GEOG0114: PRINCIPLES OF SPATIAL ANALYSIS

WEEK 3: SUITABILITY MAPPING (PART I)

Dr Anwar Musah (a.musah@ucl.ac.uk)
Lecturer in Social and Geographic Data Science
UCL Geography

Structure of Today's Lecture

1. **What is suitability analysis (or mapping)**

2. **Case study: LF transmission in Kenya**
   - **Principles for using Suitability Analysis**
   - **Binary and Score-based mapping for suitability**
   - **Analytical Hierarchy Process (AHP)**

3. **Methodology for performing AHP analysis**
   - **Standardization of raster risk factors to a single scale**
   - **Pairwise comparison**
   - **Matrix calculations for Normalization and estimation of weights**
   - **Weighted Linear Combination Model**
   - **Validation**

4. **AHP Calculator tool and practical labs session**

Definition of Suitability Analysis:

Suitability Analysis (or Suitability Mapping) is a type of spatial analytical technique used often in GIS to **determine the best place or site suitable for an outcome**.

This is part of another spatial modelling technique referred as a **knowledge-driven approach**, which uses decision rules from existing knowledge to identify areas potentially suitable for a particular phenomena.

The key focus are as follows:

1) Performing **GIS-based multicriteria decision analysis (MCDA)** on an array of raster data to generate a composite for suitability
2) For suitability analysis, **GIS-MCDA** can be thought of as a process that transforms and combines geographical data (i.e., inparticular raster) and value **judgements (i.e., the decision-maker's preferences)** to obtain information for decision making

Important note: this is especially useful in data-sparse situations, or when first exploring the potential geographical extents or limits of a certain outcome

NOTES

NOTE 1: Up to this point, we have learnt spatial dependence and autocorrelation, and we focus on analysis of discrete data.

Here is quite different… we'll explore spatial analysis from a different lens using knowledge-driven approaches.

NOTE 2: This method is purely based on the knowledge, and the major part of this analysis is the user making judgements about certain risk factors in the spatial model and assigning levels of importance to them as weights before mapping suitability - this is done through pairwise comparison in Analytical Hierarchy Process (AHP). We will see how this done in latter slides (and in the practicals).

NOTE 3: What do I mean by data-sparse situations? Well, let me narrate a real situation where I had to implement this technique for partners in Kenya where no data is available to do initial investigation for the geographical distribution of Lymphatic Filariasis (LF), and to delineate the extents of where its transmission can occur in Kenya based on an array of raster that are risk factors for (LF).

CASE STUDY: GIS-MCDA applied to identify suitable areas for Lymphatic Filariasis (LF) transmission in Kenya.

NOTES

Description: There are no available up-to-date models to explain the occurrence of Lymphatic Filariasis (LF) in Kenya and **geospatial empirical data are scarce.** The Kenyan Ministry of Health (K-MoH), through its LF control programme, is planning to launch a public health intervention by introducing mass drug administration (MDA) of albendazole (combined with ivermectin) to infected people with LF and to remove microfilaria in their bloodstream. Mapping of suspected areas for LF must be carried out, however, due to financial constraints and limited resources, the K-MoH wishes to first **identify areas that are highly suitable for LF transmission** before spending this limited resources to survey, map and apply MDAs to these areas.

African man with heavy & chronic LF microfilaria infection, resulting in a swollen leg called 'Elephantiasis'

One many vectors, i.e. the Culex mosquito, that spreads LF by injecting microfilaria (microscopic worms) into their source of food (i.e., human) before taking its bloodmeal

Principles of using Suitability Analysis (1):

NOTES

In general**, GIS-MCDA** can be thought of as a process that transforms and combines a series of raster grids, and value **judgements of decision-maker's preferences** with regards to process being studied to churn out **new** information for decision making.

Important note: There are NO statistical modelling for this approach

MCDA, applied to spatial modelling of outcomes, includes the following stages:

1) Defining the **objective** of the modelling exercise or setting the goals
2) Defining the risk factors and potential constraints (i.e., **criteria**) that influence the outcome
3) Defining the relationship between each risk factor and suitability
4) Defining the relative importance of each risk factor in the relation to the objectives
5) Standardizing the factors to the small scale
6) Combining factors to create a composite variable (i.e., **suitability index**) through a weighted analysis to produce a final weighted estimate of suitability for each areas (i.e., pixel-by-pixel).
7) Model Validation

Principles of using Suitability Analysis (2):

Some important definitions:

Objective: The measure by which the decision rule operates. In a single multicriteria evaluation, it also considered the end goal.

Criteria: These are a set of guidelines or requirements used for deciding for what makes the outcome suitable for something. There are two types – **Factors** & **Constraints.**

Suitability Index: This measure is a function of the specified criteria which quantifies the appropriateness of a given area (i.e., pixel or grid cell) for a **specified use** or **suitability of a certain phenomena** or **for occurrence of certain events**.

NOTES

NOTE 1: In our LF case study, the **objective** is there to '**identify suitable areas for certain vector-borne transmission of LF in Kenya**'

NOTE 2: What is meant by **factors - t**hese are the specified criteria (i.e., risk factors) that enhance or reduce the suitability. How these criteria or inclusion factors are specified are entirely up to the user – usually it's done through a strict systematic literature review and expert opinion. A constraint on the other hand is meant a risk factor that serves to limit or restrict.

In LF example: rainfall, temperature, elevation, population density, vegetation, and aridity are factors that enhance or reduce suitability for LF transmission because they impact a mosquitoes' feeding habits and survival behavior etc. Desert regions is an example of a constraint – they cannot serve in this type of environment hence it limits/restrict their suitability.

NOTE 3: this technique in terms of 'specified use' could applied to identifying areas best for a particular land-use (i.e., agriculture, mineral deposit stores etc.,), 'suitability' (can refer to environments conducive for species to live, diseases to spread etc,) and occurrence (i.e., crime, disaster events etc.,).

Principles of using Suitability Analysis (3):

NOTES

Situate the problem to the LF case study

Objective: Setting the goal, or defining the problem – "**Identify potentially suitable areas for lymphatic filariasis in Kenya**"

Criteria: Defining the criteria (i.e., factors (**F**) and constraints (**C**)) and their thresholds for suitability (i.e., what positively (⬆)(or negatively (⬇)) impacts a mosquitoes' survivability when levels of factors increase).

- Temperature (> 15 degree Celsius) (**F**) (⬆)
- Precipitation (> 350 mm) (**F**) (⬆)
- Vegetation Index (> 0.5) (**F**) (⬆)
- Population Density (> 0) (**F**) (⬆)
- Elevation (<1,200m above sea-level) (**F/C**) (⬇)
- Aridity (> 0.2 (Semi-humid & dry environment)) (**F/C**)(⬆)

How do we map suitability? Honestly, there several ways for do this:

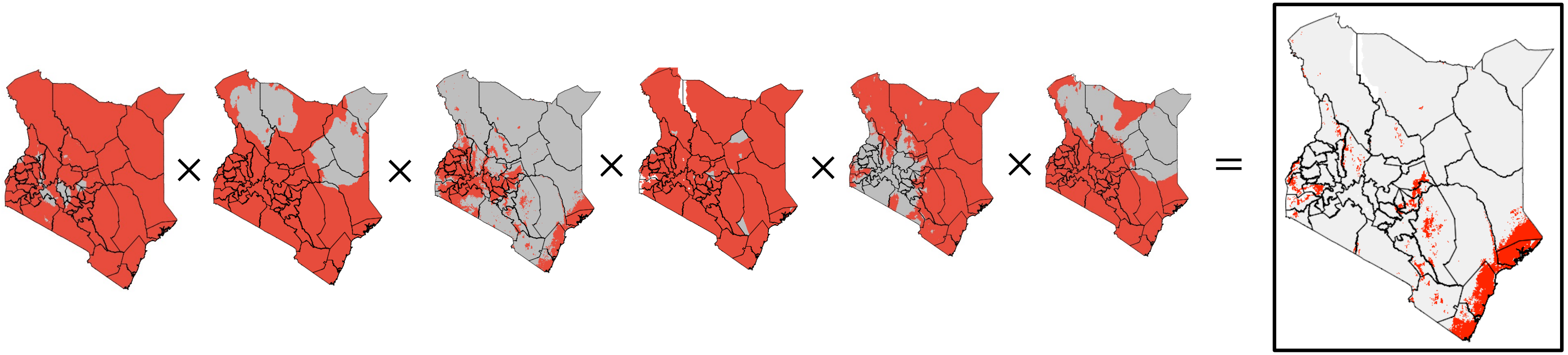
- Binary (or Boolean) maps with the factor layers
- Mapping the Ratings or Suitability scores
- Analytical Hierarchy Process (weighted analysis)

Binary (or Boolean) maps with the factor layers

We have these data as raster – for this approach requires simply reclassification of the data into binary raster and multiplication

- Temperature (> 15 degree Celsius) (**F**) (↑)
- Precipitation (> 350 mm) (**F**) (↑)
- Vegetation Index (> 0.5) (**F**) (↑)
- Population Density (> 0) (**F**) (↑)
- Elevation (<1,200m above sea-level) (**F/C**) (↓)
- Aridity (> 0.2 (Semi-humid & dry environment)) (**F/C**)(↑)

e.g., if temperature > 15 then change pixel value to 1 (good condition), else change to zero (bad condition)



Binary suitability is determined simply by multiplying the six individual layers that where reclassified to 0's and 1's.

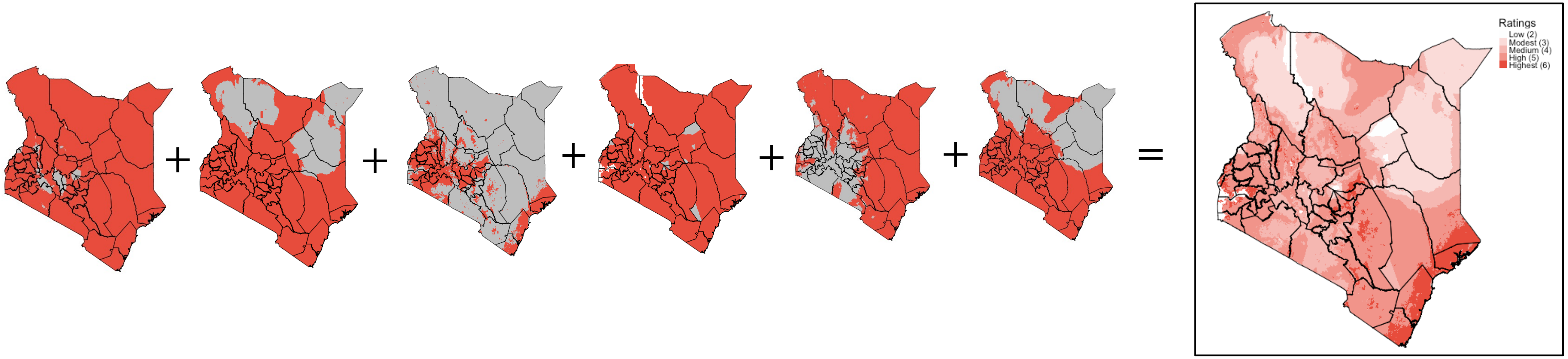
NOTES

Rankings/Ratings maps with the factor layers

We have these data as raster – for this approach requires simply reclassification of the data into binary raster and summation to get scores

- Temperature (> 15 degree Celsius) (**F**) ()
- Precipitation (> 350 mm) (**F**) ()
- Vegetation Index (> 0.5) (**F**) ()
- Population Density (> 0) (**F**) ()
- Elevation (<1,200m above sea-level) (**F/C**) ()
- Aridity (> 0.2 (Semi-humid & dry environment)) (**F/C**)()

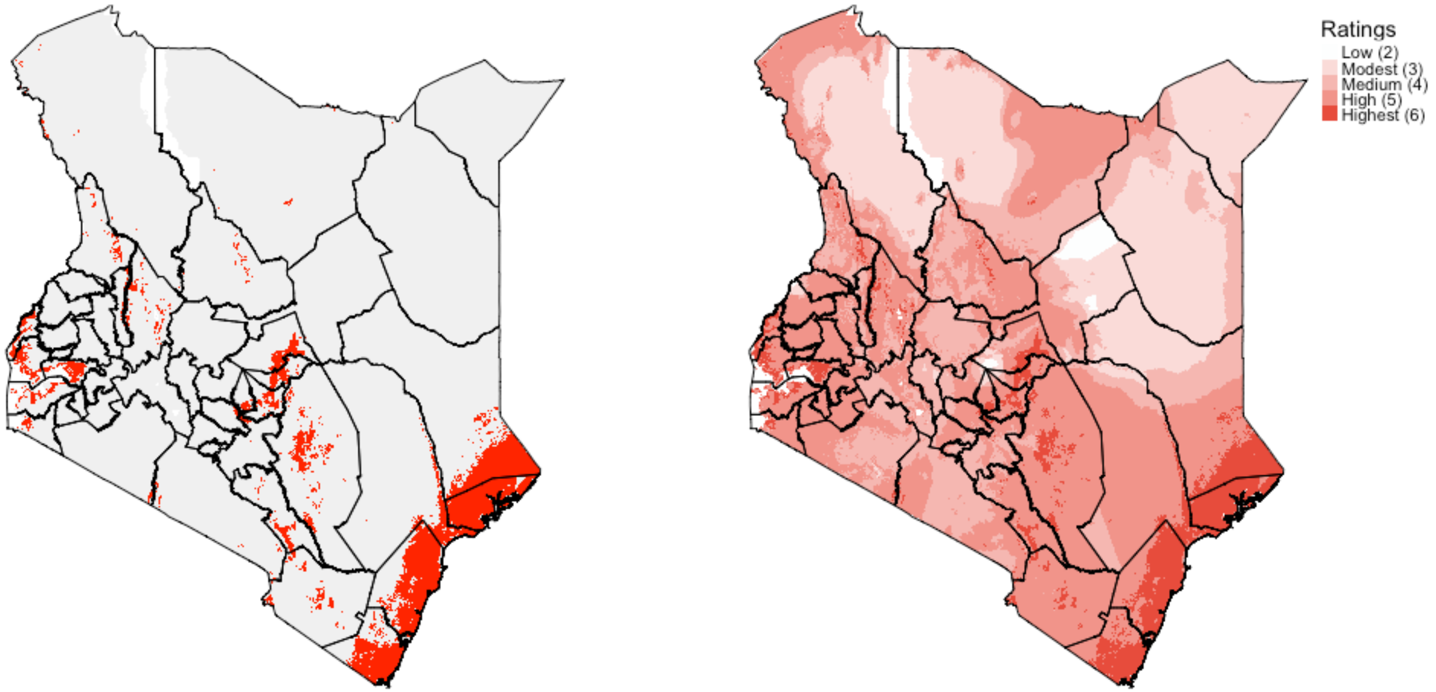
e.g., if temperature > 15 then change pixel value to 1 (good condition), else change to zero (bad condition)



The ratings or suitability scores are determined simply by summing the values across the six individual layers that where reclassified (minimum value = 2, and maximum = 6).

NOTES

Boolean versus Ratings map:



- Both approaches are accessible and are very great for rapid and descriptive analysis for the potential areas for LF suitability
- However, the maps with ratings/score provides more details with regards to the intensity of LF suitability
- Both maps have major issues:
  - Does not take in to account the fact that certain variables have more importance (or dominance) in terms of influence over other variables. For instance: LF (highest importance – precipitation >> population density >> vegetation – least importance)

NOTES

NOTE 1: We can do this by applying weights to each variable in order to show level of importance for the MCDA process. This approach is referred to as Saaty's Analytical Hierarchy Process (AHP)

Analytical Hierarchy Process (AHP) (1):

The **AHP** method is a pairwise comparison approach used on the criteria regarding the objective. These pairwise comparisons are carried out for all relevant factors under consideration.

Developed by Thomas L. Saaty (in 1970) and was primarily used for organizing and analyzing complex groups of decisions – and applied within the context of mathematics and psychology.

- Based on the pairwise comparison between two factors – its main purpose are to estimate priority weights for each factor, which in turn, are used in a model called the Weighted Linear Combination (**WLC**) model.
- The **WLC** model is what we use to make pixel-point predictions for suitability.

$$\boldsymbol{S = \left\{\sum_{i=1}^{n} w_i\, x_i\right\} \prod_{j=1}^{m} c_j}$$

Important note: This approach is very easy – the most difficult part is performing the pairwise comparison where you must make a judgement about which variables in question to determine its importance over the others. Here, it requires you to have knowledge about the such phenomena or carried out some extensive literature review as well as seeking expert opinion from experts on the matter!

NOTES

Analytical Hierarchy Process (AHP) (2):

Before, implementing the **AHP** method, there is one crucial step that must be done:

- Standardization of ALL variables to the same scale in order to make the comparison and estimation of the WLC possible
  - Jenks Natural Breaks algorithm (best approach to standardizing all continuous raster variables (10-, 20– 50-, or 100-point scale) (highest value mean most suitable)
  - Reclassification of raster to using a scale from 1 to 255, where 1 means it “less suitable”; to 255 “most suitable”.

  Use this formula:

$$\textbf{Standardization} = \frac{R - R_{min}}{R_{max} - R_{min}} (255)$$

NOTES

NOTE 1: Jenks is the best because it calculates the optimum break points while seeking to minimize the variance within categories, and at the same time minimize variance again between categories.

The 255 point scale is just a simple transformation. In fact, if you want the scale to be up to 1, multiply it to 1. If you want this to be up to 10 – multiply this by 10 etc.

Analytical Hierarchy Process (AHP) (3):

This methods is based on pairwise comparison – therefore a matrix is constructed, where each criterion is compared with the other criteria, relative to its importance, on a scale from 1 to 9 in which 1 = equal preference between two factors & 9 = a particular factor that is extremely favored over the other.

Let us a simple example with 3 factors (precipitation (P), vegetation (V) and population density (PD))

	Extremely Favours (+VE)	Very Strong Favours (+VE)	Strongly Favours (+VE)	Slightly Favours (+VE)	Equal	Slightly favours (-VE)	Strongly favours (-VE)	Very Strong favours (-VE)	Extremely favours (-VE)	
	9	7	5	3	1	3	5	7	9	
P	X									V
P				X						PD
V							X			PD

Now, convert this into a square matrix to estimate the priority weights!

NOTES

NOTE 1: Create a pairwise table

NOTE 2: To know number of pairs is

= n(n-1)/2
= 3(2)/2
= 3 pairs

P,V
P,PD
V,PD

Note 3: Interpretation are as follows
P,V = +9, based on judgement and what we know, we have assigned a highest preference for precipitation over vegetation

P,PD = +3, Slight preference for precipitation over population density.

V,PD = -5, the preference is going against vegetation and is strongly in favour of population density instead.

Analytical Hierarchy Process (AHP) (4):

Since we are dealing with 3 factors (three variables) in this example – we use the information from the pairwise table to create a 3-by-3 square matrix.

	Extremely Favours (+VE)	Very Strong Favours (+VE)	Strongly Favours (+VE)	Slightly Favours (+VE)	Equal	Slightly favours (-VE)	Strongly favours (-VE)	Very Strong favours (-VE)	Extremely favours (-VE)	
	9	7	5	3	1	3	5	7	9	
P	X									V
P				X						PD
V							X			PD

	P	V	PD
P	1	9	3
V		1	1/5
PD			1

NOTES

Note 1: Interpretation are as follows
P,V = +9, based on judgement and what we know, we have assigned a highest preference for precipitation over vegetation

P,PD = +3, Slight preference for precipitation over population density.

V,PD = -5, the preference is going against vegetation and is strongly in favour of population density instead.

Note 2:

V,PD = -5, the preference is going against vegetation and is strongly in favour of population density instead. This is reciprocal (1/5)

Analytical Hierarchy Process (AHP) (5):

Since we are dealing with 3 factors (three variables) in this example – we use the information from the pairwise table to create a 3-by-3 square matrix.

	Extremely Favours (+VE)	Very Strong Favours (+VE)	Strongly Favours (+VE)	Slightly Favours (+VE)	Equal	Slightly favours (-VE)	Strongly favours (-VE)	Very Strong favours (-VE)	Extremely favours (-VE)	
	9	7	5	3	1	3	5	7	9	
P	X									V
P				X						PD
V							X			PD

	P	V	PD
P	1	9	3
V	1/9	1	1/5
PD	1/3	5	1

How to populate the matrix, and prepare to simple matrix calculations

For V,P we take the value for P,V and use the reciprocal = 9 to 1/9
For PD,P we take the value for P,PD and use the reciprocal = 3 to 1/3
For PD,V we take the value for V,PD and use the reciprocal = 1/5 to 5

We need to normalize the matrix and calculate the weights (i.e., priority weights) for the WLC model

NOTES

Note 1: Interpretation are as follows
P,V = +9, based on judgement and what we know, we have assigned a highest preference for precipitation over vegetation

P,PD = +3, Slight preference for precipitation over population density.

V,PD = -5, the preference is going against vegetation and is strongly in favour of population density instead.

Note 2:

V,PD = -5, the preference is going against vegetation and is strongly in favour of population density instead. This is reciprocal (1/5)

Analytical Hierarchy Process (AHP) (6):

NOTES

Normalization and weight calculations (1):

	Extremely Favours (+VE)	Very Strong Favours (+VE)	Strongly Favours (+VE)	Slightly Favours (+VE)	Equal	Slightly favours (-VE)	Strongly favours (-VE)	Very Strong favours (-VE)	Extremely favours (-VE)	
	9	7	5	3	1	3	5	7	9	
P	X									V
P				X						PD
V							X			PD

	P	V	PD
P	1	9	3
V	1/9	1	1/5
PD	1/3	5	1
SUMS	1.444	15	4.20

	P	V	PD
P	1/1.444	9/15	3/4.20
V	(1/9)/1.444	1/15	(1/5)/4.20
PD	(1/3)/1.444	5/15	1/4.20
SUMS	1.444	15	4.20

We need to normalize the matrix and calculate the weights (i.e., priority weights) for the WLC model

Take the column sums first

Then normalise the cell value by dividing it by its correspond column total.

Analytical Hierarchy Process (AHP) (7):

NOTES

Normalization and weight calculations (2):

	P	V	PD
P	1/1.444	9/15	3/4.20
V	(1/9)/1.444	1/15	(1/5)/4.20
PD	(1/3)/1.444	5/15	1/4.20
SUMS	1.444	15	4.20

Then normalise the cell value by dividing it by its correspond column total.

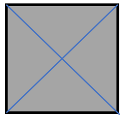
	P	V	PD	Priority Weights
P	0.6923	0.6000	0.7143	(0.6923+0.6+0.7143)/3 = 0.6689
V	0.0769	0.0667	0.0476	(0.0769+0.0667+0.0476)/3 = 0.0637
PD	0.2308	0.3333	0.2381	(0.2308+0.3333+0.2381) = 0.2674
SUMS	1.444	15	4.20	

Finally, for the normalised matrix, calculate the average to estimate the priority weights for the WLC model. IMPORTANT NOTE: The weights when summed should equal to 1 (or approximately close to 1 like 0.99XXX), if not, then there is a mistake somewhere in the calculations!

- The **WLC** model is what we use to make pixel-point predictions for suitability.

$$S = \left\{\sum_{i=1}^{n} w_i \, x_i\right\} \prod_{j=1}^{m} c_j$$

$$\mathbf{S = 0.6689prec + 0.0637veg + 0.2674popl}$$



Suppose (on a scale 1 to 10) - Precipitation = 7, Vegetation = 2 and Population Density = 5

Suitability LF = 0.6689(7) + 0.0637(2) + 0.2674(5) = 6.146 (Medium suitability for LF)

Analytical Hierarchy Process (AHP) (8):

NOTES

Model Validation (1):

This involves calculating an estimate called the **Consistency Ratio (CR), it tells us how consistent our judgements (i.e., Consistency Index) have been relative to large samples of purely random judgements (RI).**

We first determine our consistency index (CI) = $\frac{\lambda_{max} - n}{(n - 1)}$, λ_{max} is an eigenvalue calculated from our matrix and n (number of factors), **which is the summed product between the priority weights and column sums**

	P	V	PD	Priority Weights
P	0.6923	0.6000	0.7143	(0.6923+0.6+0.7143)/3 = 0.6689
V	0.0769	0.0667	0.0476	(0.0769+0.0667+0.0476)/3 = 0.0637
PD	0.2308	0.3333	0.2381	(0.2308+0.3333+0.2381) = 0.2674
SUMS	1.444	15	4.20	

λ_{max} = 0.6689(1.444) + 0.0637(15) + 0.2674(4.20) = 3.0445

CI = $\frac{\lambda_{max} - n}{(n - 1)}$ = (3.0445 – 3)/(3 – 1) = 0.0222

Analytical Hierarchy Process (AHP) (9):

NOTES

Model Validation (2):

Saaty (1980) created a system through random simulation and provided statistics to help us determine whether the judgements made for our pairwise comparison are reasonable. We call this the Random Consistency Index (RI).

N	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49

The **N is basically the number of factors, which is 3 (in this example). Therefore, our RI = 0.58**

CI = 0.0222
RI = 0.58

CR = $\frac{CI}{RI}$ < 0.1 (acceptable): 0.0222/0.58 = 0.04 < 0.1

Important note: If the CR > 0.1, then some pairwise values need to be reconsidered and the process (pairwise comparison) is repeated again until the desired value of CR < 0.1.

Practicals in Computer Labs:

- **Combination of using RStudio geoprocessing and analysis of raster data & excel spreadsheet for conducting AHP calculations.**

Decision table using Saaty's scale

Factor(s)	Extreme Favours 9	Very Strong Favours 7	Strongly Favours 5	Sligthly Favours 3	Equal 1	Factor(s)
Precipitation					1	Temperature
Precipitation				3		Population Density
Precipitation			5			Elevation
Precipitation	9					NDVI
Precipitation		7				Aridity
Temperature				3		Population Density
Temperature			5			Elevation
Temperature		7				NDVI
Temperature					1	Aridity
Population Density				3		Elevation
Population Density			5			NDVI
Population Density			5			Aridity
Elevation				3		NDVI
Elevation					1	Aridity
Aridity	9					NDVI

STEP 2: Create a matrix and populate the pairwise values (in the green section of matrix). On the row side (shaded "yellow") of matrix list the factors of interest, on the column (shaded "grey") are the factors compared against. A factor facing itself is given value of "1" along the main

	Precipitation	Temperature	Population Density	Elevation	NDVI	Aridity

NOTES

Provided a spreadsheet with full data for the LF in Kenya. Extra self practical exercise for predicting areas for landslides hazards.