

The Integrated Economic- Environmental Modeling (IEEM) Platform: IEEM Platform Technical Guides. User Guide for the IEEM-enhanced Land Use Land Cover Change Model Dyna-CLUE

Peter H Verburg
Žiga Malek
Sean P. Goodwin
Cecilia Zagaria

Environment, Rural
Development and Risk
Management Division

TECHNICAL NOTE N°
IDB-TN- 02284

The Integrated Economic- Environmental Modeling (IEEM) Platform: IEEM Platform Technical Guides. User Guide for the IEEM-enhanced Land Use Land Cover Change Model Dyna-CLUE

Peter H Verburg
Žiga Malek
Sean P. Goodwin
Cecilia Zagaria

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

The Integrated Economic-Environmental Modeling (IEEM) Platform: IEEM Platform
technical guides: user guide for the IEEM-enhanced land use land cover change model
Dyna-CLUE / Peter H. Verburg, Žiga Malek, Sean P. Goodwin, Cecilia Zagaria.
p. cm. — (IDB Technical Note ; 2284)

Includes bibliographic references.

1. Environmental economics-Mathematical models. 2. Land use-Mathematical models.
3. Land cover-Mathematical models. 4. Ecosystem services. I. Verburg, Peter H. II.
Malek, Žiga. III. Goodwin, Sean. IV. Zagaria, Cecilia. V. Inter-American Development
Bank. Environment, Rural Development and Risk Management Division. VI. Series.
IDB-TN-2284

JEL Codes: Q24; R14; C33; Q15.

Keywords: Integrated Economic-Environmental Modeling (IEEM), land use land cover
(LULC) change modeling

<http://www.iadb.org>

Copyright © [2021] Inter-American Development Bank. This work is licensed under a Creative Commons IGO 3.0 Attribution-NonCommercial-NoDerivatives (CC-IGO BY-NC-ND 3.0 IGO) license (<http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) and may be reproduced with attribution to the IDB and for any non-commercial purpose. No derivative work is allowed. Any dispute related to the use of the works of the IDB that cannot be settled amicably shall be submitted to arbitration pursuant to the UNCITRAL rules. The use of the IDB's name for any purpose other than for attribution, and the use of IDB's logo shall be subject to a separate written license agreement between the IDB and the user and is not authorized as part of this CC-IGO license.

Any dispute related to the use of the works of the IDB that cannot be settled amicably shall be submitted to arbitration pursuant to the UNCITRAL rules. The use of the IDB's name for any purpose other than for attribution, and the use of IDB's logo shall be subject to a separate written license agreement between the IDB and the user and is not authorized as part of this CC-IGO license.

Note that link provided above includes additional terms and conditions of the license.

The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.



The Integrated Economic-Environmental Modeling (IEEM) Platform: IEEM Platform Technical Guides

User Guide for the IEEM-enhanced Land Use Land Cover Change Model Dyna-CLUE

Peter H Verburg¹, Žiga Malek^{1,2}, Sean P. Goodwin¹ and Cecilia Zagaria¹

¹Institute for Environmental Studies (IVM), Vrije Universiteit Amsterdam
VU University Amsterdam
De Boelelaan 1111,
1081 HV Amsterdam, The Netherlands
T: + 31 20 5989555
E: Info@ivm.vu.nl

² Corresponding author
T: +31 20 59 85 670
E: ziga.malek@vu.nl

Abstract

The Conversion of Land Use and its Effects modeling framework (CLUE) was developed to simulate land use change using empirically quantified relations between land use and its driving factors in combination with dynamic modeling of competition between land use types. Being one of the most widely used spatial land use models, CLUE has been applied all over the world on different scales. In this document, we demonstrate how the model can be used to develop a multi-regional application. This means, that instead of developing numerous individual models, the user only prepares one CLUE model application, which then allocates land use change across different regions. This facilitates integration with the Integrated Economic-Environmental Modeling (IEEM) Platform for subnational assessments and increases the efficiency of the IEEM and Ecosystem Services Modeling (IEEM+ESM) workflow. Multi-regional modelling is particularly useful in larger and diverse countries, where we can expect different spatial distributions in land use changes in different regions: regions of different levels of achieved socio-economic development, regions with different topographies (flat vs. mountainous), or different climatic regions (dry vs humid) within a same country. Accounting for such regional differences also facilitates developing ecosystem services models that consider region specific biophysical characteristics. This manual, and the data that is provided with it, demonstrates multi-regional land use change modeling using the country of Colombia as an example. The user will learn how to prepare the data for the model application, and how the multi-regional run differs from a single-region simulation.

JEL Codes: C33: Panel Data Models • Spatio-temporal Models; Q15: Land Ownership and Tenure • Land Reform • Land Use • Irrigation • Agriculture and Environment; Q24: Land; R14 Land Use Patterns.

Keywords: Integrated Economic-Environmental Modeling (IEEM), land use land cover (LULC) change modeling.

Table of contents

1.	Introduction.....	1
1.1.	Background	1
2.	Model description	2
2.1.	Model structure	2
2.2.	Spatial policies and restrictions	4
2.3.	Land use type specific conversion settings	4
2.4.	Land use requirements (demand)	5
2.5.	Location characteristics.....	6
2.6.	Allocation procedure	7
3.	Preparing a new application	9
3.1.	Data collection	9
3.2.	Spatial data processing and formats	11
3.3.	Definition of land use demands and conversions	17
3.4.	Logistic regression analysis.....	20
4.	Using the model	27
4.1.	User interface	27
4.2.	Main parameters	28
4.3.	Regression results.....	32
4.4.	Conversion matrix	33
4.5.	Area restrictions	34
4.6.	Demand scenario	35
4.7.	Initial land use configuration	36
4.8.	Driving factors.....	37
4.9.	Location specific preference addition.....	38
4.10.	Calculating probability maps.....	40
4.11.	Incorporating multiple regions.....	41
4.12.	Running the model	42
4.13.	Running the model through batch mode	44
5.	Analyzing the results.....	45
5.1.	Output file description	45
6.	Problems	46
6.1.	File checker	46
6.2.	Frequently asked questions / errors	46
	References	48
	Appendices.....	49
	Appendix 1 – Location factor data.....	49
	Appendix 2 – Data suggestions	51

1. Introduction

1.1. Background

Spatial models of land use change are important tools to analyze the possible trajectories of land use change in the near future. The results of land use models are important to evaluate policy options and assess the impact of land use change on natural and socio-economic conditions.

The Conversion of Land Use and its Effects modeling framework (CLUE) (Veldkamp and Fresco, 1996; Verburg et al., 1999) was developed to simulate land use change using empirically quantified relations between land use and its driving factors in combination with dynamic modeling of competition between land use types. The model was developed for the national and continental level and applications for Central America, Ecuador, China and Java, Indonesia are available. For study areas with such a large extent the spatial resolution for analysis was coarse and, as a result, each land use is represented by assigning the relative cover of each land use type to the pixels.

Land use data for study areas with a relatively small spatial extent is often based on land use maps or remote sensing images that denote land use types respectively by homogeneous polygons or classified pixels. This results in only one dominant land use type occupying one unit of analysis. Because of the differences in data representation and other features that are typical for regional applications, the CLUE model cannot directly be applied at the regional scale. Therefore, the modelling approach has been modified and is now called CLUE-S (the Conversion of Land Use and its Effects at Small regional extent). CLUE-S is specifically developed for the spatially explicit simulation of land use change based on an empirical analysis of location suitability combined with the dynamic simulation of competition and interactions between the spatial and temporal dynamics of land use systems. The most recent version of the CLUE model includes autonomous developments through bottom-up simulation. This version is called Dyna-CLUE and it has also been applied at detailed local scales all around the world: Austrian Alps, the Philippines, Iran, China... to name a few.

To date, 133 English-language publications from different authors all around the world have used the Dyna-CLUE model based on 'web of knowledge' and 'google scholar' searches. In total, 1100 studies cite the CLUE modeling framework. The model has been used in different scientific domains (Figure 1), from environmental to soil science. This makes it one of the most widely used and versatile land use models available. More information on the development of the CLUE-S model can be found in Verburg et al. (2002), Verburg and Veldkamp (2003) and Verburg and Overmars (2009).

Note: This technical note (modified from the CLUE-S model description) begins with a description of the model and its functionalities and is followed by a guided description of how to prepare and run an application. We focus on a case study of Colombia and specifically on the implementation of a regionalized functionality, whereby users of the Dyna-CLUE model can run a single application in which different regressions, demands

and other model functionalities can be differentiated and specified for different regions. This is an improvement from the previous Dyna-CLUE set-up, whereby users were required to implement separate model applications for each of the different sub-regions within a broader area of interest. Such a multi-regional set-up of Dyna-CLUE presents a big improvement for use within the Integrated Economic-Environmental Modeling (IEEM) and Ecosystem Services Modeling (IEEM+ESM) workflow due to several reasons. First, this enables better coupling with IEEM outputs, which can be subnational, meaning that land use changes are projected on the same level as IEEM. Secondly, this also results in more representative land use change simulations, that consider specific regional characteristics (socio-economic and biophysical). Region specific land use changes also allow for quantifying ecosystem services and assessing their changes on a regional level using ecosystem services modeling. Finally, a multi-regional model application saves time, as the user only prepares one application instead of individual models for each region.

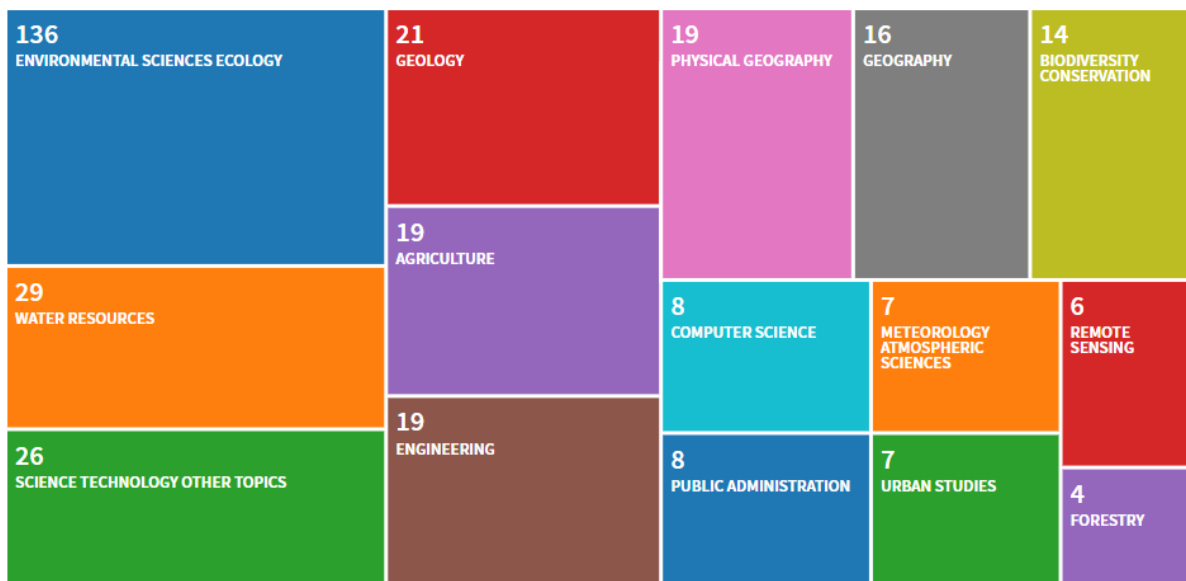


Figure 1 Research categories where the Dyna-CLUE model has been applied or referred to. Source: webofknowledge.com

2. Model description

2.1. Model structure

The model is sub-divided into two distinct modules, namely a non-spatial demand module and a spatially explicit allocation procedure (Figure 2). The non-spatial module calculates the area change for all land use types at the aggregate level. Within the second part of the model these demands are translated into land use changes at different locations within the study region using a raster-based system. The user-interface of the Dyna-CLUE model only supports the spatial allocation of land use change. For the land use demand module different model specifications are possible ranging from simple trend extrapolations to complex economic models. In the case of this ToR, the land use demands are provided by IEEM.

The choice for a specific model is very much dependent on the nature of the most important land use conversions taking place within the study area and the scenarios that need to be considered. The results from the demand module need to specify, on a yearly basis, the area

covered by the different land use types, which is a direct input for the allocation module.

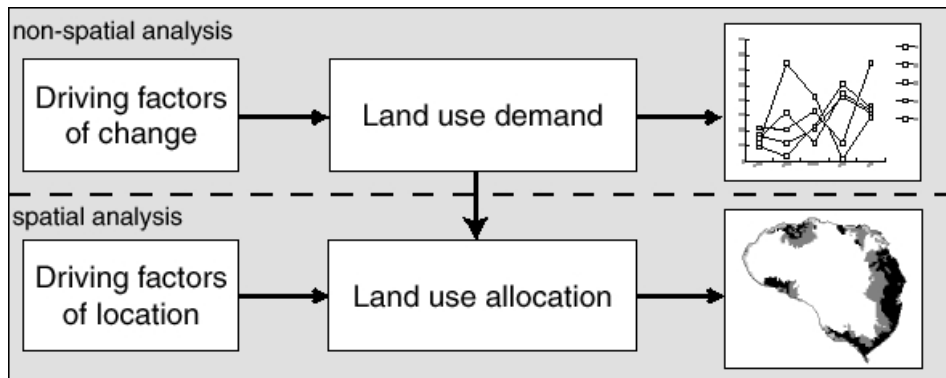


Figure 2 Overview of the modelling procedure

The allocation is based upon a combination of empirical, spatial analysis and dynamic modelling. Figure 3 gives an overview of the information needed to run the Dyna-CLUE model. This information is subdivided into four categories that together create a set of conditions and possibilities for which the model calculates the best solution in an iterative procedure. The next sections introduce each of the boxes: spatial policies and restrictions, land use type specific conversion settings, land use requirements (demand) and location characteristics.

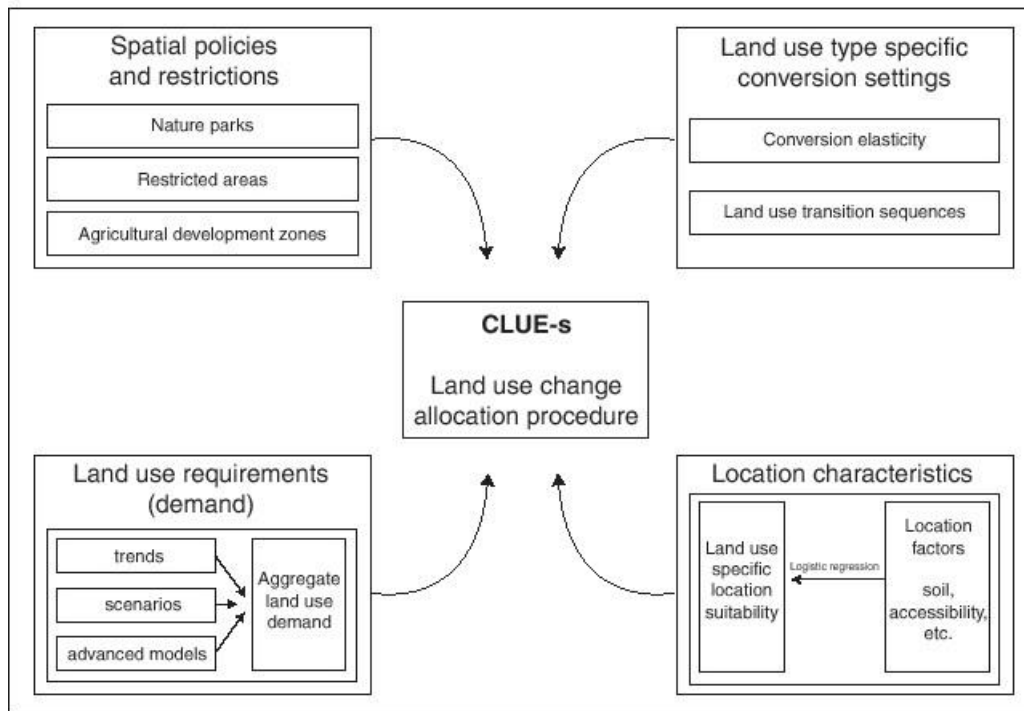


Figure 3 Overview of the information flow in the Dyna-CLUE/CLUE-S model

2.2. Spatial policies and restrictions

Spatial policies and land tenure can influence the pattern of land use change. Spatial policies and restrictions mostly indicate areas where land use changes are restricted through policies or tenure status. Besides that, spatial policies can imply stimulation arrangements for a certain land use on a location. For the simulation, maps that indicate the areas for which the spatial policy is implemented must be supplied. Some spatial policies restrict all land use change in a certain area, e.g., no deforestation in a national park. Other land use policies restrict a set of specific land use conversions, e.g., residential construction in designated agricultural areas or permanent agriculture in the buffer zone of a nature reserve. The conversions that are restricted by a certain spatial policy can be indicated in a land use conversion matrix (section 2.3): for all possible land use conversions it is indicated if the spatial policy applies.

2.3. Land use type specific conversion settings

Land use type specific conversion settings determine the temporal dynamics of the simulations. Two sets of parameters are needed to characterize the individual land use types: conversion elasticities and land use transition sequences. The first parameter set, the conversion elasticities, is related to the reversibility of land use change. Land use types with high capital investment will not easily be converted in other uses as long as there is sufficient demand. Such land use types are therefore more 'static' than other land use types. Examples are residential locations and plantations with permanent crops (e.g., fruit trees), that are unlikely to be converted to another use in the years after their establishment. Other land use types easily shift location when the location becomes more suitable for other land use types.

Arable land often makes place for urban development while expansion of agricultural land

occurs at the forest frontier. An extreme example is shifting cultivation: for this land use system the same location is mostly not used for periods exceeding two seasons because of nutrient depletion of the soil.

These differences in behavior towards conversion can be approximated by conversion costs. However, costs cannot represent all factors that influence the decisions towards conversion such as nutrient depletion, esthetic values etc. Therefore, for each land use type a value needs to be specified that represents the relative elasticity of change, ranging from 0 (easy conversion) to 1 (irreversible change). The user should decide on this factor based on expert knowledge or observed behavior in the recent past.

The second set of land use type characteristics that needs to be specified are the land use type specific conversion settings and their temporal characteristics. These settings are specified in a conversion matrix which defines:

- To what other land use types the present land use type can be converted or not.
- In which regions a specific conversion is allowed to occur and in which regions it is not allowed.
- How many years (or time steps) the land use type at a location should remain the same before it can change into another land use type. This can be relevant in case of the regrowth of forest. Open forest cannot change directly into closed forest. However, after a number of years it is possible that an undisturbed open forest will change into closed forest because of regrowth.
- The maximum number of years that a land use type can remain the same. This setting is particularly suitable for arable cropping within a shifting cultivation system. In these systems the number of years a piece of land can be used is commonly limited due to soil nutrient depletion and weed infestation.

It is important to note that only the minimum and maximum number of years before a conversion can or should happen is indicated in the conversion table. The exact number of years depends on the land use pressure and location specific conditions. The simulation of these interactions combined with the constraints set in the conversion matrix will determine the length of the period before a conversion occurs.

2.4. Land use requirements (demand)

Land use requirements (demand) are calculated at the aggregate level (the level of the case-study as a whole) as part of a specific scenario. The land use requirements constrain the simulation by defining the totally required change in land use. All changes in individual pixels should add up to these requirements. In the approach, land use requirements are calculated independently from the Dyna-CLUE model itself. The calculation of these land use requirements is based on a range of methods, depending on the case study and the scenario. The extrapolation of trends in land use change of the recent past into the near future is a common technique to calculate land use requirements. When necessary, these trends can be corrected for changes in population growth and/or diminishing land resources. For policy analysis it is also possible to base land use requirements on advanced models of macro-economic changes, which can serve to provide scenario conditions that relate policy targets to land use change requirements.

2.5. Location characteristics

Land use conversions are expected to take place at locations with the highest 'preference' for the specific type of land use at that moment in time. Preference represents the outcome of the interaction between the different actors and decision-making processes that have resulted in a spatial land use configuration. The preference of a location is empirically estimated from a set of factors that are based on the different, disciplinary, understandings of the determinants of land use change. The preference is calculated following:

$$Rki = a_k X_{1i} + b_k X_{2i} + \dots$$

where R is the preference to devote location i to land use type k , $X_{1,2}, \dots$ are biophysical or socio-economic characteristics of location i and a_k and b_k the relative impact of these characteristics on the preference for land use type k . The exact specification of the model should be based on a thorough review of the processes important to the spatial allocation of land use in the studied region.

A statistical model can be developed as a binomial logit model of two choices: convert location i into land use type k or not. The preference Rki is assumed to be the underlying response of this choice. However, the preference Rki cannot be observed or measured directly and has therefore to be calculated as a probability. The function that relates these probabilities with the biophysical and socio-economic location characteristics is defined in a logit model following:

$$\text{Log} \left(\frac{P_i}{1 - P_i} \right) = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_n X_{n,i}$$

where P_i is the probability of a grid cell for the occurrence of the considered land use type on location i and the X 's are the location factors. The coefficients (β) are estimated through logistic regression using the actual land use pattern as dependent variable. This method is similar to econometric analysis of land use change, which is very common in deforestation studies. In econometric studies the assumed behavior is profit maximization, which limits the location characteristics to (agricultural) economic factors. In the study areas is assumed that locations are devoted to the land use type with the highest 'suitability'. 'Suitability' includes the monetary profit but can also include cultural and other factors that lead to deviations from (economic) rational behavior in land allocation. This assumption makes it possible to include a wide variety of location characteristics or their proxies to estimate the logit function that defines the relative probabilities for the different land use types.

Most of these location characteristics relate to the location directly, such as soil characteristic and terrain. However, land management decisions for a certain location are not always based on location specific characteristics alone. Conditions at other levels, e.g., the household, community or administrative level can influence the decisions as well. These factors are represented by accessibility measures, indicating the position of the location relative to important regional facilities, such as the market and by use of spatially lagged variables. A

spatially lagged measure of the population density approximates the regional population pressure for the location instead of only representing the population living at the location itself. An overview of example location factors and their respective (open access) spatial data sources is described in Appendix 1.

2.6. Allocation procedure

When all input is provided the Dyna-CLUE model calculates, with discrete time steps, the most likely changes in land use given the before described restrictions and suitabilities. The allocation procedure is summarized in Figure 4. The following steps are taken to allocate the changes in land use:

1. The first step includes the determination of all grid cells that are allowed to change. Grid cells that are either part of a protected area or presently under a land use type that is not allowed to change are excluded from further calculation. Also the locations where certain conversions are not allowed due to the specification of the conversion matrix are identified.
2. For each grid cell i the total probability ($TPROP_{i,u}$) is calculated for each of the land use types u according to:

$$TPROP_{i,u} = P_{i,u} + ELAS_u + ITER_u$$

where $P_{i,u}$ is the suitability of location i for land use type u (based on the logit model), $ELAS_u$ is the conversion elasticity for land use u and $ITER_u$ is an iteration variable that is specific to the land use type and indicative for the relative competitive strength of the land use type. $ELAS_u$, the land use type specific elasticity to change value, is only added if grid-cell i is already under land use type u in the year considered. $P_{i,u}$ consists of a part based on the biophysical and socio-economic factors, and a neighbourhood interaction part.

3. A preliminary allocation is made with an equal value of the iteration variable ($ITER_u$) for all land use types by allocating the land use type with the highest total probability for the considered grid cell. Conversions that are not allowed according to the conversion matrix are not allocated. This allocation process will cause a certain number of grid cells to change land use.
4. The total allocated area of each land use is now compared to the land use requirements (demand). For land use types where the allocated area is smaller than the demanded area the value of the iteration variable is increased. For land use types for which too much is allocated the value is decreased. Through this procedure it is possible that the local suitability based on the location factors is overruled by the iteration variable due to the differences in regional demand. The procedure followed balances the bottom-up allocation based on location suitability and the top-down allocation based on regional demand.

Steps 2 to 4 are repeated as long as the demands are not correctly allocated. When allocation equals demand the final map is saved and the calculations can continue for the next time step. Some of the allocated changes are irreversible while others are dependent on the changes in earlier time steps. Therefore, the simulations tend to result in complex, non-linear changes in land use pattern, characteristic of complex systems.

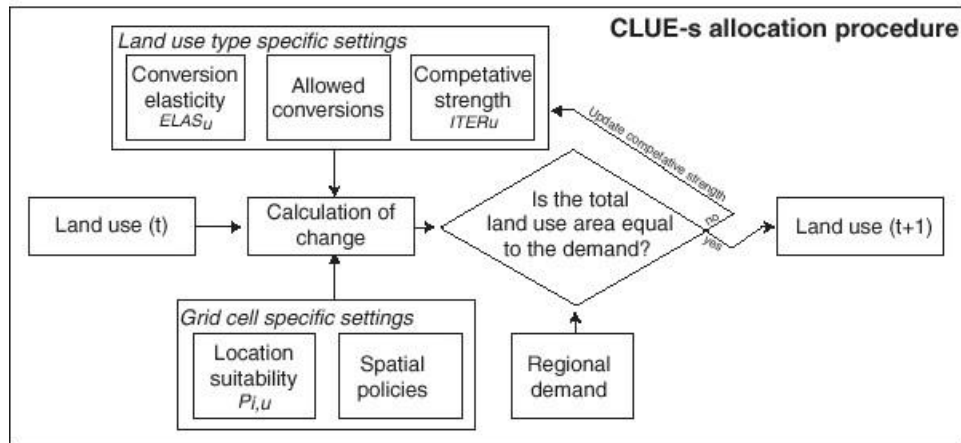


Figure 4 Flow chart of the allocation module of the CLUE modeling framework

3. Preparing a new application

Creating a new application of Dyna-CLUE begins with a data collection phase which addresses all four components of the model: spatial policies and restrictions, land use type specific conversion settings, land use requirements (demand) and location characteristics. Data collection involves both the gathering of quantitative (spatial) information, as well as the more qualitative definition of land use change conversion rules or demand scenarios. The data collection phase is followed by the processing of spatial datasets into the format required for the model; this concerns the preparation of an initial land cover file, an area restriction file and files for each of the identified location factors. This step is undertaken alongside the definition of land use demands and conversion rules, as these must refer to the same land use classes delineated in the respective spatial land use file. The final step of the data preparation phase addresses the logistic regression analysis, which makes use of the identified location factors and initial land use grids to determine land use suitability. Figure 5 provides an overview of the sequential steps which comprise the data preparation phase. Details on each of these steps is provided in the following sections.

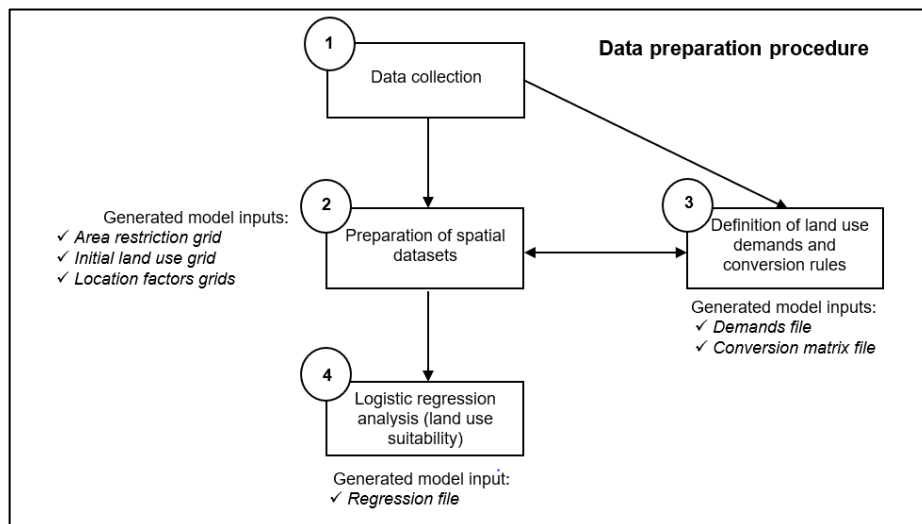


Figure 5 Flow chart of the required steps for data preparation when creating a new application of Dyna-CLUE

3.1. Data collection

For the simulation of spatial dynamics of land use change using the Dyna-CLUE model, (spatial) data are primarily needed for determining (1) land use distribution within the area of interest and existing area restriction rules, (2) the biophysical and socio-economic parameters that are considered as important potential drivers of the land use pattern, and (3) the definition of land use demands and conversion rules.

Land use distribution and area restrictions

The model requires, at a minimum, spatially explicit land use data for 1 year (the initial land use/land cover). However, to allow model calibration and validation, it is necessary to have data for another 2 years, preferably about 10 years apart. This is often not possible, as in many countries and regions, multitemporal spatially explicit land cover data is not available.

In such cases, validation is done by looking at any past data available (e.g. tree cover loss), or by expert judgement to identify whether the projected locations and spatial pattern of change are logical. Global and regional freely available spatial land use datasets are listed in **Appendix 2**. Area restriction (or boundary) data is also required in a spatially explicit format. This is likely to be found within regional or national publicly available zoning plans based on administrative boundaries, as well as global open-source data, e.g., addressing watershed boundaries or the distribution of protected areas (**Appendix 2**).

Location factors

Driving parameters are also required in a spatially explicit format. The Dyna-CLUE model allows for a maximum of 30 different driving factors. These refer most commonly to variables that describe the demography, soil, geomorphology, climate and infrastructural situation of an area. A list with examples and description of spatial data used in the Colombian case study can be found in **Appendix 1**. In general, a large number of data on driving factors can be collected from existing population maps and data from the agricultural census. However, a considerable portion of data is more commonly available from global or regional environmental models, such as soil and climate data; or from other scientific studies, such as data on market accessibility. Two broad categories of data can be used. The first category are maps (like soil, climate, altitude etc.). This type of data can be procured in digital format, or, if not available in digital form, it can be digitized. GIS programs, like ArcMap or QGIS, can be used for digitizing maps. Each polygon of a map contains a code describing the feature. In a GIS it is possible to make selections and reclassifications of the map (e.g. texture, drainage). Besides vector, line and point data, digital maps can also be of raster type, for each cell in the map a value is presented (e.g. Digital Elevation Models). It is, of course, important that the spatial resolution of the different data sources is comparable.

The second category of data are statistical data (e.g. census). These data are structured by administrative units, like district or provinces. By linking these data to maps with administrative boundaries the data are presented in a spatially explicit manner. When relating the tabular data to a map both sources should be of the same year, or one must be sure that the administrative boundaries correspond to the year the census data were collected.

Land use demands and conversion rules

The land use demands must refer to the classes and hectares defined in the identified spatial land use file. The land use demands provide information for each land use class on the amount of hectares, or share of land use, which is expected to undergo change under certain conditions. There are multiple sources of data that can be used to calculate relative demand for each land use type. First, data can be based on past trends, where statistics on past land use/land cover and crop and livestock production need to be collected. Same or similar trends can then be continued in the future, if modelling a so called 'business as usual' scenario. Second, future land cover demand can be calculated based on government plans, or any other expectations on future crop and livestock production and trends. One example are trade agreements, that specify how many more crops or livestock products can be exported due to trade liberalization. Finally, land cover scenarios can be calculated using external economic models, for example by using IEEM. However, there are also other models, with projections on future crop production being publicly available (for example the SSP database, providing

national or regional scale data on future land cover or agricultural production). For the determination of conversion rules, users can rely on expert knowledge as well as any available observations of past land use transition (e.g., on patterns of shifting cultivation).

3.2. Spatial data processing and formats

Preparing ASCII output files

In order to enable analysis and modelling in Dyna-CLUE, spatial data on (1) the starting land cover configuration, (2) the regional extent of the case study area and any area restrictions, and (3) the location factors need to be provided in an ASCII raster format. ASCII files are text files in which values for individual cells are stored in rows and columns with a header that describes the format, i.e., the number of columns and rows, the cell size and the coordinates. Data can alternatively be stored in one single column containing all the cells. The grid data is read from left to right, starting at the top left corner. Grid size depends on the scale and resolution of the original data. Bear in mind that smaller resolutions result in slower running times and large data storage requirements.

For the preparation of these ASCII files, many GIS packages can be used. In general, the vector-based data has to first be converted in raster based (grid) data. These grids can be exported directly to ASCII files following the nomenclature required by Dyna-CLUE (see the following section). All grids should have the same (equal-area) projection and be georeferenced with exactly the same extent and cell size. Additionally, all data cells that are NODATA should be aligned across all input data. Competent understanding of the chosen GIS programs is expected. Any dedicated software can be used to accurately prepare the spatial files in the correct ASCII format, however, from our experience, ArcGIS is the most suitable, and is therefore used to explain the data formatting steps in the following section.

Preparatory steps using ArcGIS

- **Preparation of the area restriction file (model input: *region*. *)*** – we recommend beginning with the preparation of the area restriction file, i.e. the file outlining the extent of the case study region, any eventual sub-regions of interest (if running a multi-region application), and the existence of any area-specific restrictions.
 - Load your original case study extent raster file into ArcMap and make sure it is at your desired resolution (by checking the cell size under Properties). If you would like to modify the resolution, use the [Resample tool](#) in the Data Management → Raster → Raster Processing Toolbox (Figure 6).
 - If you are running a single region application of Dyna-CLUE, the area restriction file must indicate a value of 0 for all active cells, a value of -9999 for the no-data cells, and a value of -9998 for any cells which are under special restrictions, i.e. will not face any changes in land use. In order to change the cell values as required for Dyna-CLUE, the [Re-classify tool](#) in the Spatial Analyst → Reclass Toolbox can be used (Figure 7). If you are running a multi-region application of Dyna-CLUE, cell values must reflect unique codes for different regions. The coding must start with 0 and increase by 1 with each additional region.

- Once the correct coding has been established, the file can be converted to the correct ASCII format by using the Raster to ASCII tool in the Conversion → From Raster Toolbox. Make sure to save the file under the correct nomenclature (i.e. *region*. **) (Figure 8).
 - Make a note of the number of cells in the ASCII file by accessing *Properties* → *Symbology* → *Classified* and then selecting the Classify button. Under Classification Statistics – Count you can identify the number of cells in the file (Figure 9). Each of the remaining location and land cover ASCII files to be prepared must match this number of cell values. Similarly, open the ASCII file outside of ArcGIS and through a text editor and make a note of the values stated in the header (number of rows, columns, etc.); these must also be kept equal across all ASCII files used in the same Dyna-CLUE application.
- **Preparation of a “mask” layer (intermediate file)** – the best way to ensure that all ASCII files have the same required characteristics is to create a single reference raster file covering the case study area extent, where a value of 1 represents active areas and NODATA is set to -9999. This file can be created from the previously generated area restriction file and modified using the Re-classify Tool in the Spatial Analyst → Reclass Toolbox to set values of 1 to active areas (Figure 7). By multiplying this region extent file with further datasets we can ensure the extents line up (Figure 10).
- **Preparation of the starting land cover file (model input: *cov_all.0*)** – only one land use type is allowed within each grid cell in the Dyna-CLUE model. In case more land use types are found within a grid cell, a finer resolution should be implemented (by using the Resample tool in the Data Management → Raster → Raster Processing Toolbox) (Figure 6) or the dominant land use type should be chosen for the grid cell (the land use coding can be modified following using the Re-classify Tool in the Spatial Analyst → Reclass Toolbox) (Figure 7). This can unfortunately mean larger uncertainties of the used land use/cover maps, but often it also means harmonization with external economic models that calculate the land use demand. Following any modifications to the land cover raster data file, it can be imported to ArcMap and modified as follows:
 - Multiply the mask layer by the land cover file by using the Times Tool in the Spatial Analyst → Math → Trigonometric Toolbox (Figure 10).
 - Use the Re-classify tool in the Spatial Analyst → Reclass Toolbox to ensure that the land cover classes begin with a code 0 and increase by 1 with each additional land cover class (Figure 7). Make a note in your documentation of what class each code corresponds to.
 - Once the correct coding has been established, the file can be converted to the correct ASCII format by using the Raster to ASCII tool in the Conversion → From Raster Toolbox (Figure 8). Make sure to save the file under the correct nomenclature (i.e. *cov_all.0*).
 - Check that the number of cells in the resulting ASCII file match those of the area restriction file by accessing *Properties* → *Symbology* → *Classified* and then selecting the Classify button (Figure 9). Similarly, you can check the ASCII file header through a text editor and make sure these align with the area restriction file. If the number of cells is not the same then it is likely that the land cover file had NODATA values where the mask file did not. You

may rectify this by choosing to either re-classify the NODATA cells in the land cover raster to a value of 0, if appropriate, or to the averages or median values from surrounding cells by using the Raster Calculator tool. You may alternatively create a new mask which contains the NODATA values from both layers and modify your spatial extent file to fit this new extent.

- **Preparation of the location files (model input: *sc1gr #.fil*)** – the location files in raster format can be imported to ArcGIS and modified as follows:
 - Multiply the mask layer by each location file by using the Times Tool in the Spatial Analyst → Math → Trigonometric Toolbox (Figure 10).
 - Convert each resulting file to the correct ASCII format by using the Raster to ASCII tool in the Conversion → From Raster Toolbox (Figure 8). Make sure to save the file under the correct nomenclature (i.e. *sc1gr #.fil*, where the hashtag refers to the number codes which must begin with 0 for the first location factor and increase by 1 with each additional factor). Make a note in your documentation of what location factor each code corresponds to.
 - Check that the number of cells in the resulting ASCII file match those of the area restriction file by accessing *Properties* → *Symbology* → *Classified* and then selecting the Classify button (Figure 9). Similarly, you can check the ASCII file header through a text editor and make sure these align with the area restriction file. If the number of cells is not the same then it is likely that the location file had NODATA values where the mask file did not. You may rectify this by choosing to either re-classify the NODATA cells in the location raster to a value of 0, if appropriate, or to the averages or median values from surrounding cells by using the Raster Calculator tool. You may alternatively create a new mask which contains the NODATA values from both layers and modify your spatial extent file and land cover file to fit this new extent.
- **Preparation of land use data for statistical analysis (intermediate file)** – for the statistical analysis, individual binary (presence / absence) land cover maps need to be produced for each of the land covers included in the Dyna-CLUE application.
 - Load the starting land cover ASCII file into ArcMap.
 - Use the Raster Calculator tool in the Spatial Analyst → Map Algebra Toolbox and create a conditional function whereby all pixels within the starting land cover grid which correspond to a specific value are assigned the value 1, while the remaining pixels which hold the remainder land use codes are assigned a value of 0 (Figure 11).
 - Repeat this procedure for each of the land cover classes present in the starting land cover file.
 - As an alternative, you may also use the Re-classify Tool in the Spatial Analyst → Reclass Toolbox)

Note: while we do not require or advocate for any particular GIS software, from personal experience we have found that using the ArcGIS suite performs these tasks with higher precision, while using QGIS for these tasks can sometimes result in NODATA values not aligning (though this could be data dependent). If you are using QGIS to clip your data, we advise you to double check the results to ensure that the clipping/raster calculator has performed the task correctly. The exact numbers should be checked as sometimes the issue lies only within a few raster cells and is therefore not immediately visible. This is

particularly the case when it comes to the location layers that are often a result of processing global layers prepared by different institutions which can use different country or coastline masks. Unaligned data or having one or more layers with a different NODATA cell number and spatial distribution is, in our experience, the most common issue when running Dyna-CLUE. If using other software, make sure you additionally export to the ArcGIS header type of ASCII file.

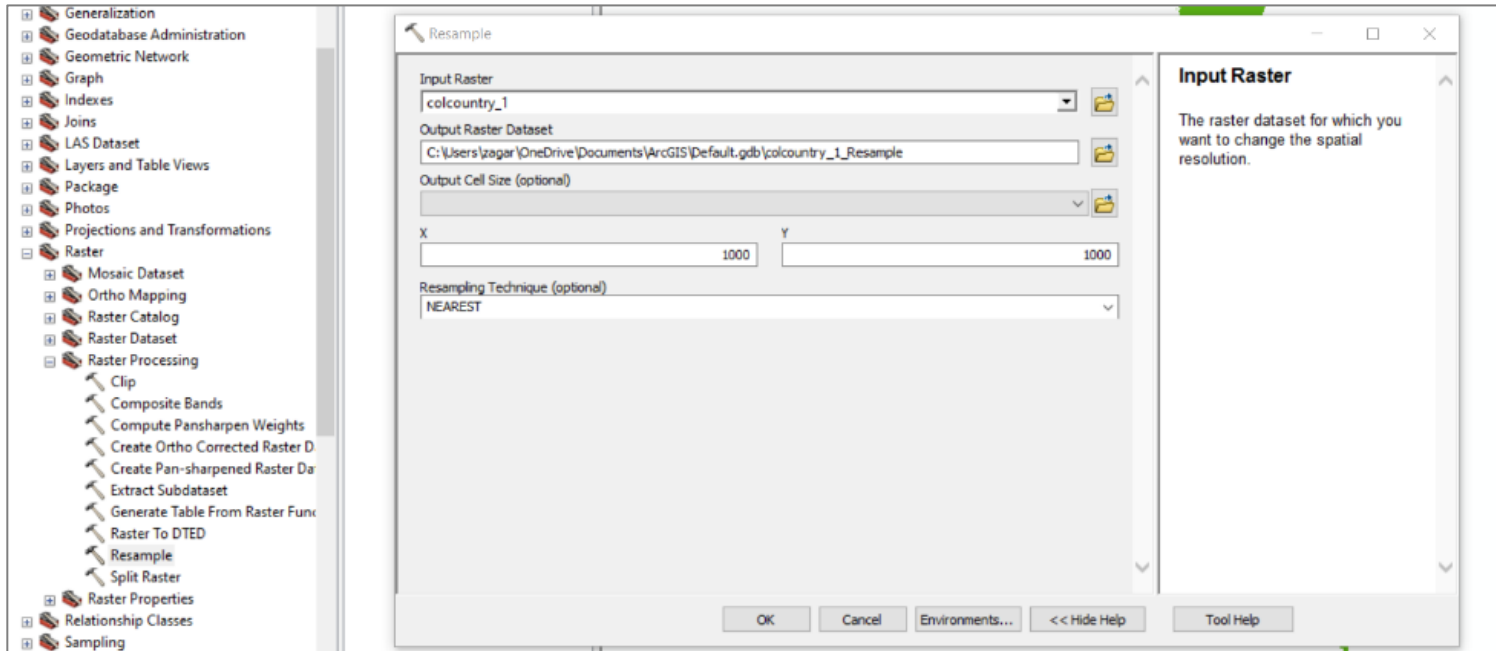


Figure 6 Changing the resolution of raster data by using the Resample Tool in ArcMap (accessed within the Data Management Toolbox). The Nearest resampling technique should be selected for discrete data (e.g. region classes), while Bilinear should be used for continuous data

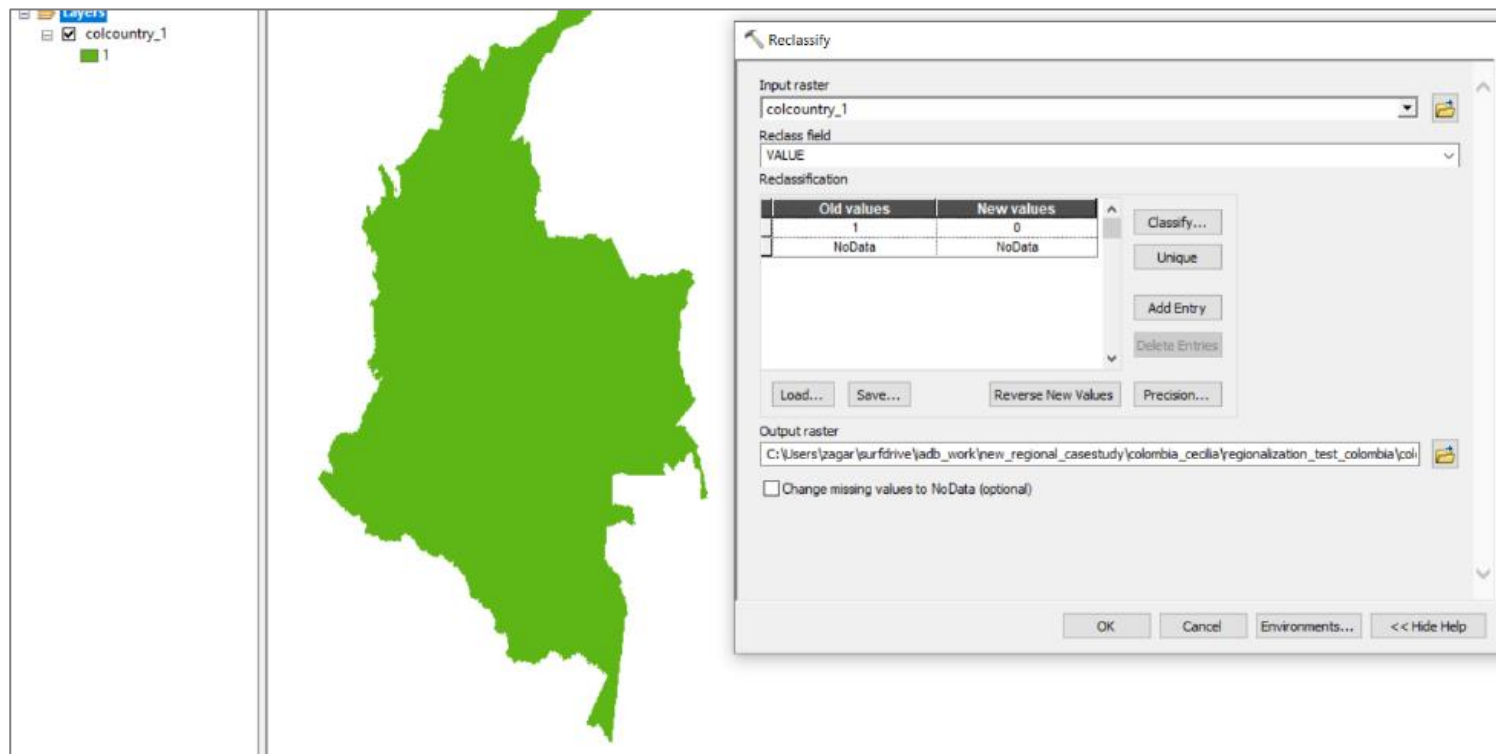


Figure 7 Changing the value of the area restriction file by using the Reclassify Tool in ArcMap (accessed within the Spatial Analyst Toolbox). In the illustrated example, the whole country of Colombia represents the active area to be modelled (indicated by the new value of 0), and no special regional restrictions are in place (otherwise identified by the value -9998)

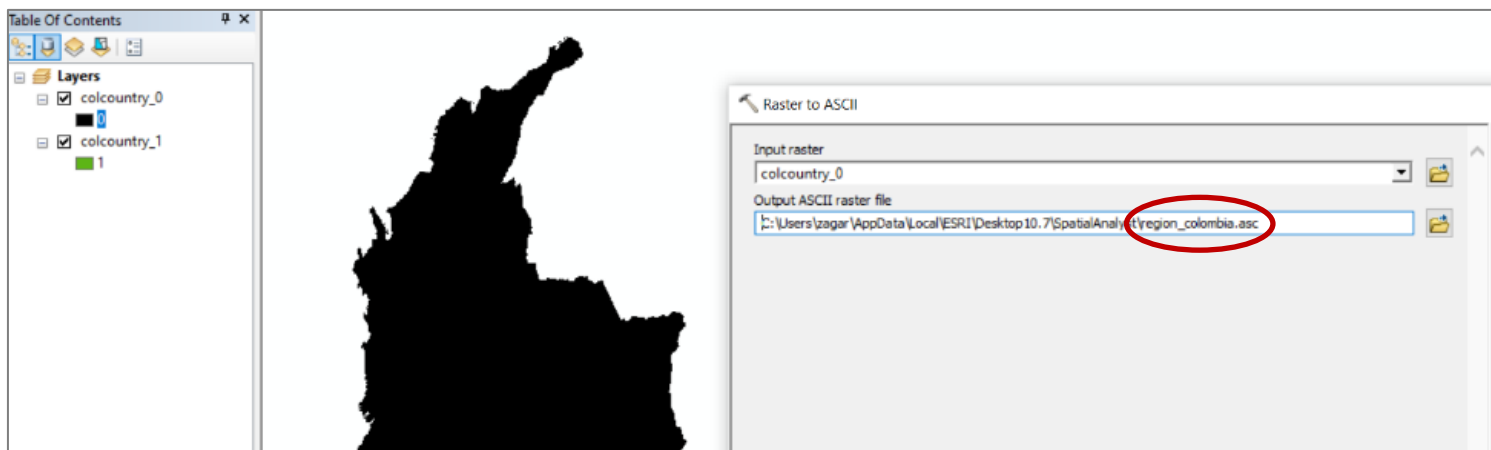


Figure 8 Converting the area restriction file from raster to the required ASCII format by using the Raster to ASCII Tool in ArcMap (accessed within the Conversion Toolbox). The ASCII file should be saved in the correct nomenclature as required by the Dyna-CLUE model, hereby following the “region*.asc” format

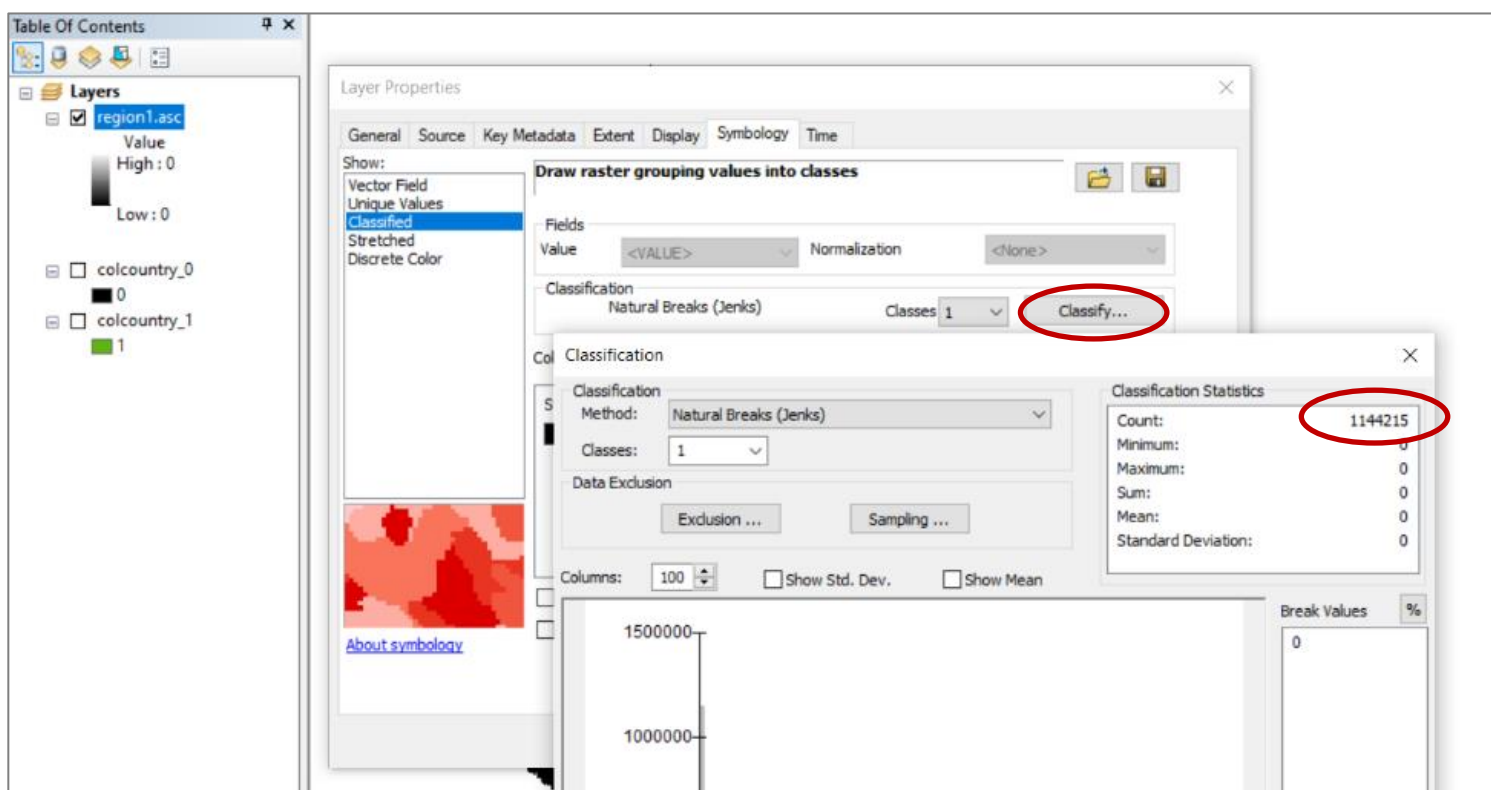


Figure 9 Checking whether the number of cells in the prepared ASCII file matches across all data. The count values in the classification statistics should be equal in all cases

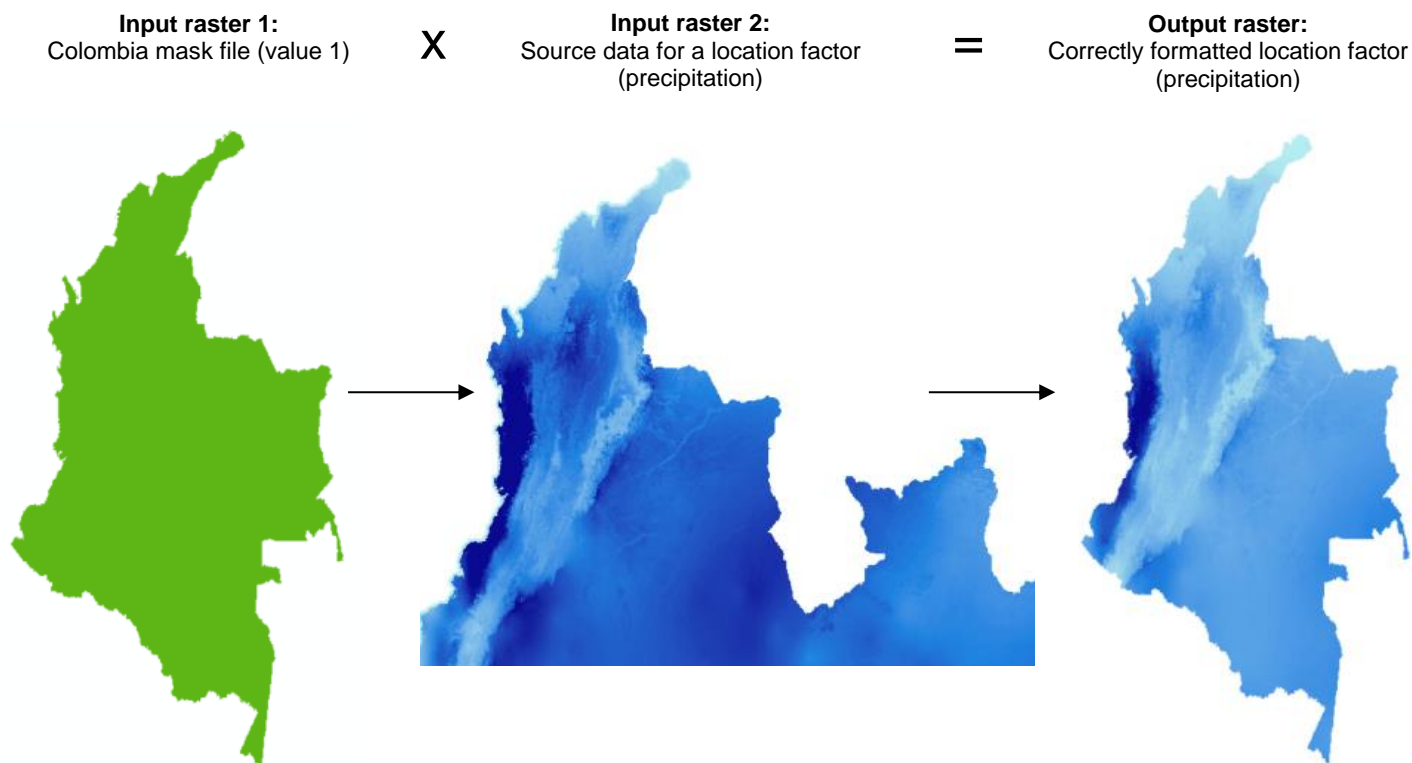


Figure 10 Using a “mask” file corresponding to the desired spatial extent (as outlined in the area restriction file) to multiply all other required spatial input files (i.e. location factors and initial land use data) in order to make sure all files have the same number of cells and spatial extent. This step makes use of the Times Tool in ArcMap (accessed within the Spatial Analyst Toolbox)

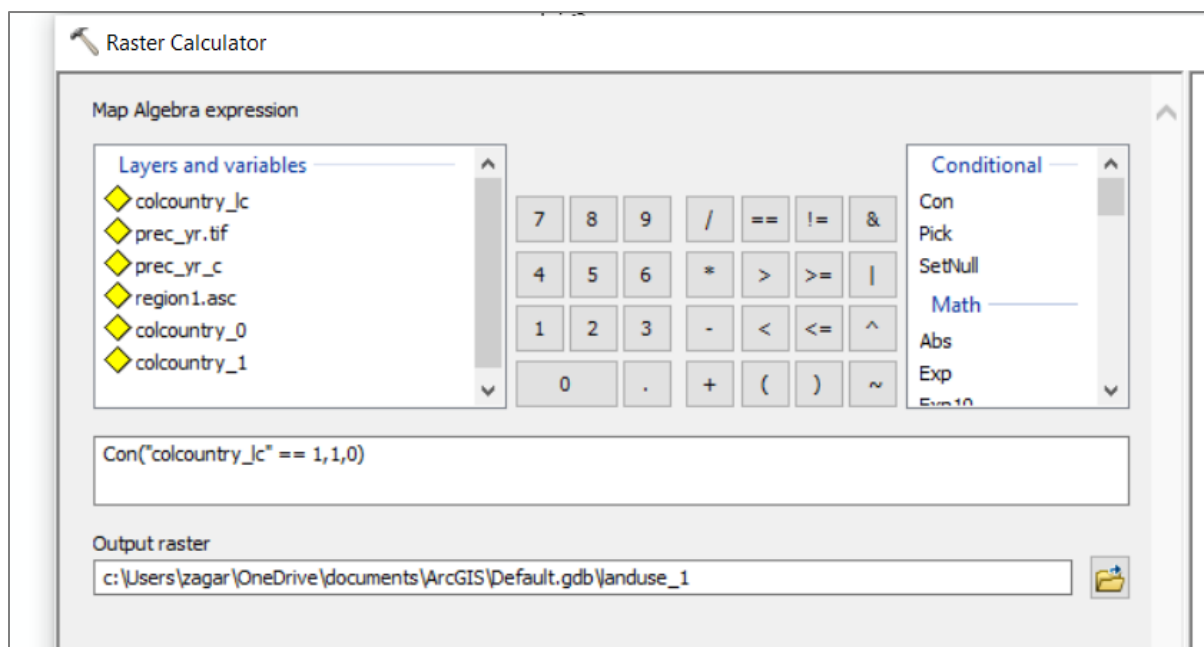


Figure 11 Using the Raster Calculator in ArcMap (accessed within the Spatial Analyst Toolbox) to create binary presence / absence maps for each land cover layer. The above example illustrates the creation of a binary map from the national land cover file (“colcountry_lc”) for the land cover type with code 1

3.3. Definition of land use demands and conversions

Defining land use demands

Demand must be calculated on a yearly basis for each land cover class stated in the initial land use cover ASCII file. The total land cover area must remain constant throughout each year and therefore the course of the application. In practice, this means that if the amount of cropland and grazing area is expected to increase in a region, these amounts have to be subtracted from other land cover classes. These can be from forests, but also from natural grasslands and other (semi)natural areas. These conversions are region and context specific, as in some parts of the world deforestation (conversions from forest to agricultural land) is common, and in others it is not. Additionally, these conversions and subtracted areas need to make sense. New cropland will not be converted from ice areas, or cities. Below is a table demonstrating an example demand file preparation, where for each year demands per class either stay constant, increase, or decrease. Notice that year by year, the total land cover units do not change.

The demand file(s) should be present in the installation directory as a text file named “demand.in”. The first line in the text file should specify the number of years for which demand is being considered in this application. After that, every line contains the demands for the respective land use types for each consecutive year, beginning with the starting land cover configuration. The demand should be specified in hectares. The order of the land use types should follow the numerical order stipulated in the initial land cover ASCII file. In case of multiple regions, the demand for all years is first given for the first region, and then repeated consecutively in the same format for the remaining regions, all within the same demand file. The number of years at the first line has only to be indicated once.

Table 1. Example of future land cover demands in hectares. Land cover types such as bare areas, cities and water and wetlands are static, as they do not change. Cropland and grazing increase on the account of forest and shrubland, which experience decreases each year. The total amount is the same, as the space in the region is limited and does not expand.

Year	Bare and other areas	Cities	Cropland	Grazing	Forest	Shrubland	Water and wetlands	Total
2015	1884	3084	343572	1296040	6580016	676536	109324	9010456
2016	1884	3084	343574	1296111	6579992	676486	109324	9010456
2017	1884	3084	343576	1296163	6579968	676456	109324	9010456
2018	1884	3084	343578	1296198	6579944	676444	109324	9010456
2019	1884	3084	343580	1296264	6579919	676399	109324	9010456
2020	1884	3084	343584	1296352	6579893	676334	109324	9010456
2021	1884	3084	343588	1296451	6579865	676259	109324	9010456
2022	1884	3084	343592	1296557	6579835	676179	109324	9010456
2023	1884	3084	343596	1296667	6579803	676097	109324	9010456
2024	1884	3084	343600	1296781	6579769	676013	109324	9010456
2025	1884	3084	343605	1296899	6579733	675926	109324	9010456

Defining land use conversion rules

For each land use type, a value needs to be specified that represents the relative elasticity of change, ranging from 0 (easy conversion) to 1 (irreversible change):

- 0: Means that all changes for that land use type are allowed, independent from the current land use of a location. This means that a certain land use type can be removed at one place and allocated at another place at the same time, e.g. shifting cultivation.
- >0...<1: Means that changes are allowed, however, the higher the value, the higher the preference that will be given to locations that are already under this land use type. This setting is relevant for land use types with high conversion costs.

We can use different elasticity values in different scenarios. For example, we could lower the conversion resistance of forest to 0, thus simulating a complete lack of forest protection. This way we can observe the result of an absence of any forest protection.

In addition, a matrix needs to be created where conversion rules are specified for each potential land use conversion. Alongside defining which conversions between land uses are allowed, the matrix may also specify in which regions a specific conversion is allowed to occur or not, how many years the land use type should remain the same before converting, or the maximum number of years a land use can remain the same. Figure 12 provides an example of the use of a conversion matrix for a simplified situation with only three land use types, with rows presenting present land use and columns presenting future land use.

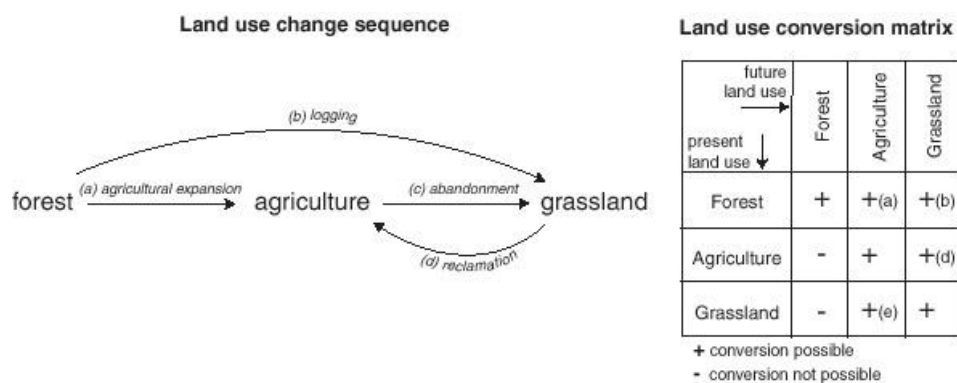


Figure 12 Illustration of the translation of a hypothetical land use change sequence into a land use conversion matrix

In this relatively simple example, where users are only concerned with which land use conversions are allowed throughout the whole area, and not with any temporal aspects or regional restrictions, the matrix should only indicate:

- A value of 0 where no conversions are allowed
- A value of 1 where conversions are allowed

This matrix is illustrated below for an example case study area with three land use types. In this case all land use changes are allowed except the conversion of land use type 0 into land use type 2:

	LU0	LU1	LU2
LU0	1	1	0
LU1	1	1	1
LU2	1	1	1

Additionally, users may wish to specify the location for which special conversion rules apply, or the maximum and minimum number of years an area can remain under the same land use type (examples provided in Figure 13).

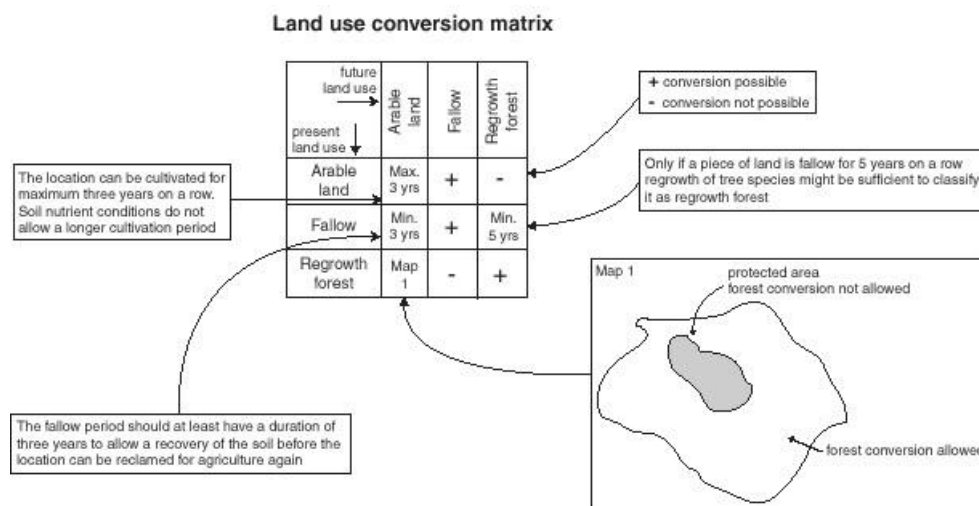


Figure 13 Example of a land use conversion matrix with the different options implemented in the model

In order to indicate special areas for conversion (e.g. if crop rotation is only valid in a certain part of the region), a number referring to a map that indicates where the changes are allowed should be provided within the matrix. The map should be named *sc1gr*.fil*, with * denoting a unique number which does not correspond to any of the other driving factor files generated in the spatial processing phase. In the example below, the conversion of land use type 0 into land use type 2 is only allowed in areas indicated in file *sc1gr20.fil*. This file has a similar format as the area restriction file with a value 1 in areas where this conversion is allowed and a value 0 for areas where this change is not allowed. If areal restrictions are provided within the conversion matrix, then make sure that you also increase the number of driving factor files (line 4) in the main parameter file (*main.1*) according to this setting.

	LU0	LU1	LU2
LU0	1	1	20
LU1	1	1	1
LU2	1	1	1

In other cases, the conversion matrix may be used to indicate the minimum or maximum number of time-steps that a land use type, once established, has to remain stable before a new conversion takes place. Two situations can be distinguished:

- A minimum number of time-steps is needed before the next conversion takes place. This option is relevant for e.g. regrowth of secondary forest after abandonment of farm fields. It can be assumed that the land remains under grassland for at least 5 years before it can be classified as secondary forest. The minimum number of time steps will be indicated in the conversion matrix by specifying the value 100 + the number of time steps required.
- A maximum number of time-steps is specified for crop rotations or, e.g., shifting cultivation. It can be indicated that the conversion shifting cultivation to shifting cultivation (i.e., no change) is only allowed if shifting cultivation is no longer at that location than two cropping seasons. The maximum number of time steps will be indicated in the conversion matrix by specifying the value -100 + the number of time steps required.

In the example provided below, the conversion of land use type 0 to land use type 1 is only possible if the location is already under land use type 0 for 5 time-steps. Land use type 2 may not stay under the same land use type for longer than 5 time-steps. The conversion matrix should be provided in the installation directory as a text file containing only the conversion codes for each land use combination (i.e. it should not include the land use type coding).

	LU0	LU1	LU2
LU0	1	105	1
LU1	1	1	1
LU2	1	0	-105

3.4. Logistic regression analysis

A statistical analysis is used to reveal and quantify the relations between the locations of land use and a set of explanatory factors. The set of explanatory factors is based upon the user's knowledge of the dominant factors causing land use change in the study area. For example, altitude, population density and distance to roads can be considered to be the most important determinants of the forest distribution in a specific study area. Therefore these are the

explanatory factors in the statistical analysis. Based on the biophysical and socio-economic conditions of a location the relative probability of finding the different land use types at that location is defined. A logit model is used to relate the probabilities and the location characteristics.

To estimate the coefficients (β) of the logit model a logistic regression procedure is used with the actual land use pattern as dependent variable. A statistical package, like SAS, SPSS or R, can be used to regress land use upon its explanatory factors. A stepwise regression procedure can be used to select the relevant factors from a larger set of location characteristics. Variables that have no significant contribution to the explanation of the land use pattern are excluded from the final regression equation. More information about logistic regressions can be found in Menard (2002). In the next section the logistic regression analysis is explained for the SPSS package.

Where in ordinal least squares regression the R^2 gives a measure of model fit, there is no equivalent for logistic regression. Instead, the goodness of fit can be evaluated with the ROC method (Pontius and Schneider, 2001), which evaluates the predicted probabilities by comparing them with the observed values over the whole domain of predicted probabilities instead of only evaluating the percentage of correctly classified observations at a fixed cut-off value. This is an appropriate methodology for the Dyna-CLUE model, because a wide range of probabilities is used within the model calculations.

File conversion for statistical analysis

For the statistical analysis, standard statistical packages can be used. Since most of these packages do not support the use of GIS data directly, a number of conversions are needed to translate the GIS data into a data format supported by the statistical packages. With the Dyna-CLUE model a small application is supplied (*convert.exe*) that helps to make this conversion. Basically, it is able to translate ArcInfo/ArcView/ArcGIS ASCII grids into a tabular format that can easily be imported in most statistical packages.

In order to run the conversion application, binary (presence / absence) land cover grids need to be provided for each of the land use types to be included in the statistical analysis (preparation steps outlined in section 3.2). As with all other grids, these should be in ASCII format and have the same number of cells as the location files (*sc1gr #.fil*). Areas where the land use type is present should be indicated by the value 1, while remaining areas within the region of interest should be indicated by a value of 0.

Secondly, a text file entitled *names.txt* should be provided in the directory. This file simply contains a list with the names of the files which need to be converted by the program i.e. the name of one land use file and the name of the explanatory variables to be included in the statistical analysis. In the program (left in Figure 14 below) the 'Edit file list' button can also be used to open and edit the names text file. The conversion needs to be repeated for each of the land use types to be included in the analysis.

Additional options for the conversion include the exclusion of all no-data values, a balanced sample based on the division 0/1 in the first file indicated, or a structured sampling indicating a minimum difference between the observations (to minimize spatial autocorrelation). It is

also possible to indicate the % of the total number of observations that need to be included. This program can also be run in Batch mode by calling the executable followed by the settings for the program. These should include, in the following order: balanced (1) or unbalanced (0); exclusion of NODATA (1) or inclusion (0); the minimal distance; the sample size (%).

The output of the conversion program is saved in a newly created *stat.txt* file; this file provides the tabular format which can be imported in SPSS or other statistical programs. This conversion can also be done using any GIS software. In this case, make sure that the sampling will be balanced (equal number of absence and presence of a land cover map), that you will maintain a minimum distance between sample points and that the sample is geographically spread.

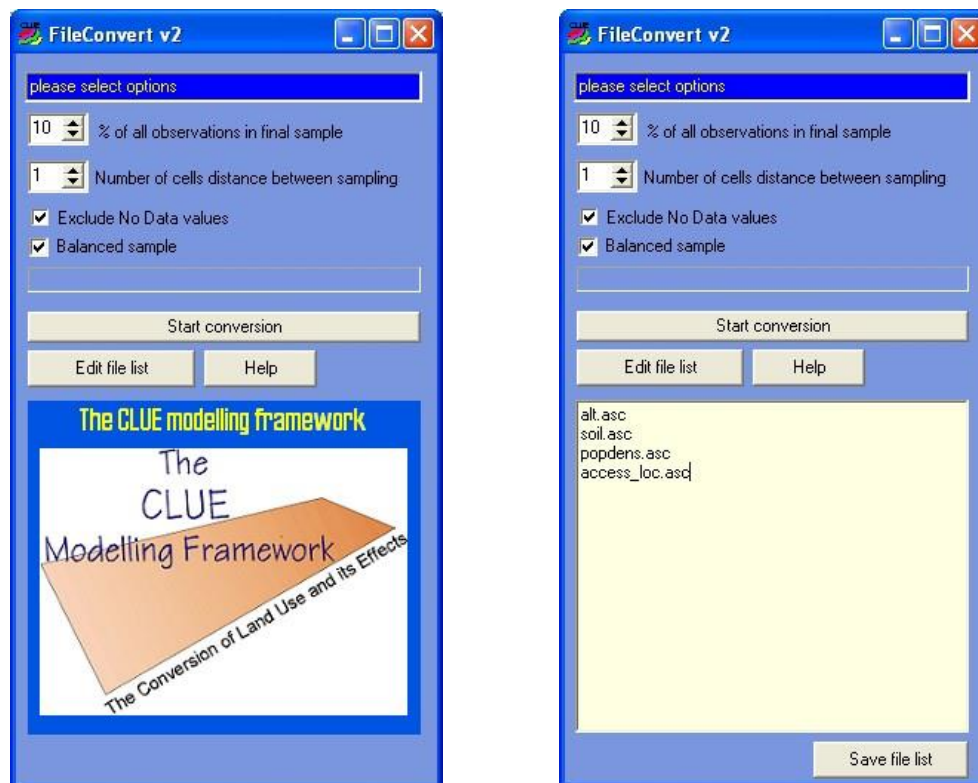


Figure 14 Conversion tool within Dyna-CLUE

Logistic regression analysis with SPSS

Any software can be used to perform logistic regression, as Dyna-CLUE needs only resulting regression coefficients as input. In this document, we however only present how to perform regression using SPSS. To perform the logistic regression analysis, import the sampled data in SPSS (i.e. the *stat.txt* file created through the conversion application) and click 'Analyze - Regression - Binary logistic'. Select a land use type as dependent variable and the location factors that might influence the suitability of a location for that land use type as independent variables/covariates (Figure 15). In this example cropland is the dependent

variable and numerous factors serve as the independent variables. Next, a method for regression has to be selected: 'Enter' if all selected variables will be included or another method, like 'Forward: Conditional' for stepwise regression. If stepwise regression is used the **significance values for Entry and Removal during the stepwise procedure can be indicated under 'Options - probability for stepwise'.** Default values for Entry and Removal for large data sets are respectively 0.01 and 0.02. As last the 'Probabilities' option under 'Save' should be checked to save the predicted probabilities for the evaluation of the goodness of fit with the ROC method. Using other software, we also recommend using forward conditional regression, as it is best to only include the variables that significantly contribute to the spatial distribution of the analyzed land cover type.

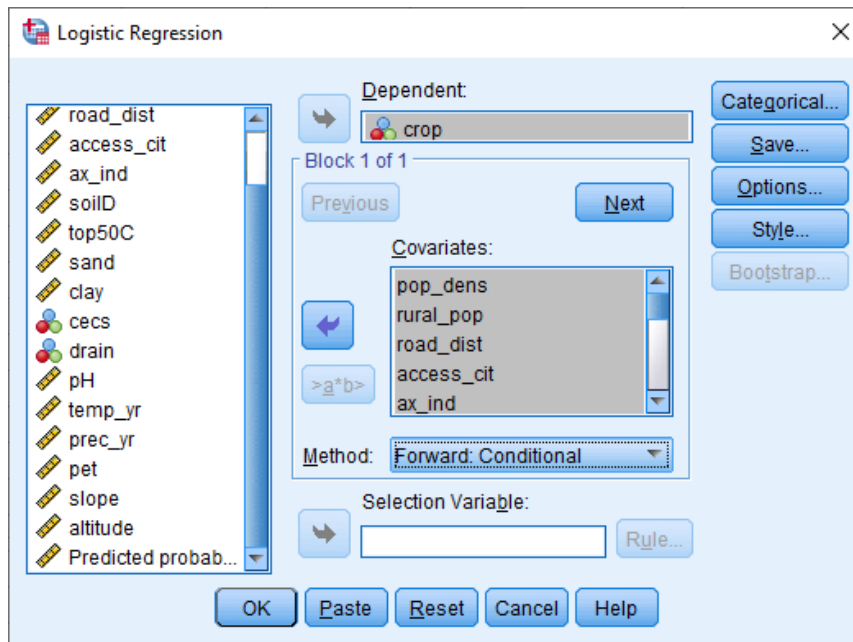


Figure 15 Logistic Regression window of SPSS

The output of the logistic regression is divided in two blocks with several tables. The table 'Variables in the Equation' of Block 1 contains the resulting β -values in the second column (Figure 16). The coefficients of the last step should be used. The logistic regression has to be performed for every land use type. This analysis should result in a separate regression equation for every land use type considered. The resulting coefficients can be entered in a text file with the regression results in the Dyna-CLUE model. In the case of Figure 16 the resulting logit model for cropland in Colombia is:

$$\begin{aligned} \text{Log} \left(\frac{Pi1}{1 - Pi} \right) = & 0.21704 - 0.00465 * \text{top50C} - 0.01276 \\ & * \text{soilD} + 0.00345 * \text{rural} - 0.00033 * \text{pop}_d \\ & + 0.03884 * \text{cecs} + 2.78217 * \text{ax_in} \end{aligned}$$

Variables in the Equation							
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	clip_ax_in	3.817	.253	228.357	1	.000	45.471
	Constant	-.595	.058	105.199	1	.000	.552
Step 2 ^b	clip_cecs	.042	.005	80.847	1	.000	1.043
	clip_ax_in	3.109	.251	153.394	1	.000	22.396
	Constant	-1.236	.094	173.464	1	.000	.291
Step 3 ^c	clip_soilD	-.014	.002	32.598	1	.000	.986
	clip_cecs	.035	.005	53.435	1	.000	1.036
	clip_ax_in	2.853	.252	127.767	1	.000	17.342
	Constant	.332	.288	1.324	1	.250	1.393
Step 4 ^d	clip_soilD	-.013	.002	29.148	1	.000	.987
	clip_rural	.003	.001	11.731	1	.001	1.003
	clip_cecs	.033	.005	45.837	1	.000	1.033
	clip_ax_in	2.665	.255	109.087	1	.000	14.362
	Constant	.249	.290	.737	1	.391	1.283
Step 5 ^e	clip_soilD	-.013	.002	28.572	1	.000	.987
	clip_rural	.003	.001	12.304	1	.000	1.003
	clip_pop_d	.000	.000	5.845	1	.016	1.000
	clip_cecs	.033	.005	46.440	1	.000	1.034
	clip_ax_in	2.808	.265	112.018	1	.000	16.580
	Constant	.229	.291	.622	1	.430	1.258
Step 6 ^f	clip_top50	-.00465	.00239	3.80404	1.00000	.05113	.99536
	clip_soilD	-.01276	.00248	26.48093	1.00000	.00000	.98732
	clip_rural	.00345	.00099	12.26380	1.00000	.00046	1.00346
	clip_pop_d	-.00033	.00013	6.08804	1.00000	.01361	.99967
	clip_cecs	.03884	.00566	47.12990	1.00000	.00000	1.03960
	clip_ax_in	2.78217	.26540	109.89265	1.00000	.00000	16.15412
	Constant	.21704	.29083	.55694	1.00000	.45550	1.24240

Figure 16 Output of logistic regression with SPSS

Evaluation of the logistic regression

The ROC characteristic is a measure for the goodness of fit of a logistic regression model similar to the R^2 statistic in ordinary least square regression. A completely random model gives a ROC value of 0.5, while a perfect fit results in a ROC value of 1.0. The ROC curve can be defined under 'Analyze – Classify – ROC curve'. Select the observed land use as 'State Variable' with value 1 for occurrence of this land use type and the 'Predicted probability' as 'Test Variable' (Figure 17). In this example cropland is the state variable and the predicted probability for forest of the logistic regression is the test variable. The resulting ROC curve is shown in Figure 18. The area under the curve is the ROC value, in this case 0.868. The closer this value is to 1, the better is the fit of the logistic regression model.

Note: We usually observe higher ROC values for land cover types such as forest or built-up areas, and lower for more indefinite or multifunctional land cover types, such cropland (that can consist of numerous crops with different requirements for soil, climate...) or agroforestry (as it partly resembles forests, and partly cropland/grazing). The AUC can also be related to the quality of the land cover map, where some classes could be mapped with higher uncertainties.

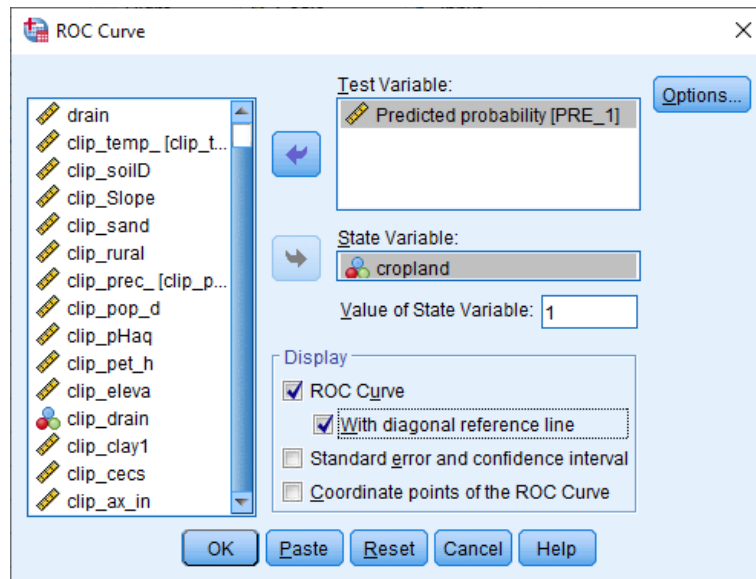


Figure 17 ROC curve definition window of SPSS for cropland

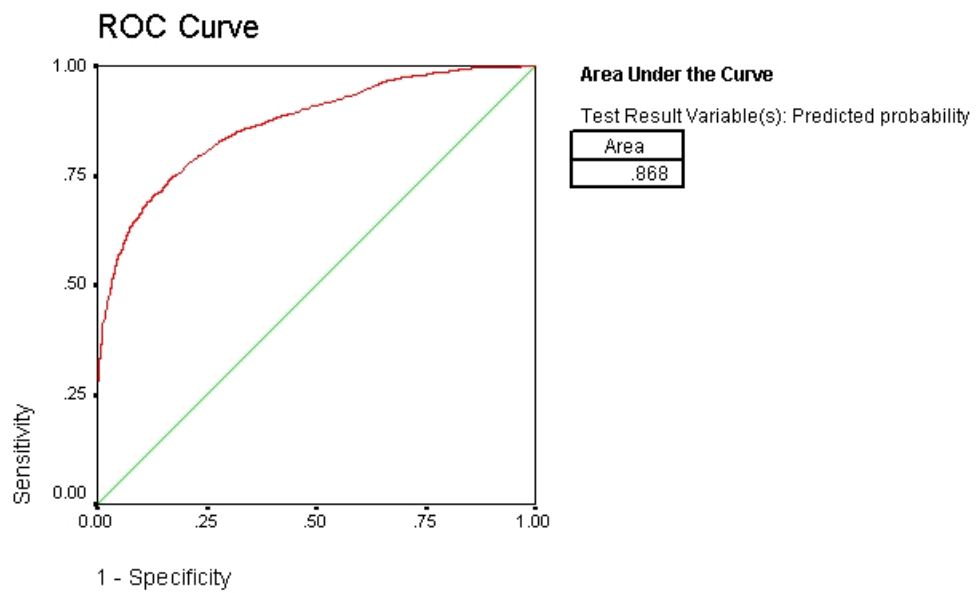


Figure 18 ROC curve for a logistic regression for cropland

Preparing the regression file

The results of the regression analysis have to be provided in a text file entitled *alloc1.reg* following the specific sequence outlined below which has to be repeated for each consecutive land use type.

Line 1:	Number code for land use type (e.g. forest).
Line 2:	Constant of regression equation for land use type (i_0).
Line 3:	Number of explanatory factors (<i>sc1gr#.fil</i> files) in the regression equation for that land use type. For example population density, distance to roads and volcanic rocks are the explanatory factors for the land use type forest, so the number of explanatory factors is 3.
Line 4 and further:	On each line the beta coefficients (i_1 , i_2 , etc.) for the explanatory factor and the number code of the explanatory factor.

If more than one region is considered and different regressions are derived for the different regions, the complete sequence above with all land use types should be replicated for the different regions within the same *alloc1.reg* file.

4. Using the model

In this next section we describe the process of running an application of Dyna-CLUE. Figure 19 illustrates the different files which must be present in the installation directory (at a minimum) in order for the Dyna-CLUE model to run correctly. We begin with an explanation of the model' interface and of the "main parameters" text file which should be adjusted to reflect the new application. We then briefly iterate the required format for the ASCII grids and text files prepared following the steps outlined in section 3, as well as some optional features of the Dyna-CLUE model (location-specific preference addition, probability maps and regionalization).

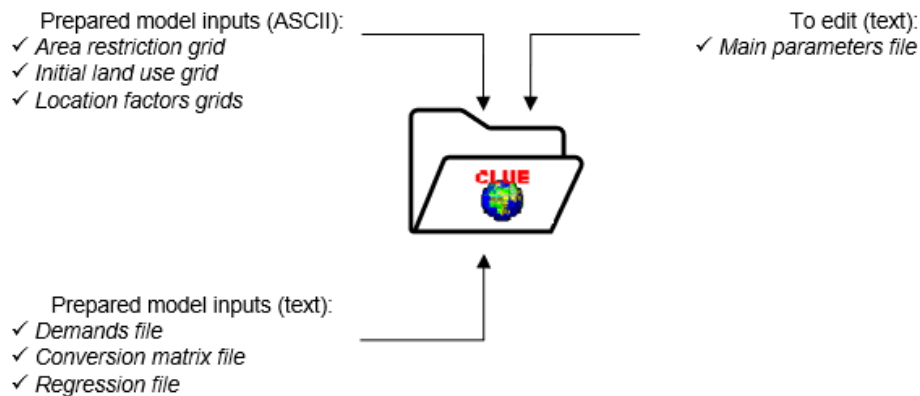


Figure 19 Components of the installation directory required for running an application of Dyna-CLUE

4.1. User interface

Open *clues.exe* in the installation directory to open the Dyna-CLUE model. The user-interface of the Dyna-CLUE model consists of two parts (Figure 20). The dark-gray panel in the upper part supports the specification of a number of the input-variables of the model. The bottom part of the user-interface provides runtime information during the simulation. The edit input menu gives a number of entries needed to edit the text files that can be edited in order to configure the model settings. Under 'Area restrictions' and 'Demand scenario' one should choose one of the scenario files for a specific model run. Only if all these variables are entered correctly the simulation can be started. The Mode menu gives the possibility to either make a normal simulation (none of the items checked), a simulation where only the probability maps for the first year are generated (including neighborhoods and location specific addition): calculate probability maps; a simulation where the probability maps for all years are prepared for the meta-model runs: prepare probability maps; and a simulation based on already calculated probability maps (much faster for repeated runs, calibration or sensitivity analysis): meta-model run. After selecting (highlighting) your desired area restriction and demand scenario files, you can run the model by pressing the run button and the simulation will begin. A progress bar will indicate the progress of the simulation based on the number of years of the simulation. The lower bar shows the iteration count during the simulation. The graph indicates the difference between allocated land use quantity and the land use quantity indicated in the land requirements file.

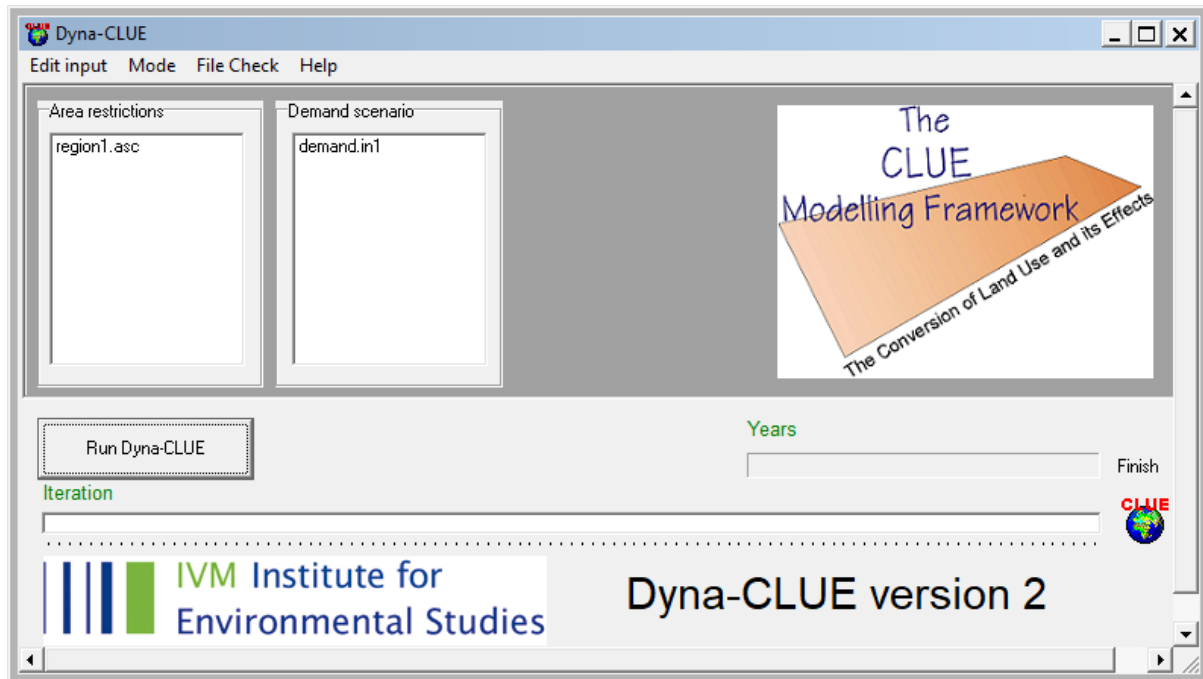


Figure 20 User interface of the Dyna-CLUE model.

4.2. Main parameters

After pressing the button 'Main parameters' a text editor will appear (Figure 21). This file contains all important parameters that determine the configuration of the simulation. The parameters can be edited in the text editor and saved by pressing the Save button. Only after pressing the Save button the changes will affect the simulations.

The parameters (Table 2) specified in the text editor are saved in a file called *main.1* that is located in the installation directory. This file can also be edited by Notepad or any other text editor. When the file *main.1* is not present in the installation directory it cannot be edited in Dyna-CLUE and the model will not be able to run. Each parameter in the file is discussed below.

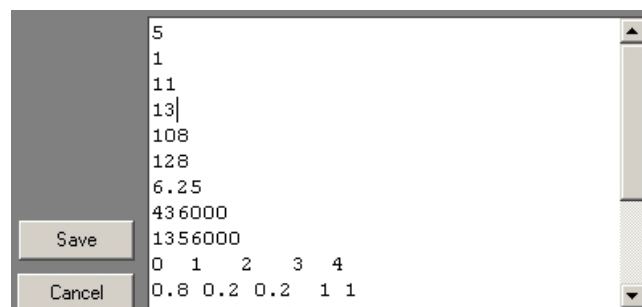


Figure 21 Text window to edit the main parameters

Table 2 Main parameters for Dyna-CLUE main file (main.1)

Line	Description	Format
1	Number of land use types	Integer
2	Number of regions	Integer
3	Max. number of independent variables in a regression equation	Integer
4	Total number of driving factors	Integer
5	Nr. of rows	Integer
6	Nr. of columns	Integer
7	Cell area (in ha)	Float
8	xll coordinate	Float
9	yll coordinate	Float
10	Number coding of the land use types	Integers
11	Codes for conversion elasticities	Float
12	Iteration variables	Float
13	Start and end year of simulation	Integers
14	Number and coding of explanatory factors that change every year	Integers
15	Output/input file choice	1, 0, -2 or 2
16	Region specific regression choice	0, 1 or 2
17	Initialization of land use history	0, 1 or 2
18	Neighborhood calculation choice	0, 1 or 2
19	Location specific addition	Integers
20	Optional: iteration parameter	Float

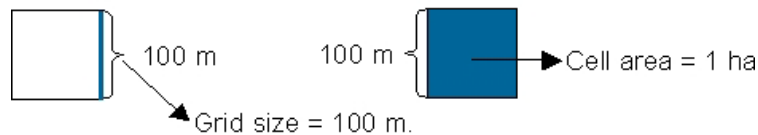
Detailed description of the main parameters

1. The number of land use types that are distinguished for the simulation. Maximum 12 different land use types can be identified in this version*.
2. The number of regions, default is 1. Normally a study area consists of only one region, but for study areas that are divided in parts with a very different behavior of land use types, more regions might be used. For example for areas with a highland and lowland region or for areas with more islands.
3. The maximum number of independent variables in a regression equation. This is the number of driving factors of the regression equation with most variables. The model can handle 20 variables as maximum*.
4. The total number of driving factors that are addressed, this means the number of driving factor files per year. The model can handle 30 files as maximum.*
5. Number of rows of the input grids. The maximum extent is 1000 rows in the demo

version, which means a maximum northing of 1000 times the grid size.*

6. Number of columns of the input grids. The maximum extent is 1000 columns in the demo version, which means a maximum easting of 1000 times the grid size.*

7. The cell area (e.g. 25 ha) of the grid cells. The cell area should be expressed in hectares.* The picture gives an explanation of the terms "cell area" and "grid size":



8. X-coordinate of the lower left corner, this is also indicated in the header of the ArcView and ArcGIS ASCII files.

9. Y-coordinate of the lower left corner, this is also indicated in the header of the ArcView and ArcGIS ASCII files.

10. The number coding of the land use types. The coding should start with 0 for the first land use type, 1 for the second land use type, etc. If two or more land cover types have a 'joint' land requirement which may be the case when 'autonomous developments' are implemented only unique numbers are provided for land cover types that have a 'unique' land requirements. Land cover types the join the same land requirements are given the same coding

11. The codes for the allowed changes and behaviour of the land use types (conversion elasticities) have to be indicated for each land use type, following the order of line 10. This must be a value between 0 and 1.

- 0: Means that all changes for that land use type are allowed, independent from the current land use of a location. This means that a certain land use type can be removed at one place and allocated at another place at the same time, e.g. shifting cultivation.
- >0...<1: Means that changes are allowed, however, the higher the value, the higher the preference that will be given to locations that are already under this land use type. This setting is relevant for land use types with high conversion costs.
- 1: Means that grid cells with one land use type can never be added and removed at the same time. This is relevant for land use types that are difficult to convert, e.g., urban settlements and primary forests. A value of one stabilizes the system and prevents that in case of deforestation other areas are reforested at the same time.

12. Iteration variables: three numbers should be specified:

- Iteration mode. Options include:

* The full version of the model supports also a larger configuration. Both versions are available on the CLUE model website and have otherwise the same functionalities.

- 0, which means that convergence criteria are expressed as a percentage of the demand;
 - 1, which means that convergence criteria are expressed as absolute values (units of demand, so hectares)
 - First convergence criterion:
 - Average deviation between demanded changes and actually allocated changes (default for %: 0.35; for the absolute iteration mode at least the cell area divided by the number of land use types).
 - Second convergence criterion:
 - Maximum deviation between demanded changes and actually allocated changes (default for %: 3). The percentage should minimally equal 1 cell difference between the allocated area and the demand for the land use type with the smallest demand, e.g. for a land use type with a demand of 200 ha and a cell area of 4 ha, the value should be minimally $(4/200) \cdot 100\% = 2\%$. For the absolute iteration mode, the value should be at least the area of one cell.
13. Start year and end year of the simulation.
14. Number of dynamic driving factors, e.g. population density, elevation. This number should be followed by the coding of these explanatory factors, e.g. 2 9 11. This means that there are 2 dynamic driving factors: driving factors 9 and 11 are dynamic and have input files for each year.
15. Choice for the type of output file:
- 1: ArcView headers will be printed in output files;
 - 0: No headers in output files (suited for e.g., Idrisi).
 - -2: No headers in output files, suppress information in log-file on iteration procedure.
 - 2: ArcView headers will be printed in output files, suppress information in log-file on iteration procedure.
 - 3: ArcGIS/ArcMap extensions (*.asc) and headers are added to the output files for easy conversion in ArcGIS.

Please note that running the model in this mode requires this extension also for all input files that contain map data. The .asc extension is added to the standard file names (e.g. region1.fil.asc)

Note: If the ArcView-ArcGIS file type is chosen, all input grid files should contain an ArcView-ArcGIS header.

16. Choice for a region specific regression:
- 0: No different regressions for different regions
 - 1: Equal or different regressions for different regions; demands for each region specifically defined
 - 2: Different regressions for different regions, demand for all regions aggregated

The default is value 0.

17. When temporal dynamics are taken into account it is necessary to know the land use history. Often this is not known and a random number has to be assigned. One can choose between three options:

- 0: The initial land use history will be read from file `age.0`, which should contain a grid with, for every pixel, the number of years that the pixel is used for the current land use type.
- 1: A random number will be assigned to all pixels to represent the number of years that the current land use is already found at that location according to the standard seed for the random number generator. With this option two runs will result in the same random number, which is useful for comparison.
- 2: A random number will be assigned to all pixels to represent the number of years that the current land use is already found at that location with a different random number generator. Two consecutive runs will start with different land use histories. For option 1 or 2 an additional number should be added that indicates the maximum number of years that can be generated by the randomizer (5 in the example below).

18. Choice for using the neighborhood function:

- 0 Neighbourhood function is not used.
- 1 Neighbourhood function is used in simulation.
- 2 Only the influences are calculated, the influence files are saved directly, no simulation.

19. Variables for location specific preference addition. The first number is a switch: activate the function (1) or not (0). If the switch is set to 1, it should be followed by, for each land use type, the fraction of the preference addition that is added to the probability as calculated in the regression.

20. Iteration parameter. If no value is supplied 0.05 is assumed which works well with most applications. The value of this parameter should be between 0.001 and 0.1. With a somewhat higher value (e.g. 0.06) the iteration is more stable and more likely to find a solution. A lower value (e.g. 0.01) will give a faster convergence although instability is more likely.

An example of the main parameter file can be found in the data package.

4.3. Regression results

After pressing the 'Regression results' button a text editor for the file that contains the results of the regression analysis for the different land use types appears. The file can also be edited directly with any text editor by opening the file `alloc1.reg`, which is located in the installation

directory. When the file *alloc1.reg* is not present in the installation directory it cannot be edited in Dyna-CLUE. The example of Figure 22 shows for land use type 0 a regression with three explanatory factors (line 1 = land use code, line 2 = constant of regression equation for land use code 0, line 3 = number of explanatory factors for land use code 0, and lines 4-6 = the beta coefficients for the explanatory factor and the number code of the explanatory factor, in this case *sc1gr10*, *sc1gr7*, *sc1gr1*).

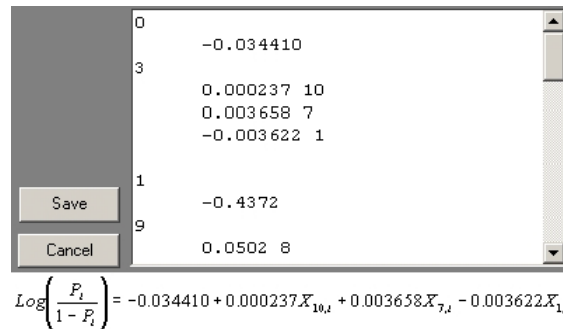


Figure 22 Text window to edit the regression parameters

The above sequence has to be repeated for each land use type. If more than one region is considered and different regressions and demands are derived for the different regions, the complete sequence above with all land use types should be replicated for the different regions. The switch on line 16 of the main input file should in this case be set to 1 or 2. If more regions are considered with the same regression functions, the switch on line 16 should be set to 0 (if regions have the same demands) or 1 (if regions have separate demands); in this latter case, the exact same regression sequence has to be replicated for the number of different regions within the same regression file.

4.4. Conversion matrix

After pressing the 'Change matrix' button a text editor for the file *allow.txt* appears (Figure 23). This file is a matrix indicating the allowed land use conversions. It is a $y \times y$ matrix where y equals the number of land use types, e.g., in case of 8 land use types it will be an 8x8 matrix. Rows denote the present land use type and columns the potential future land use type. Values indicate the following:

- 0: no conversion allowed
- 1: conversion is allowed
- 10x: Conversion allowed after a specified number of years as the 'from' land use class. 102 means a conversion after 2 years at the 'from' land use class, 110 means a conversion after 10 years, etc. For example, shifting cultivation switches to arable plantation after 2 years.
- -10x in the diagonal (e.g. -102 defines the matrix relationship in the same land use class, so shifting cultivation/shifting cultivation): The class cannot stay the same for more than x years. In this example, the conversion shifting cultivation to shifting cultivation (i.e. no change) is only allowed if shifting cultivation is no longer at that location than two cropping seasons and is changed to another land use type after that.

Figure 23 illustrates how values of the matrix (with in this case 5 different land use types) should be presented in the *allow.txt* file present in the installation directory. In this case, all conversions are allowed between land uses with no temporal or aerial restrictions. If regional restrictions are implemented (see section 3), make sure that you also change the number of driving factor files (line 4) in the main parameter (*main.1*) according to this setting.

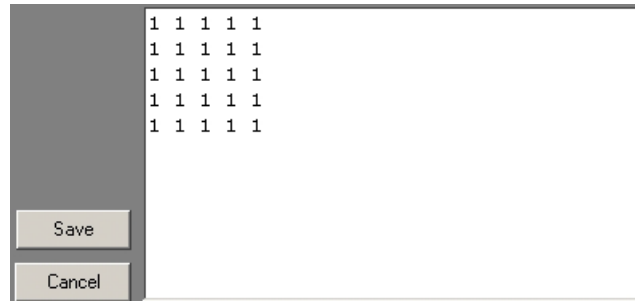
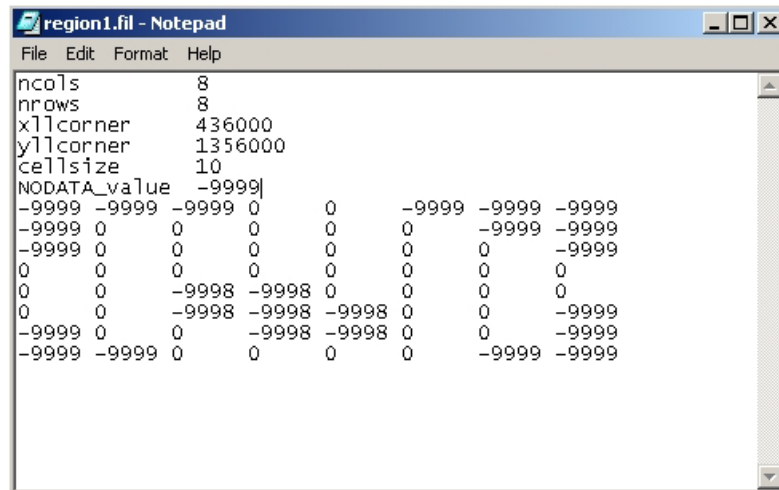


Figure 234 Text window with the conversion matrix

4.5. Area restrictions

In order for a Dyna-CLUE application to be able to run, an area file (titled *region*. **) must be provided in the installation directory. This file should contain a grid with the same cell size and extent as the other grids, indicating for which grid cells land use change calculations should be made. Only active cells are used for calculations; these active cells are indicated by the value 0, while no data cells should have the value -9999. Additionally, the grid can specify for which part of the active area, if any, special conditions are valid, for example as a result of spatial policies, restrictions or tenure status. These restricted area cells should have the value -9998. Grid cells that are defined to be with restricted area are considered to face no changes in land use, e.g. because of the protection of nature in a nature reserve. Figure 24 shows an example for an area of 8 rows by 8 columns with a restricted part in its center. Note: the land use area within the restricted area should be included in the demand (for example, forests in a protected area should still be considered in the demand for forest, and not excluded). All files in the installation directory called *region*. ** (e.g. *region1.fil*) will appear in the 'Area restriction' selection box of the Dyna-CLUE model interface.



```

region1.fil - Notepad
File Edit Format Help
ncols      8
nrow       8
xllcorner  436000
yllcorner  1356000
cellsize   10
NODATA_value -9999
-9999 -9999 -9999 0 0 -9999 -9999 -9999
-9999 0 0 0 0 0 -9999 -9999
-9999 0 0 0 0 0 0 -9999
0 0 0 0 0 0 0 0
0 0 -9998 -9998 0 0 0 0
0 0 -9998 -9998 -9998 0 0 -9999
-9999 0 0 -9998 -9998 0 0 -9999
-9999 -9999 0 0 0 0 -9999 -9999

```

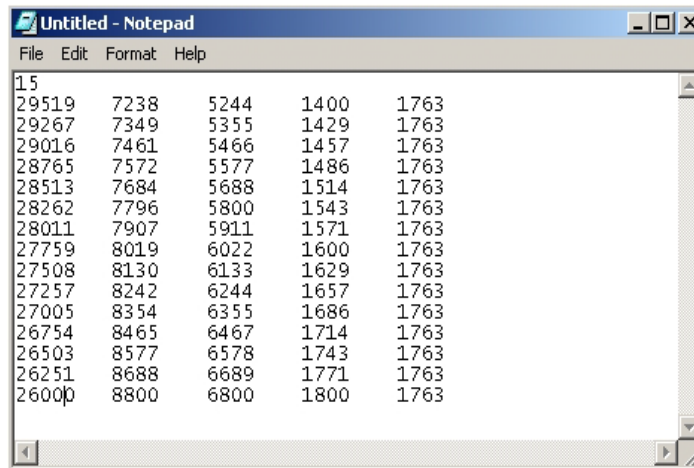
Figure 24 Example of an area restriction file in the form of a ASCII file. Values -9998 define a protected area, whereas values -9999 define the 'No data' area.

4.6. Demand scenario

All files names in the installation directory that are named *demand.in** will be listed in the selection window. These files contain for each year that needs to be simulated the demand for each land use type. The demand should be specified in hectares. The first line specifies the number of years for which demand is included in this file. This should at least be the number of years for which the simulations are to be made. After that, every line contains the demands for the respective land use types for each consecutive year, beginning with the starting land cover configuration. The order of the land use types should be the same as in the main parameter file. In case of multiple regions, the demand for all years is first given for the first region, and then repeated consecutively in the same format for the remaining regions, all within the same demand file. The number of years at the first line has only to be indicated once.

Note: the total demand for all land use types together should be equal to the total land area from the region file and should remain constant for each year.

Figure 25 shows an example for a simulation with 5 land use types and a demand scenario for 15 years. The demand for land use type 0 decreases, while the demand for land use types 1, 2 and 3 increases. The demand for land use type 4 remains constant.



	1	2	3	4	5
15					
29519	7238	5244	1400	1763	
29267	7349	5355	1429	1763	
29016	7461	5466	1457	1763	
28765	7572	5577	1486	1763	
28513	7684	5688	1514	1763	
28262	7796	5800	1543	1763	
28011	7907	5911	1571	1763	
27759	8019	6022	1600	1763	
27508	8130	6133	1629	1763	
27257	8242	6244	1657	1763	
27005	8354	6355	1686	1763	
26754	8465	6467	1714	1763	
26503	8577	6578	1743	1763	
26251	8688	6689	1771	1763	
26000	8800	6800	1800	1763	

Figure 25 Example of a demand scenario for 5 land use types

4.7. Initial land use configuration

The initial land use has to be provided in order for the simulation to start. An ASCII file called *cov_all.0* should contain the grid values indicating the dominant land use type for each grid cell at the start of the simulation (year 0). The grid values should contain the code of the land use type (beginning with 0 and increasing consecutively); these should match the codes indicated in the respective main parameter file line (10). For the areas outside the simulation area, e.g. sea or lake, the grid values should be set to -9999. The file should have the same extent and cell size as the other input grids. **Note:** if Arcview-ArcGIS option in *main.1* (line 15) is chosen the file should contain an Arcview-ArcGIS header.

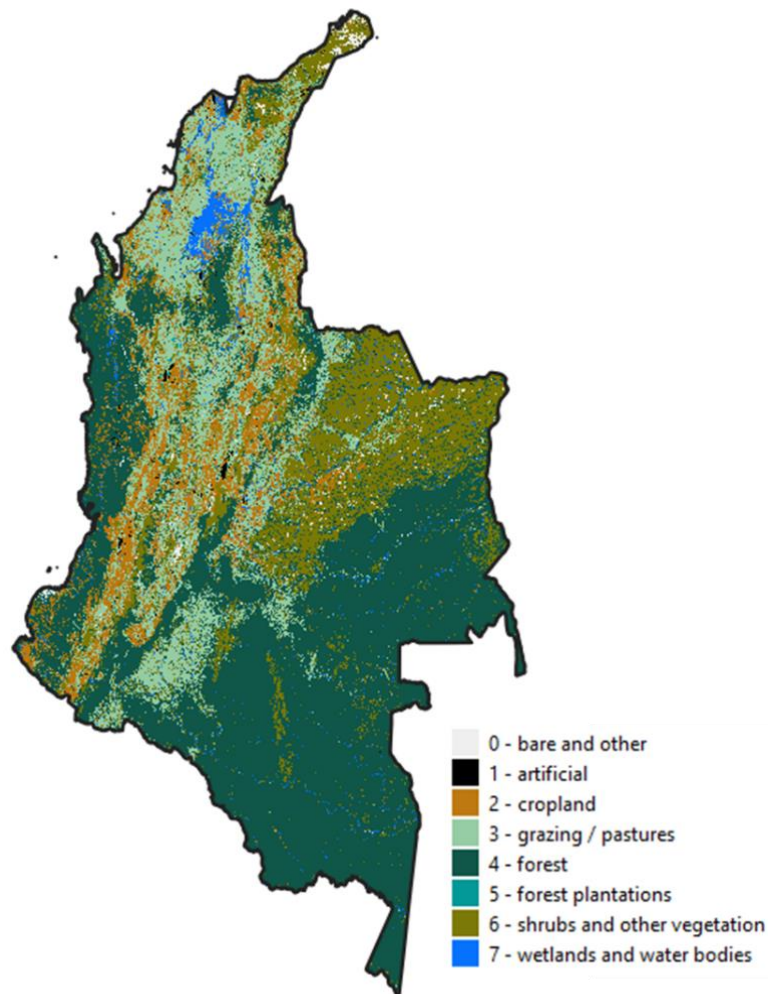


Figure 26 Initial, aggregated land use for Colombia used as input for the Dyna-CLUE model

4.8. Driving factors

Two types of files that specify the driving factors of land use change are distinguished. The first type of driving factors remains constant during the simulation period (e.g. altitude) while the second type changes (e.g. climate).

Stable driving factors

The stable driving factors are represented by a file called *sc1gr#.fil*, where # indicates the code for the explanatory factor. For each explanatory factor that does not change during the simulation, e.g., altitude, a file should be prepared.

Dynamic driving factors

The dynamic driving factors are represented by a file called *sc1gr#. **, where # indicates the code for the explanatory factor and * the year. For each explanatory factor that changes during the simulation, e.g., population density or precipitation, a file should be prepared for each year of the scenario to be simulated. Besides that, a file named *sc1gr#.0* is needed that is identical to *sc1gr#.0*.

Both types of files should be ASCII files containing the grid values of the explanatory factor.

For areas outside the simulation area the code -9999 should be used. The files should have the same extent and cell size as the other input grids. Figure 27 shows an example of a driving factor.

Note: if the Arcview-ArcGIS option in *main.1* (line 15) is chosen the file should contain an Arcview-ArcGIS header. Please check that all files have a value for each grid cell for which simulation will take place, as defined in the region file.



Figure 27 Example of a location factor (annual precipitation) for Colombia

4.9. Location specific preference addition

Spatial policies can result in an increase of the preference for a certain land use on a location. As an example, when financial support is given for maintaining agriculture in mountainous areas, the probability that agricultural land use is conserved in these areas is higher than can be expected based on the regression between land use and (biophysical and socio-economic) driving factors that apply for the study area as a whole.

To take these factors into account, the probability can be increased at specific locations. If this option is used, for each land use type a file named *locspect#.fil* (# indicates the land use type number) should be supplied in the installation directory that contains a map with for each cell the preference addition. **NOTE:** these additions should preferably be between 0-1 in order to fit the range of the regression results.

The switch on line 19 of *main.1* should be set at 1 and weight factors should be specified. If the switch is set at 0 the function is not activated and locspec#.fil files are not needed.

Figure 28 shows an example of a locspec#.fil map for Colombia. In the example, the probability for silvopastoral systems is increased in areas with high erosion risk. In locspec0.fil, the probability increase is 0.2 or 0.5 (depending on the erosion risk), but in line 19 of *main.1* the probability increase can be changed by specifying the weight factor.

Please note that for each land use type a locspec#.fil file needs to be defined. If the locspec#.fil is only used for one or two land use types the weight factor for the other land use types should simply be set to 0.

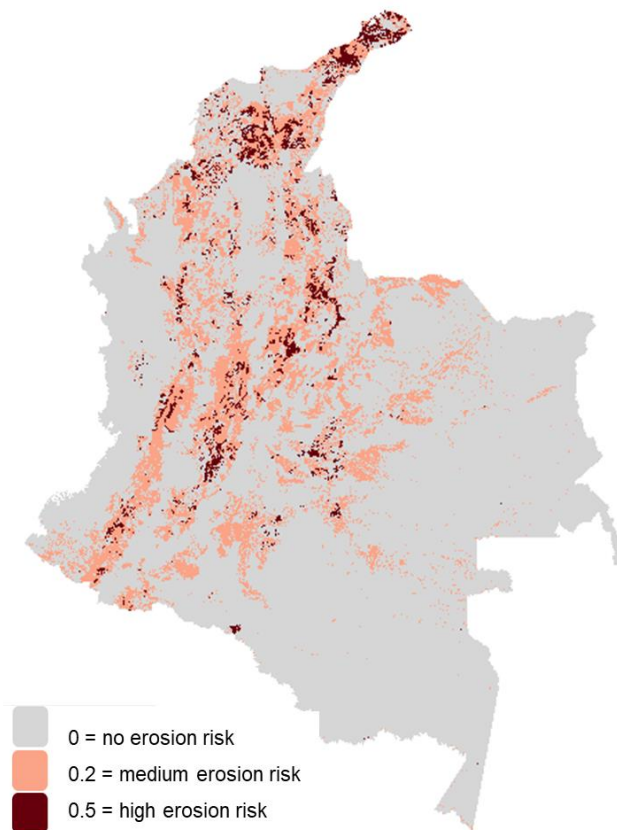


Figure 28 Locspec file for silvopastoral system based on the erosion risk severity in Colombia. Numbers present the value assigned as location specific addition for the spatial probability of the silvopastoral system.

4.10. Calculating probability maps

The Dyna-CLUE model has an option to calculate probability maps from the logistic regression results. The maps can be calculated by clicking on the 'Calculate probability maps' button. Note that before clicking the calculate probability button it is needed to select a scenario (demand and area restriction). This does not influence the calculated probabilities. Probabilities will only be calculated for grid cells outside restricted areas.

The results are saved in ASCII files that are called *prob1_*.1*, where * indicates the land use type number. A file is created for each land use type. These files can be imported by a GIS package to see the result (Figure 29). Please make sure that floating point data (not integer data) are imported. The probability maps are useful to test whether the hypothesis for the driving factors of each land use type where correct. The probability for that land use type should then be high on the current locations.

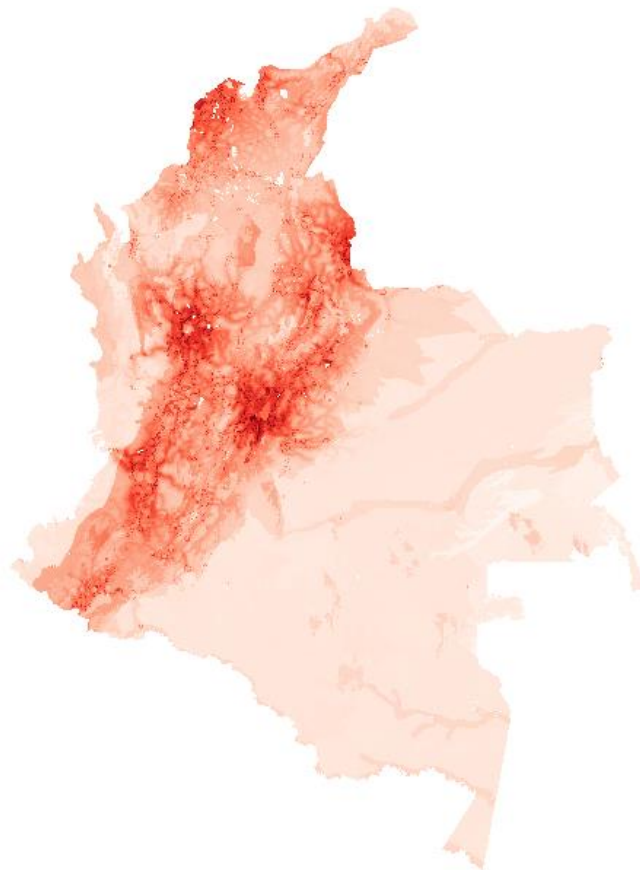


Figure 29 Example of a probability map for cropland in Colombia. Areas with dark red present areas with high probability, and bright colors are areas with low probability.

4.11. Incorporating multiple regions

The Dyna-CLUE model has the possibility to include more than one region. This might be useful for study areas consisting of two or more regions that react differently regarding land use change. Regionalization can be necessary when different driving factors or specific demands for land use are found in different parts of the study area. Examples include the subdivision of a country in different regions with distinct biogeography (e.g. coastal flats vs. mountain regions). Also on a smaller scale it might be useful to divide an area in two or more regions. For example, an area with two different ethnicities that both have a different type of land use management, e.g., livestock vs. cropland. In this case, different driving factors are important for land use allocation and it is wise to prepare a specific regression for each of these regions. In case of a multi-country study, it might be wise to specify demands on a country-by-country basis since the production of agricultural goods will most likely be based on national demands and export/import conditions.

To include multiple regions, the following changes have to be made in the model input files:

- (1) Area restriction file (*region*. **): a new file needs to be created with a unique value code for each region (beginning with 0 and increasing consecutively). An example of the region file for Colombia and 5 of its natural sub-regions is provided in Figure 30.
- (2) Main parameters file (*main.1*): line 2 should be updated to indicate the number of regions. Line 16 should also be updated to define the regressions and demands conditions for the different regions. A value of 1 should be indicated whereby multiple demands are used with equal or different regressions across regions, while a value of 2 is indicated whereby different regressions but aggregate demands are provided for the regions.
- (3) Regression file (*alloc1.reg*): if the regressions are different for different regions, these are simply listed consecutively for each of the regions within the same file and must be listed in the order of the coding provided in the area restriction file. If the regressions are the same, the same regression parameters are listed consecutively for each of the regions within the same file, thus making sure that the same regression is used in all regions. An example of such a regional regression file is provided in the data package.
- (4) Demand file (*demand.in**): when different demands are specified for different regions, the demands are also listed consecutively within the same file and in the same order of the coding provided in the area restriction file. The number of years at the first line has only to be indicated once. One should be sure that the sum of the demand does not change and that the demand for the first year coincides with the initial land use map (*cov_all.0*) for each region. In the case of aggregate demands, the *demand.in** file only specifies one set of demands which are to be applied across the whole case study area (i.e. these demands represent the sum of the areas of the individual regions). We provide both examples in the data package.

Regionalization is also possible for other optional components of the Dyna-CLUE model. One example is for the establishment of regionalized neighborhood rules (optional *alloc2.reg* file) via which neighboring land uses are also considered as driving factors of land use conversions. We invite you to read the CLUE help documentation for detailed explanations of additional functionalities.

Note: Different land uses might be present across the different regions. When running a

simulation where different demands are specified for different regions, the *demand.in** file should be prepared so that (for each region) a column is provided following the complete land use coding of the *cov_all.0* file and *main.1* file (line 10). If a specific land use is present at the broader scale but missing for a specific region, a value of 0 is simply provided for the first year of this region's land use demands, representing the starting configuration for the region. As a result of this, the iteration variable (*main.1* file, line 12) should be set to provide absolute rather than percentage values.

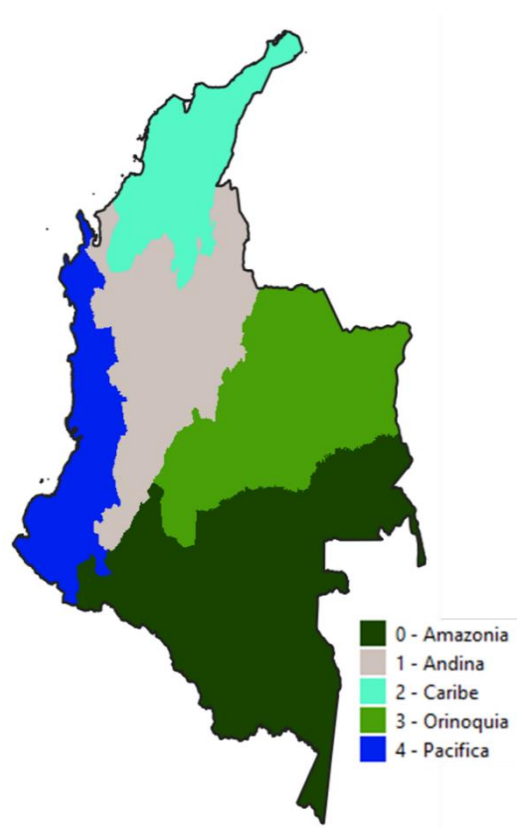


Figure 30 Example of the region file for a regionalized application of Colombia referring to 5 of its natural sub-regions. The required coding is illustrated, starting with 0 and increasing by one with each additional region

4.12. Running the model

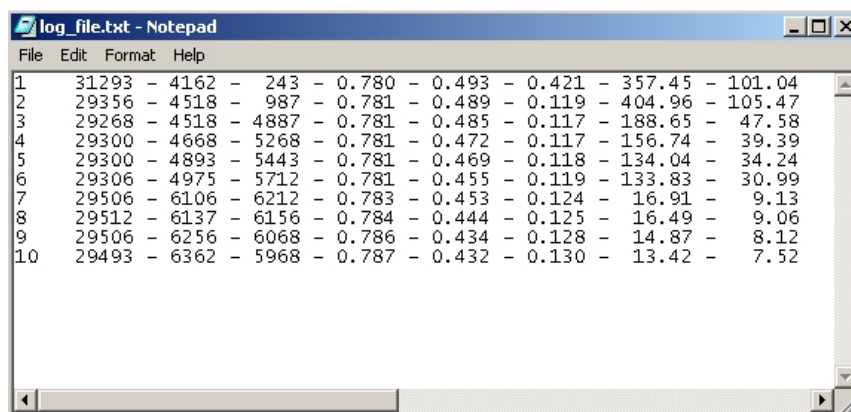
The model can be run by selecting the desired region file and demand scenario and by simply pressing the 'Run Dyna-CLUE' button. The 'Years' bar will indicate the progress while the 'Iteration' bar indicates the iteration count at a logarithmic scale. This bar shows the relative difference between the allocated area of the different land use types and the required allocation of the different land use types. The simulation of one year is finished if the allocated area deviates less than the specified maximum allowed deviation (line 12 of the main parameters). After the simulation a 'View LOG-file' button will be displayed to show the iteration parameters. This is only useful when problems during the simulation have occurred. The log file is stored in the installation directory and can also be opened by any text editor.

The log file (*log.fil*) contains in addition to general run-time information, like the regression

parameters, conversion matrix and demand, also some iteration parameters. These are listed for each step made, respectively:

1. Iteration steps
2. Allocated area for land use types
3. Iteration parameters for land use types
4. Maximal deviation between demand and allocated land use (% or absolute area)
5. Average deviation between demand and allocated land use (% or absolute area)

Figure 31 shows an example of the iteration parameters from a log file for 10 iteration steps and 3 land use types. The first column shows iteration steps, the next three columns show the allocated area for each land use type, the next three columns show the iteration parameters for each land use type, the penultimate column shows the maximal deviation between demand and allocated land use and the last column the average deviation between demand and allocated land use. With more land use types more columns will be added for point 2 and 3. The iteration will continue until the convergence criteria (line 12 of the main parameter file) have been reached, e.g. the value of the penultimate column has to become lower than 3% and the value of the last column should be lower than 0.35%. If the convergence criteria are not reached within 20,000 iterations, Dyna-CLUE terminates without a solution having been found for the desired demands.



Iteration Step	Land Use Type 1 Area	Land Use Type 2 Area	Land Use Type 3 Area	Land Use Type 1 Param	Land Use Type 2 Param	Land Use Type 3 Param	Max Deviation	Avg Deviation
1	31293	4162	243	0.780	0.493	0.421	357.45	101.04
2	29356	4518	987	0.781	0.489	0.119	404.96	105.47
3	29268	4518	4887	0.781	0.485	0.117	188.65	47.58
4	29300	4668	5268	0.781	0.472	0.117	156.74	39.39
5	29300	4893	5443	0.781	0.469	0.118	134.04	34.24
6	29306	4975	5712	0.781	0.455	0.119	133.83	30.99
7	29506	6106	6212	0.783	0.453	0.124	16.91	9.13
8	29512	6137	6156	0.784	0.444	0.125	16.49	9.06
9	29506	6256	6068	0.786	0.434	0.128	14.87	8.12
10	29493	6362	5968	0.787	0.432	0.130	13.42	7.52

Figure 31 Example of iteration parameters from a log file

4.13. Running the model through batch mode

The Dyna-CLUE model can be run through the MS-Dos command prompt. To do this, go to the Dyna-CLUE installation directory and give the name of the model (clues.exe), the demand file and the region file. After pressing enter the model starts. In Figure 32 an example of the command line is given. An example use of the batch mode is when working with a larger number of independent regions. While Dyna-CLUE does enable allocating land use in different regions in the study area, we do not recommend using this function in the following cases:

- Running Dyna-CLUE in a considerable large number of regions (e.g. 30)
- Working in different independent study regions (e.g. neighboring countries) at the same time
- Working in regions with different demands, different land cover maps and different spatial factors

In such cases, it is best to prepare individual applications for each region, and run the model through the batch mode.

The following arguments can be given when the model is run in batch mode:

clues <land requirements file> <area restrictions file> {mode}

The {mode} argument is optional.

For {mode} the following arguments are valid:

no

specification:

normal run

1: meta model run

2: all probability maps are prepared (preparation for meta-model run) 3: calculate probability maps for year 1

If during a batch run the model encounters an error a file called error*.txt where * indicates the time of the error will be stored which contains the time and directory of the error. In the log.fil file the error can be studied in more detail. During normal simulation such a file will not be displayed and an error box will appear on the screen.

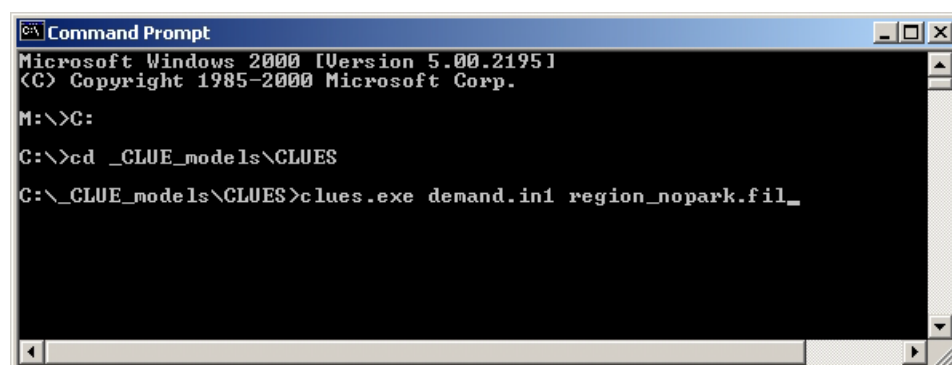


Figure 32 Command-line function to run Dyna-CLUE through batch mode.

5. Analyzing the results

5.1. Output file description

All output files are saved in the same directory where the Dyna-CLUE model is installed. Besides the log file two other types of files are created during the simulation. For each year a file *cov_all.** is created, where * stands for the year of the simulation. It contains the resulting land use type distribution after the simulation for that year. This is an ASCII file that can be imported by a GIS package.

The other file is called *age.**, where * stands for the year of the simulation. This file indicates for each grid cell the number of time-steps after the last land use change at that location. At the start of the simulation the land use history can be set randomly or specified in file *age.0*. In the main parameter file (line 17) it is possible to choose between these options. Also, this file is an ASCII file that can be imported by a GIS package. Every time the model is run, any previously present *age.** and *cov_all.** files in the directory will be over-written by the results of the new simulation. Opening the *age.** file in the end might be useful in cases, when it is necessary to identify how long ago land use changes have occurred before the final year. For example, one might be interested in when a forest was deforested, or how long has a location been intensively cultivated.

6. Problems

6.1. File checker

After checking the 'Check input' option the model will check for potential errors in the driving factor and initial land use files. Errors are listed in the log file.

6.2. Frequently asked questions / errors

Error messages

Formal 'Error' messages

Most common errors that are identified by a window stating: "Error, for more information see 'log.fil'" are errors for which some information on the cause of the error is included in the 'log.fil'. Open the log.fil file in the model directory and try to solve this problem. In case the error reads "Maximum dimensions exceeded" the model configuration does not support the size/configuration of your study area.

Frequent error messages not documented in the 'log.fil'

1. 'exp: OVERFLOW error' followed by 'Invalid Floating Point Operation'

error description:

If this error occurs directly after loading, both during a normal run and during the calculation of probability maps it is most likely that there is either:

- an error in the regression coefficients entered in the 'alloc1.reg' file or:
- an error in one of the files with driving factors (sc1gr#.fil)



Solutions:

- check the 'alloc.reg' file carefully
- check the sc1gr#.fil by:
 - try to import the files in a GIS: does the file import correctly? If not, correct the file
 - make sure that for all pixels for which a value > -9999 is defined in the 'region#.fil' a value is defined, i.e., for pixels for which a simulation is made it is not allowed to have 'No Data' as one of the pixel values in the 'sc1gr#.fil' files. An easy

check is to use a map calculator to calculate the sum of all driving factor grids and check if the result has a value for all cells with a value in the 'region#.fil' grid. Correct the files and try again.

- if you cannot find the error, replace the regressions in the 'alloc.reg' file with very simple regressions that only include one driver with the same beta value. Repeat this until you find a run that does not work. Most likely this identifies the grid/file with some error in one of the data values. Make the corrections identified and run it again.

2. No solution found

After 20000 iterations the model will assume that it is not possible to achieve convergence of the iteration and will quit.

Possible causes may be:

- the demand does not correspond with the available area. Check if the demand for all land uses together corresponds, for each year, with the count of all files that are assigned to a region multiplied by the cellsize. Also check the units. Units of demand should correspond with units of cell area specified in the main.1 file.
- the demanded changes cannot be allocated due to strict restrictions, e.g., demanding a lot of deforestation that exceeds the amount of forest outside the restricted areas.
- contradictions in the conversion matrix. Not enough options for conversion remain, e.g., LUT2 may only replace LUT1 but the demand for LUT2 exceeds the amount of LUT1 corrected for the demand for LUT1.
- not sufficient cells of the minimum 'age' are available given the settings in the conversion matrix. This can be resolved by changing the initialization of the age.

References

- Menard, S. 2002. Applied logistic regression analysis, second edition. Series: Quantitative applications in the social sciences. London, Sage Publications.
- Pontius, R.G. and Schneider, L.C. 2001. Land use change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems and Environment*, 85: 239-248.
- Veldkamp, A. and Fresco, L.O. 1996. CLUE-CR: an integrated multi-scale model to simulate land use change scenarios in Costa Rica. *Ecological modelling*, 91: 231-248.
- Verburg, P.H. and Overmars, K.P., 2009. Combining top-down and bottom-up dynamics in land use modeling: exploring the future of abandoned farmlands in Europe with the Dyna-CLUE model. *Landscape ecology*, 24(9), p.1167.
- Verburg, P.H., Veldkamp, A., de Koning, G.H.J., Kok, K. and Bouma, J. 1999. A spatial explicit allocation procedure for modelling the pattern of land use change based upon actual land use. *Ecological modelling*, 116: 45-61.
- Verburg, P.H., Soepboer, W., Veldkamp, A., Limpiada, R. and Espaldon, V. 2002. Modeling the spatial dynamics of regional land use: the CLUE-S model. *Environmental Management*, 30(3): 391-405.
- Verburg, P.H. and Veldkamp, A. 2003. Projecting land use transitions at forest fringes in the Philippines at two spatial scales. *Landscape Ecology*, 19(1): 77-98.
- Verburg, P.H., T. de Nijs, J. Ritsema van Eck, H. Visser, K. de Jong. 2004. A method to analyse neighbourhood characteristics of land use patterns. *Computers, Environment and Urban Systems*, 28(6): 667-690.

Appendices

Appendix 1 – Location factor data

All files are provided in the data package. If you wish to use the location data files for your own research purposes, please check the respective data use requirements and eventually contact the relevant corresponding authors for authorization.

Variable name	Description	Unit	File name	Source
Population density	Distribution of human population	People/km ²	Sc1gr0	CIESIN and SEDAC 2015
Rural population	Distribution of rural population	People/km ²	Sc1gr1	CIESIN, IFPRI, and CIAT 2011
Market access	Indicator for the accessibility to markets	Index	Sc1gr2	Verburg, Ellis, and Letourneau 2011
Organic content	Organic carbon content in the top 50 cm of soil	g /kg of soil	Sc1gr3	Stoorvogel et al. 2016
Soil drainage	Internal drainage of soils	class	Sc1gr4	ISRIC 2018
Clay soil	Share of clay in soil	%	Sc1gr5	Stoorvogel et al. 2016
Cation Exchange Capacity (CEC)	Proxy for nutrient retention capacity	cmol/kg	Sc1gr6	ISRIC 2018
Soil depth	Soil depth	cm	Sc1gr7	Stoorvogel et al. 2016
Sand	Share of sand in soil	%	Sc1gr8	Stoorvogel et al. 2016
Soil pH	pH index measured in water solution	1-7	Sc1gr9	ISRIC 2018
Elevation	Elevation above sea level	m	Sc1gr10	Provided by IADB
Slope	Derived from altitude	Slope degrees	Sc1gr11	Derived from elevation
Precipitation	Annual precipitation	mm	Sc1gr12	Hijmans et al. 2005
Temperature	Average temperature (mean of monthly means)	°C	Sc1gr13	Hijmans et al. 2005
PET	Annual evapotranspiration (combination of water content lost through soil evaporation and transpiration)	mm	Sc1gr14	Trabucco and Zomer 2009
Distance to roads	Distance to roads	m	Sc1gr15	Meijer et al. 2018
Distance to cities	Travel time to cities	m	Sc1gr16	Weiss et al. 2018

CIESIN, IFPRI, and CIAT. 2011. "Global Rural-Urban Mapping Project, Version 1: Population Density Grid. Center for International Earth Science Information Network - CIESIN - Columbia University, International Food Policy Research Institute - IFPRI, The World Bank, and Centro Internacional de Agricultura Tropical - CIAT." <https://doi.org/10.7927/H4R20Z93>.

CIESIN, and SEDAC. 2015. "Gridded Population of the World, Version 4 (GPWv4): Population Density. Center for International Earth Science Information Network - CIESIN - Columbia University. NASA Socioeconomic Data and Applications Center (SEDAC)." <https://doi.org/10.7927/H4DZ068D>.

Hijmans, Robert J., Susan E. Cameron, Juan L. Parra, Peter G. Jones, and Andy Jarvis. 2005. "Very High Resolution Interpolated Climate Surfaces for Global Land Areas." *International Journal of Climatology* 25 (15): 1965–78. <https://doi.org/10.1002/joc.1276>.

ISRIC. 2018. "ISRIC SoilGrids1km Visualisation and Distribution Website." SoilGrids1km. 2018. <http://soilgrids1km.isric.org/>.

Meijer, J. R., Mark A. J. Huijbregts, Kees C. G. J. Schotten, and Aafke M. Schipper. 2018. Global patterns of current and future road infrastructure. *Environmental Research Letters*, 13, 6, 164006. <https://doi.org/10.1088/1748-9326/aabd42>

Stoorvogel, Jetse J., Michel Bakkenes, Arnaud J. A. M. Temme, Niels H. Batjes, and Ben ten Brink. 2016. "S-World: A Global Soil Map for Environmental Modelling." *Land Degradation & Development*, 1–12. <https://doi.org/10.1002/ldr.2656>.

Trabucco, A., and R.J. Zomer. 2009. "Global Aridity Index (Global-Aridity) and Global Potential Evapo-Transpiration (Global-PET) Geospatial Database." CGIAR Consortium for Spatial Information.

Verburg, Peter H, Erle C Ellis, and Aurelien Letourneau. 2011. "A Global Assessment of Market Accessibility and Market Influence for Global Environmental Change Studies." *Environmental Research Letters* 6 (3): 034019. <https://doi.org/10.1088/1748-9326/6/3/034019>.

Weiss, D. J., A. Nelson, H.S. Gibson, W. Temperley, S. Peedell, A. Lieber, M. Hancher, E. Poyart, S. Belchior, N. Fullman, B. Mappin, U. Dalrymple, J. Rozier, T.C.D. Lucas, R.E. Howes, L.S. Tusting, S.Y. Kang, E. Cameron, D. Bisanzio, K.E. Battle, S. Bhatt, P.W. Gething. 2018. A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, 553, 333-336, 10.1038/nature25181

Appendix 2 – Data suggestions

This section presents the most useful and suitable data examples used in CLUE. Land use and land cover data, together with other data used to map land systems in any region or globally, are described in the subsection 'Land systems mapping', and data for setting up a new application in the subsection of the same name.

Land cover in general

Used to identify the composition of the landscape in terms of land use/land cover. Data can be used to derive specific systems, that can be defined by dominant land cover, such as deserts, urban areas (where data on urban areas is poor). Usually, national data is better (in terms of spatial resolution and mostly being more recent), however in some cases it might not be available.

Description	Original resolution	Spatial and temporal coverage	Unit	Source
Global Land Cover SHARE	1 km	Global, 2014	%	Latham et al. (2014), http://www.glcn.org/databases/lc_glcshare_en.jsp
Land cover 30	30 m	Global, 2000 and 2010	class	Jun et al. (2014), http://www.globallandcover.com/GLC30Download/index.aspx
CORINE land cover	depends, from 100 m	Europe; 1990, 2000, 2006, 2012	class	European Environment Agency, http://land.copernicus.eu/pan-european/corine-land-cover/view
GlobCover	300 m	Global, 2005, 2009	class	ESA (2011), http://dup.esrin.esa.int/page_globcover.php
Land cover CCI	300 m	Global, 2000, 2005, 2010	class	ESA (2016), http://maps.elie.ucl.ac.be/CCI/viewer/

Forests

Used to identify woodlands and forests, and the intensity of forest management (if available - usually scarce data on global scale).

Description	Original resolution	Spatial and temporal coverage	Unit	Source
Tree cover, gains and losses	30 m	Global, 2000-2014	%	Hansen et al. (2013), https://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.2.html
Tree cover, vegetation continuous fields (MODIS)	from 250 m	Global, 2010	%	Di Micelli et al. (2011), http://glcf.umd.edu/data/vcf/
Global percent tree cover	500 m	Global	%	ISCGM et al. (2008), http://www.iscgm.org/gm/ptc.html
Global forest classes and management types	1 km	Global, 2000	forest type and use	Schulze, Malek and Verburg (2019), http://www.environmentalgeography.nl/site/data-models/data/forest-classes-and-uses
Global forest canopy height map	1 km	Global, 2005	canopy height	Giglio et al. (2013), http://lidarradar.jpl.nasa.gov
Annual burned area	27 km	Global, from 1997 onwards	% of burned area	Giglio et al. (2013), http://www.globalfiredata.org/data.html
Net Primary production	28 km	Global, 1995 (v1)	amount of NPP	Imhoff et al. (2004), http://sedac.ciesin.columbia.edu/data/set/hanpp-net-primary-productivity

Urban areas

To map the extent of settlements, or the extent of urban activities.

Description	Original resolution	Spatial and temporal coverage	Unit	Source
Impervious areas	25 m	Europe;2002, 2009, 2012	%	European Environment Agency, http://land.copernicus.eu/pan-european/high-resolution-layers/imperviousness/view
Global urban extent map	from 500 m	Global,2001-2002	%	Schneider et al. (2009), https://nelson.wisc.edu/sage/data-and-models/schneider.php
Global building density	cca. 30x30km (hexagonal grid)	global, 2018	nr of buildings	Silva et al. (2018). Global Earthquake Model (GEM) Exposure Map https://maps.openquake.org/map/global-exposure-map/

Cropland and livestock

Data on the extent of croplands, type and yields of crops at a specific location, presence of irrigation and livestock density.

Description	Original resolution	Spatial and temporal coverage	Unit	Source
IIASA FIPRI Cropland extent	1 km	Global, 2014	%	Fritz et al. (2015), http://geo-wiki.org/downloads/
Global field size map	1 km	Global, 2014 and 2018	class	Fritz et al. (2015), Lesiv et al. (2018), http://geo-wiki.org/downloads/
European	1 km	Europe, 2010	size	Kuemmerle et al. (2013), Emma van der Zanden has the data

Description	Original resolution	Spatial and temporal coverage	Unit	Source
cropland field size map				
Global distribution of crops, harvest data, yield gap	10 km	Global, 2000	various	Ramankutty et al. (2008), http://www.earthstat.org/data-download/
Fertilizer application	10 km	Global, 2000	fertilizer application	Mueller et al. (2012), http://www.earthstat.org/data-download/
Fertilizer intensity	1 km	EU27, 2000	fertilizer application	Temme and Verburg (2011)
Global map of irrigated areas	10 km, some parts 1 km	Global, 2006	ha, or %	Siebert et al.(2013), http://www.fao.org/nr/water/aquastat/irrigationmap/index.stm
European crop type map	vector (based on CORINE)	EU27, 2006	% or ha	Britz and Witzke (2014), http://www.capri-model.org/dokuwiki/doku.php?id=capri:faq_e
SPAM Crop extent and crop production systems	10 km	Global, 2014	ha, tons, %, intensity	You et al. (2014), http://mapspam.info/data/
Livestock density (bovines, goats, sheep, pigs, poultry)	1 km	Global, 2014	nr/km2	Robinson et al.(2014), http://www.livestock.geo-wiki.org/home-2/

Nature

Data used to identify areas under protection, natural and semi-natural areas, and future priorities for protected areas.

Description	Original resolution	Spatial and temporal coverage	Unit	Source
Wetlands and lakes	250 m	Global, 2004	class	WWF (2004), http://www.worldwildlife.org/publications/global-lakes-and-wetlands-database-lakes-and-wetlands-grid-level-3
Terrestrial ecoregions	vector	Global, 2001	class	Olson et al. (2001), http://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world
Protected areas	vector	Global, from 2001 (updated)	class	IUCN (2015), https://www.protectedplanet.net/
Priorities for protected areas	1-10 km	Global, National, 2000	% of priority	Pouzols et al. (2014), http://avaa.tdata.fi/web/cbig/gpan/data-download

Socio-economic

Socio-economic data helps to characterise the type of activities expected in an area, as well as the degree of human impact.

Type	Unit/Description	Original resolution	Spatial and temporal coverage	Source
Population density	People/km ²	1 km	Global, 2010	CIESIN (2015), http://sedac.ciesin.columbia.edu/data/collection/gpw-v4
Rural	Rural population/km ²	1 km	Global, 2000	CIESIN (2011), http://sedac.ciesin.columbia.edu/data/set/grump-v1-

Type	Unit/Description	Original resolution	Spatial and temporal coverage	Source
population				population-density
Population	different data on population	varies	global	https://www.popgrid.org/
Markets	Market accessibility and influence, Index or USD/person	1 km	Global,2000	Verburg et al. (2011).
Accessibility	Distance to roads in m	depends	local to global	depends on which road map is used, most recent GRID data, available on http://www.globio.info/download-grip-dataset
WorldPop 100	High-resolution population	100m and 1km	Global	doi: 10.1038/sdata.2017.1 (2017), http://www.worldpop.org.uk/data/
Accessibility 2015	Distance to cities larger than 50000	1km, 2015	global	doi: 10.1038/nature25181 (2018), https://map.ox.ac.uk/research-project/accessibility_to_cities/
Travel time	Friction map of travel time to cities	1km, 2015	global	doi: 10.1038/nature25181 (2018), https://map.ox.ac.uk/research-project/accessibility_to_cities/

Climate

Climate data is especially useful when dealing with agricultural systems, as climatic variables limit (or encourage) vegetation growth. Climate data can/should be dynamic in your CLUMondo application, using the projections on different Representative Concentration Pathways (RCP). Some downscaled data available at worldclim.org (see below).

Type	Unit/Description	Original resolution	Spatial and temporal coverage	Source
Precipitation	mm	from 1 km	Global, 2000-2100	Hijmans et al. (2005), www.worldclim.org
Temperature	degree Celsius	from 1 km	Global, 2000-2100	Hijmans et al. (2005), www.worldclim.org
Solar radiation	Horizontal surface irradiation (kWh/m ²)	1.5 arc minute	Global, some only regional (EU, Africa)	Zomer et al. (2008) - global: http://www.cgiar-csi.org/data/global-aridity-and-pet-database ; Huld et al. (2012) - Europe, Africa: http://re.jrc.ec.europa.eu/pvgis/
Potential Evapotranspiration	mm	from 1 km	Global, 2007	Zomer et al. (2008), http://www.cgiar-csi.org/data/global-aridity-and-pet-database
Aridity index	Index	from 1 km	Global, 2007	Zomer et al. (2008), http://www.cgiar-csi.org/data/global-aridity-and-pet-database
Climate and water balance	temperature, precipitation, vapor pressure, wind speed, solar radiation, evapotranspiration, water runoff, soil moisture	4 km	Global, 1958-2015	Abatzoglou et al. (2018), https://www.nature.com/articles/sdata2017191

Soil

Soil characteristics describe the suitability for growing crops, encouraging and constraining agricultural activities.

Type	Unit/Description	Original resolution	Spatial and temporal coverage	Source
Soil drainage	Drainage class	1 km	Global, 2010	Hengl et al. (2014), https://soilgrids.org/ ,
Sand content	Sand mass in %	1 km	2010	Stoorvogel et al. (2016), need to contact authors
Clay content	Clay mass in %	1 km	2010	Stoorvogel et al. (2016), need to contact authors
Cation Exchange Capacity	cmol/kg	1 km	2010	Hengl et al. (2014), https://soilgrids.org/ ,
pH	log(h+)	1 km	2010	Hengl et al. (2014), https://soilgrids.org/ ,
Organic carbon content	g of organic content in kg of soil	1 km	2010	Stoorvogel et al. (2016), need to contact authors
Soil depth	cm	1 km	2010	Stoorvogel et al. (2016), need to contact authors
Soil water balance	Soil evapotranspiration and water deficit	1 km	2000	Trabucco and Zomer (2010), https://cgiarcsi.community/data/global-high-resolution-soil-water-balance/

Physical

Data on terrain and derived characteristics. Usually, national institutions have better data on terrain, particularly when it come to the spatial resolution.

Type	Unit/Description	Original resolution	Spatial and temporal coverage	Source
Altitude	Elevation in m above sea level	1 km	Global, 2005	Hijmans et al. (2005), www.worldclim.org
Slope	Slope degrees or % (whatever you derive)	1 km	Global, 2005	derived in GIS from Hijmans et al. (2005), www.worldclim.org

Crop production

Data on existing and future crop production and efficiency.

Type	Unit/Description	Original resolution	Spatial and temporal coverage	Source
Growing period	Climate (temperature and precipitation) and the length of growing period	10 km	Global, 2012	http://www.fao.org/nr/gaez/en/
Crop production capacity	Current efficiency	10 km	Global, 2012	http://www.fao.org/nr/gaez/en/
cropland suitability	locations that can host cropland	10 km	Global, 2000	Either GAEZ: http://www.fao.org/nr/gaez/en/

Other

Other data that might help to describe the distribution of potential land use and land cover.

Type	Unit/Description	Original resolution	Spatial and temporal coverage	Source
potential vegetation	Class	10 km	Global, 2000	Ellis and Ramankutty (2008), http://ecotope.org/anthromes/data/