To: Appalachian Basin Geothermal Play Fairway Analysis Group

From: Calvin Whealton, Jery Stedinger, Frank Horowitz, and Jared Smith

Date: September 17, 2015

Subject: Risk Analysis and Required Risk Factor Descriptions

In preparation for the analysis of the combined risk factors, it is wise to resolve what information will be needed to construct the final 4 risk-factor maps and the resultant summaries, and to conduct appropriate sensitivity and uncertainty analyses. Having this discussion now will allow more time for implementation later, and for those generating critical results to include in their analysis the ability to generate the needed information.

This memo focuses on required map data and formats, including methodologies for uncertainty analysis and sensitivity analysis of the final risk matrix. Methods of scaling each risk factor to a 3-point or 5-point scale are also described. Two critical issues are addressed: (i) completing the risk-matrix tasks as they are described in the SOPO, and (ii) being creative and innovative in methodologies for the analysis of the risk matrix information, and visualization of the results.

This memo extends the 1 April 2015 memo entitled "Combining Risk Factors." One difference here is an emphasis that, when possible, map colors for 3-color or 5-color maps should be related to the *actual acceptability* of a location measured on that risk index at the scale of the analysis. Using this point of view, they are not relative metrics providing just a comparison to other locations or projects, but absolute evaluations of project acceptability. If color ranking represents the scale of project acceptability, then it is reasonable, for example, to consider the minimum value across the four risk factors as a criterion for project acceptability. Other options for representing the combined risk maps are presented below.

We claim that it is critical for groups to submit correctly formatted maps to 1) construct the risk matrices, 2) efficiently complete the effort which requires analysis of alternative projects, and 3) conduct uncertainty analysis and visualization of results.

List of Appendices:
1: Methods for creating the standardized raster grid for all risk factors, and converting vectors to the standardized raster grid.

Map Data and Formats

This section outlines the required data for each risk factor map. There should be two or three raster maps submitted for each risk factor:

1) the value of a continuous risk variable (one raster), and
2) the uncertainty, described by either the standard deviation or coefficient of variation (1 raster), or 5-95% CI (2 rasters), of the estimated risk value at each point.

The raster should include a value for "no information," which is equivalent to stating that we do not have data to estimate the risk. The selected value for "no information" is "-9999" for this project (a numerical value is used because raster formatting does not allow text).

"No information" should be distinguished from areas that have information, but are not suitable for a geothermal project. One may wish to consider development in regions for which no data exists; however we would not consider development where a risk factor has an infeasible value. When mapping, no information (-9999) may be represented as white pixels for individual risk factors. These pixels would remain white when combining risk factors into a summary statistic. For example, if no information about reservoirs exists, but data regarding seismic, thermal, and utilization risk factors are available, then a summary statistic considering all risk factors would receive a white pixel; whereas a map that considers only seismic, thermal, and utilization risk factors would be colored according to the selected color scheme.

Groups must also submit *thresholds* for a 3-color map (2 thresholds) or 5-color map (4 thresholds). When possible, threshold values should be based on previous studies of what is considered high, medium, and low risk (or performance levels) for each risk factor. Level 3 for a 3-color map and level 5 for a 5-color map are considered to be excellent conditions. Level 1 in both color schemes represents unsatisfactory levels, indicating a project would not be successful, regardless of the values of the other risk factors.

Finally, for the purpose of scaling or transforming the continuous risk variables into composite or summary values, we request for each risk factor a *maximum* and a *minimum* value be specified. We do not anticipate that these are the maximum and minimum values within this dataset across the map, but rather the maximum and minimum values that might be generated (or values we wish to employ for scaling to a 3-point or 5-point scale). The maximum and minimum values may also be viewed as extreme thresholds. For example, all values below the minimum useful temperature of 50 °C (scaled to a value of 0.0) also receive a scaled value of 0.0 because the temperature below 50 °C is not useful for this project's consideration.

Required deliverables:

1. Raster map* of the continuous risk factor (in original units) on the grid developed for the project (Raster grid name: GridNAD.tif).
2. Raster map(s)* representing the uncertainty of the continuous risk factor.
3. A set or sets of 2-thresholds for defining a 3-color map, including a third value for no information.
4. A set or sets of 4-thresholds for defining a 5-color map, including a fifth value for no information.
5. Maximum and minimum values for each risk factor.
* *See Appendix 1 for a description of how the project grid was developed, and how vector files (e.g. reservoirs) were converted into rasters.*

It may be unclear how to adequately represent one or more of the risk factors as a single raster set (risk metric and uncertainty rasters). In this case, we recommend that the risk factor be submitted in a manner that captures the complexity. The values of the ranking thresholds do not have to change between submitted sets of rasters. For example, a single depth at which the thermal risk factor should be represented is unclear because many depths could be selected and justified as reasonable (maximum economical drilling depth, minimum depth to reach the minimum use temperature, average depth between these two extremes, etc.). It is plausible that each of these depths could be evaluated using the same thresholds to arrive at a composite thermal risk. Alternatively, because the temperature corresponding to, for instance, "unfavorable" will change depending on the depth, temperatures in depth slices can be evaluated separately in the risk matrix analysis to provide depth-specific thermal risk factor contributions to the composite project risk. Therefore, when complexities exist with depth, we recommend that raster sets be submitted in incremental depth slices of the risk factor to allow for flexibility in computing the composite project risk. Thermal and reservoir risk factors have been submitted in this manner.

Proposed Simple Risk Standardization

A first computational task is to convert the continuous variable for each risk factor into the play fairway color scheme. The conversion into the play fairway ranking system (e.g. 0 to 3) will be accomplished using the thresholds, unique to each risk factor. For example, for a 3-color scheme with thresholds at 15 and 25, a risk metric value ≤ 15 would be plotted as red, a value between 15 and 25 would be yellow, and a value > 25 would be plotted as green. The map will not be colored using a continuous color bar; however values of the risk variable will be continuous.

Using the continuous values of each risk variable with its specified thresholds allows plotting the 3- or 5-levels of each variable across the map. But, how to combine these variables when they have completely different units, and perhaps very different scales is unclear. We propose to use the thresholds and min/max values with linear interpolation to generate standardized values for each risk variable.

For a 3-color system, one expects that 1 and 2 represent the thresholds. Zero is appropriate to represent a score for the minimum value of the variable, and three is appropriate to represent the best value of the variable. (This may require reversing the scale of some variables so that 3 is good and 0 is bad.)

Thus the ranges are:

0 – 1   Red = Bad, unacceptable

1 – 2   Yellow = okay, marginally acceptable

2 – 3   Green = good, advantageous

For a 5-color scheme, thresholds are located at 1, 2, 3, and 4, with 0 representing the minimum, and 5 representing the maximum value.

Very Unfavorable	Unfavorable	Neutral	Favorable	Very Favorable
0.0 1.0	1.0 2.0	2.0 3.0	3.0 4.0	4.0 5.0
Threshold 1	Threshold 2	Threshold 3	Threshold 4	

After linear interpolation of continuous risk variables to this standardized scheme, all of the variables will be on a 0-3 score for 3-colors, and a 0-5 score for 5-colors. Using these scores it would be reasonable to compute the minimum value of the standardized scores, their product, or a weighted average (functional forms, below). Perhaps of special interest would be the average score, $\bar{s}$, for all pixels whose minimum score exceeds s_{min} for each risk factor, where one might take s_{min} to be any of the thresholds (e.g. 1 or 2, or an intermediate value such as 1.6). Thus one would consider among all of the acceptable projects, those that appeared to do best overall.

Uncertainty Analysis and Visualization

A final step will be to select a few interesting locations for visualization of the four individual project risk factors, and consideration of uncertainty in the computed values.

Map-level

The SOPO deliverable is to combine all of the risk factors using a single objective function (risk matrix) to calculate the combined risk. The combined risk factor (CRF) for pixel (i,j) may be calculated using one of these suggested functions in Table 1.

Table 1: Functional forms for the combined risk factor calculation, along with advantages and disadvantages of each function.

Functional Form	Advantages and/or Disadvantages
Average of Risk Factors	Equal weighting. Not dependent on the scale (min/max values to represent color)
Product of Risk Factors	Penalizes areas where there is one especially low value. Highlights areas with one especially high value. Therefore the min/max values representing 0, and 3 or 5 become very important.
Minimum of Risk Factors	Focuses on the most unfavorable risk factor at a location because all of the other risk factors are at least as good as the minimum. Unclear how much better a location is in other risk factors.

Combinations of the approaches listed in Table 1 are also feasible. For example, one might first compute the minimum risk factor, and accept only those location above a selected threshold (say 2.0), and then plot the average risk factor for acceptable locations.

Examples of the functional forms listed in Table 1 are provided in the equations below. The utilization risk factor has a tilde accent because it will be calculated as the maximum of the utilization values within a certain distance *d* of the pixel of interest. The distance *d* can be thought of as a pumping distance to reach the utilization target. In this calculation of the CRF map, each pixel is where to develop the resource, not necessarily where surface utilization is present. Using a utilization distance avoids potential problems with sharp edges to reservoirs. Areas without reservoir information would be "whited out" in the risk factor maps and also in the CRF map because of no information, which might lead to abrupt transitions from high CRF areas to no information CRF areas. The utilization distance would alleviate this problem because areas on the no information side of such a boundary would be able to utilize the resource a few pixels away.

$$\text{CRF}(i,j)$$

$$= \begin{cases} \dfrac{\text{RF}_{\text{thermal}}(i,j) + \text{RF}_{\text{reservoir}}(i,j) + \text{RF}_{\text{seismic}}(i,j) + \widetilde{\text{RF}}_{\text{utilization}}(i,j,d)}{4}, & \text{Average} \\[2mm] \text{RF}_{\text{thermal}}(i,j) * \text{RF}_{\text{reservoir}}(i,j) * \text{RF}_{\text{seismic}}(i,j) * \widetilde{\text{RF}}_{\text{utilization}}(i,j,d), & \text{Product} \\[2mm] \min\{\text{RF}_{\text{thermal}}(i,j), \text{RF}_{\text{reservoir}}(i,j), \text{RF}_{\text{seismic}}(i,j), \widetilde{\text{RF}}_{\text{utilization}}(i,j,d)\}, & \text{Minimum} \end{cases}$$

For the commensurate function calculated using any of the above equations, a combined play fairway map is defined. The advantage of using a continuous variable on the play fairway scheme is that it allows for more precision in numerical values used to compute the final map, and the uncertainty analyses as compared to a discretized metric (e.g. 3.2 vs. 3). The risk factor map coloring should only be discrete 3- or 5- colors as opposed to a continuous coloring scheme that could cause slight color differences in areas to appear meaningful.

Project-level

Once the final combined risk map is generated, a few project sites should be selected for more detailed presentation of results. For these individual sites, and with the uncertainty for each risk factor map, a Monte Carlo or first-order Taylor Series analysis can be used to arrive at the distribution of the commensurate risk statistic for that site. An example of such a plot is given in Figure 1. Using such methods will allow for:

1. Evaluation of if a project might really be unsatisfactory
2. Test of statistically significant differences in the play fairway metric among project sites
3. Easy visualization of sites that seem to be more or less certain
4. Overall comparison of play fairway metrics across the most promising sites

Technically, we have a four-objective problem that was commensurated into the single play fairway metric. Once a group of a few potential project sites is selected, the problem can be represented again as a four-objective problem. One method of comparing sites is with a parallel axis plot, such as shown in Figure 2. Once the initial map has shown some likely good areas for development, a parallel axis plot will show four objectives for the same resource and will convey more information to a decision maker. It could be that there are large tradeoffs in the objectives, for instance the highest utilization might be near seismically active areas. Conveying more information for a few cases can be informative for a decision maker.
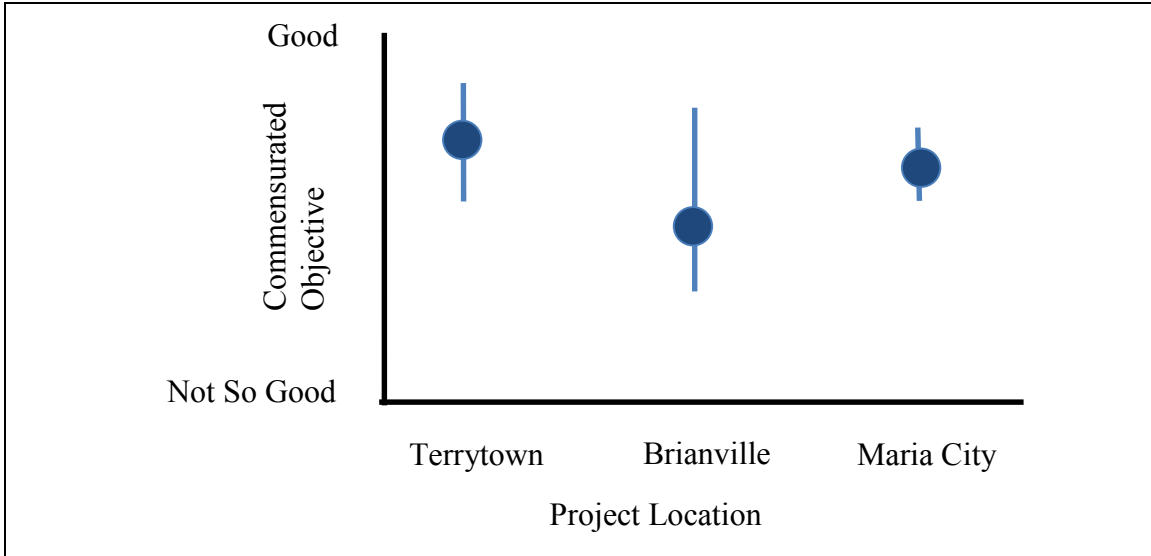
Figure 1: Example of plots comparing uncertainty in the computed commensurate play fairway objective for selected project locations. The circle would represent the combined risk factor map value and the bars could be the 5 and 95% limits based on Monte Carlo or first-order Taylor series analysis.
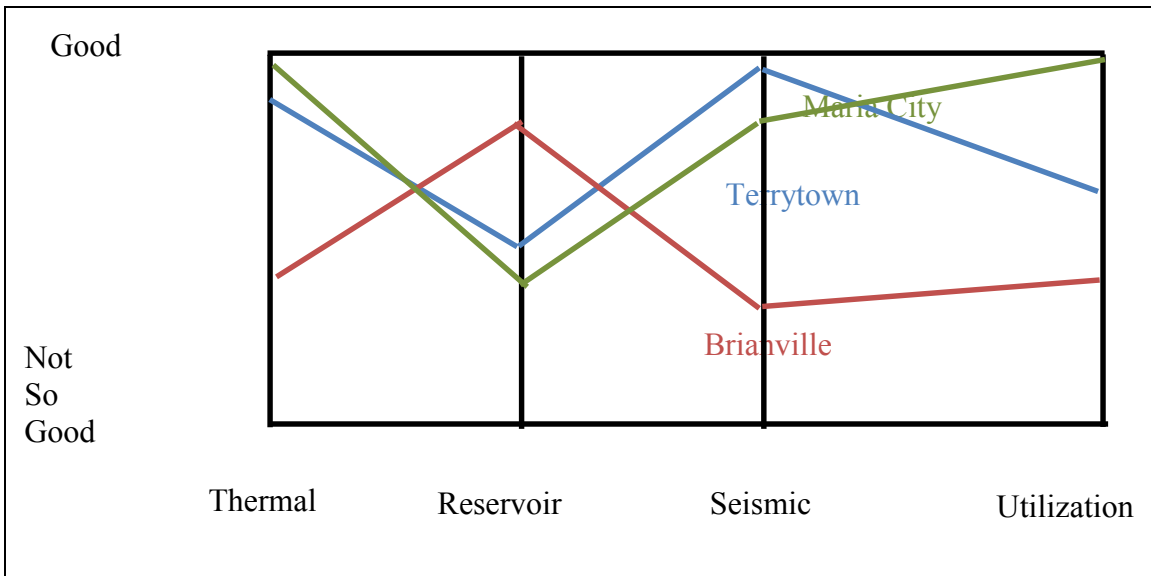

Figure 2: Example of a parallel axis plot where the objectives are plotted separately for each site. Such a plot would allow a decision-maker to look at the various objectives for a few sites and select places for further study.

Sensitivity Analyses

Sensitivity analysis of any CRF map will proceed by considering the impact of:

(i) Varying the thresholds for the individual risk factor maps, and
(ii) Varying the pumping distance d used to represent distance to utilization.

Other variables may also be explored.

This sensitivity analysis of the map does not address the uncertainty of inputs or of the economic model. The analysis in (i) addresses the thresholds used for green/yellow/red classification on the output summary. If we changed our thresholds, would the output map be very similar or drastically different? (ii) This set of results looks at the impact of the selected value of d describing feasible pumping distance.

Discussion of Weighting

The Cornell SOPO mentions weighting when combining the risk maps. One can assign weights, but if one does not have a justification for the weighting scheme, then it is arbitrary and could be (unknowingly) manipulated to favor certain sites or general conclusions. Also, weights can easily be manipulated by specifying special cases for certain geographic areas.

One justifiable method of weighting would be based on an economic model of the system for each pixel. In this case the weights would be derived economically because the cost of the project could be divided into portions associated with each risk factor. For instance, the levelized cost of heat (LCOH) would reflect the thermal quality of the reservoir, the natural quality of the reservoir, and the utilization infrastructure and pricing. Seismic hazard could be reflected by insurance premiums.

A problem with the use of an economic model is that, although it might be related to the cost of the project, it does not model the preferences of different agents. It is possible that willingness-to-pay for geothermal heat will be different across the agents, and this would be difficult to capture accurately in a model without clear agent preference data. Additionally, accurately specifying all inputs, or input distributions, would be very time consuming.

We recommend that weights initially be assigned equally to all risk factors in calculating the commensurate play fairway metric. Time permitting, other options can be explored.

Closing Thoughts

Using simple metrics will provide easy to interpret results for decision makers. The map itself should not be considered a complete analysis of whether developing a geothermal resource at that site is appropriate because a much more detailed analysis should be done

before a major decision is made. Completing a holistic analysis is outside the scope of this initial portion of the project. Realistic economic and risk analyses would include a range of factors and considerations, which would be unique for each developer. The individual risk factor (RF) maps and composite risk factor (CRF) maps described here support and allow an initial or exploratory analysis of the development of geothermal resources in the Appalachian Basin.

Appendix 1: Methods for creating the risk factor grid, and converting vectors to rasters

Creating the raster grid

From a computational perspective, there is a need to have a single grid on which to plot all risk factors so that calculations of the combined risk factor metric may be easily performed without manipulation of the original data to a different grid via resampling of the risk factor products. Two main considerations are needed for the grid: 1) what should the spatial extent be, and 2) what should the resolution of the grid be?

The spatial extent of the grid was determined based on the rectangular area enclosing the 50 km buffered states of New York, Pennsylvania, and West Virginia (State_NADBuff.shp). Much of the grid area will not have data for most risk factors, but rasters must be rectangular, so this is the smallest possible extent of the grid.

In this project, a 1 km^2 pixel size is selected for the grid. This size was selected based on the minimum expected size of a single risk factor output. In this case, seismic risk and reservoir risk are determined based on buffered points. The minimum reservoir extent is slightly greater than 1 km^2, so the pixel size needed to capture these reservoirs is 1 km^2.

Using this information, the project grid was created in three formats: points (cell centers) and polygons using the Create Fishnet tool in ArcGIS. The resulting files are Fishnet2_label.shp and Fishnet2.shp, respectively. It is important to note that this grid was created in coordinate system NAD83 UTM Zone 17N so that cells were all of equal size (1 km on each side). The input cell size in UTM coordinates is 1000 m x 1000 m, with a "template" plotting extent of the buffered states. The polygons were converted into a raster called GridNAD.tif using the Polygon to Raster tool in ArcGIS. A combination of points, polygons, and the raster grid were used in extracting individual risk factor data to this standardized grid format.

For final maps, this grid was projected into WGS84, so the cell sizes may warp, but they will occupy a 1km^2 area.

Converting vector files to rasters

The thermal risk factor is computed directly on the grid and does not require any conversion from vector to raster; however some processing is required (see Thermal section below). All other risk factors are computed as vector files that must be converted to rasters on this standardized grid. A simple tool called PolygonToRaster_wPolygons was developed in ArcGIS Model Builder for this conversion process. The steps of the tool are outlined here.

1) Spatial Join of the risk factor polygons to the standard grid polygons, ensuring that many risk factor attributes can be joined to a single grid cell. This is important when overlapping occurs within a single risk factor (for example, reservoirs at different depths in the same surface location).
2) Convert the specified field in the joined data files to a raster using the Polygon to Raster tool. The cell size of the resulting raster should be specified as the cell size of the GridNAD.tif raster. The result is a raster of the risk factor.

This tool was sufficient for converting the utilization vectors to rasters. Reservoir and seismic risks required additional processing.

Reservoirs:

Specific to reservoirs, a model was built to deal with overlapping reservoirs in the same spatial location. This model was called ReservoirConversion_FINAL. The steps of this model are summarized below.

Prior to processing, reservoirs must be sorted into depth slices from 1000 m to 4000 m in 500 m increments, plus one file for reservoirs shallower than 1000 m, for a total of 7 vector files.

1) Call the PolygonToRaster_wPolygon tool, with inputs as the depth slice of interest and the standardized grid polygons. This resulted in a raster of the risk factor and an intermediate file.
2) The intermediate file from this tool is the polygons containing the joined risk factor data. The uncertainty of the risk factor is a field in the attribute table of this file, so this field is converted to a raster using Polygon to Raster tool in ArcGIS. This resulted in a raster of the uncertainty in the risk factor.
3) These files contain data in the spatial extent of the reservoir polygons and ArcGIS default NoData elsewhere. NoData values are not desirable for calculations, so all of these NoData points are converted to -9999 to indicate no information using the Raster Calculator tool. This processing takes place in a separate model developed for this project, called FullRegionGrid, described below.

The FullRegionGrid model takes an input raster and converts NoData locations to -9999. The steps are as follow:

1) Use the Extract Values to Points tool to extract the raster with NoData values to the cell centers of the standardized grid (Fishnet2_label.shp). This field will be named the Arc default RASTERVALU

2) Use the Add Field tool to add a field name to the attribute table. This will be the field for the resulting raster.

3) Use the Calculate Field tool to populate the values in the added field with the RASTERVALU.

4) Convert the points to a raster using the Point to Raster tool.

5) Use Raster Calculator to convert all NoData (IsNull) fields to -9999.

Seismic:

Earthquake-Based Maps

A tool called SeismicEQ_ToRaster was created to convert the distance to the nearest earthquake information to a raster of distance to nearest earthquake.

1) Use the Spatial Join tool to join the distance to nearest earthquake attribute table to the standardized grid points.

2) Use the Point to Raster tool to convert these points into a raster.

A tool called New_EQJoin was created to make rasters of the

1) Clip the raster of distance to nearest earthquake to the buffered gravity worm points (buffering is described in the Seismic Risk Map Creation Methods memo).

2) Clip the raster of distance to nearest earthquake to the buffered magnetic worm points.

3) Use Extract Multi Values to Points tool to add this clipped raster information to the standardized grid points.

This results in a point file that has distance to the nearest earthquake for the buffered gravity and magnetic worm locations. These are processed to obtain the risk metric, as described in the Seismic Risk Map Creation Methods memo. Post processing, the risk metric and the uncertainty are converted into rasters on the standardized grid using the Polygon to Raster tool.

Stress Field-Based Maps

A tool called SeisStressMagGrav was developed to convert the information about angle to normal into a raster.

1) Buffer and dissolve the magnetic or gravity points by a selected amount (2 km in this project).

2) Use the Polygon to Raster twice: once to convert the prediction angle, and once to convert the uncertainty. The priority field should be set to the Weight field.
3) Use the Extract Multi Values to Points tool to add the prediction and uncertainty information to the standardized grid points.

This results in a point file that has the angle to the critical stress and the uncertainty in that angle. These points are processed to obtain the risk metric as described in the memo "Conversion of Seismic Risk Data to Risk Maps". Post processing, the risk metric and the uncertainty are converted into rasters on the standardized grid using the Polygon to Raster tool.

Thermal:

Converting separate raster files to a single raster

The thermal risk factor maps are created using laterally stratified boundaries. This means that the resulting rasters do not occupy the original extent of the standardized grid. In order to convert these files into a single raster in standardized grid format, two tools were developed.

The Final_MosaicPred tool takes all of the resulting rasters for the thermal risk factors and combines them into a single raster.

1) Make a copy of one of the rasters to be combined, save it in a different directory, and rename it. This will be the combined raster file.
2) Use the Mosaic tool in ArcGIS to combine the individual rasters. The target raster is the raster that was copied.

The resulting raster is on the standardized grid, but it does not occupy the full extent of the grid. The ThermalFiles_Final tool converts the mosaicked raster onto the full extent of the standardized grid.

1) Use the Extract Multi Values to Points tool to extract the prediction map and the uncertainty map raster information to the standardized grid cell centers.
2) Convert these points to 2 rasters: one for the prediction map, and the other for the uncertainty map.
3) Use Raster Calculator to convert all of the locations with values less than 0 to the no information value of -9999.