Using Spatially-Explicit Uncertainty and Sensitivity Analysis in Spatial Multicriteria Evaluation

Habitat Suitability Analysis

Arika Ligmann-Zielinska

Last updated: July 9, 2014
Based on Ligmann-Zielinska et al. (2012), Ligmann-Zielinska and Jankowski (2014)

<u>OBJECTIVES (OBJ)</u> You have been awarded a small grant from the National Science Foundation to conduct field work on an endangered plant called *Fictitia Flosculus* (FF) in Elysium.

Due to budget limitations, you decided to perform habitat suitability analysis to **find candidate sites**^{OBJ1} where *FF* could potentially be found. Given that you are a well-educated scientist, you also want to make sure that your candidate sites **carry a certain level of certainty** ^{OBJ2}, so that you do not have to spend too much time wandering in the field (after all, Elysium is a place of fun, not work). Finally, you want to **identify which of the habitat characteristics** ^{OBJ3} influence the uncertainty of the locations of *FF* populations the most, so that people of Elysium have enough information to restore FF's habitat and return the plant to its previous glory. The decision situation involves a total of 1369 locations. Experts from Elysium determined that the new locations of *FF* depend on distance to the last known population, forest cover, and soil suitability. Unfortunately, they greatly disagree on the degree of influence of these three habitat criteria on *FF* occurrence. Consequently, they provided distributions of criteria weights rather than distinct constant values.

Figure 1 depicts three standardized habitat criteria used in your research, which are expressed as raster surfaces of 37 rows and 37 columns (hence 1369 cells) with value range from 0.0 (worst) to 1.0 (best). The analysis is based on the premise that it would be easier to find *FF* in areas that are closer to its last known population (DIST), which are also more forested (FCOV), and which have higher values of soil suitability for *FF* (SOIL). Consequently, darker colors in Figure 1 represent the preferred locations for potential *FF* populations. The rasters are provided in *data/ascii* (text) and *data/grids* (ESRI raster) folders.

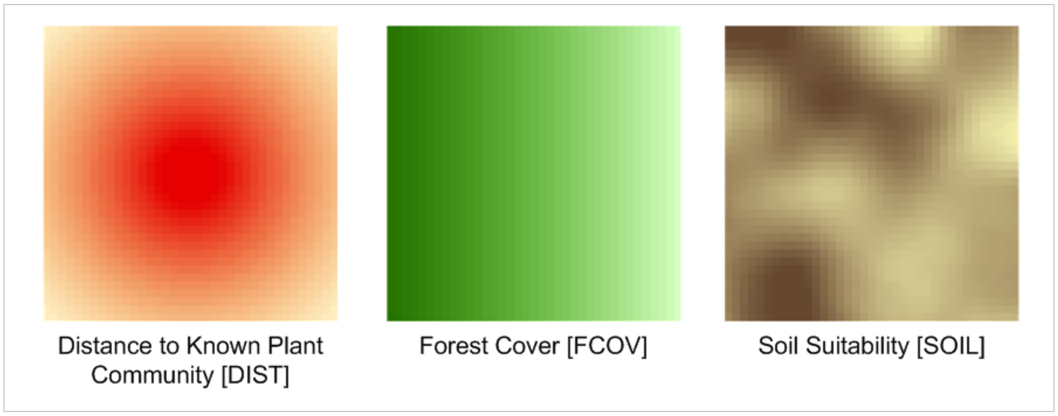


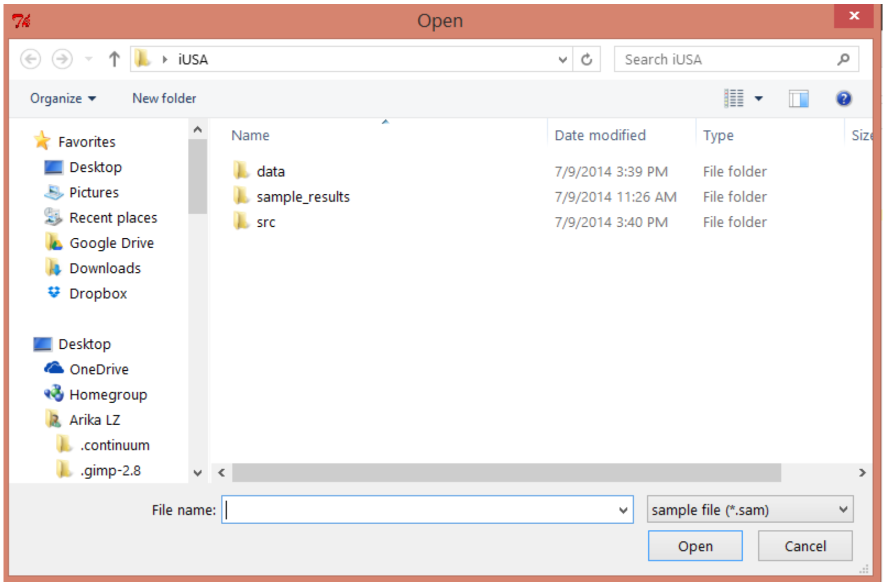
Figure 1. Input Criteria Maps. Darker colors indicate higher values.

METHODS/TASKS Your analysis consists of the following stages:

1. Employ Monte Carlo simulation to sweep through the uncertain weight space, where weights are expressed using probability density functions (PDFs) derived from expert opinion. Weight PDFs have already been approximated and sampled. The samples are provided in the form of **.sam* files located in *data/sample_files* folder. For interested: the samples were generated using quasi-random Sobol' experimental design with the radial sampling technique (Saltelli, Annoni et al. 2010).
2. Perform uncertainty analysis (UA): generate multiple output suitability maps, which are then summarized by calculating an average suitability surface (AVG for **OBJ1**) and an uncertainty surface (using a standard deviation of suitability maps – STD – for **OBJ2**).
3. Employ a model-independent method of sensitivity analysis (SA) based on output variance decomposition, in which the variability of suitability maps is broken down and apportioned to every input weight, generating one first-order (S) and one total-effect (ST) sensitivity index per criterion weight. You will apply variance decomposition to every pixel of the suitability map, producing a separate sensitivity map per every input weight (**OBJ3**). Details of this computational algorithm are presented in Saltelli et al. (2010).
4. Analyze the results and provide recommendations.

To account for possible criteria correlations, you will employ the Ideal Point (IP) aggregation function to calculate the habitat suitability scores.

SIMULATIONS Steps 1 through 3 will be carried out using a Python script called *iUSA.py* located in U-SA_LAB folder. The script requires Python 2.x with the numpy package installed on your computer. To execute the script, double click it. An 'open file' dialog box opens:



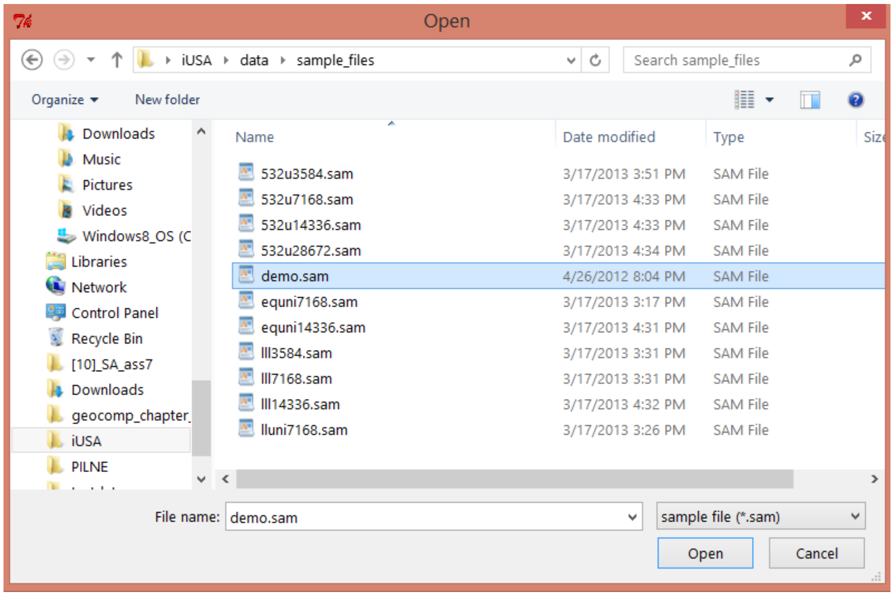
Name	Date modified	Type
data	7/9/2014 3:39 PM	File folder
sample_results	7/9/2014 11:26 AM	File folder
src	7/9/2014 3:40 PM	File folder

File name:

sample file (*.sam)

The script requires the *.sam* sample file as an input. Multiple *.sam* files were generated for this exercise using SimLab 2.2 software for uncertainty and sensitivity analysis developed by the European Commission's Joint Research Centre. In case you would like to generate *.sam* files yourself you can download a free copy of SimLab 2.2 from the Web site: http://ipsc.jrc.ec.europa.eu/index.php?id=756#c2907. Use *Sobol* design to generate the

.sam files. The raster data, in the form of ascii grid files, have already been coded into the script. For now, navigate to the sample files folder and select *demo.sam* to execute the script.

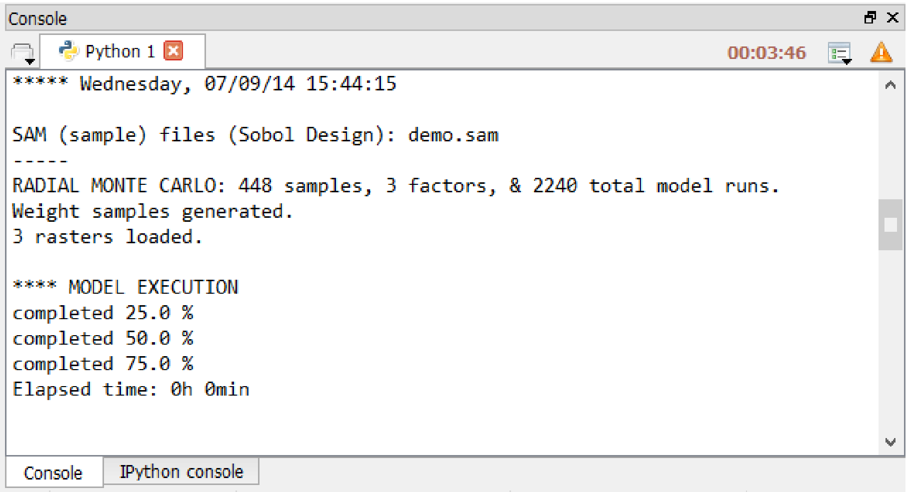


Name	Date modified	Type
532u3584.sam	3/17/2013 3:51 PM	SAM File
532u7168.sam	3/17/2013 4:33 PM	SAM File
532u14336.sam	3/17/2013 4:33 PM	SAM File
532u28672.sam	3/17/2013 4:34 PM	SAM File
demo.sam	4/26/2012 8:04 PM	SAM File
equni7168.sam	3/17/2013 3:17 PM	SAM File
equni14336.sam	3/17/2013 4:31 PM	SAM File
lll3584.sam	3/17/2013 3:31 PM	SAM File
lll7168.sam	3/17/2013 3:31 PM	SAM File
lll14336.sam	3/17/2013 4:32 PM	SAM File
lluni7168.sam	3/17/2013 3:26 PM	SAM File

File name: demo.sam

sample file (*.sam)

Since the sample is small (~2240 Ideal Point executions) the script should finish within seconds. Look into the model progress window:



```
***** Wednesday, 07/09/14 15:44:15

SAM (sample) files (Sobol Design): demo.sam
-----
RADIAL MONTE CARLO: 448 samples, 3 factors, & 2240 total model runs.
Weight samples generated.
3 rasters loaded.

**** MODEL EXECUTION
completed 25.0 %
completed 50.0 %
completed 75.0 %
Elapsed time: 0h 0min
```

It is useful to copy the content of this window for analysis. Simply right-click anywhere in the command window, select "select all", and click Enter (or Ctrl+C). The content is automatically copied to clipboard and can be pasted into a text document, e.g.

```
***** Wednesday, 07/09/14 15:44:15

SAM (sample) files (Sobol Design): demo.sam
-----
RADIAL MONTE CARLO: 448 samples, 3 factors, & 2240 total model runs.
Weight samples generated.
3 rasters loaded.

**** MODEL EXECUTION
completed 25.0 %
completed 50.0 %
```

```
completed 75.0 %
Elapsed time: 0h 0min


S min max -2.77 1.99
INEXACT S values for max: cells with val of >  1.00 over all factors (k): 33 [ 0.8 ] %
INEXACT S values for min: cells with val of < -0.05 over all factors (k): 671 [ 16.3 ] %



ST min max 0.09 0.8
INEXACT ST values for max: cells with val of >  2.00 over all factors (k): 0 [ 0.0 ] %
INEXACT ST values for min: cells with val of <  0.00 over all factors (k): 0 [ 0.0 ] %
Negatives converted to 0.0. 0.0 % of raster area had negative values
Output ASCII grid C:\Users\Arika\Desktop\iUSA/results_D2014_07_09T15_44_18/UA_avg_map.asc saved.
Negatives converted to 0.0. 0.0 % of raster area had negative values
Output ASCII grid C:\Users\Arika\Desktop\iUSA/results_D2014_07_09T15_44_18/UA_min_map.asc saved.
Negatives converted to 0.0. 0.0 % of raster area had negative values
Output ASCII grid C:\Users\Arika\Desktop\iUSA/results_D2014_07_09T15_44_18/UA_max_map.asc saved.
Negatives converted to 0.0. 0.0 % of raster area had negative values
Output ASCII grid C:\Users\Arika\Desktop\iUSA/results_D2014_07_09T15_44_18/UA_std_map.asc saved.
Negatives converted to 0.0. 19.87 % of raster area had negative values
Output ASCII grid C:\Users\Arika\Desktop\iUSA/results_D2014_07_09T15_44_18/S_dist.asc saved.
Negatives converted to 0.0. 0.0 % of raster area had negative values
Output ASCII grid C:\Users\Arika\Desktop\iUSA/results_D2014_07_09T15_44_18/ST_dist.asc saved.
Negatives converted to 0.0. 15.49 % of raster area had negative values
Output ASCII grid C:\Users\Arika\Desktop\iUSA/results_D2014_07_09T15_44_18/S_fcov.asc saved.
Negatives converted to 0.0. 0.0 % of raster area had negative values
Output ASCII grid C:\Users\Arika\Desktop\iUSA/results_D2014_07_09T15_44_18/ST_fcov.asc saved.
Negatives converted to 0.0. 22.43 % of raster area had negative values
Output ASCII grid C:\Users\Arika\Desktop\iUSA/results_D2014_07_09T15_44_18/S_soil.asc saved.
Negatives converted to 0.0. 0.0 % of raster area had negative values
Output ASCII grid C:\Users\Arika\Desktop\iUSA/results_D2014_07_09T15_44_18/ST_soil.asc saved.
Elapsed time: 0h 0min

*****Wednesday, 07/09/14 15:44:18

Simulations ended. Press any key to close the app...
```

Let’s analyze the content of this log file step-by-step.

Lines 1-14 summarize inputs and report % completed:

```
***** Wednesday, 07/09/14 15:44:15

SAM (sample) files (Sobol Design): demo.sam
-----
RADIAL MONTE CARLO: 448 samples, 3 factors, & 2240 total model runs.
Weight samples generated.
3 rasters loaded.

**** MODEL EXECUTION
completed 25.0 %
completed 50.0 %
completed 75.0 %
Elapsed time: 0h 0min
```

The details on sample size i.e. 448 samples (R) and 2240 (N) model executions are specific to the radial experimental design we adopted in the simulations (Saltelli, Annoni et al. 2010). In your analysis, simply focus on the number of samples provided in the sam file (i.e. 448 in the example above). Note, however, that the suitability calculations are actually performed 2240 times. The N number can increase dramatically even with a modest increase of R and k (*k* – number of factors/criteria).

Next, the accuracy of the Monte Carlo simulation is reported:

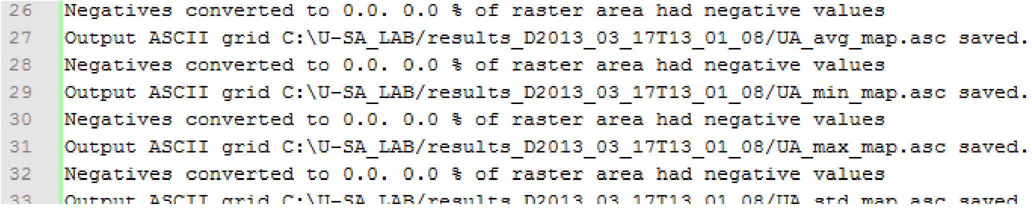
```
S min max -2.77 1.99
INEXACT S values for max: cells with val of >  1.00 over all factors (k): 33 [ 0.8 ] %
INEXACT S values for min: cells with val of < -0.05 over all factors (k): 671 [ 16.3 ] %



ST min max 0.09 0.8
INEXACT ST values for max: cells with val of >  2.00 over all factors (k): 0 [ 0.0 ] %
INEXACT ST values for min: cells with val of <  0.00 over all factors (k): 0 [ 0.0 ] %
```

The SA procedure provides, by definition, an approximation of the values of sensitivity indices (it is estimation of multi-dimensional integrals). You can assess the quality of your S/ST estimation by considering the following properties of S/ST indices:

1. Each S should be within the 0 to 1 range so that the sum of S is <= 1. If large negative values or values higher than one are obtained, the Monte Carlo sample was probably too small and hence the approximation is inadequate (negative values close to zero are OK).
2. The Sum of ST should be >= 1 and < 2.0. If ΣS= ΣST then the model is perfectly linear (an unlikely scenario in geographical modeling).

Since the *demo.sam* is based only on 448 samples, there are quite a few inexact values for S (a total of 17.1% of cells). In your experimentation, you are encouraged to select sample files with larger R.

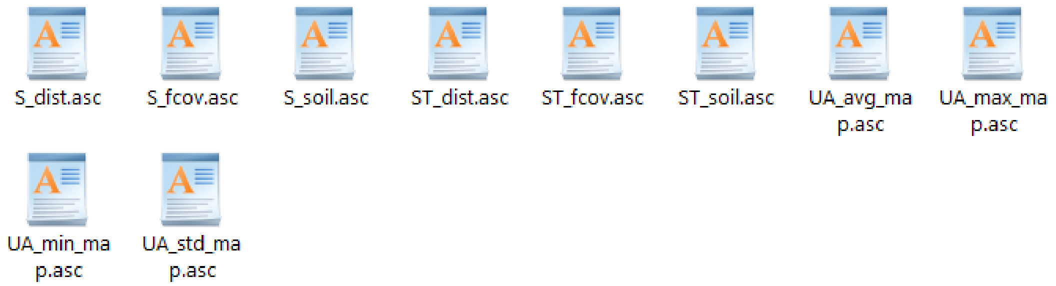
Finally, the log file lists the output:



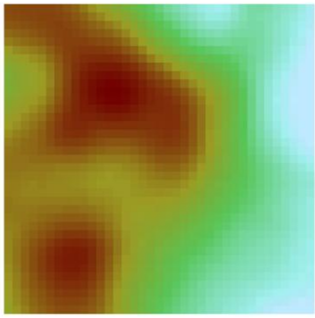
```
Negatives converted to 0.0. 0.0 % of raster area had negative values
Output ASCII grid C:\U-SA_LAB/results_D2013_03_17T13_01_08/UA_avg_map.asc saved.
Negatives converted to 0.0. 0.0 % of raster area had negative values
Output ASCII grid C:\U-SA_LAB/results_D2013_03_17T13_01_08/UA_min_map.asc saved.
Negatives converted to 0.0. 0.0 % of raster area had negative values
Output ASCII grid C:\U-SA_LAB/results_D2013_03_17T13_01_08/UA_max_map.asc saved.
Negatives converted to 0.0. 0.0 % of raster area had negative values
Output ASCII grid C:\U-SA_LAB/results_D2013_03_17T13_01_08/UA_std_map.asc saved.
```

All negative values are automatically converted to 0.0 and the % of "converted" area is reported. Observe that the number of negative values can be different for every criterion.

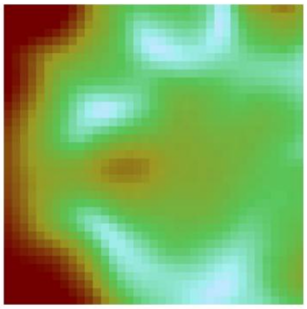
RESULTS The script generates a total of 10 maps in the form of ascii grid (txt) files. They are located in the *results_DayTime* folder, e.g *results_D2014_07_09T15_44_18* located in the iUSA directory:



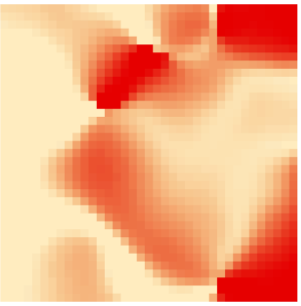
In particular, four uncertainty analysis maps are calculated: average, std, min, and max suitability scores per cell, and six sensitivity maps are generated i.e. three S maps (one per criterion) and three ST maps (one per criterion). You can view the results in any GIS software that provides *ascii to raster* conversion. Save the results of *ascii to raster* conversion as *float* rasters. For example, for the sample above and using ArcGIS:



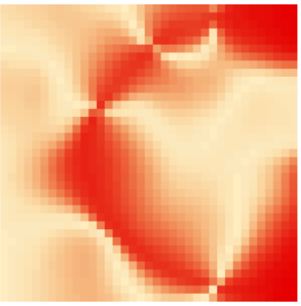
Average Suitability



Uncertainty (STD) Map



S for soils



ST for soils

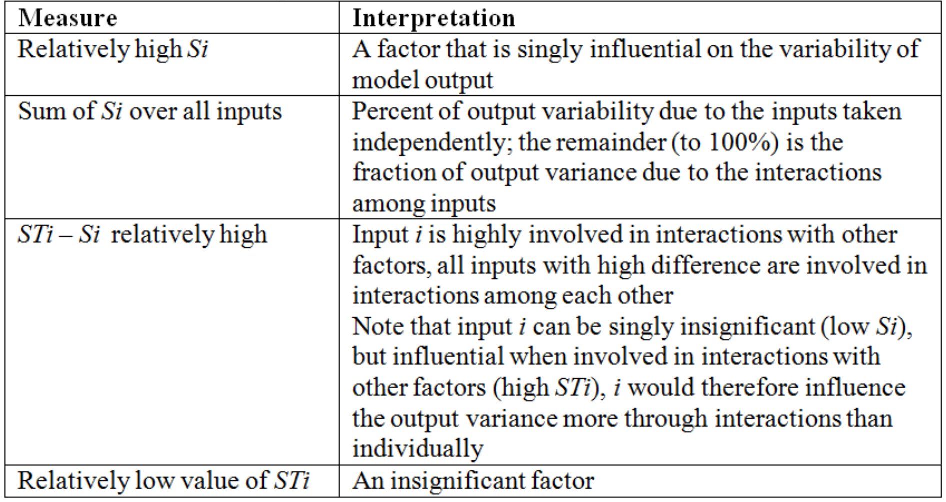
When reporting your maps, make sure to provide value ranges (e.g. min score and max score in the average suitability map etc.).

<u>SELF STUDY</u> Perform the analysis twice using different weight PDFs provided in the *./data/sample_files* folder.

Your report:

1. Input: the selected *sam* files. The files (the PDFs used and the size of sample – R) are described below. Note: using larger R gives more accurate results but large R requires more processing memory leading sometimes to memory errors - that's the tradeoff between accuracy and computational feasibility!
2. Results of Monte Carlo simulations (habitat suitability surface with suitability score stats: average suitability, min, and max, and your areas of interest i.e. of high suitability)
3. Results of UA (uncertainty surface; stdv raster, analyzed in combination with suitability surface)
4. Results of SA – sensitivity maps (criterion weights and locations where the evaluation criteria influence the variability of FF habitat suitability). Analyze both the first order and the total effect indices. Hint: to obtain regions of dominating weights you can use the HIGHEST POSITION tool in ArcGIS Spatial Analyst. Use the table below to guide your interpretation (Ligmann-Zielinska and Sun 2010).
5. Your final recommendation.

Table 1 Si and STi interpretation



Measure	**Interpretation**
Relatively high *Si*	A factor that is singly influential on the variability of model output
Sum of *Si* over all inputs	Percent of output variability due to the inputs taken independently; the remainder (to 100%) is the fraction of output variance due to the interactions among inputs
STi – *Si* relatively high	Input *i* is highly involved in interactions with other factors, all inputs with high difference are involved in interactions among each other Note that input *i* can be singly insignificant (low *Si*), but influential when involved in interactions with other factors (high *STi*), *i* would therefore influence the output variance more through interactions than individually
Relatively low value of *STi*	An insignificant factor

APPENDIX

Sample File (SAM) Descriptions (files located in ./iUSA/data/sample_files/)

All three criterion weight distributions (soil, distance, and forest cover) are uniform with value range from [0.0, 1.0]. There are two uniform distribution sample files:

equni7168.sam	R=7168
equni14336.sam	R=14336

Next, you can experiment the criterion weight distributions for narrower ranges of criterion weight values. Note that the ranges correspond to relative importance (bias) of the evaluation criteria.

Weight distribution for DIST Uniform with range [0.0, 0.5]
Weight distribution for FCOV Uniform with range [0.0, 0.3]
Weight distribution for SOILS is Uniform with range [0.0, 0.2]

532u3584.sam	R=3584
532u7168.sam	R=7168
532u14336.sam	R=14336
532u28672.sam	R=28672

Finally, you experiment with LogUniform probability density function for weight distributions. Distribution for DIST & FCOV is LogUniform with range [0.1, 0.5]
Distribution for SOILS is Uniform with range [0.0, 1.0]

lluni7168.sam	R=7168

Distribution for DIST is LogUniform with range [0.1, 0.3]
Distribution for FCOV is LogUniform with range [0.1, 0.6]
Distribution for SOILS is LogUniform with range [0.1, 1.0]

Lll3584.sam	R=3584
lll7168.sam	R=7168
lll14336.sam	R=14336

REFERENCES

Ligmann-Zielinska, A. and P. Jankowski (2014). "Spatially-Explicit Integrated Uncertainty and Sensitivity Analysis of Criteria Weights in Multicriteria Land Suitability Evaluation." Environmental Modelling & Software **57**: 235-247.

Ligmann-Zielinska, A., P. Jankowski, et al. (2012). Spatial Uncertainty and Sensitivity Analysis for Multiple Criteria Land Suitability Evaluation. Extended Abstract, Seventh International Conference on Geographic Information Science, Columbus, OH, U.S., September 18-21, 2012.

Ligmann-Zielinska, A. and L. Sun (2010). "Applying Time Dependent Variance-Based Global Sensitivity Analysis to Represent the Dynamics of an Agent-Based Model of Land Use Change." International Journal of Geographical Information Science **24**(12): 1829-1850.

Saltelli, A., P. Annoni, et al. (2010). "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index." Computer Physics Communications **181**(2): 259-270.