High-resolution modelling of land use and management decisions

Michael Wögerer, IIASA

Funded by the European Union

This project has received funding from the European Union's Horizon Europe Research and Innovation programme under Grant Agreement No 101060423. The information and views set out in this deliverable are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

LAMASUS
Land Management for Sustainability

Overview

Intro to downscalR within iBIOM

Role of the prior module

MNL in theory

MNL in land use economics

Coding our own MNL prior module

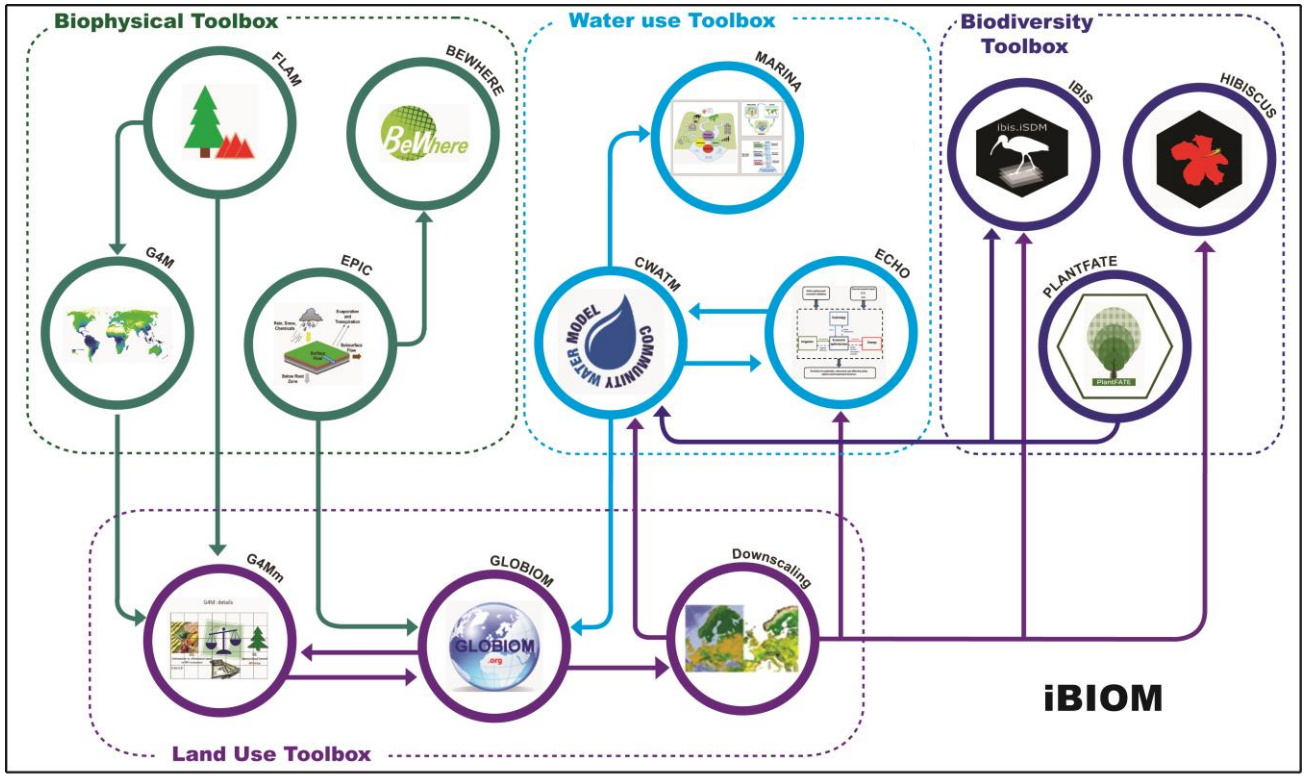
Objective of this session

GET CONVINCED THAT R IS A USEFUL TOOL NO MATTER WHAT

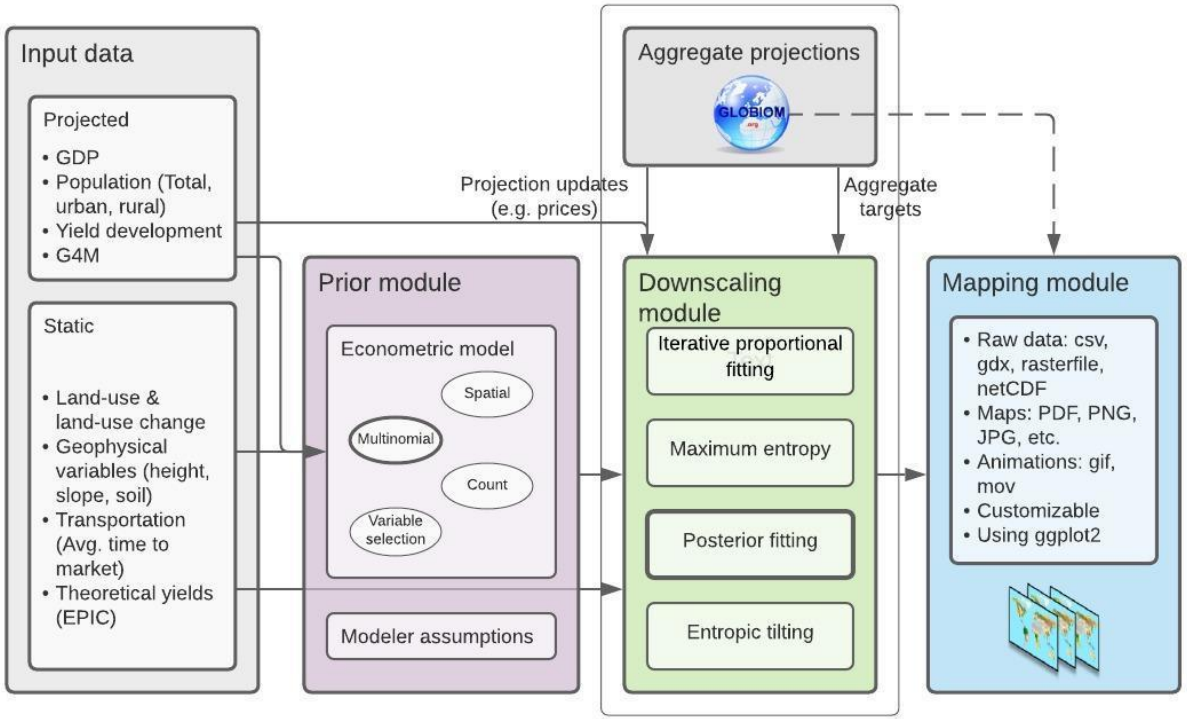
UNDERSTAND WHEN TO USE AN MNL MODEL

BE ABLE TO USE R TO ESTIMATE AN MNL MODEL WITH OBSERVED DATA

IIASA modelling ecosystem



The downscalR toolbox



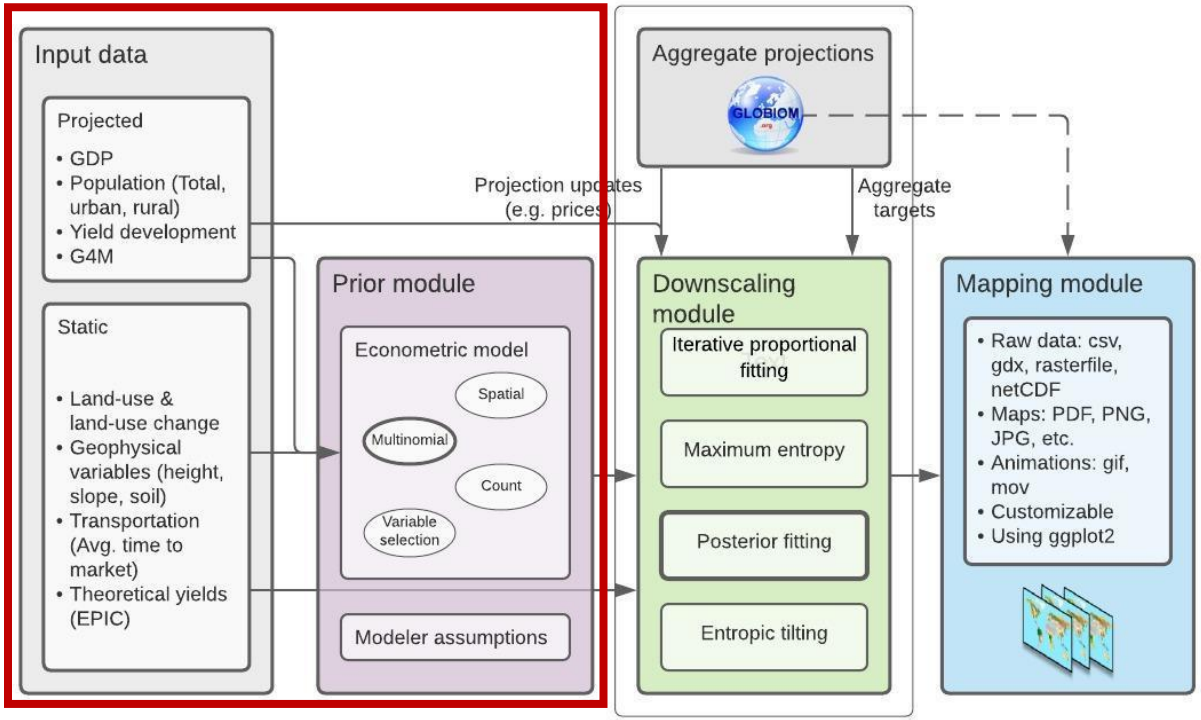
Source: https://github.com/tkrisztin/downscalr

The downscalR toolbox as a joint effort with Tamas & Leopold.

Consisting of:

- Database
- Prior module
- Downscale module
- Mapping module

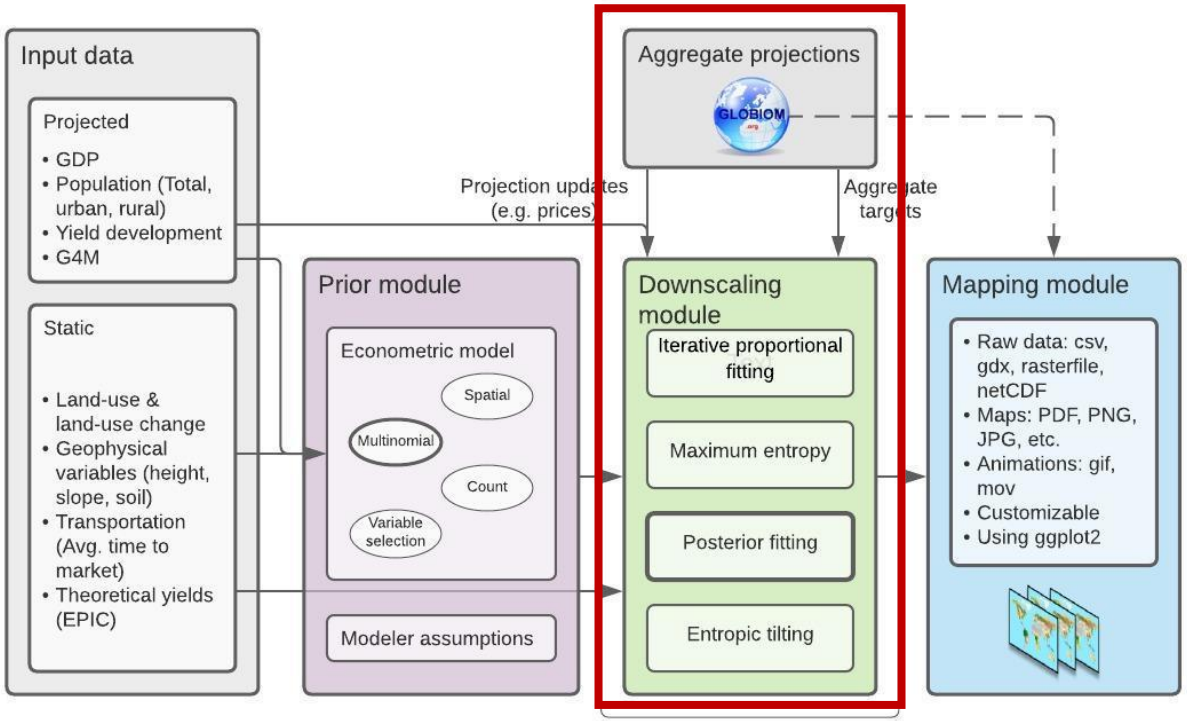
The prior module



Source: https://github.com/tkrisztin/downscalr

- Multinomial logit model to understand the location of occurring land use change based on inputs.
- Provides projections of possible land-use change maps based on future scenarios.

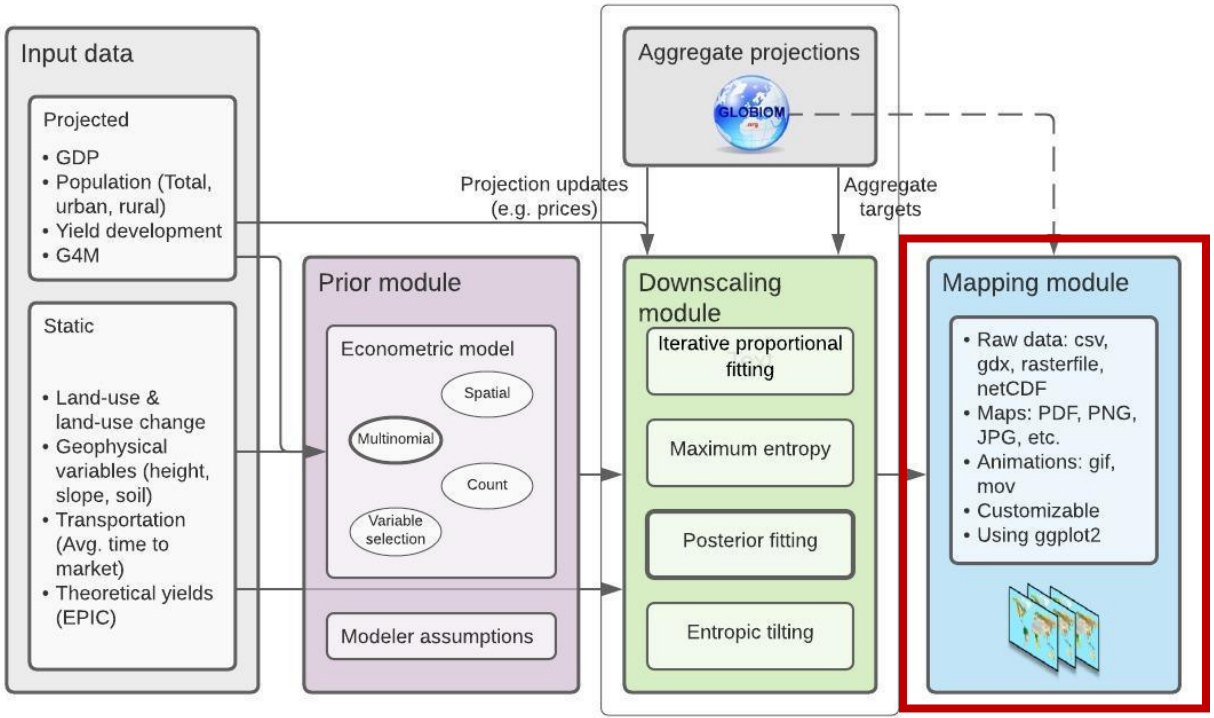
The downscaling module



Source: https://github.com/tkrisztin/downscalr

- Takes regional projections from GLOBIOM which is designed to get the intensity of land-use change right.
- Downscaling increases or decreases the prior maps to match these targets.

The mapping module



Source: https://github.com/tkrisztin/downscalr

A set of functions and routines to either:

- provide graphical feedback in terms of plots about the results
- Prepare the results for ex-post models that use spatially explicit land-use as an input (biodiversity, CO2, food security)

Let’s start as simple as possible

- We want to measure the **effect of one variable on another**.
- In economics such relationships are usually explained by linear regressions.
- **Linear regressions** helps us understand and predict relationships between variables.

- Example:

How does fertilizer use affect crop yield?

Crop Yield = a + b × **Fertilizer** + **e**

a = starting yield (if no fertilizer used)

b = how much yield increases per extra unit of fertilizer

e = everything we didn’t measure or explain

The linear regression

- The dashed line is the **linear regression** — our best estimate of the trend.
- The deviation from the trend line “the error” is from unobserved characteristics that change the individual observations yield. (weather, soil, …)
- We could control for those too. (multiple linear regression)

Let‘s have a quick look at R

Linear model

Funded by
the European Union

LAMASUS
Land Management for Sustainability

From Linear to Logit Models

What's the probability that a plot of land is used for agriculture based on soil quality?

P(Agricultural Land) = something between 0 and 1

- **Linear model:**

P(Agriculture) = a + b × Soil Quality

...can give nonsense results like:

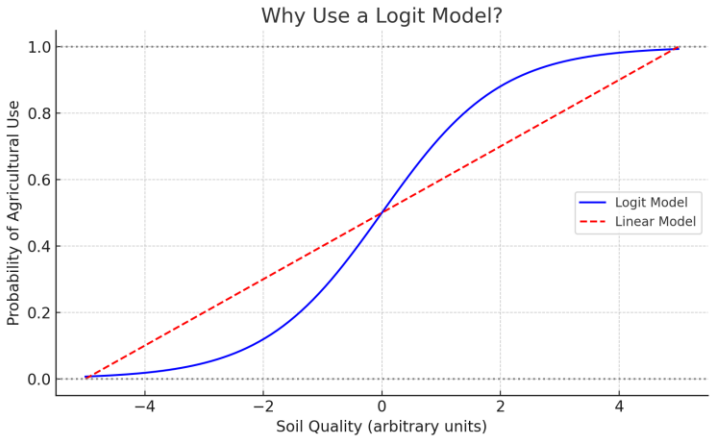
- Negative probabilities (< 0)
- Probabilities over 100% (> 1)

- **Logistic model:**

$$P(y = 1) = \frac{1}{1 + e^{-(a + b \cdot x)}}$$

- Always stays between 0 and 1
- Models how probabilities change non-linearly
- Makes most change near the middle (around 50%)

The logistic regression



Logistic regression models the probability that a binary outcome variable equals 1, using the logistic (sigmoid) function.

It assumes a linear relationship between the input variables and the log-odds of the outcome.

It’s widely used for binary classification tasks like spam detection, disease diagnosis, or customer churn prediction.

Let‘s have a quick look at R

Logit model

Funded by
the European Union

LAMASUS
Land Management for Sustainability

From Logit Models to MNL models

What's the probability that a plot takes one of J different land-cover categories?

P(Grassland) = between 0 and 1
P(Cropland) = between 0 and 1
P(Forest) = 1 - P(Grassland) - P(Cropland)

- **Logistic model:**

$$P(y=1) = \frac{1}{1+e^{-(a+b\cdot x)}}$$

- Only gives two probabilities (yes/no)

- **MNL model:**

$$P(Y=j) = \frac{\exp\left(a_j X_1 + b_j X_2\right)}{\sum_{k=1}^{J} \exp\left(a_k X_1 + b_k X_2\right)}$$

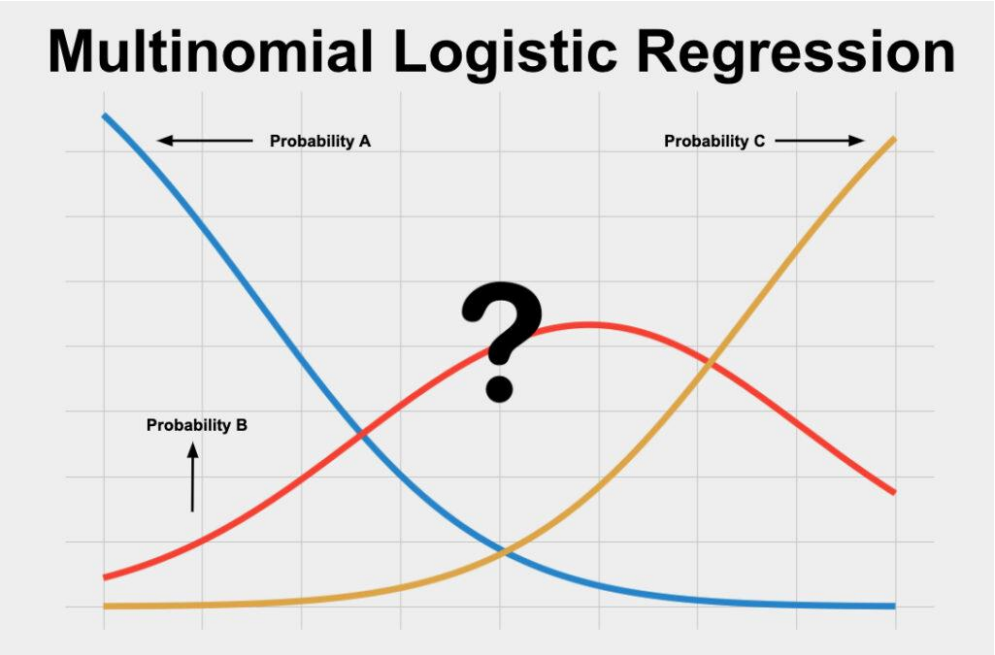
- Always stays between 0 and 1
- Models how probabilities change non-linearly
- Differentiates between J classes

The multinomial logistic regression

MNL is used when the outcome is one of three or more categories, like choosing between forest, grassland, or cropland.

Each choice is assigned a linear utility function based on features (e.g., soil quality, slope, subsidies), and the probability of choosing an option depends on its utility relative to others.

The model assumes that the relative odds between any two choices are unaffected by the presence or characteristics of a third option.



Let‘s have a quick look at R

MNL model

Funded by
the European Union

LAMASUS
Land Management for Sustainability

In a nutshell MNL for land use modelling

Each land-use class has a latent (hidden) utility:

Cropland to grow food and fodder

Grassland for livestock grazing (biodiversity)

Forest for carbon storage and biodiversity

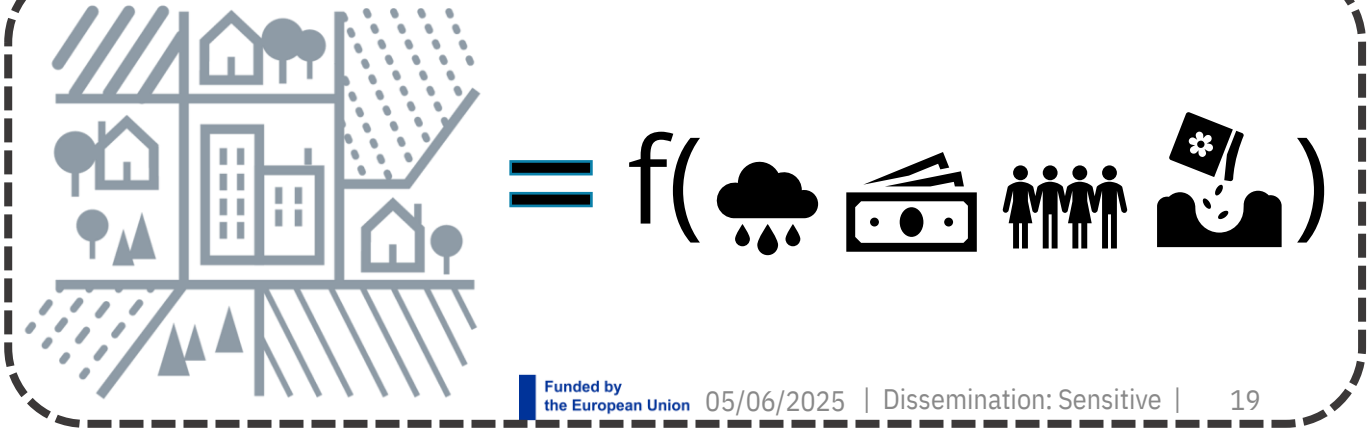
Urban to provide space to live

In a nutshell MNL for land use modelling

Given the observed patterns of land-use and variables assumed to influence land-use, we can produce estimates for the latent utility.

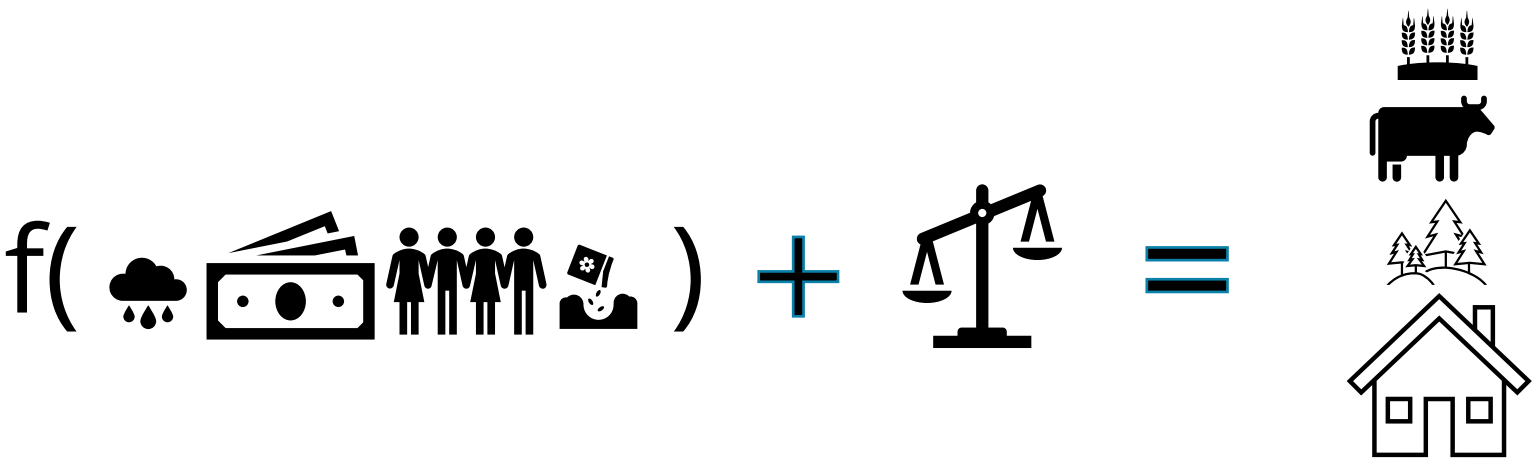
MNL:

Dependent = f(independent)



In a nutshell MNL for land use modelling

Knowing or assuming future behavior of drivers then allows to make predictions about the future landscape (e.g. SSP scenarios)



MNL for land use modelling

Main interest centers on estimating **gross** land-use change transition.

The MNL model:

$$conversion_{ik} = \frac{\exp(utility_{ik})}{\sum_{j=1}^{J} \exp(utility_{ij})}$$

Where: i = (1,…,N) are N regional observations and
j = (1,…,J) are J different land use categories
k is an element of j

$conversion_{ik}$ is the share of area that is converted to class k in observation i.
$utility_{ik}$ is based on p = (1,…,P) drivers and parameter. $utility_{ik} = \sum_{p=1}^{P} x_{ip} * \beta_{pk}$

Land use model in R

Plus exercise

Funded by
the European Union

LAMASUS
Land Management for Sustainability

Thanks for your attention!

Funded by the European Union

This project has received funding from the European Union's Horizon Europe Research and Innovation programme under Grant Agreement No 101060423. The information and views set out in this deliverable are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.