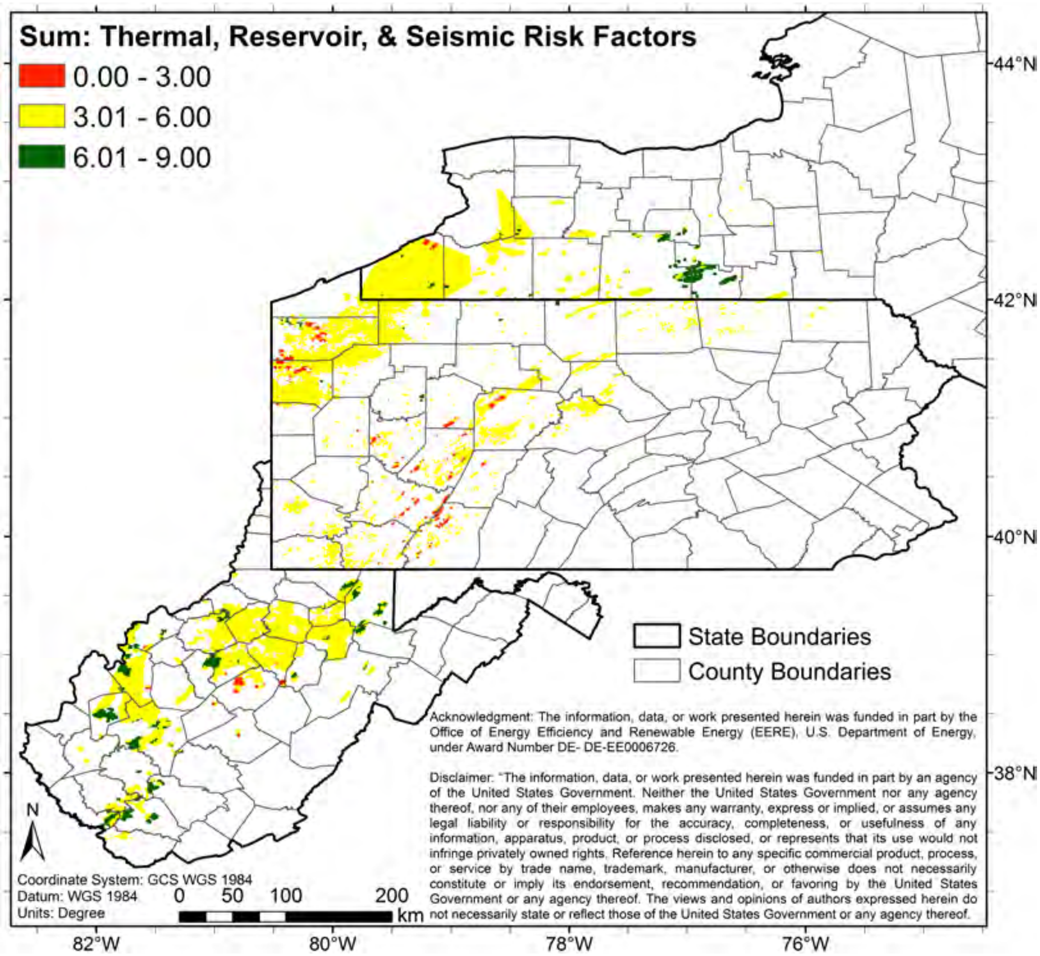


Figure 40: Map of the combined risk map with a 3-color scheme using sums for geology only. Red areas are unfavorable and green areas are favorable values of the play fairway metric.

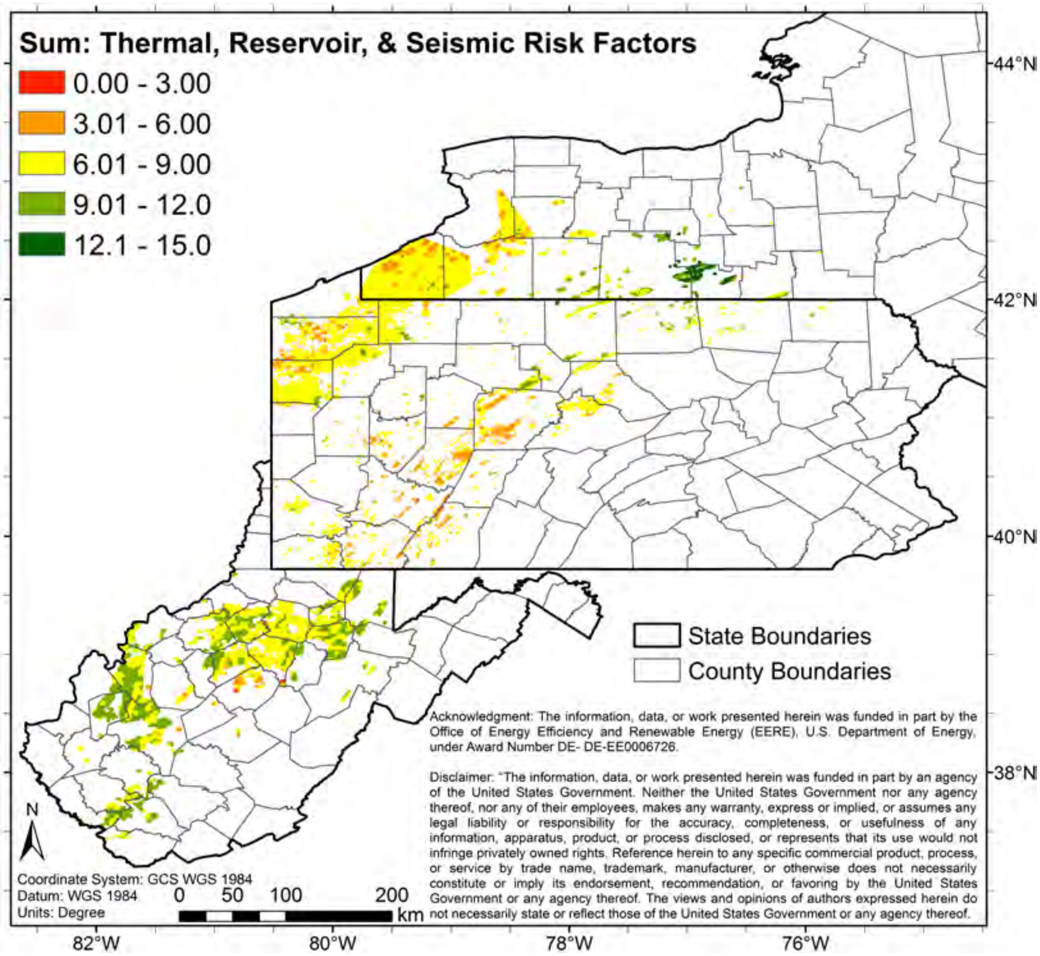
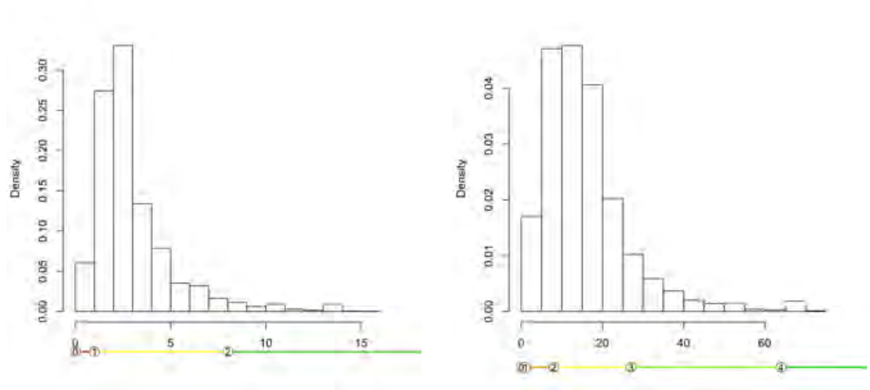


Figure 41: Map of the combined risk map with a 5-color scheme using sums for geology only. Red areas are unfavorable and green areas are favorable values of the play fairway metric.



(a) Three colors.

(b) Five colors.

Figure 42: Histograms of the combined risk metric using a product with 3- and 5-color schemes for geology only. Density is proportional to the frequency or count in the bins, but the values have been rescaled so the total area of the histogram is 1.

2.2.2 Product

The following results are for taking the product of the risk factors. The thresholds are given in Table 5 with $n = 3$. The histograms and maps are given in Figure 42 and the 3- and 5-color maps are given in Figures 43 and 44, respectively.

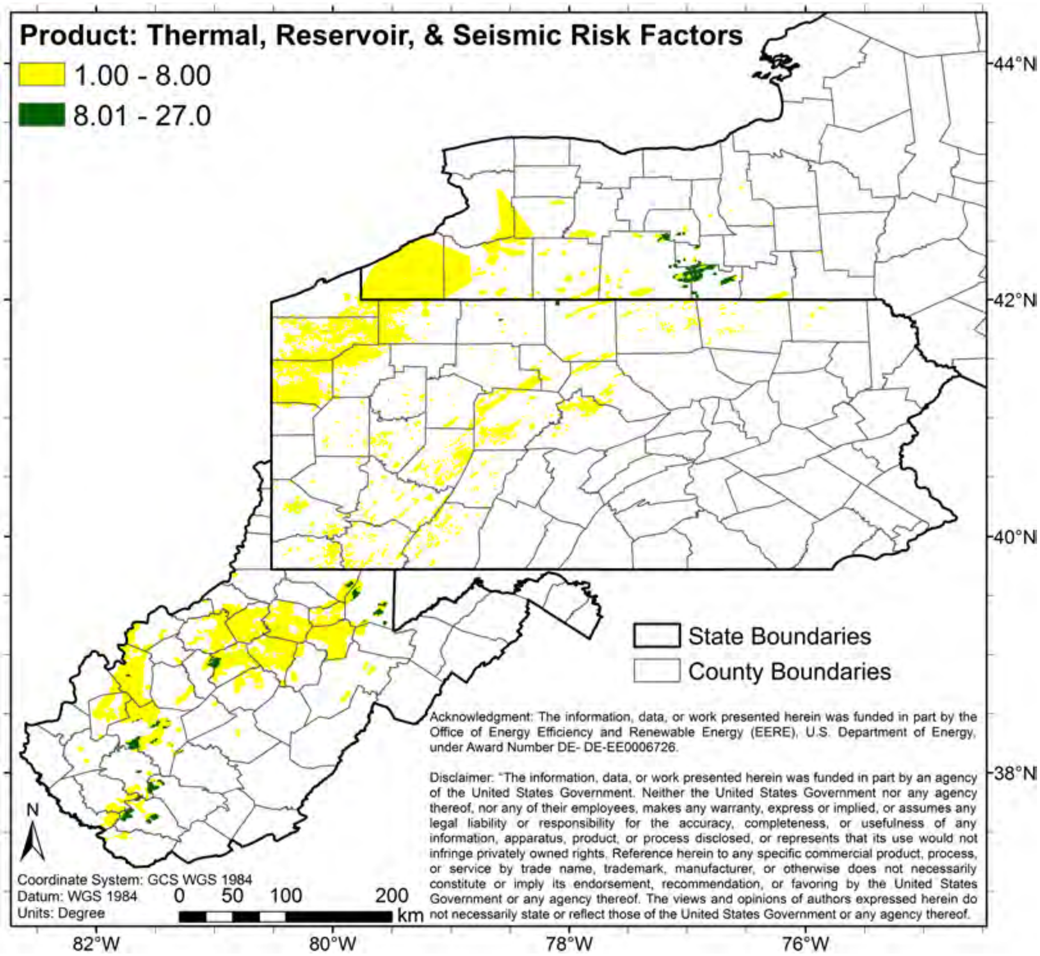


Figure 43: Map of the combined risk map with a 3-color scheme using products for geology only. Red areas are unfavorable and green areas are favorable values of the play fairway metric.

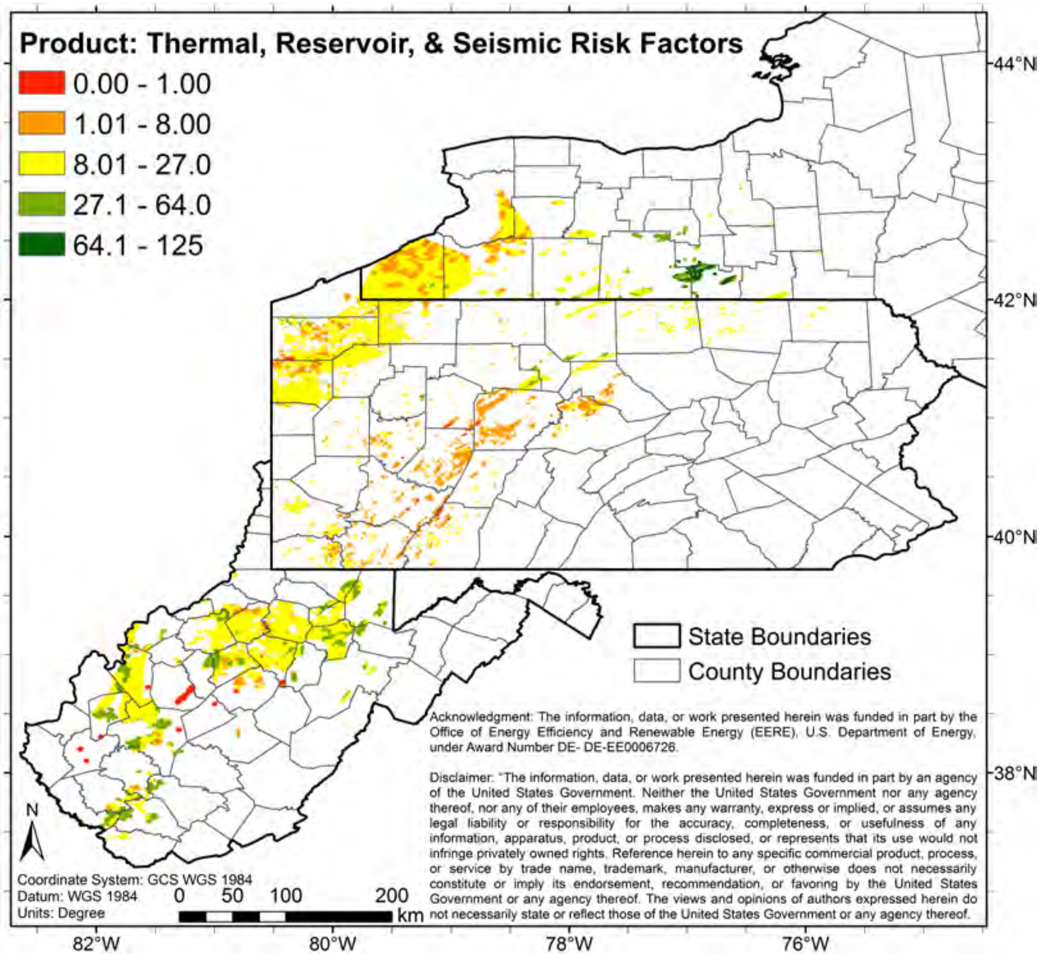
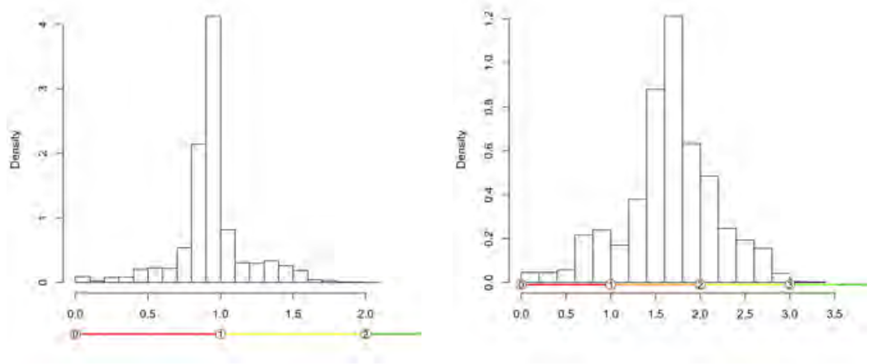


Figure 44: Map of the combined risk map with a 5-color scheme using products for geology only. Red areas are unfavorable and green areas are favorable values of the play fairway metric.



(a) Three colors.

(b) Five colors.

Figure 45: Histograms of the combined risk metric using a minimum with 3- and 5-color schemes for geology only. Density is proportional to the frequency or count in the bins, but the values have been rescaled so the total area of the histogram is 1.

2.2.3 Minimum

The following results are for taking the minimum of the risk factors. The thresholds are given in Table 5. The thresholds are based on the original scale of of the colors, so green would mean that the lowest risk metric is green at that location (very good). Because of the choice of scale, the area is dominated by reds. Figure 45 plots the histograms for the 3- and 5-color maps and Figures 46 and 47 plot the 3- and 5-color maps, respectively.

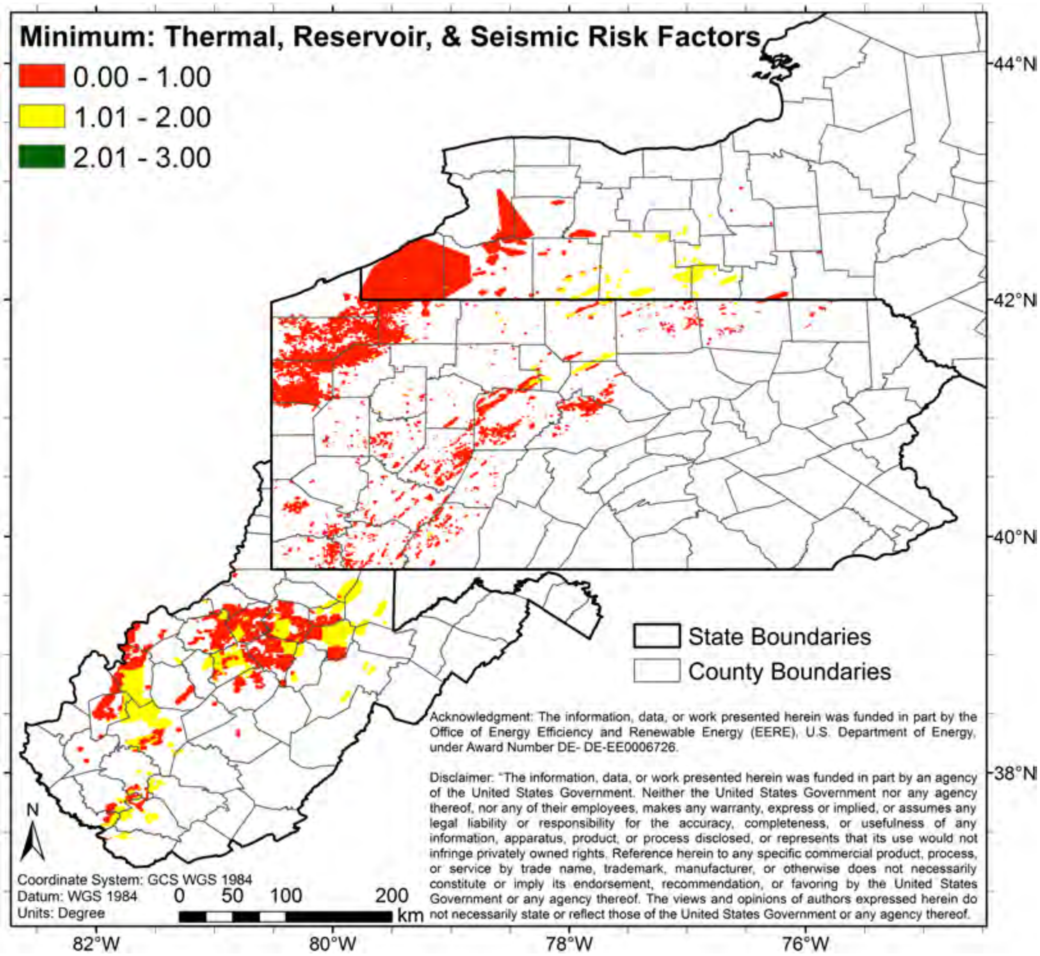


Figure 46: Map of the combined risk map with a 3-color scheme using minimums for geology only. Red areas are unfavorable and green areas are favorable values of the play fairway metric.

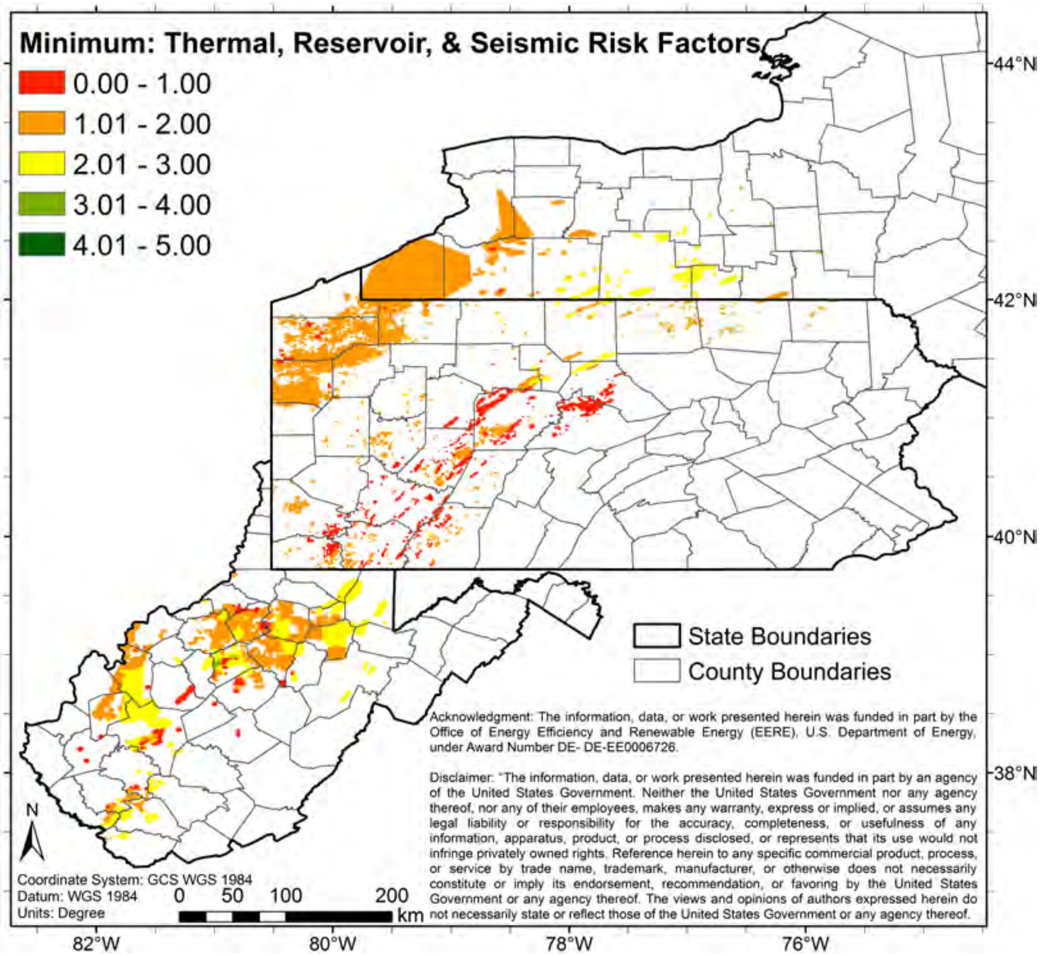
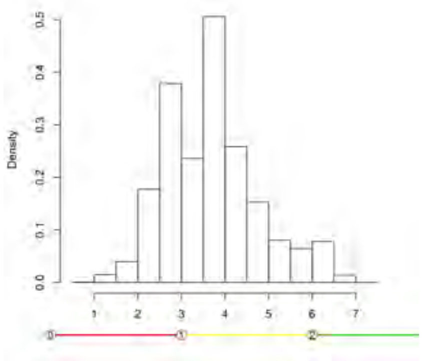
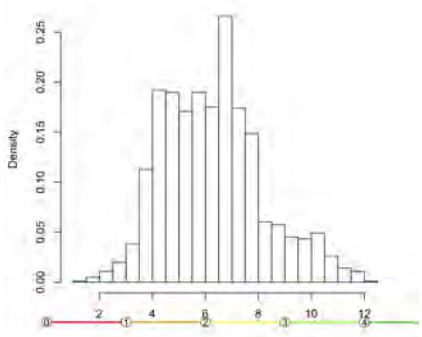


Figure 47: Map of the combined risk map with a 5-color scheme using minimums for geology only. Red areas are unfavorable and green areas are favorable values of the play fairway metric.



(a) Three colors.



(b) Five colors.

Figure 48: Histograms of the combined risk metric using a sum with 3- and 5-color schemes for all risk factors except reservoirs. Density is proportional to the frequency or count in the bins, but the values have been rescaled so the total area of the histogram is 1.

2.3 No Reservoirs

This section presents the results for the variables omitting reservoirs. This could be the case where one was considering enhanced geothermal systems (EGS) and engineering the reservoir.

2.3.1 Sum

The following results are for summing the risk factors. The thresholds are given in Table 5 with $n = 3$. Figure 48 shows the histograms for the 3- and 5-color maps, respectively, of the resulting metric and Figures 49 and 50 show the 3- and 5-color maps, respectively.

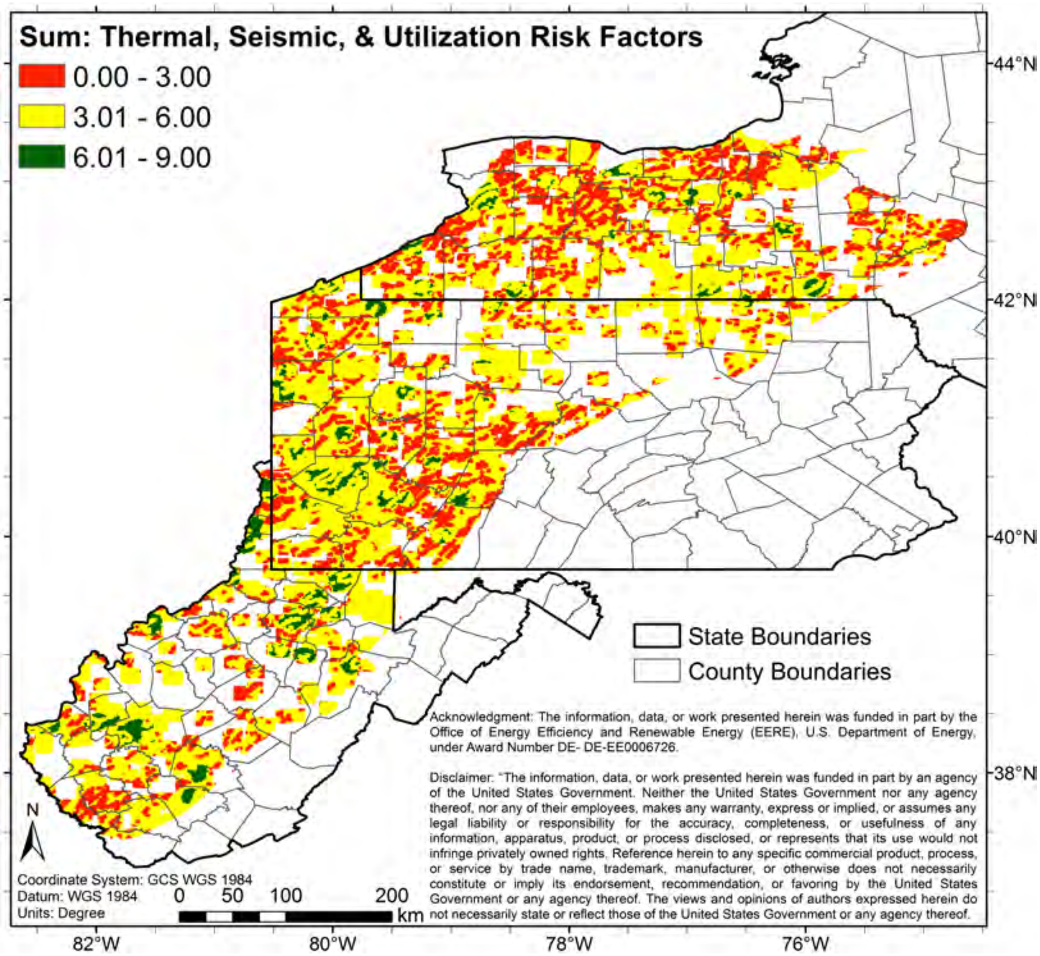


Figure 49: Map of the combined risk map with a 3-color scheme using sums for all risk factors except reservoirs. Red areas are unfavorable and green areas are favorable values of the play fairway metric.

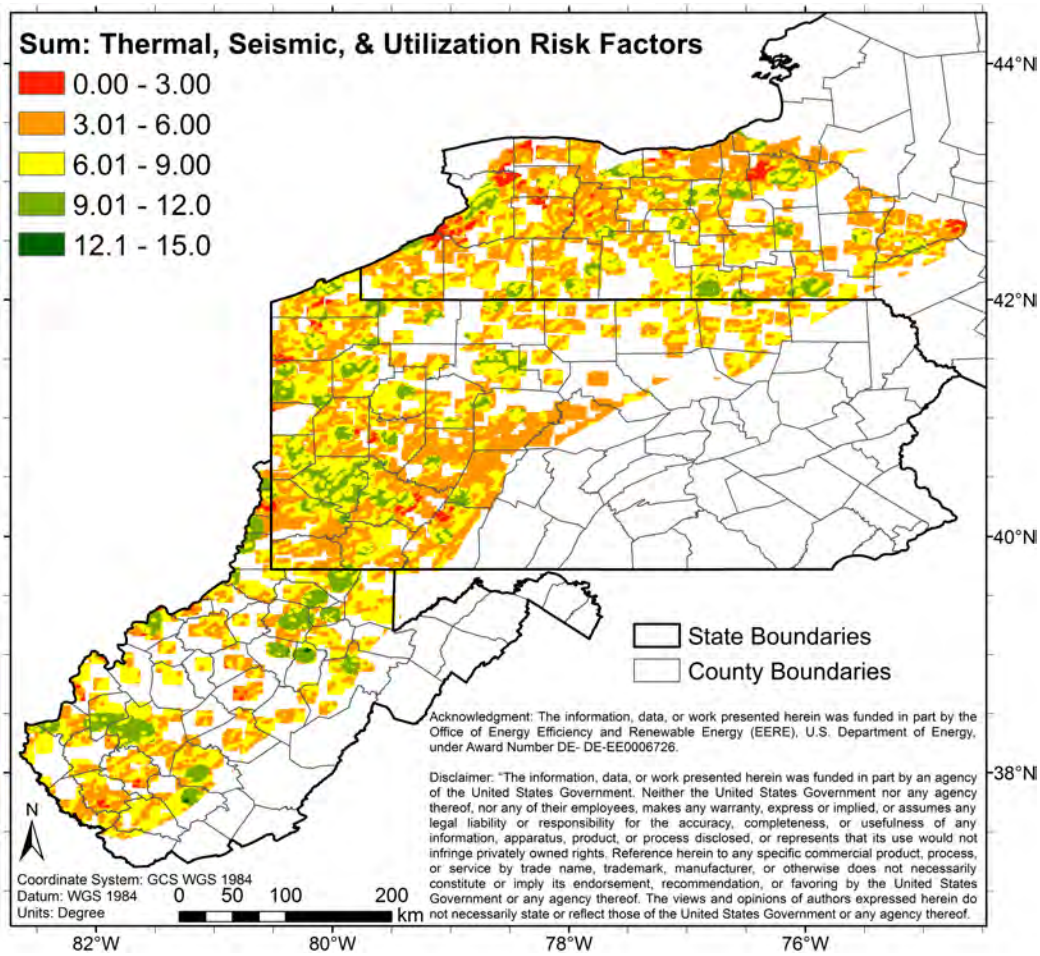
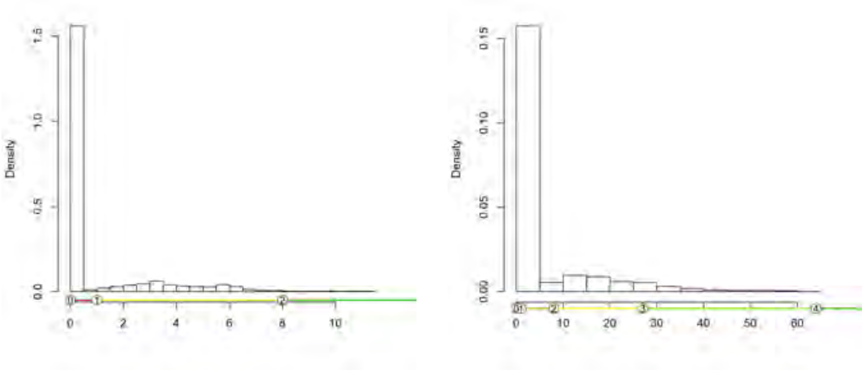


Figure 50: Map of the combined risk map with a 5-color scheme using sums for all risk factors except reservoirs. Red areas are unfavorable and green areas are favorable values of the play fairway metric.



(a) Three colors.

(b) Five colors.

Figure 51: Histograms of the combined risk metric using a product with 3- and 5-color schemes for all risk factors except reservoirs. Density is proportional to the frequency or count in the bins, but the values have been rescaled so the total area of the histogram is 1.

2.3.2 Product

The following results are for taking the product of the risk factors. The thresholds are given in Table 5 with $n = 3$. The histograms and maps are given in Figure 51 and the 3- and 5-color maps are given in Figures 52 and 53, respectively.

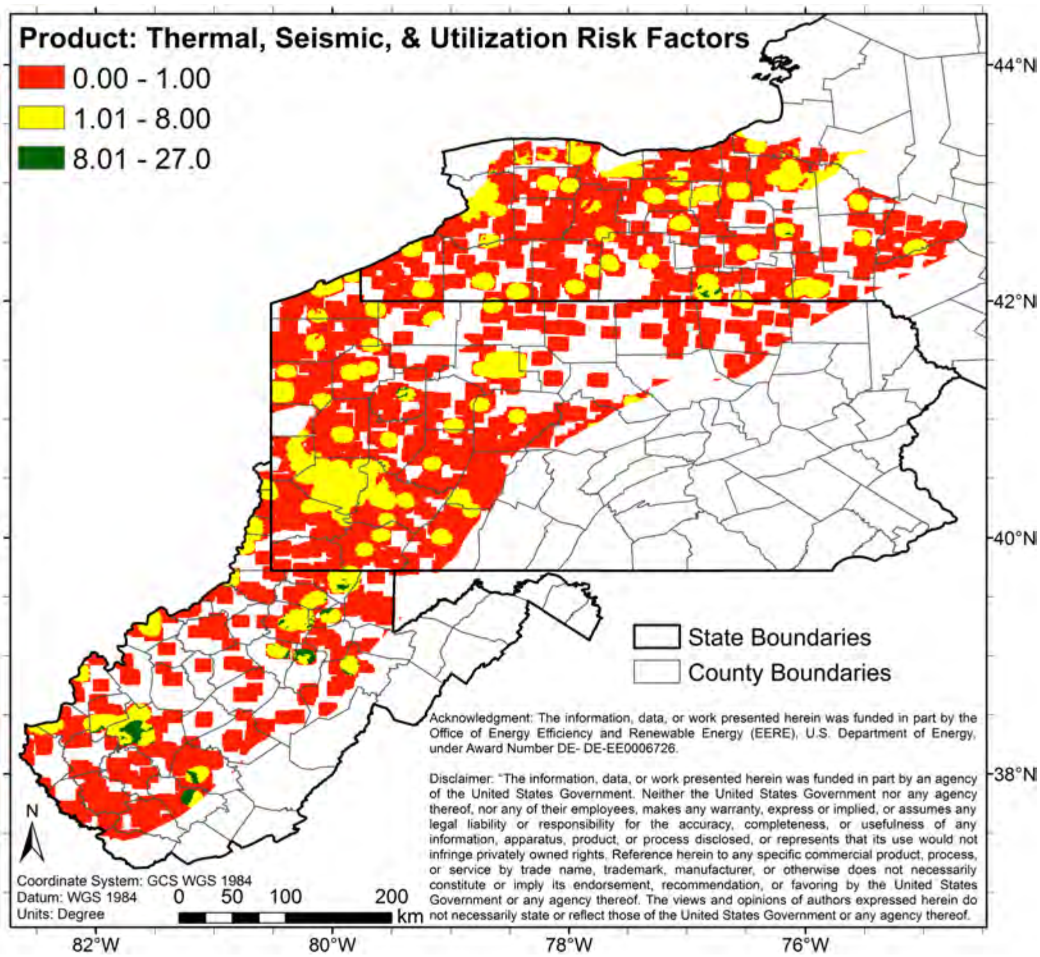


Figure 52: Map of the combined risk map with a 3-color scheme using products for all risk factors except reservoirs. Red areas are unfavorable and green areas are favorable values of the play fairway metric.

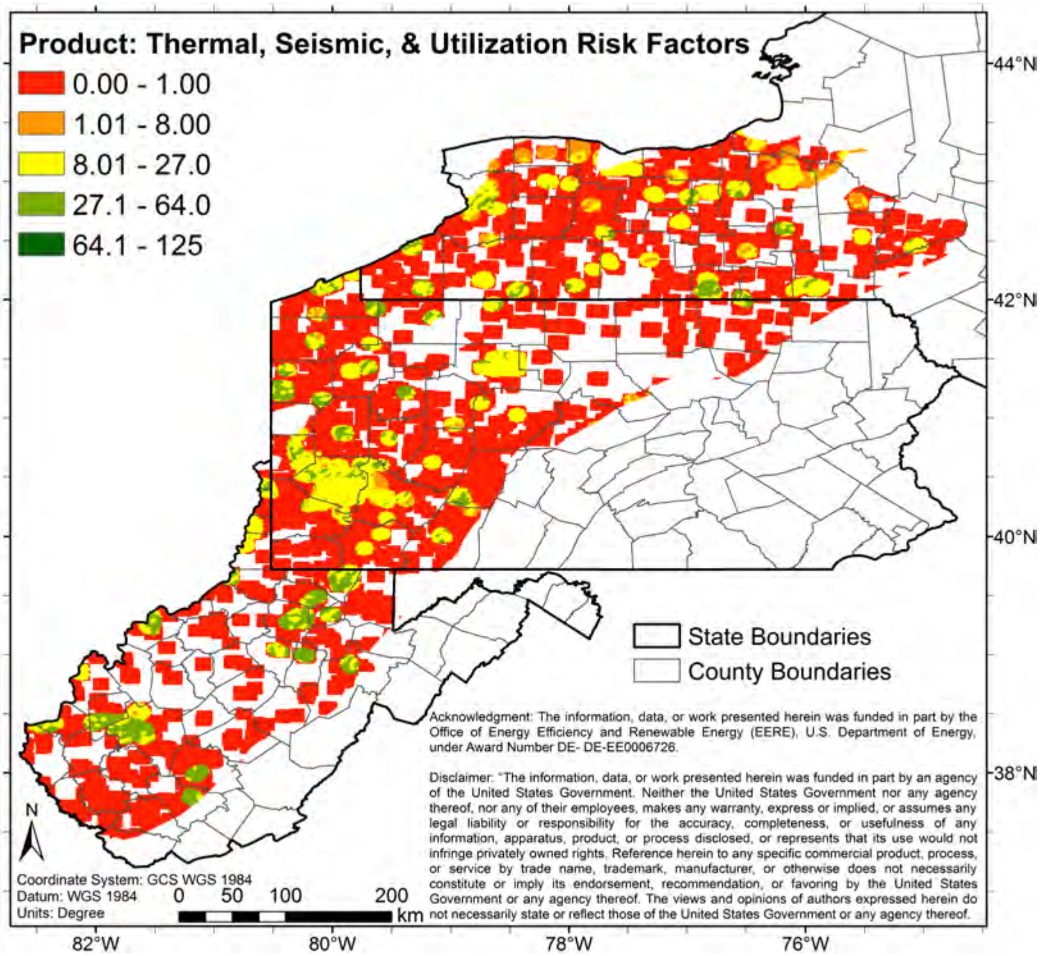
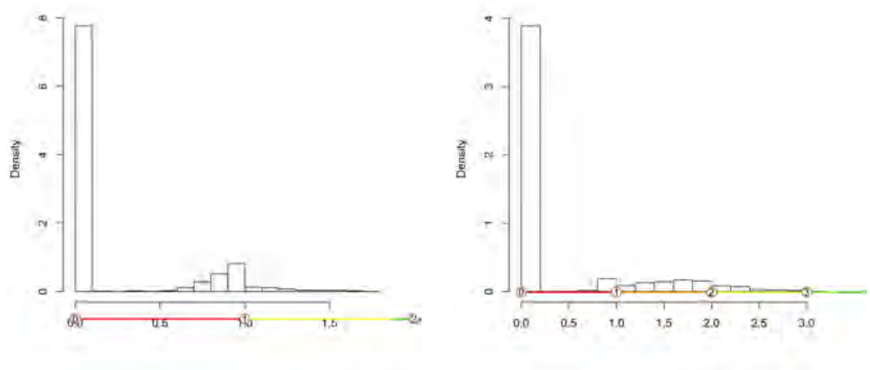


Figure 53: Map of the combined risk map with a 5-color scheme using products for all risk factors except reservoirs. Red areas are unfavorable and green areas are favorable values of the play fairway metric.



(a) Three colors. (b) Five colors.

Figure 54: Histograms of the combined risk metric using a minimum with 3- and 5-color schemes for all risk factors except reservoirs. Density is proportional to the frequency or count in the bins, but the values have been rescaled so the total area of the histogram is 1.

2.3.3 Minimum

The following results are for taking the minimum of the risk factors. The thresholds are given in Table 5 and $n = 3$ for this situation. The thresholds are based on the original scale of of the colors, so green would mean that the lowest risk metric is green at that location (very good). Because of the choice of scale, the area is dominated by reds. Figure 54 plots the histograms for the 3- and 5-color maps and Figures 55 and 56 plot the 3- and 5-color maps, respectively. The large number of zeros is because many of the utilization locations did not meet the requirement of 4,000 people, so they had the high SLCOH assigned, which in turn means that their scaled risk factor value was zero.

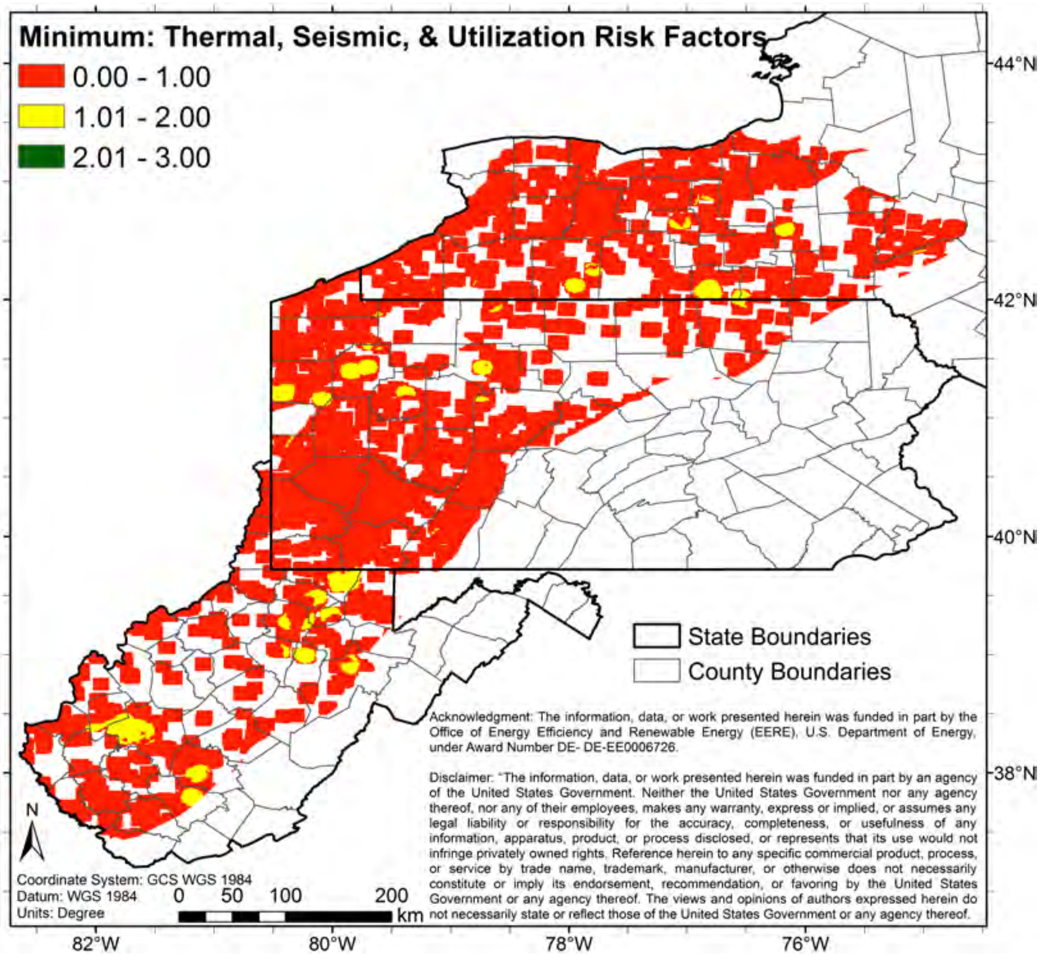


Figure 55: Map of the combined risk map with a 3-color scheme using minimums or all risk factors except reservoirs. Red areas are unfavorable and green areas are favorable values of the play fairway metric.

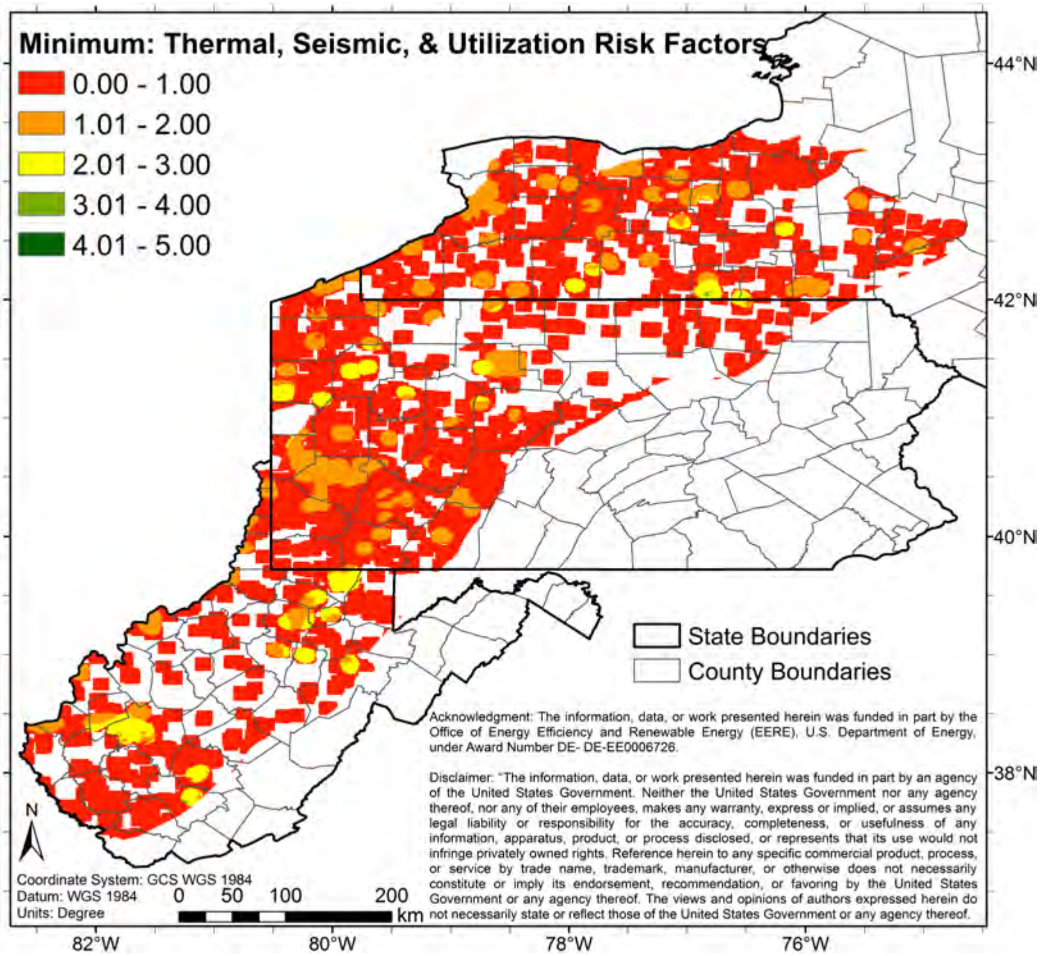
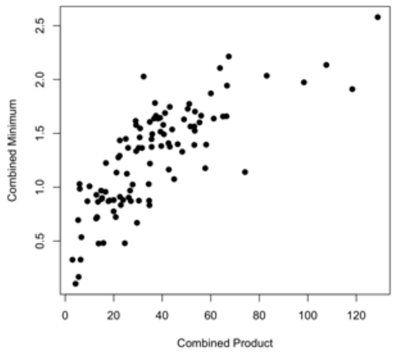


Figure 56: Map of the combined risk map with a 5-color scheme using minimums for all risk factors except reservoirs. Red areas are unfavorable and green areas are favorable values of the play fairway metric.

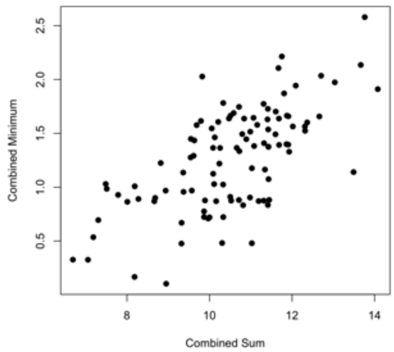
3 Robustness of Combining Functions

It is not clear which method of combining the risk factors is best. Therefore, the consistency of the different methods of combining the risk factors should be investigated. If the methods are fairly similar, then locations with a high value for the sum should also have a high value for the product, for example. The locations of US Census places were found and then the 5-color sum, product, and minimum PFMs were sampled at those locations using the function `extract` from the Raster package in R. The combined play fairway metrics were for the 5-color scheme. Figure 57 plots the comparison of the metrics provided that all points were where all risk factors were defined and the minimum value was greater than zero (105 total).

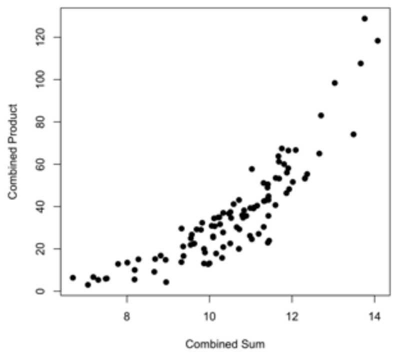
Generally, the metrics seem to match fairly well provided that all of the values are greater than zero. This could be a reasonable indication that the metrics would tend to give the same relative rankings. The coloring scheme is discrete, so the maps could look significantly different even if the relative rankings are the same.



(a) Minimum and product.



(b) Minimum and sum.



(c) Product and sum

Figure 57: Scatter plots comparing the play fairway metric for the three different measures (sum, product, minimum) when combining all four risk factors for areas with non-zero utilization scaled risk factor values.

4 Uncertainty of Risk Factors and Uncertainty for Combined Risk Maps

We are not only interested in the mean value of the play fairway metric, but also the uncertainty. Each group working on a geologic risk factor submitted an uncertainty map. The uncertainty sometimes is statistically based, and in other cases based on the professional judgment of members in the group. The first objective was to convert the uncertainty in a risk factor into the uncertainty of the scaled risk factor (risk factor converted into the play fairway scheme), and display these results in a map. The second objective was to use the uncertainties in the scaled risk factors to estimate the uncertainty in the combined play fairway metric.

There are two main methods of approximating the variance (uncertainty) of a function given the uncertainty in the inputs: Taylor series approximations and Monte Carlo simulation. The Taylor series approximation is given in Equation 6, where PFM is the combined play fairway metric (e.g. sum or product), SRF_i is the scaled risk factor (scaled to the play fairway system) for risk factor i, RF_i is risk factor i (e.g. thermal or seismic), Var(.) is the variance, μ is the mean or mean vector of values denoted in the subscript. This approach is useful for many applications, but in our case the function used to map RF_i to SRF_i, based on the thresholds, has many kinks and is often constant over large ranges. The kinks make the partial derivative discontinuous and the constant areas make the derivative equal to zero; however, we still anticipate some uncertainty in SRF_i. One advantage of this approach is that it could be completed with the existing rasters fairly easily [Note: Equation 6 assumes that there is no covariance between the risk factors.]

$$\begin{aligned}\mathrm{Var}(\mathrm{PFM}) &\approx \sum_{i=1}^{n}\left[\frac{\partial \mathrm{PFM}(\mu_{\mathrm{SRF}})}{\partial \mathrm{SRF}_i}\right]^2 \mathrm{Var}(\mathrm{SRF}_i) \\ &\approx \sum_{i=1}^{n}\left[\frac{\partial \mathrm{PFM}(\mu_{\mathrm{SRF}})}{\partial \mathrm{SRF}_i}\right]^2 \left[\frac{\partial \mathrm{SRF}_i(\mu_{\mathrm{RF}_i})}{\partial \mathrm{RF}_i}\right]^2 \mathrm{Var}(\mathrm{RF}_i) \qquad (6)\end{aligned}$$

The Monte Carlo simulation approach to calculating the variance of PFM is to generate a random sample from the distribution of each risk factor at a pixel (raster cell) and then calculate the variance of the resulting distribution of PFM. In principle this should be more accurate, but using a separate Monte Carlo trial for each pixel becomes challenging computationally. It is also limited in accuracy by the distribution used to generating replicates (random values) of each risk factor.

Because of the considerations above, a mixed method was selected. For each individual risk factor, RF_i, the variance of its SRF_i was calculated using the Monte Carlo methods for different fixed values of the mean and the uncertainty (e.g. standard error or coefficient of variation). Bounds of this table were selected so that it encompassed all of the values seen in the distributions present. Each cell in the table represented a value of the mean and a value of the uncertainty for the risk factor, and the variance of SRF_i was determined based on 100,000 Monte Carlo replicates drawn from the distribution with the specified parameters.

After the table was developed, for an individual risk factor, RF_i, the variance of SRF_i could be approximated by linear interpolation in the table. Using this method avoids problems with zero derivatives and kinks in the mapping function from RF_i to SRF_i. The function `interp2` in the R package 'pracma' was used for the interpolation.

The Taylor approximation in Equation 6, line 1, can then be used to approximate the variance of a PFM. If the PFM is a sum, the equation degenerates to an accurate value because the variance of a sum is the sum of the variances. We are assuming that values are independent, so there will be no covariance among terms. For the product PFM, the result will be approximate.

The method of combining the maps using the minimum should not use the Taylor series approximation for the variance because the minimum function is very poorly approximated by the Taylor series if the risk factor that defines the minimum is not always the same (e.g. sometimes utilization and sometimes thermal). The distributions of all the scaled risk factors is likely different and closed form solutions are not readily available. One might derive the uncertainty using Monte Carlo, but this is infeasible for this project because calculations would have to be performed thousands of times at each pixel (raster cell).

5 Potential Project Locations

In addition to the maps, we can also consider a specific set of project locations and show the risk factors in more detail for these sites. After discussion with Prof. Teresa Jordan, the project locations were selected as listed in Table 6, mainly because the geological factors are fairly favorable with a reasonable population or other demand for heat.

The processing used to obtain values was that the values at the center of each raster cell were obtained using the `extract` function in the R package 'raster'. The raster was defined on the UTM 17N projection. The locations of each place in Table 6 were converted into the NAD83, UTM17N projection, and the distance between each of the raster cell centers and the project locations was calculated. Next, the cells within 10 km (10,000 m) were considered as representing reasonable locations near the project location,and the maximum value of the summed map of all four risk factors was used to select the values for the project location. Ithaca, NY, is not on the plots because the nearest raster cell with all risk factors defined was more than 10 km away.

Figure 58 plots the combined metrics and their uncertainty. The uncertainty for all of the geologic factor combinations is represented by boxplots, which are a representation of the data that indicate the center and general spread of the data. The distributions show clear differences is symmetry and spread.

Figure 59 is a parallel axis plot of the project locations. The lines are color-coded by location. This plot illustrates that there can be trade-offs between locations, so a more favorable value in one risk factor might mean you have to accept a less favorable value in another.

Table 6: Table of project locations.

Town/City/Area	State
Corning/Elmira	NY
Ithaca	NY
Jamestown	NY
Mansfield College	PA
Meadsville	PA
Sayre	PA
Charlestown	WV
Morgantown	WV
Pineville	WV

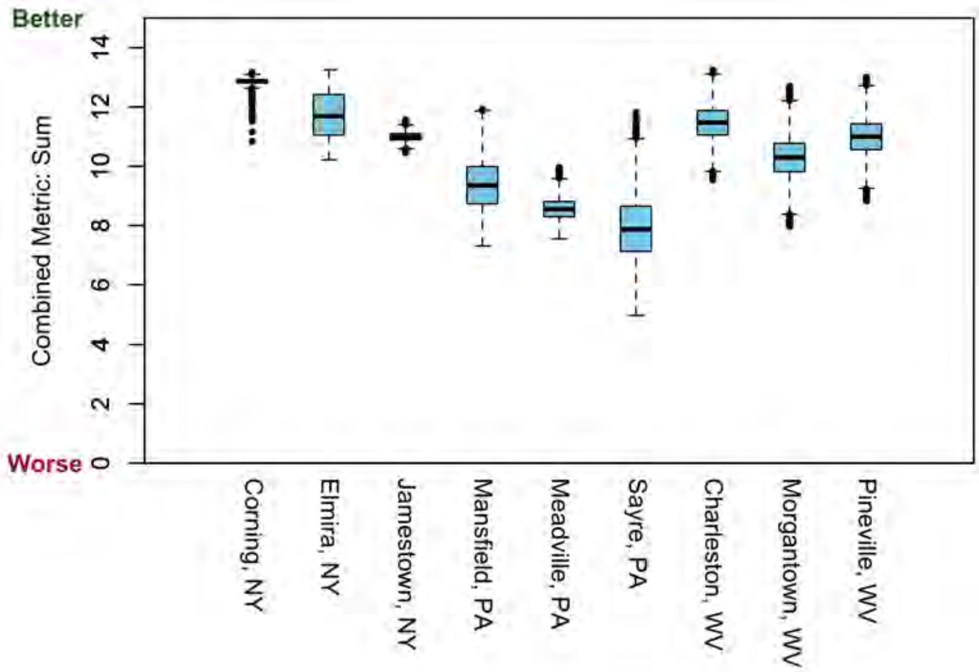


Figure 58: Boxplots showing the Monte Carlo distribution of the combined play fairway metric (PFM) for the summed geologic risk factors. The box extends from the 25th to 75th percentiles with a line at the median. The whiskers extent to the most extreme point that is within 1.5 times the interquartile range (25th to 75th percentiles distance). Points beyond the whiskers are plotted individually. High values are favorable.

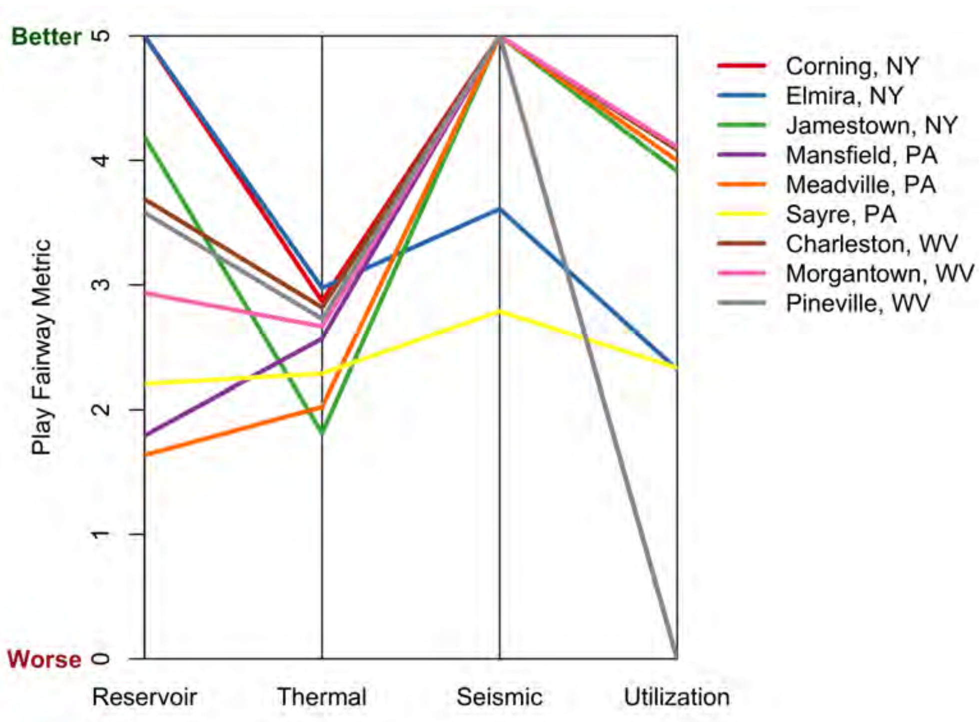


Figure 59: Parallel axis plot for nine illustrative site locations. The lines are color-coded by location. High values are favorable and low values are unfavorable.