Virtual Neanderthals : a study in agent-based modelling Late Pleistocene hominins in western Europe

Scherjon, F.

PART TWO: CREATING THE ACTORS

“Van Helmont, in the midst of odd experiments to obtain mice from junk and sawdust, made one rather intelligent experiment. He grew a willow twig weighing 5 lb in a large clay pot containing 300 lb of soil and irrigated it with rainwater. After 5 years, he harvested a willow tree of 164 lb, with a loss of only 2 oz of soil. Van Helmont concluded from this that water was condensed to form plants.”

(Lieth 1973, 305)

4. THE MODELLED ENVIRONMENT

4.1 Introduction

HomininSpace models the environment of the Late Pleistocene for part of western Europe for simulations with modelled hominins. The topography is reconstructed for a delimited area using current day topographical and bathymetry data taking a fluctuating sea level into account (Section 4.2). HomininSpace further uses values for past temperatures and precipitation levels (described in Section 4.3) in order to compute how much energy there is in the landscape available for hominins who sustain themselves mostly with the consumption of large ungulates (Section 4.4). Ultimately this gives a carrying capacity for every grid cell for every time step in a simulation that will guide mode and speed of hominin dispersal.

Within the HomininSpace simulation system climate and topography are modelled to study the effect that temperature and precipitation can have on the behaviour of Late Pleistocene hominins in the landscape (Banks, d'Errico, Peterson, Kageyama, *et al.* 2008; cf. Eriksson *et al.* 2012). It is suggested that these are probably the two most important environmental variables that influenced changes in the abundance of fauna and flora through time (Kelly 1983, 279; Krohne 2001, 47-51; Monserud and Leemans 1992; Thackeray 2013). These variations drive availability and prevalence of food resources and were among the most important factors determining whether hominins could survive and thrive in any particular area (Hamilton *et al.* 2007; Freeman and Anderies 2015).

It is however notoriously difficult to reconstruct Pleistocene environments and to isolate individual parameters that would have attracted or deterred hominins, if only because ancient hominins lived in a wide diversity of landscapes and were more or less successful in each one of them (Roebroeks *et al.* 1992; Zuk 2013). Using only a few variables to model biomes as different as those of the British Isles, the western European mainland, the Mediterranean coast or the diverse mountainous areas of the Pyrenees and Alps invariably leads to simplification. HomininSpace makes use of the fact that all these environments have in common that they can exist only by virtue of both the influx of energy and the amount of water supplied to each square meter (Guthrie 1990, 214), which are in turn influenced by topography (Boivin *et al.* 2013, 38).

Hominins, like all animals, fulfil their energy requirements by extracting it from the landscape around them. Energy in general is provided by the sun delivered onto the earth in the form of heat and light, and converted into primary biomass using water and nutrients from soil and air. This process is referred to as primary production and is mainly done by plants and marine algae (some bacteria also contribute). All energy in the environment is ultimately generated by the available sunlight[13]. The amount of solar energy is described by the Milankovitch cycles, the orbital forcing which is the combined effects of obliquity (axis tilt), eccentricity of the orbit of the earth around the sun and the precession of the orbit (Hays *et al.* 1976). This fluctuating solar radiation energy or insolation is further subject to atmospheric losses due to scattering and absorption and latitudinal differences between absorbed and emitted radiation and therefore varies throughout the days and months.

Insolation is the most important source of energy in primary productivity, the production of organic materials from atmospheric carbon dioxide. The produced vegetation forms the base of every food chain. The most important factor defining the net production is solar energy but nutrient availability, water management in the area and local temperature also play a part (Hawkins *et al.* 2003). Primary production decreases with latitude (cf. Damuth 1991; Silva *et al.* 2001) and the amount of evaporation and variation in geography cause most deviation from expected solar radiation production (Binford 2001, 82).

Herbivores or primary consumers directly consume plant material and carnivores, situated at higher levels of the food chain, consume other animals. This accumulation of living secondary biomass by heterotrophs is referred to as secondary production (Allaby 2005; Benke and Huryn 2006). There is a direct relation between terrestrial primary productivity and the availability of ungulates (Coe *et al.* 1976; Janis *et al.* 2000; Oesterheld *et al.* 1992), which are the prime targets for Pleistocene hominins. The productivity hypothesis states that the primary production will limit the secondary production of herbivores, and the amount of herbivores (secondary biomass or SB) will limit the number of predators (Hawkins *et al.* 2003, 3106).

Hominins can sustain themselves with both plant materials and meat. The archaeological record is so far not sufficiently detailed to allow determination of exact proportions (Kaplan *et al.* 2000, 180). The underlying model in this thesis however assumes that

[13] The contribution of internal heat from the earth's core is relatively low (less than one percent) and mainly visible in the movement of the continents.

hominins primarily feed on large terrestrial game to satisfy their energy needs (Salazar-García *et al.* 2013; Snodgrass and Leonard (2009) but see Power *et al.* (2018)), as summarized in the schema below.

Primary Production ==➔	*Secondary Production* ==➔	*Neanderthals*
(Vegetation)	(Herbivores)	(Predators)
Limited by solar energy and water availability.	Limited by Primary Production.	Limited by Secondary Production.

The amount of energy available as herbivores is often predicted using empirical data from different climates and vegetation types to infer a formula with one or more climatic parameters. Examples include Coe *et al.* (1976) who are modelling large herbivores on the east African grasslands using rainfall, or Eisenberg and McKay (1974) who also used rainfall to infer the amount of herbivores in tropical forests. To model secondary biomass in different biomes a single climatic parameter is however generally not sufficient (Binford 2001, 106). HomininSpace uses precipitation as well as temperature and implements two different model variants to calculate the available energy in the landscape as the availability of large edible ungulates as food for hominins in the area of western Europe during the Late Pleistocene (Section 4.4).

4.2 Reconstructing topography with fluctuating global sea levels

Paleobathymetric variation can influence early hominin dispersals by exposing or hindering likely routes (Nakazawa and Bae 2017). The model underlying HomininSpace uses reconstructed sea levels which are compared to modern day topography data for land masses and upon current day bathymetry (bed rock data) for marine areas. This latter is rationalized by the fact that the bed-rock level of the sea bottom represents the maximum depth where erosion and sedimentation of sand layers constantly change the local conditions. The model ignores further the effects of suppression of land surfaces by glaciers, rebounce when the ice retreats (glacio-hydro-isostatic contributions to local sea-levels (Lambeck and Chappell 2001)), and the formation of natural sea defences like dunes and swamp marshes that might influence or hinder (local) sea level rise or fall.

The simulation area is divided into grid cells with the Spatial Analysis in Macroecology tool (SAM) (Rangel *et al.* 2006; Rangel *et al.* 2010). The area used in the Neanderthal case study is delimited in the North by latitude 51.5, in the South by 41.3, to the West the limit is longitude -6.3 and to the East the edge is defined at longitude 8.5 (see Figure 5). This area includes the South of England but excludes Sardinia and little of Spain is included. The area encompasses all of the Pyrenees mountain range and the western part of the Alpine regions.

A grid cell is defined as 0.1 x 0.1 degrees (roughly 10 x 10 kilometers, depending on the latitude) with the area composed of 14.948 grid cells (Figure 5). The surface area of 10 x 10km has been selected after extensive performance testing, and represents a compromise between detail of climate and topographical data and simulation duration. The area of a single grid cell is generally insufficient to support a hominin group. Foraging radii of modern hunter gatherers hardly ever fall below ten kilometer. For the enviromental settings of today, 4394 of the cells in the simulation area are under water. This number reduces considerable when the sea level drops in colder periods and more land emerges. The dimensions of a grid cell can mask the real reasons why Neanderthals were somewhere in the past. They select places to stay but since a modelled group only forages in an area and varying motivations for specific site selection are not implemented (for instance a preference for locations near other-than-food-resources).

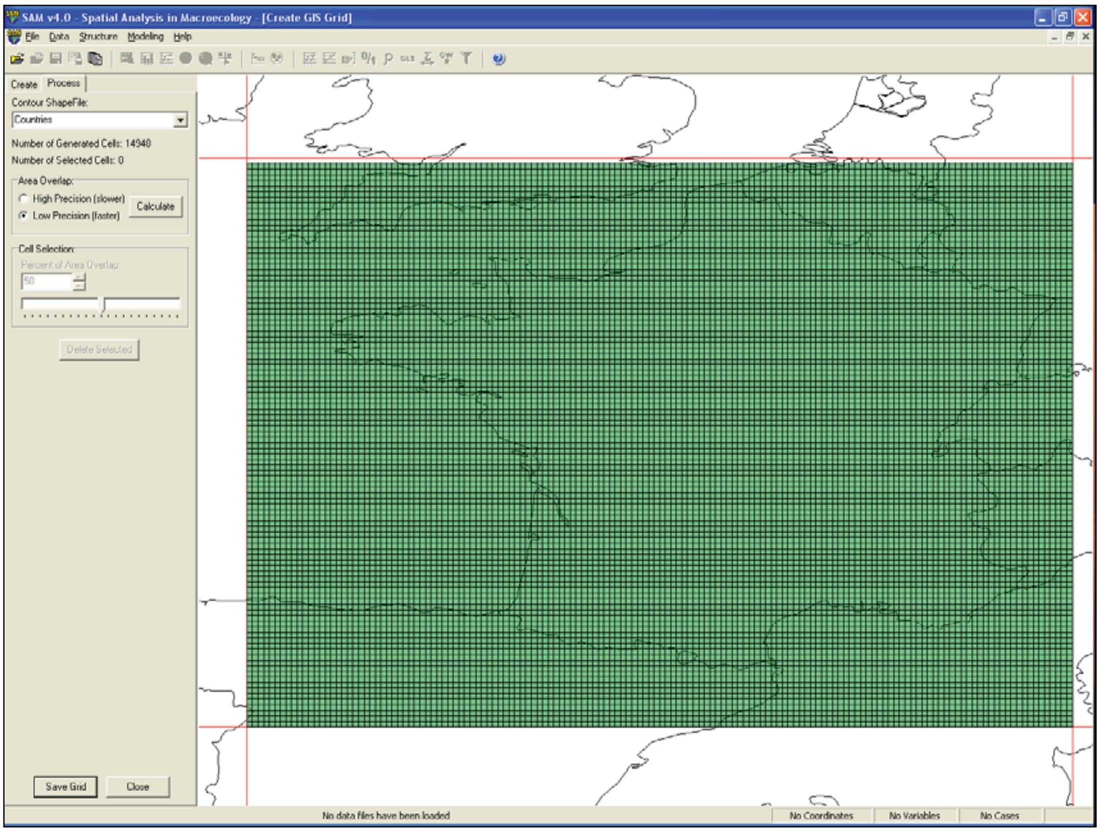


Figure 5: The simulation area divided into 14.948 grid cells in the SAM tool.

SAM is also used to distribute topographical and climate parameters onto these grid cells. Next to the topographical data (altitude or elevation) these parameters include 19 bio-values from http://www.worldclim.org (both present and past climate data). In each grid cell all the datapoints with information (there can be more than one datapoint in a grid cell depending on the underlying dataset) for each parameter are collected, and they are converted into aggregrate values per grid cell: mean, min, max, std, etc. For most variables HomininSpace uses the mean value. However, for bedrock the maximum value is used (see below). When data is missing the values are automatically interpolated from neighbouring cells by SAM.

Altitude information is derived from the modern day dataset downloaded from http://www.worldclim.org (verified 16 April 2013). Unfortunately this dataset does not contain depth information for areas currently located under sea level but which during the simulations can become dry land. The dataset contains only elevation data for land areas. For the sea grid cells bedrock data from a different dataset, the ETOPO1 dataset is used (https://www.ngdc.noaa.gov/mgg/global/global.html, accessed November 2016). Bedrock data details the altitude information for current day submerged landmass. The model assumes that whenever a single bedrock datapoint in a grid cell emerges (because the water level drops below this height), the total grid cell becomes useable land with the altitude of that point. Thus, for the bathymetry data the aggregated maximum is used (which is the minimum depth). Emerged land cells then take missing values for temperature and precipitation extrapolated from nearby coastal areas (the worldclim current day data does not contain values for water cells).

Altitude data in WorldClim.org is derived from the Shuttle Radar Topography Mission (SRTM) dataset, which contains elevation (above sea level) data on a near global scale in very high resolution. The SRTM radar system flew onboard the Space Shuttle Endeavour in February 2000 (http://www2.jpl.nasa.gov/srtm/, address verified 1 October 2013 with links to the original datasets). WorldClim.org provides datasets with aggregrated data in different spatial resolutions. HomininSpace uses the 2.5 arc minutes distribution (5-8 km), for modern-day data and for past climate data sets. This is the highest uniform resolution for which all datasets are available. The data can be downloaded in ZIP file format (compressed data set) with one data layer for the altitude besides the bioclimate variables.

Bathymetry data for parts of the Channel area and Atlantic Ocean and other sea grid cells were obtained from the National Geophysical Data Center (NOAA). Among the data that this organization makes available is the ETOPO1 dataset (Amante and Eakins 2009). Figure 6 presents the data taken from the website (http://maps.ngdc.noaa.gov/viewers/wcs-client/, accessed 13 April 2013). This data is retrieved in XYZ format (geocentric coordinates) (Frakes 2003). SAM can read a matrix of XYZ formatted data (Rangel *et al.* 2010), and because these files are georeferenced it can map the data into the grid cells. Those grid cells that have no altitude value from WorldClim (sea and lake cells) take the altitude from the ETOPO1 dataset. The ETOPO1 dataset compares well against other datasets (see http://vterrain.org/Elevation/Bathy/bathy_comp.html, with more bathymetry data available at http://vterrain.org/Elevation/Bathy/, both accessed 12 April 2013).

Figure 6: Bathymetry data for the simulation area (the red square), taken from the ETOPO1 dataset.

Mean global sea level values are taken from Bintanja and van de Wal (2008) (Figure 7). This data matches rather well the more recent sea level reconstructions made by Grant *et al.* (2012) (Figure 8). The latter argue that ice-core data chronology as used by Bintanja and van de Wal (2008) is not fit for sea level reconstruction since this assumes that the ice volume is in a systematic phase with either Antarctic or Greenland climate, which may not be the case (Grant *et al.* 2012, 744). They connect dated cave speleotherms from the eastern Mediterranean to a high resolution foraminiferal record using the Red Sea basin isolation concept.

However, although Grant *et al.* (2012) claim a high resolution continuous record, their data for the LGM for instance contains only four values from 13,857 to 26,055, a period of 13,000 years, with a minimum value for Red Sea level of -115.74m. Bintanja and van de Wal (2008) provide a continuous record with values every 100 years and a minimum sea level value for 19,800 years ago of -123.41m. They also provide their level as a relative global sea level, where Grant *et al.* (2012) provide a Red Sea relative level. The results from Bintanja and van de Wal (2008) closely match the benthic foraminiferal d^{18}O curve (Waelbroeck *et al.* 2002). For further comparison and an overview of reconstruction methods see Medina-Elizalde (2013).

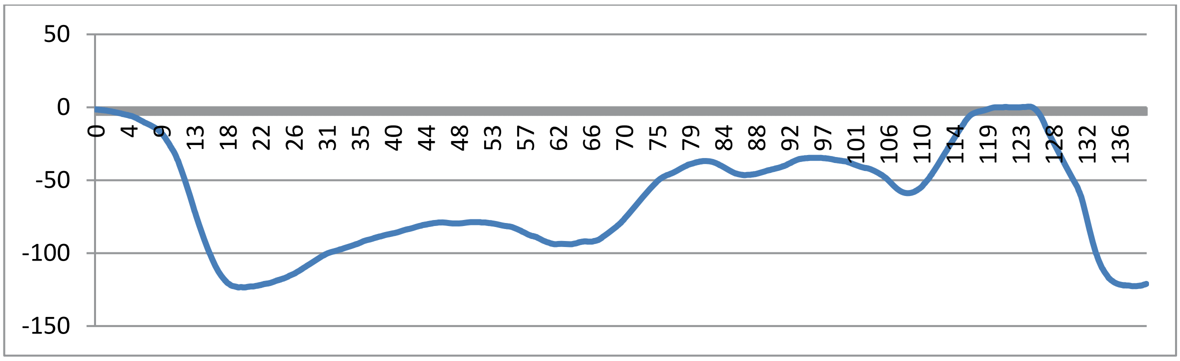


Figure 7: Reconstructed global sea level in meters (Y-axis), from today to 140 ka (from Bintanja and van de Wal (2008)).

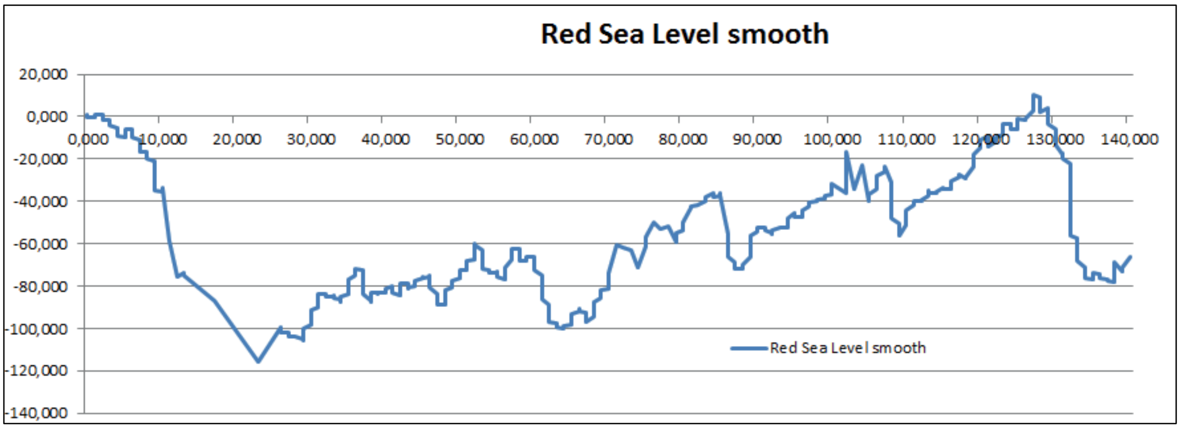


Figure 8: Red Sea level until 140 ka adapted from the supplementary materials from Grant *et al.* (2012) converted to equidistant horizontal axis.

Global sea level values taken from Bintanja and van de Wal (2008) are interpolated between time steps where needed. In each grid cell the height (altitude or depth) is stored where 0 (zero) is the modern day sea level. In each time step the reconstructed value of the sea level is added to all height levels. A resulting value of -10 meters then exposes all land mass with a depth between 0 and 10 meters. A value of +2 meters sea level rise will drown all current day landmasses with an altitude below two meters.

4.3 Distribution of reconstructed temperature and precipitation

The spatial and temporal distribution of the temperature and precipitation in HomininSpace is derived from two different datasets: (1) the current distribution of climatic variables over the world and (2) their distribution as reconstructed in high resolution for the Last Glacial Maximum (LGM, ~21 ka). The LGM is a very cold period with enlarged ice sheets and low atmospheric CO_2 levels. Modern day data is considered to represent a very warm period. The aim for HomininSpace is to have reconstructed environmental data for each time step for the simulation period (roughly 131-50 ka, with each time step representing one year in real time). Such continuous data is not available and expensive to compute. Time series for HomininSpace must be obtained through other means. Local conditions are estimated using a reconstructed global yearly mean temperature record as an index for interpolation between the possible local climatic extremes.

Since past climate variables cannot be measured directly, local paleoclimates are generally reconstructed through interpretation of a diverse range of proxy indicator patterns (Isarin and Bohncke (1999); Lowe and Walker 2014). These include data from lake sediments, tree rings, ice cores and peat deposits. Most of the terrestrial archives are however short and/or discontinuous which is why instead often the longer and more continuous marine data sets are used (Sanchez Goñi *et al.* 2008; Rasmussen *et al.* 2014). Most bioclimatic indicators respond to more than just climate and must be used with care when reconstructing and interpreting past conditions to compare simulation results against (Kageyama *et al.* 2013). Correlation between marine and terrestrial sequences are notoriously difficult but occasionally possible, for instance by using common event markers (Sier *et al.* 2011). The European continental climate records for instance correlate rather well with the global oxygen isotope record from marine environments (Mosbrugger *et al.* 2005).

Temperatures reconstructed from ice volumes are yearly global mean values. Local temperatures (and precipitation values) can differ enormously with strong regional expression, and between locations and continents the yearly means can be far apart (Consortium 2013). As such, these nearly continuous data sets from the stable isotope data analysis are very good *relative* indicators of the temperature fluctuations through time but less suitable for *absolute* local reconstructions. Within the HomininSpace modelling system such a record is therefore used to scale between high resolution local spatial reconstructions for two extreme climatic phases (Lawing and Polly 2011).

To reconstruct the yearly mean temperature Bintanja and van de Wal (2008) use the integrated LR04 marine stack of benthic $\delta^{18}O$ from 57 globally distributed sediment cores (Lisiecki and Raymo 2005). They quantify and separate the two main components affecting the $\delta^{18}O$ values using models that describe the relationship between surface temperature and deep-water temperature and between surface temperature and the storage in ice sheets (see also Moum *et al.* (2013) or Shen *et al.* (2005), 379). They then build a continuous reconstruction of atmospheric temperature, ice volume and global sea level for the last three million years. Figure 9 presents their data for the period of interest in the Neanderthal case study. In their supplementary information section Bintanja and van de Wal (2008) compare their results favourably against different and independent proxy records of variable length. Reconstructed values become more uncertain further back in time, especially before 400 ka (Bintanja *et al.* 2005).

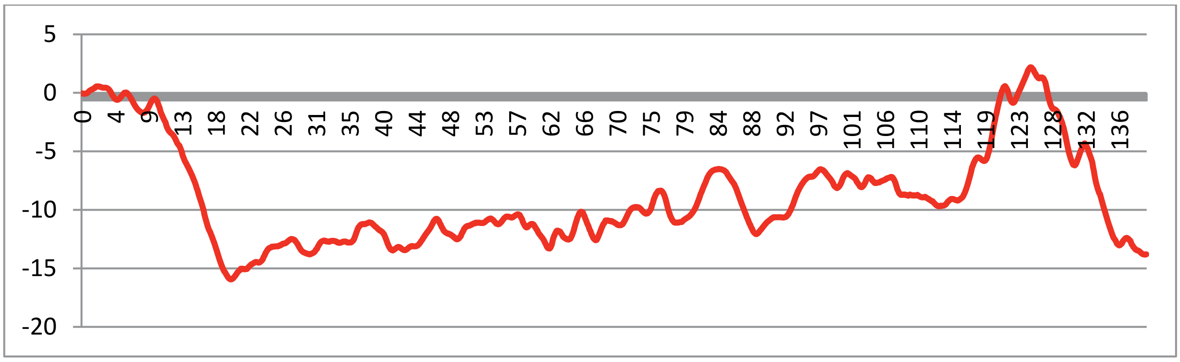


Figure 9: Overview of reconstructed global mean temperatures (Y-axis) for the last 140ky, drawn according to Bintanja and van de Wal (2008).

In the temperature record global and many local minima and maxima can be found, with some of the more prominent ones given in Table 2 (these points in time will be used to monitor environment reconstructions). Dating of the LGM climatic minimum matches estimates from other areas (Barrows *et al.* 2002; Yokoyama *et al.* 2000). A yearly temperature of minus 15.9 ºCelsius is an absolute minimum for the period considered. It is a mean value, suggesting that colder periods throughout the year were present. The yearly temperature during the Mid Holocene is hardly higher than modern day temperatures, albeit differently distributed throughout the year, which is why this study uses current day data as maximum. The reconstructed data[14] is available as an Excel database in the supplementary information of Bintanja and van de Wal (2008) (see subsection 8.5.2 for a detailed description of this input file). This high resolution temperature record is used in

[14] Global 3Ma Temperature, Sea Level, and Ice Volume Reconstructions collected by R. Bintanja and R.S.W. van de Wal. *IGBP PAGES/World Data Center for Paleoclimatology Data Contribution Series* # 2011-119. NOAA/NCDC Paleoclimatology Program, Boulder CO, USA.

this study to interpolate between climatic extremes since it represents the climate more accurately than the isotope data itself and since it is a very good relative indicator of global conditions through time.

Table 2: Local and global minimum and maximum temperatures from Bintanja and van de Wal (2008), supplementary information. Temperatures are offset in degrees Celsius relative to modern day yearly mean values.

Description	*Date (years ago)*	*Temperature (in Celsius)*
A local maximum around 2 ka	2,100	0.56
A local maximum around 6 ka (Mid Holocene)	5,800	0.03
A global minimum around 19 ka (LGM)	19,600	-15.94
A local minimum around 61 ka	61,500	-13.31
A local minimum around 78 ka	78,100	-11.85
A local maximum around 83 ka	83,800	-6.51
A local minimum around 88 ka	88,700	-12.07
A local maximum around 97 ka	97,200	-6.53
A local minimum around 112 ka	112,800	-9.70
A local maximum around 120 ka	121,400	0.57
A global maximum around 124 ka	124,800	2.19

Data on climatic conditions is provided by WorldClim, one of many sources of climate data and a database of global climate layers with a very high resolution (Hijmans *et al.* 2005). It was developed at the University of California, Berkeley and is freely available through http://worldclim.org (first accessed 25 July 2013). The data for modern day climate parameters are collected from weather stations and other sources from around the world and originally sampled in 30 second resolution. Coarser datafiles have been created by calculating the mean of the original variable over larger grid cells (Hijmans *et al.* 2005). Bioclimatic variables are derived from the monthly temperature and rainfall values in order to generate more biologically meaningful variables. The bioclimatic variables represent annual trends (e.g., mean annual temperature, annual precipitation), seasonality (e.g., annual range in temperature and precipitation) and extreme or limiting environmental factors (e.g., temperature of the coldest and warmest month, and precipitation of the wet and dry quarters) and are used in numerous biogeographical studies (Elith *et al.* 2006; Broennimann *et al.* 2012) and widely recognized for their practical value (Fick and Hijmans 2017). A list of all variables is given in Appendix 1.

The used data for LGM conditions at WorldClim is calculated using a coupled general circulation models: Community Climate System Model (CCSM), available from http://www.worldclim.org/past, verified 10 January 2018 (Collins *et al.* 2006). It is favourably compared against other modelling results in Appendix 2. Such models integrate mathematical functions describing atmospheric and oceanic circulation, sea ice, land surface characteristics and properties of the atmosphere (Hijmans *et al.* 2005). Running such simulations is time consuming, computationally intensive and expensive. Data is only available for a limited number of time steps. That is why this study interpolates between high resolution extreme local conditions. Results farther back in time are progressively more uncertain. Data from the WorldClim database was selected due to its availability of similar data sets for current and past periods, the easily usable format in high resolution (30 arc second or 1-km resolution), and the included level of detail (19 bioclimatic parameters for all datasets).

4.3.1 Reconstruction via interpolation between climate extremes

The WorldClim database contains data on modern day environmental conditions, together with future projections and past climate reconstructions. However, intermediate results between reconstructed points in time are not available, especially not at a high resolution temporal scale. The model underlying the current research uses a linear interpolation between two known data points of extreme climates to generate continuous values for climatic parameters in the past.

A straight linear interpolation between current climate and a general circulation model of the LGM using stable oxygen isotope ratios provides a good estimate of paleoclimate parameters for other time periods, including extrapolation for points in time outside the range of the two measuring points (Lawing and Polly 2011). Lawing and Polly (2011, 8) compare the results of two general circulation models with their interpolated dataset of temperature and precipitation. They find that the differences between their data and each of these modelled reconstructions are less than or equal to the differences between these two models. This suggests that their interpolated model is of equal quality as the general circulation models.

Lawing and Polly use an interpolation that is governed by stable oxygen isotope ratios which unfortunately are not only influenced by oceanic $\delta^{18}O$ but also by local deep-water temperatures (Imbrie *et al.* 1984). The underlying research therefore instead uses reconstructed surface temperatures in which those effects are compensated (Bintanja and van de Wal 2008). HomininSpace uses an similar interpolation based on isotopic values measured from within ice cores, but with the isotope ratios converted into a mean global temperature reconstruction that compensates for effects from upwelling deep water that represent earlier climatic influences (Bintanja and van de Wal 2008). It is expected that this compensation provides more realistic results for reconstructing temperature and precipitation at any point in time.

For each grid cell values for LGM (global temperature offset at -16) and for current day (global temperature offset is 0) local precipitation and temperature are available[15]. The reconstructed global temperature from Bintanja and van de Wal (2008) is used as an index in each step to interpolate in each grid cell between these two values to obtain interpolated values for precipitation and temperature. This interpolation method is illustrated in Figure 10. For a hypothetical location two temperature values are given: the value for the LGM which is 5, and for today a value of 12. Horizontally in Figure 10 the reconstructed yearly temperature (a global mean value) is used to index the interpolated value. Thus, when at any time during the simulation this yearly temperature takes the value -8, the interpolated temperature will be calculated at 8.5. In a similar manner this global mean temperature is used to index the line between precipitation values for LGM and today.

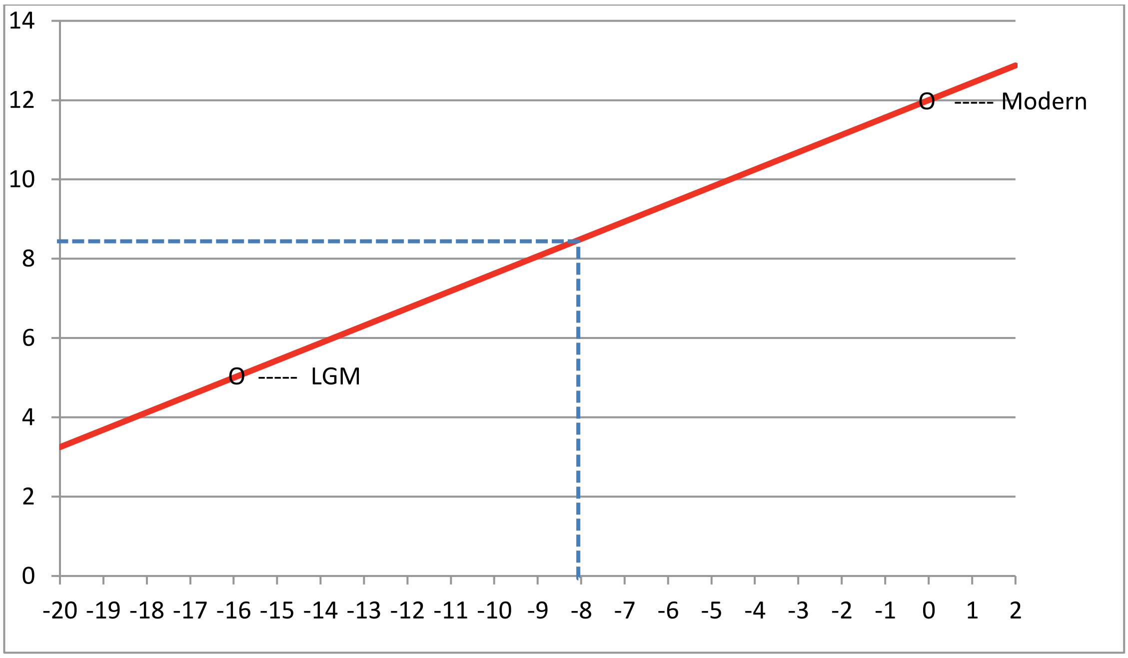


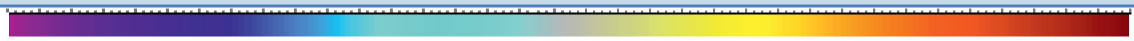
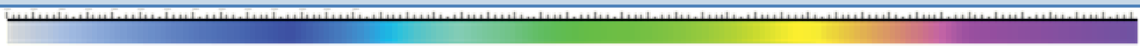
Figure 10: Interpolation of temperature values between two given data points: LGM and modern day values, with horizontally the reconstructed global mean temperature used as index for the linear interpolation.

The distribution of the climatic values for the interpolation for mean annual temperature and precipitation are given in , Figure 12, Figure 13 and Figure 14. For the warmest and coldest months Figure 15 presents the values for modern day and LGM. A frequency

[15] For some grid cells, notably current day sea areas, values are calculated after interpolation from nearby land cells.

distribution for three LGM parameters is given in Figure 16. Temperatures are in degrees Celsius, rainfall is in mm per year. For those grid cells where current day data is not available (for instance for exposed shelves in the Channel area) values for temperature and precipitation are extrapolated from nearby grid cells (compare with the similar extrapolation of soil properties in Hoogakker *et al.* (2016, 55)). Color coding of the grid cells in these and other figures is according to the values in Table 3.

Table 3: Color coding of grid cells for temperature and precipitation.

Color coding for temperature (top) in Celsius, precipitation (bottom) in mm per year.					
-30	**-15**	**0**	**15**	**30**	**40**
					
300			**1400**		**2400**
					

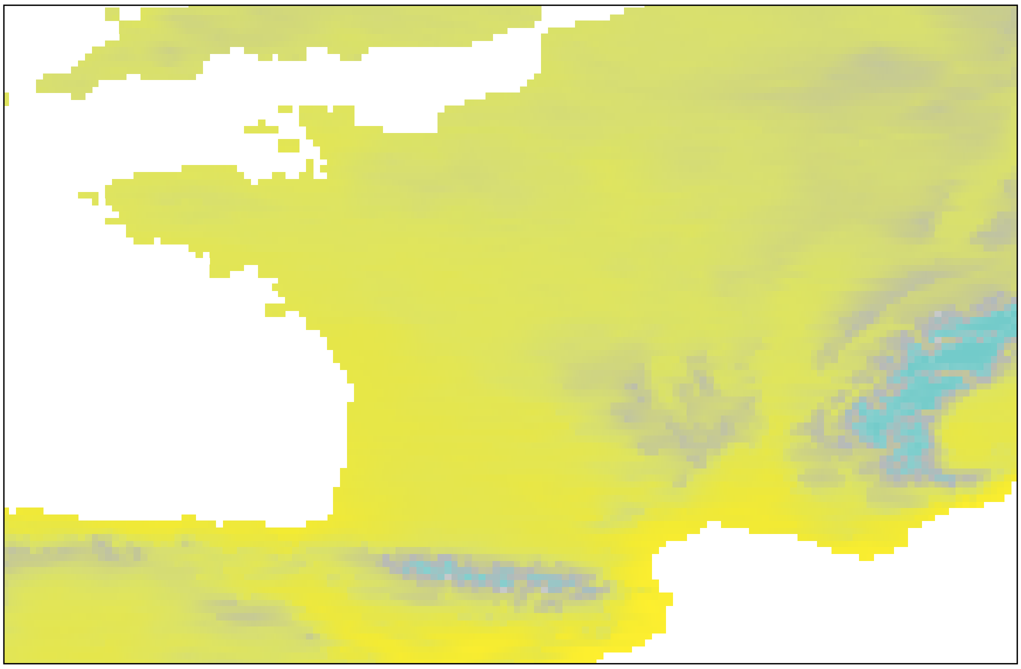


Figure 11: Mean annual temperature distribution for modern day conditions.

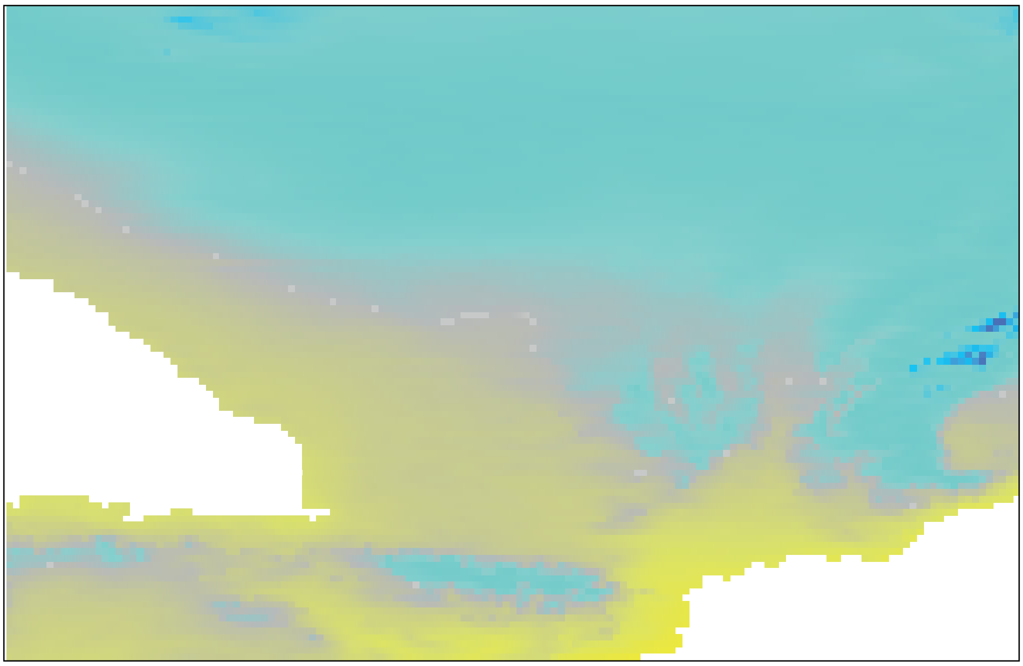


Figure 12: Mean annual temperature distribution during the LGM (source CCSM).

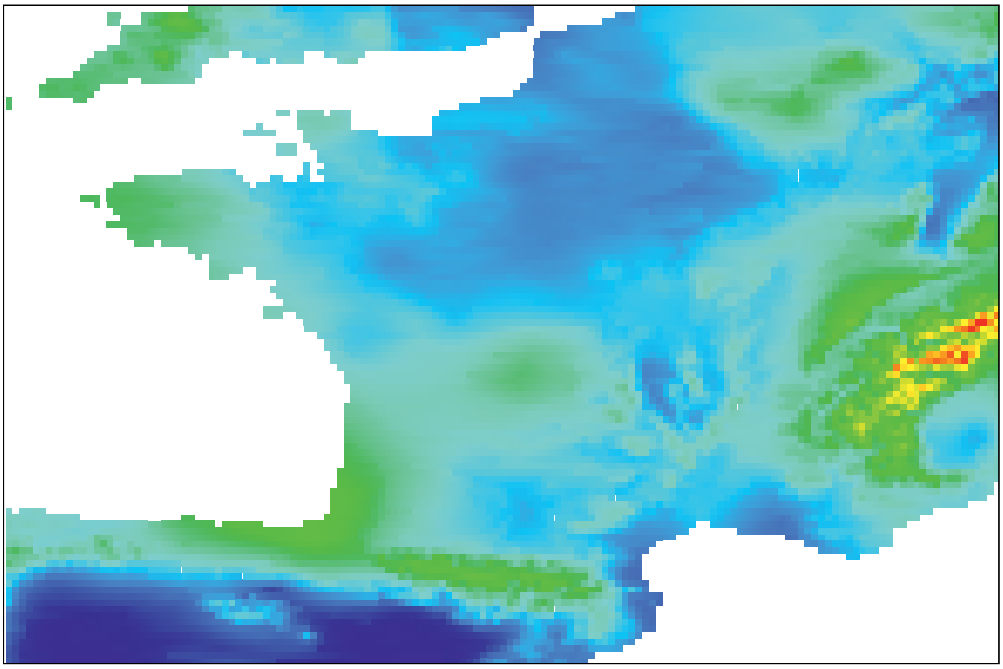


Figure 13: Annual precipitation distribution for the modern day conditions.

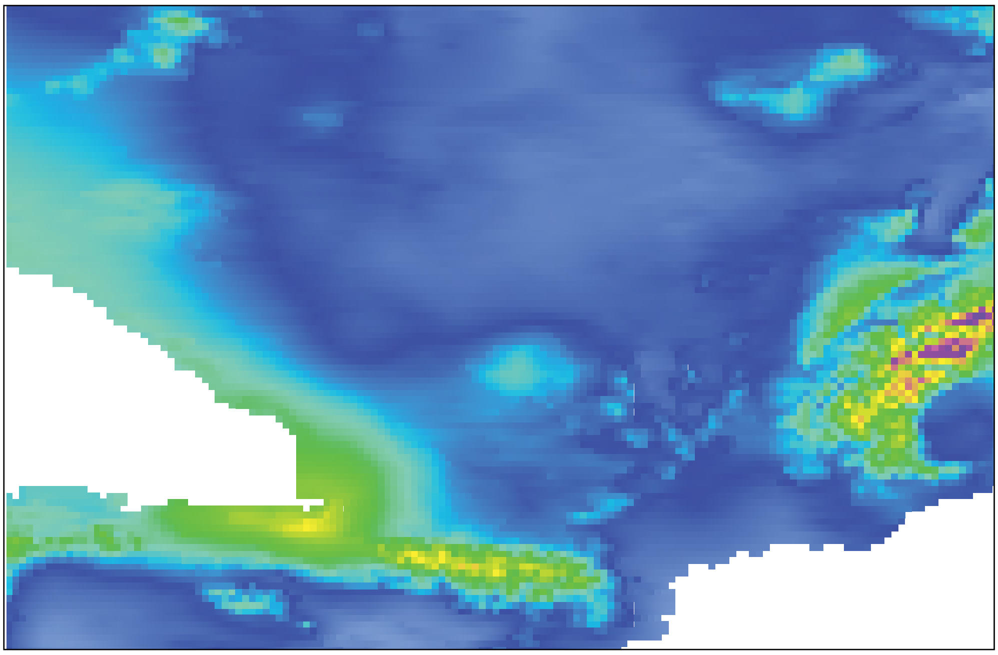


Figure 14: Annual precipitation distribution during the LGM (source CCSM).

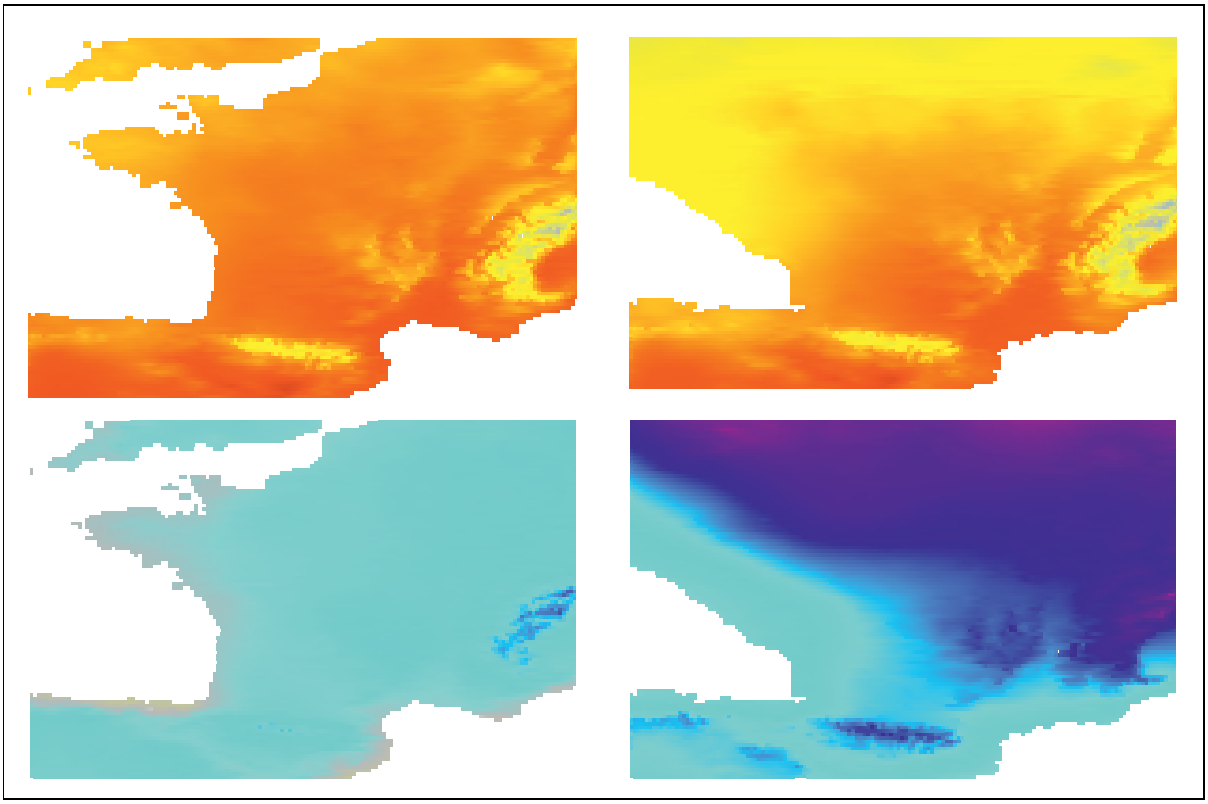
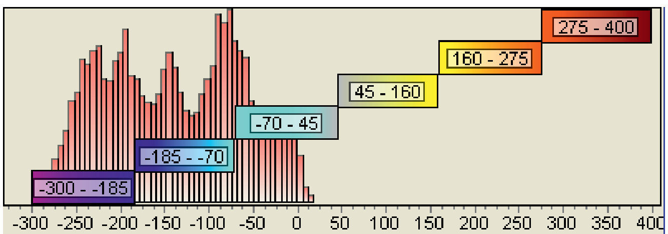


Figure 15: Overview of modern day (left) and LGM temperatures of the warmest month (top) and coldest month (bottom).



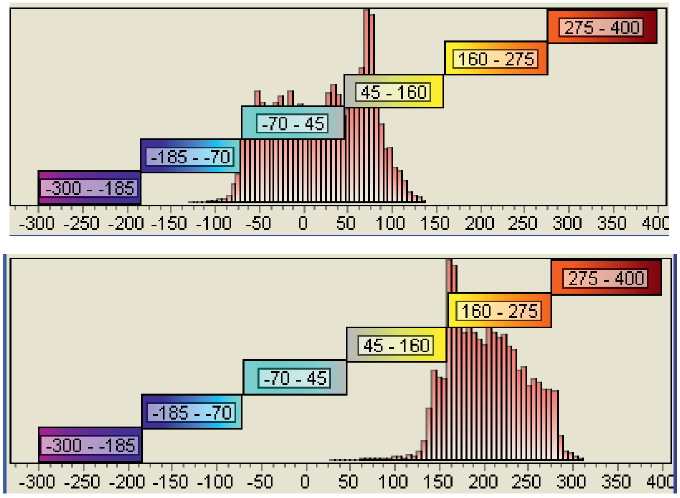


Figure 16: Relative frequency distribution for the three LGM temperature climate parameters: coldest month (top), yearly mean, and hottest month (bottom).

4.4 An energy landscape in two forms

HomininSpace uses the two most important climatic parameters, temperature and precipitation, as defining factors for the reconstructed climate in the simulations, and computes from these the amount of energy in the landscape. Reconstructing the available energy levels within the HomininSpace modelling system involves several steps. The aim and final prediction of these calculations is an amount of energy available in the form of edible secondary biomass (large ungulates). This is recalculated for every year, with results stored in each grid cell, and forms the basis of the carrying capacity per cell, taking fluctuating sea levels into account.

HomininSpace explores and compares two different methods to calculate secondary biomass given the primary production. The first one (Subsection 4.4.2), henceforth referred to as the *continuous energy model*, uses data on primary production and available secondary biomass from a wide range of climates and vegetation types to infer a single direct relationship between the two important climatic parameters (precipitation and temperature) and herbivore availability across different biomes (McNaughton *et al.* 1989). The second variant (Subsection 4.4.3) will use these same variables to first identify the prevailing climate type, then determine the biome type and only then use empirical data for each biome to calculate available secondary biomass (Binford 2001; Kelly 1983). Since this last method first models the habitat for an area this method is referred to as the *habitat model* in this work.

Both approaches start with primary productivity and use data from linear regression on empirical data to infer relationships between climate parameters and biomass. In every time step during a simulation several reconstructions are done in each grid cell, where reconstruction of the secondary biomass depends on the chosen variant. The schema in Figure 17 indicates the pathways for the two variants with steps one and four the same in both approaches.

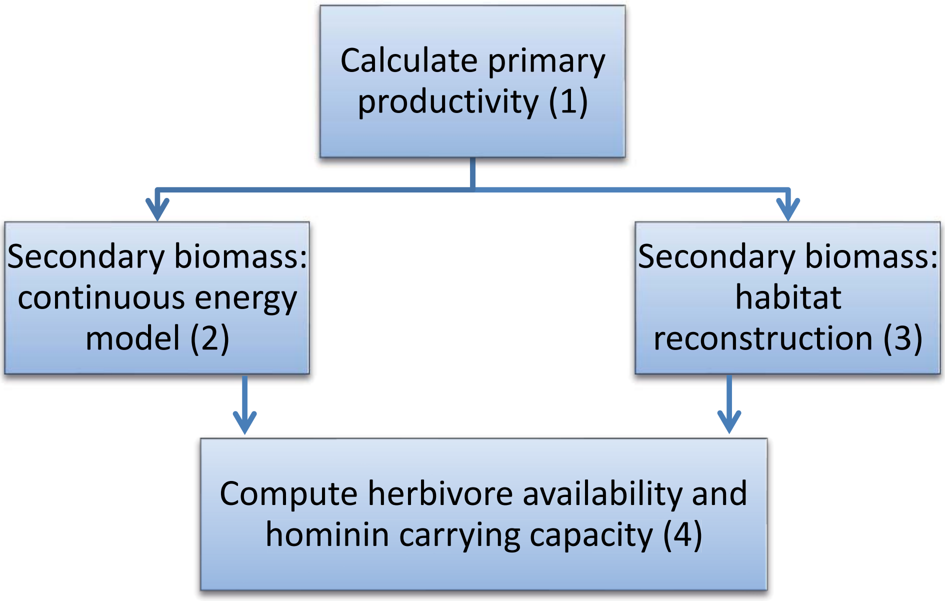


Figure 17: Schema illustrating both pathways to compute carrying capacity.

Each of these steps are summarized here and detailed in the following sections.

1. Primary Productivity (PP) is based on mean yearly temperature and precipitation using formulas designed by Lieth (1973). This gives the PP mass in kg per grid cell per year.
2. In the *continuous energy* variant the secondary biomass that is required in the final step is computed based on empirical data. This data was collected by McNaughton *et al.* (1989) and presents values for SB that are found in a wide variety of habitats for measured PP levels. A linear regression formula has been derived and is implemented in the HomininSpace simulation system.
3. In the *habitat reconstruction* model the type of climate is determined using temperature ranges based on data collected by Binford (2001). For each cell Effective Temperature (ET) is computed, a composite variable using temperature of the coldest and hottest months of the year. This value specifies which kind of climate (polar, boreal, cool temperate, etc.) is realized. Then the type of vegetation is considered per kind of environment. For several types of environment more than one type of vegetation can be present, depending on the amount of precipitation. For instance, cool temperate areas can be covered by deciduous forest or open woodland. Primary Biomass (PB) can be calculated for the chosen type of vegetation using the PP mass computed in step 1, with formulas based on research by Kelly (1983) including the mean annual precipitation. The amount of secondary

biomass that can be sustained by the computed PB can now be calculated and is based on empirical key values from Kelly (1983), the reconstructed vegetation type and the calculated PB. Using empirical data takes into account the changing flow of energy between trophic levels.

4. The herbivores (SB) consuming the primary production are not prefabricated meat slices, they are not all ungulates and certainly do not all have a useful size fit to be hunted or consumed by hominins. To convert the SB values into available energy several modifiers are applied, effectively reducing the amount of secondary biomass present in the landscape into packages of meat available to hominins.

4.4.1 Calculating primary productivity

The amount of energy remaining in vegetation from the process of photosynthesis after respiration is defined as the (net) primary production. This is the plant material (the primary producers) that is potentially available to secondary consumers, the herbivores. Both temperature and precipitation govern primary productivity. The relation between temperature (T) and primary productivity (PP), derived from empirical data, is given in Equation 1, with T in degrees Celsius[16] (Lieth 1973, 325):

$$\boldsymbol{PP = \frac{3000}{(1 + e^{1.315 - 0.119 * T})}} \ \textbf{in} \ \frac{\frac{\mathbf{g}}{\boldsymbol{m^2}}}{\textbf{year}}$$

Equation 1: Relation between Temperature and Primary Productivity.

The constant *e* equals 2.71828182845904, the base of the natural logarithm. The formula is derived from meteorological data taken from the Climate Diagram World Atlas and assumes that the net productivity does not exceed three kg per square meter per year and that the relation between temperature and productivity is sigmoid shaped (a logistic growth equation). An example of the calculations for different mean annual temperatures is shown in Table 4, where a grid cell is defined as a square area of 10 x 10 kilometres.

Table 4: Example values for primary productivity for given temperatures.

Temperature	*Denominator in Formula (1)*	*Primary Productivity in g/m²/year*	*Primary Productivity in kg/gridcell/year*
5	3,054433211	982,1789	98.217.894,88
10	2,133148453	1406,372	140.637.187,99
15	1,625002268	1846,151	184.615.126,92
20	1,344727855	2230,935	223.093.467,53
25	1,19013898	2520,714	252.071.400,92

[16] Throughout this thesis all temperatures are given in degrees Celsius.

Lieth (1973) found that precipitation in a similar fashion steers productivity. The formula for this process integrates an empirically found relationship for more arid climates, a saturation curve for yield factors and an exclusion threshold for lower temperatures (Equation 2 with P in mm per year). Again it is assumed that productivity is less than three kg per square meter per year and the graphic representation of the formula has a saturation curve form. Figure 18 presents both datasets and their derived formula in one graph.

$$\boldsymbol{PP = 3000 * (1 - e^{-0.000664*P}) \quad \text{in } \frac{g}{m^2}/\text{year}}$$

Equation 2: The effect of Precipitation on Primary Productivity.

Both formulas assume an exclusion line for low precipitation and low temperature levels respectively. For any combination of temperature value and precipitation level the minimum factor controls production (according to von Liebig, referenced in Lieth (1973, 326)). Thus, for each grid cell in each turn both values are computed and the smaller of the two is used. It is further assumed that these formulas are valid for all values occurring in the simulations. Alternative formulas have been derived from empirical data (Wisiol 1984, 472) but they often include the parameter evapotranspiration which is unavailable for the simulation period, or they are based on data for a single climate type (only for desert, etc.). For problems with modelling NPP in open vegetation areas versus forested areas see Del Grosso et al. (2008).

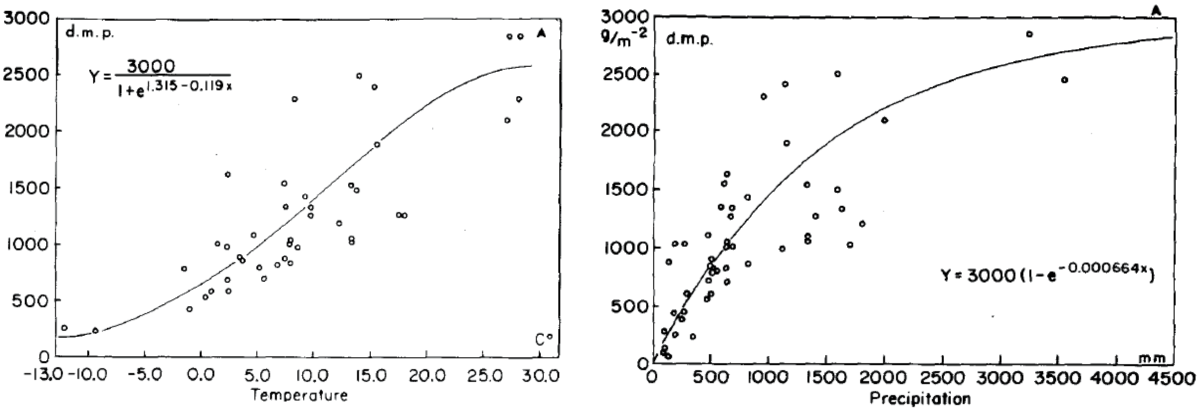


Figure 18: Datasets and derived equations for productivity based on temperature (left) and precipitation values. Figures reproduced from Lieth (1973), Figures 5 and 6.

4.4.2 Secondary biomass: the continuous energy model

This first variant to calculate herbivore availability is based on empirical data collected by McNaughton *et al.* (1989). In landscape ecology is has been recognised for some time that environments are not very homogeneous, and form patchworks of concentrations of different plant (and animal) species (Hansson *et al.* 1995, preface). The resulting complexity in the landscape is almost impossible to reconstruct and deviates from an even distribution of stable resources (Grøn 2018) as created by the habitat reconstruction method of the next subsection. Within the continuous energy approach in every time step during the simulation the following calculations are done in each grid cell:

1. Primary Productivity (PP) is based on mean yearly temperature and precipitation using formulas designed by Lieth (1973). This gives the PP mass in kg per grid cell per year;
2. The amount of herbivore availability or secondary biomass that can be sustained by the computed PP can now be calculated and is based on empirical key values from McNaughton *et al.* (1989).

The relationship between primary productivity and secondary biomass as collected by McNaughton *et al.* (1989) from different environments is depicted in Figure 19. The dataset consists of 51 data points from desert (labelled 1 in Figure 19), tundra (2), temperate grassland (3), temperate vegetation succession of old fields (4), unmanaged tropical grassland (5), temperate forest (6), tropical forest (7), salt marsh (8) and agricultural tropical grassland (9). Few points are from anthropogenic landscapes (4, 9). NAP is the net above-ground primary productivity, here simplified to be equal to PP.

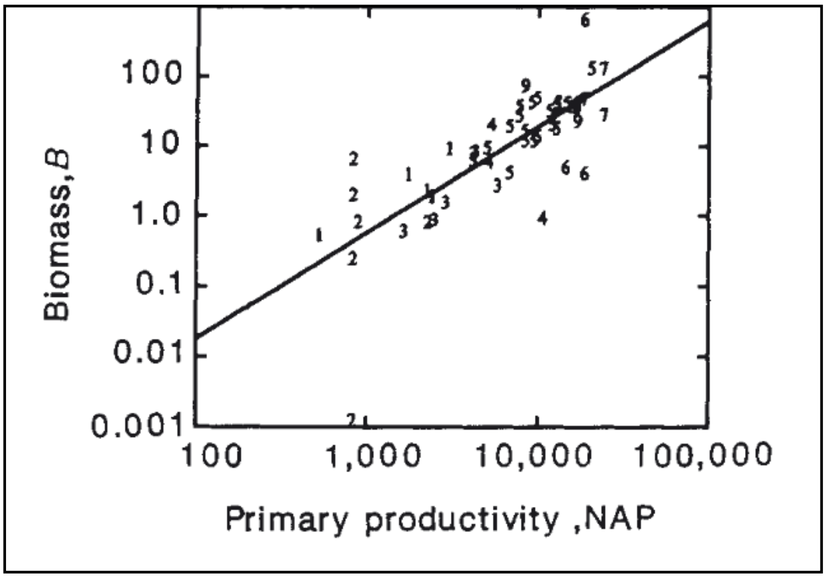


Figure 19: Relationship between herbivore biomass (B) and net primary productivity (NAP) for different types of ecosystem (from McNaughton *et al.* (1989, Figure 1)).

Regression analysis produced the following formula (Equation 3) to calculate the Secondary Biomass (SB), with log indicating common logarithms and NAP equated to PP, as implemented in HomininSpace:

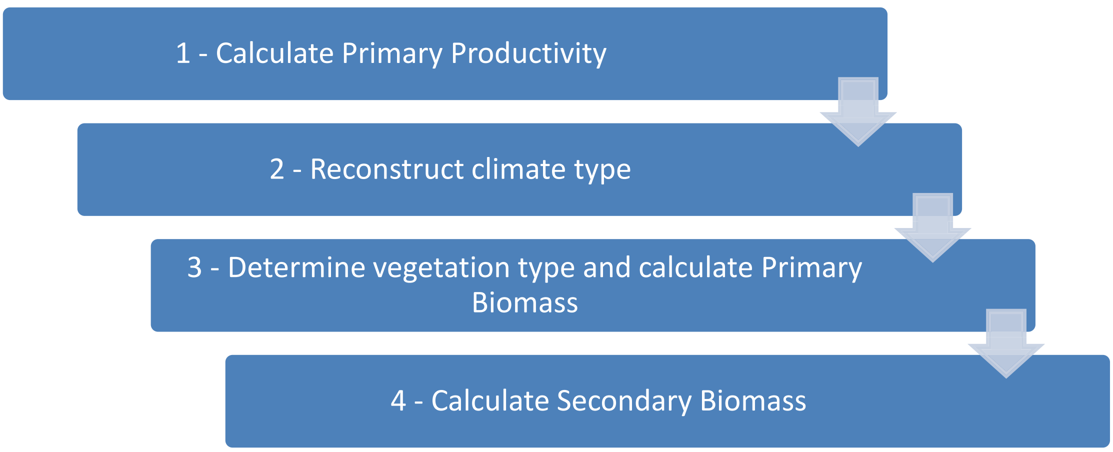
$$\mathbf{\log \boldsymbol{SB} = 1.52 * (\log \boldsymbol{NAP}) - 4.79 \quad \boldsymbol{in\ kJm^{-2}}}$$

Equation 3: Relation between Secondary Biomass and Net Primary Production.

This gives a value for secondary biomass independent of biome type and is used to construct a continuous energy landscape based on temperature and precipitation.

4.4.3 Secondary biomass: the reconstructed habitat model

Most global models of primary productivity depend on stratifying a geographical area into a number of functional units, land cover types, biomes, or vegetation classes (Ellis and Ramankutty 2008; Feddema *et al.* 2005; Haxeltine and Prentice 1996; Thomas *et al.* 2004). In such models the major biome types (desert, tundra, grasslands, savannas and the tropical, temperate and boreal forests, each identifiable primarily by their predominant vegetation) are predicted from climate parameters like temperature, precipitation and correlation with the water balance (Bond *et al.* 2005, 525; Prentice *et al.* 1992). See Appendix 1 for a list of major (mega) biome types and what they encompass. Within the habitat model of the HomininSpace simulation system identification of biome has been deferred to the second stage of the production calculation, with the computation of secondary productivity. In each time step during a simulation the following reconstruction steps are executed in each grid cell:



1. Primary Productivity (PP) is based on mean yearly temperature and precipitation using formulas designed by Lieth (1973). This gives the PP mass in kg per grid cell per year;
2. The type of climate is reconstructed using temperature ranges based on data collected by Binford (2001). Calculated is Effective Temperature (ET), a composite variable

using temperature of the coldest and hottest months of the year. This value specifies which kind of environment (polar, boreal, cool temperate, etc.) is realized;
3. Then the type of vegetation is considered per kind of environment. For several types of environment more than one type of vegetation can be present, depending on the amount of precipitation. For instance, cool temperate areas can be covered by deciduous forest or open woodland. Primary Biomass (PB) can be calculated for the chosen type of vegetation using the PP mass computed in step 1, with formulas based on research by Kelly (1983) including the mean annual precipitation;
4. The amount of Secondary Biomass (SB) that can be sustained by the computed PB can now be calculated and is based on empirical key values from Kelly (1983), the reconstructed vegetation type and the calculated PB.

Temperature and precipitation values used in the preceding steps are calculated per grid cell using linear interpolation between the modern day and LGM temperature and precipitation distributions (see Section 4.3). Within a single reconstructed habitat there are differences between cells due to the fact that precipitation and temperature are used to calculate productivity, and these parameters vary throughout the total simulation area.
To calculate the secondary biomass available for hunter-gatherers, first the Primary Biomass must be estimated. This is dependent on the type of biome in which the primary production is realized. The major biome types (desert, tundra, grasslands, savannas and the tropical forests, temperate forests and boreal forests) can be predicted from temperature and precipitation and correlate with the water balance (Bond *et al.* 2005, 525). First, the general type of climate for which the biome has to be selected is determined based on the given temperature. For this there are a few rules of thumb using a variable called Effective Temperature (ET). This calculated value was proposed by Bailey (1960) to refer to the amount of solar energy that is available at a certain location. Its usage in selecting a type of climate is based on three empirically found constants:

- 18° C - the minimal mean temperature of the coldest month of the year that will sustain tropical plant communities;
- 10° C - the minimal mean temperature expected at the beginning of the growing season along the boundary between polar and boreal environments;
- 8° C - the minimal mean temperature at the beginning and the end of the growing season;

From these constants effective temperature values can be calculated (Equation 4) relating directly to empirically found biological boundaries separating different zones of biological activity. For instance, an ET of 18° C or higher indicates that there is no killing frost[17] all

[17] Many plants can be damaged by freezing temperatures, depending on type of plant and temperatures reached (https://en.wikipedia.org/wiki/Frost, accessed December 2018).

year round, and an ET of less than 10° C suggests fewer than 30 days in the year without a killing frost. Effective Temperature is calculated using the mean temperature for the warmest month (MWM) and the mean temperature for the coldest month (MCM) (Binford 2001, 59):

$$\boldsymbol{ET} = \frac{(\mathbf{18} * \mathbf{MWM}) - (\mathbf{10} * \mathbf{MCM})}{(\mathbf{MWM} - \mathbf{MCM} + \mathbf{8})} \quad \textbf{in °C}$$

Equation 4: Calculation of Effective Temperature using the mean extreme temperatures.

With the ET the type of climate for each grid cell can be determined. Boundary values for different types of climate are given in Table 5.

Table 5: Different types of climate and their defining boundaries, derived from Binford (2001), Table 4.04, page 70.

#	Type climate	Min ET (°C)	Max ET (°C)
1	Polar	----	9.99
2	Boreal	10.00	12.49
3	Cool temperate	12.50	14.55
4	Warm temperate	14.56	16.61
5	Subtropical	16.62	18.15
6	Tropical	18.16	22.57
7	Equatorial	22.58	----

Both the Cool temperate as well as the Warm temperate climates can produce two different biome types (types of vegetation) (Kelly 1983). For both climates this depends on different amounts of rainfall. For Warm temperate environments it is possible to get Temperate Deciduous Forest (relatively humid) or Woodland/Scrubland (more arid), and for Cool temperate climates this would be Temperate Evergreen forest (humid) versus Temperate Grassland (arid). The more arid biomes are characterized by continuous grass layers with trees in a climatic regime of distinct wet and dry seasons. Besides rainfall the main additional forces maintaining such biome types are herbivory, fire and frost (Du Toit and Cumming 1999). In the grasslands, most of the biomass is reproductive tissue (blades of grass) and as such, production is relatively high and biomass low (Binford 2001, 82).

The vegetation community that is the transition between deciduous forest and more water-stressed environments is the tall grass prairie-forest steppe, with annual rainfall 717 +/- 257 mm (Binford 2001, Table 4.08, 98). For cooler temperatures Binford identifies mid-latitude short grass prairie (page 97). Rainfall for this vegetation type is 431 +/- 178 mm, with an upper value of 609mm. In the absence of similar information for other vegetation transitions, this latter value is rather arbitrarily chosen as the default value for precipitation levels for arid, cold environments. Although both the climate types qualify as arid, following Kelly (1983, 284) the difference between the biome types is defined as 900mm rainfall per year. Coe *et al.* (1976) use 700mm rainfall per year as an indication for semi-arid areas.

Not all energy stored in PP is equally available to other trophic levels. The ratio of Primary Biomass (PB) (the producers) and Secondary Biomass (SB) (the consumers) indicates both how much herbivores are available as prey as well as the availability of PB itself to herbivores. A high level indicates more potential prey as well as relatively more available plant material in the form of foliage, fruits and seeds (as opposed to wood). From PP for a given biome it is possible to calculate PB, using the empirical derived ratios as given by Kelly (1983, 284) (c.f. Keeley 1988, 379), see Table 6. PB is the total amount of standing plant material present in an area at any time. The PB can be used to calculate the SB with the given ratios, specifically the herbivores that will provide the meat that supplies the energy for the hominins in the underlying model.

Table 6: Calculating PB and SB from PP. Values derived from Table 3, page 284 (Kelly 1983, 284, Table 3). Type climate from Table 5 matched against biome type.

Biome#	Type climate (Binford 2001)	Type biome (Kelly 1983)	a = PP / PB	b = SB / PB
1	Polar	Tundra	0.2333	0.0006667
2	Boreal	Boreal forest	0.0400	0.0002500
3	Cool temperate	Temperate Evergreen forest	0.0371	0.0002857
4		Temperate Grassland	0.3800	0.0043750
5	Warm temperate	Temperate Deciduous Forest	0.0400	0.0005333
6		Woodland/Scrubland	0.1200	0.0008333
7	Subtropical	Tropical Savanna	0.2250	0.0037500
8	Tropical	Tropical Seasonal Forest	0.0457	0.0003429
9	Equatorial	Tropical Rain Forest	0.0489	0.0004222

To calculate the PB and SB values for each biome the formulas in the last two columns of Table 6 are combined. These are derived from Kelly (1983, Table 3, 284) where a linear relationship is found between PP and PB. This translates into Equation 5 for estimating SB (PP is per cell and in gram, hence the division by 1000 to get to kilograms):

$$SB = \frac{PP * b}{a} / 1000 \quad \text{in kg/cell/year}$$

Equation 5: Computing Secondary Biomass from Primary Productivity (a, b from Table 6).

That SB indeed varies with climatic variables is also empirically attested (Binford 2001, 101). Examples with calculated values according to Equation 5 for different biomes are given in Table 7.

Table 7: Calculated secondary biomass per grid cell in kilograms per year for different mean yearly temperatures (theoretical values) in different biomes[18].

Temp	*Biome1*	*Biome2*	*Biome3*	*Biome4*	*Biome5*	*Biome6*	*Biome7*	*Biome8*	*Biome9*
-10	64.73	14.16	174.44	260.80	302.01	157.30	377.54	169.97	195.58
-5	110.58	24.18	297.98	445.49	515.89	268.70	644.90	290.33	334.08
0	181.45	39.68	488.97	731.03	846.55	440.92	1,058.26	476.42	548.22
5	280.68	61.39	756.36	1,130.80	1,309.49	682.04	1,636.96	736.96	848.01
10	401.90	87.90	1,083.02	1,619.18	1,875.05	976.61	2,343.95	1,055.24	1,214.25
15	527.57	115.38	1,421.69	2,125.50	2,461.38	1,282.00	3,076.92	1,385.22	1,593.96
20	637.53	139.43	1,718.00	2,568.51	2,974.39	1,549.20	3,718.22	1,673.93	1,926.18
25	720.34	157.54	1,941.15	2,902.14	3,360.74	1,750.43	4,201.19	1,891.36	2,176.37

Since the formulas and categories are based on extant biomes and actual measured data, they are applicable only in environments that exist today. Types of landscape that no longer exist (like the Mammoth Steppe) are not given by Kelly since actual values for secondary biomass or MWM/MCM cannot be measured and productivity must therefore be modelled. It has been shown that especially the cool temperate grassland was very productive through a combination of intense sunlight and fertile loess providing mosses, lichens, grasses and shrubs feeding Mammoth Steppe herbivores (Guthrie 1990). This biome is best implemented by applying the figures from the African savannah.

Terrestrial biomes are distinguished primarily by their predominant vegetation, and are mainly determined by temperature and rainfall (Forseth 2012). The model in HomininSpace assumes a direct relation between temperature, precipitation and vegetation type, and acknowledges but ignores other factors that (could) play an important part in the resulting biome, most importantly soil type and groundwater levels (Fan *et al.* 2013). The limiting effects of temperature and precipitation levels however surpass the influence of other parameters and their availability significantly correlates with primary productivity, today (Li *et al.* 2013) and in the past (Janis *et al.* 2000).

4.4.4 The number of edible ungulates: carrying capacity

Carrying capacity in the underlying research is defined as the amount of kilocalories in the form of a number of edible ungulates within a certain area that are available for

[18] See also Supplementary Materials for the spreadsheet "Climate Calculations.xlsx" to compute these values.

consumption by the modelled hominins at a given time. The carrying capacity is a variable that is used in the hominin demographic calculations (see Chapter 1). Research has shown that in some of the most relevant biome types about half of the plant production is eaten by large (> 5 kg) mammalian herbivores (Du Toit and Cumming 1999). The remaining yearly production is eaten by small mammals, insects, decomposed by fungi, etc. The model in HomininSpace therefore assumes that 50% of the herbivore mass in any area consists of individuals from species that can be hunted and consumed.

It is further estimated that from every kilogram of secondary ungulate biomass only about 60% is consumable by hominins. The remaining 40% are bones, skin and assorted inedible items including hooves and horns. These can be useful raw materials but are not usable for caloric consumption (White 1953; White 2006, 11). It has been shown that Neanderthals can switch between different types of prey species, depending on the availability in the landscape (Hodgkins *et al.* 2016). The available secondary biomass for consumption is given by the equation below.

$$\boldsymbol{SB_{avail}} = \frac{\mathbf{1}}{\mathbf{2}} * \frac{\mathbf{6}}{\mathbf{10}} * \boldsymbol{SB} \quad \textbf{in kg}$$

Equation 6: Computing the available Secondary Biomass for hominin consumption.

The linear relationship in this study between primary production and edible large ungulate biomass is but a model of reality in which the interaction between the different components is an important element. See for other examples the work from Augustine and McNaughton (2006) or Hobbs (1996). The underlying model has to necessarily simplify and generalize matters and states that from a hypothetical secondary biomass of 100 kg, the following remains available for consumption:

- 50% is large ungulates = 50kg;
- 60% of 50kg = 30kg edible material.

When prey animals are harvested from the environment, prey populations can recover. Generally there is an optimal harvest size which still allows full recovery within the hunting cycle. Since HomininSpace calculates the available prey species based on primary productivity, harvesting actually reduces primary biomass and thus recovery of prey populations and secondary biomass must be translated into recovered primary biomass. Ecosystem responses to disturbances are complex due to species – species interactions and their relationships with the environment (Ives and Carpenter 2007). Full recovery of certain forest type environments to re-establish after disturbances for instance can take very long (Cole *et al.* 2014). Therefore recovery of the environment from harvesting by hominins is modelled after larger ungulate recovery mechanisms, and very conservatively and rather arbitrarily implemented as a 20% yearly rate of recovery and prey populations increase (Steinmetz *et al.* 2010, 44), summarized in the equation below.

$$\boldsymbol{PP_{t+1} = PP_{P,T} - PP_{consumed} + \frac{1}{5} * PP_t}\text{ , }\textbf{in kcal}$$

Equation 7: Calculating the result for Primary Productivity at time step t+1.