A Review of the Allocation Method for Historical Cropland Cover used for the HYDE Database

Ella Sikkens, 6459676
Bachelor's Thesis
Global Sustainability Science, Utrecht University
Supervised by Kees Klein Goldewijk
Submitted on 27-01-2023
Wordcount: 5812

Contents

Summary

The HYDE database is one of the most widely used scenarios for historical land use and land cover reconstruction. It is used for international assessment, for example, in IPCC's 5th Assessment Report, and in general for global climate modelling. The dataset is used for studying the environmental consequences of long-term LUCC, such as their contribution to climate change and in the calculation of carbon emissions. It is essential that the historical data is accurate, since uncertainties in the dataset can lead to uncertainties in climate and CO2 projections. This can consequently affect policy makers in establishing reasonable emission mitigation strategies. Therefore, the aim of this study is to critically reflect on the current allocation methods for historical cropland cover used by HYDE. To analyse this, a literature review was conducted and a GIS analysis was done with two of the proxies that are used as input for the allocation. The results show that the HYDE allocation methods can be improved to be completely correct. This can be achieved by changing the climate proxy and by using varied weighting for the integration of the different proxies into one historical weighting map.

1. Introduction

Over the past thousands of years land changes enforced by humans have transformed more than 75% of the Earth's ice-free land areas (Ellis & Ramankutty, 2008). Worldwide humans have changed forests, farmlands and waterways to provide for food, fibre, fuel and water (Foley et al., 2005). These land use practices are utterly essential for humanity, since they provide for crucial natural resources and ecosystem services. However, the anthropogenic alterations to the biosphere resulting from land use and cover changes (LUCC) are degrading the ecosystems and services upon which humanity depends. Moreover, the expansion of global croplands, pasture and urban areas has led to extensive global and regional environmental changes, including climatic warming, land degradation, the weakening of ecosystem services, and biodiversity loss (Díaz et al., 2019; Hua et al., 2022).

Since the recognition that large scale LUCC over the last few centuries has global climatic influences and with the arrival of the remote sensing era, there have been numerous attempts by scholars to reconstruct historical land use changes in a spatially explicit way. There are several noteworthy datasets, including the Sustainability and Global Environment (SAGE) dataset (Ramankutty & Foley, 1998), the History Database of the Global Environment (HYDE) dataset (Klein Goldewijk, 2001; Klein Goldewijk et al., 2011, 2017), the Pongratz Julia (PJ) dataset (Pongratz et al., 2008), and the Kaplan and Krumhardt (KK10) dataset (Kaplan et al., 2011). They have been widely used for studying the environmental consequences of long-term LUCC, such as their contribution to climate change and the calculation of carbon emissions at global and regional scale (Ellis et al., 2013). In order to correctly estimate greenhouse gas (GHG) emissions, it is essential that the current and historical LUCC data is accurate. Therefore, it is important to study historical land use data.

The HYDE dataset is one of the most widely used scenarios. One of the main uses is in the creation of the Land-Use Harmonization (LUH) datasets (Hurtt et al., 2020), which are used in global and regional climate models. They are specifically developed to be used for international assessment, for example, in IPCC's 5th Assessment Report and in the latest Coupled Model Intercomparison Project (CMIP6) of the World Climate Research Programme. For the CMIP6 the next generation of Earth System Models (ESM) was developed. These are used to estimate the combined effects of human activities, like land use and fossil fuel emissions, on the climate system. The goal of the LUH project is to make a harmonized set of land use scenarios that smoothly connects the historical LUCC reconstructions with the future projections in the format required for ESMs. As input for these models, the HYDE data is used for the historical period and data from Integrated Assessment Models (IAMs) is used for the future period.

Since the HYDE database is so widely used as a basis for climatic modelling, it is of high importance that the data is accurate. Present land cover data is mainly based on satellite data and is thus very accurate. For this, the ESA Land Cover consortium maps were used (ESA, 2016; Hollman et al., 2013). For the allocation of the historical data, cropland is firstly allocated and is therefore the main land cover class studied in this thesis. This allocation is done by using several proxies and weighting this to the reference map (a satellite map). Further explanation of how this is done can be read in the theory section. A key focus in research on historical land cover changes has been to improve cropland allocation methods to generate credible historical cropland cover data.

The problem with historical land cover data is that there will always be uncertainties. If there are uncertainties in quantifying the impact of land use change on the global carbon cycle, this will lead to uncertainties in atmospheric CO2 projections and climate predictions (Klein Goldewijk et al., 2017). This consequently affects policy makers who use these projections to install reasonable strategies to mitigate carbon emissions. Therefore, this thesis aims to critically reflect on the current allocation methods for historical cropland cover used for HYDE. To study this, the following research question is formulated:

Is the HYDE allocation method for historical cropland cover correct?

Subsequently, the hypothesis is drafted that the allocation method is not completely correct, since many studies, including Klein Goldewijk (2013), the founder of HYDE himself, state that there are still many uncertainties related to the process of historical land use reconstruction. Many studies have shown that these uncertainties could be attributed to the allocation algorithm (Li et al., 2016, 2019; Yang et al., 2015; Yuan et al., 2017; Zhang et al., 2022). The hypothesis will be tested by means of literature review and GIS analysis. Four regions are selected that will be researched: China, Germany, Sub-Saharan Africa and Scandinavia. Supposing that the hypothesis is true, the second research question is defined together with a number of sub-questions:

How can the allocation rules be improved to generate credible historical cropland cover data?

- Are there alternative proxies that should be considered in the allocation of historical cropland?
- Should there be other weighting and/or normalizing methods for the proxies?
- Should the proxies be weighted differently in comparison to each other?

2. Background

To provide an understanding of what this thesis is about, the important concepts and theories related to the research objective must be described. In this section, the key theory and concepts that are used in this research are explained and supported with background information. It is important to comprehend the concept of allocation and the tools that are needed for this, such as weighting proxies and normalizing methods. An explanation of the allocation rules used by HYDE is provided, which is supported by a theoretical framework.

The common method of historical cropland allocation is to calculate the land suitability for cultivation by selecting several physiogeographic factors, then allocate the cropland area into gridded fractions in order of the land suitability from higher to lower (Klein Goldewijk, 2017; Zhang et al., 2022). In general, this is determined by topography, soil, climate conditions and the presence of rivers (Li et al, 2016). HYDE allocated cropland according to six assumptions:

1. *Area*: Urban built-up areas were excluded for allocation. Protected areas were also excluded, but only from 1900 onwards since they did not exist before that time, as was the no-land-use area in Australia.
2. *Population*: A population density of less than 0.1 capita per km^2 does not allow for permanent agriculture.
3. *Suitability*: Land with better soil suitability for crops are used first. The soil suitability proxy is based on the Global Agro-Ecological Zones map developed by the Food and Agriculture Organization of the United Nations (FAO) and the International Institute for Applied Systems Analysis (IIASA).
4. *Distance to water*: Easily accessible areas such as coastlines and river plains are more favourable for early settlement.
5. *Slope*: Inaccessible terrains with steeps slopes are less promising for settlement and agriculture.
6. *Temperature (climate)*: Below an annual mean temperature of 0°C no agriculture is assumed.

These assumptions, symbolized by the input proxies in Figure 1, resulted in weighting maps that were normalized between 0 and 1 and then multiplied into a final weighting map for each time step. Then, this historical weighting map was combined with a satellite map for cropland from the year 2000. As a result, crop area is allocated to the grid cells. This is done at a high resolution of 5 arcmin, meaning that the allocation was carried out for each 5’ x 5’ grid cell. The influence of the historical weighting map decreases gradually starting from 10,000 BC to 2000 AD until the satellite map completely dominates the historical weighting map and the cropland distribution equals the present situation.

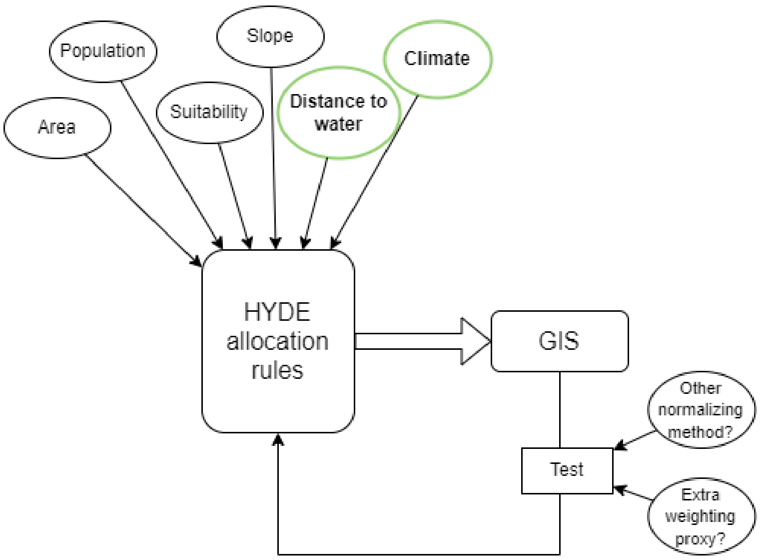

Figure 1: Theoretical framework illustrating the HYDE allocation rules and how the GIS analysis will influence this.

3. Methodology

In this section, the methods are presented that were used in this research to answer the research question. Two methods were used: literature review and land use mapping. The combination of the two methods led to new findings regarding the allocation of historical cropland cover.

3.1 Literature review

Firstly, a literature research was done to study the existing literature and to gain knowledge on what is already known about LUCC mapping, especially about the allocation of cropland cover. The databases Google Scholar, Web of Science and Scopus were employed to gather the scientific literature. This is to ensure that only peer-reviewed articles were included. By reviewing the titles, abstracts and key-words assigned by the author, the relevance of the articles was determined. On these terms, it was ensured that the collected qualitative data reflects the area of research and addresses the knowledge gaps, thus protecting the validity (Taherdoost, 2016). Since the articles were gathered from publicly accessible databases, it is possible for the literature review to be replicated by other researchers. The search terms that were used are ‘historical cropland cover’, ‘HYDE database cropland’, ‘allocation of historical cropland’ and ‘Sahara historical cropland HYDE’. Then, the articles were added to Mendeley, a software that is used to structure and share scientific papers.

The literature review is used to answer the first research question ‘*Is the HYDE allocation method for historical cropland cover correct?*’. For this, articles were gathered that studied the HYDE database and that compare different datasets. Next to that, there are numerous scientific articles that present a comparison of the approximated historical land cover data from HYDE to regional historical cropland reconstructions. These articles were reviewed to study whether the allocation rules used by HYDE are justified or whether there is room for improvement.

Secondly, the literature review tried to answer the second research question '*How can the allocation rules be improved to generate credible historical cropland cover data?*'. The articles were selected based on a few factors: 1) whether HYDE was mentioned in the article; 2) whether it provided new insights in cropland allocation; 3) the region that the study was focused on. From this, the main study areas were derived. These are China, Germany, Scandinavia and Sub-Saharan Africa, with a main focus on China because there were significantly more articles on this study area. Then, within these regions it was studied why the HYDE allocation rules are not correct and whether an explanation could be linked to one of the proxies studied in the GIS analysis. Next to that, proposed alternative methods of cropland allocation that were mentioned in the articles were thoroughly evaluated. Lastly, the proxies used by HYDE were reconsidered in their ability to accurately model the historical cropland cover.

3.2 GIS analysis

The second method that was used during this research is land use mapping, which was used to answer the second research question. This was performed using Geographical Information Systems (GIS), which offers a broad range of tools to analyse the data that is used for the allocation of cropland cover in the HYDE database.

3.2.1 Data collection

For this research, quantitative data was analysed in ArcGIS Pro. The data was obtained from the HYDE database which is openly available at the Netherlands Environmental Assessment Agency (PBL). Specific data was sourced from the founder of the HYDE database, Kees Klein Goldewijk, who is also supervising this thesis. The data that was worked with is the original input data from HYDE that is used to allocate historical cropland cover, so the weighting maps mentioned in chapter 2. What must be noted is that, originally, a different software than GIS was used to make the output data for the HYDE database. The input maps were obtained via Kees Klein Goldewijk, downloaded as .asc files and uploaded in ArcGIS to be analysed. The maps were projected on the WGS 1984 geographical coordinate system, using the 'Define Projection' tool from the ArcGIS Spatial Analyst toolbox.

3.2.2 Analysis of the weighting proxies in ArcGIS Pro

In order to try out different ways of changing the cropland allocation, some tests were done using several GIS tools; this is illustrated in Figure 1. When an allocation method was changed, it is written down and can be found in the GIS results section in chapter 4. Due to limitations of the extent of this research, it was not possible to analyse all of the allocation proxies that are mentioned in Figure 1. The proxies that were studied are climate and distance to water.

CLIMATE

In the current allocation method of HYDE, the climate proxy only consists of a temperature map. However, according to Li et al. (2016), the weight of the climate factor on land suitability for cultivation is reflected by the potential maximum productivity of climate, which is determined by precipitation and temperature. Therefore, in this research the rainfall map was added to the climate proxy next to the already adopted temperature map.

The input map that was used for weighting the temperature factor was the *climate_cr.asc* map, which was the original input map for HYDE that was used for the climate proxy when only the temperature map was taken. Originally, the temperature map ranges from -28 to 30°C. For this research, the map was weighted in order to give a higher value to the more suitable temperatures for agriculture. The areas with higher values thus have a higher probability to contain cropland area. The weighting was realized by using the tool 'Reclassify' and the input values for this tool are shown in Table 1 in the results section. Normalisation was done using the 'Raster Calculator' tool. With this tool the raster map was divided by its highest value to create a normalised raster map with values between 0 and 1.

The rainfall map, *rainfall.asc*, first had to be converted from 0.5° x 0.5° grid size to 5' x 5'. This was done with the tool 'Resample' and with the input value of 0.0833 for both X and Y. The rainfall map did not have to be weighted since the more rainfall, the higher the land suitability for cultivation. The map was normalised in the same way as the temperature map. After both the temperature and the rainfall maps were weighted and normalised, they were merged together to create the weighted climate map. The integration of the two maps is shown in the results section.

DISTANCE TO WATER

The other proxy that was studied was the distance to water proxy. The map that was used as input was the *gdistwa5.asc* map. First, the map was converted to a 5 arcmin resolution (from 0.5° x 0.5° grid size to 5' x 5') in the same way as the rainfall map. Then the map had to be weighted and normalised. The input map has a negative correlation with the cropland fractions, since the higher the value of a certain grid area, the further away from water it is. Therefore, the map was changed in order to have a map where a higher value means that the grid area is closer to water. To make this happen, the map was weighted and normalised using formula 1 in the raster calculator. This method was also used by HYDE to create the normalised distance to water map, so it is not a change to the allocation method.

$$\frac{x_{max} - r_{distw}}{x_{max} - x_{min}} \qquad (1)$$

Where r_{distw} refers to the input raster, *gdistwa5_5arcmin.asc*; x_{max} is the maximum value of the *gdistwa5.asc* map; and x_{min} is the minimum value of the *gdistwa5.asc* map.

INTEGRATION

When both of the proxies were weighted and normalized, they were integrated into a final historical weighting map. For the integration, three different weighting methods were tested. This is one of the changes that was made in the allocation rules, thus, it will be explained in the results section. The input for this was the normalised distance to water map obtained earlier and the normalised climate map obtained from the temperature and rainfall mergence.

COMPARISON TO HYDE

Then, to analyse the changes that were made to the allocation methods, they were compared to the current method of HYDE. For this, a map was created with the same two proxies and with the methods used by HYDE. The original maps from HYDE were taken; for the distance to water proxy this was the normalised map *n_distw.asc*; and for the climate proxy this was *climate_cr.asc*. The climate map still had to be normalised. This was done by first making sure that every temperature below 0 was set to 0. Then the raster map was divided by its highest value using the raster calculator. The integration of the two proxies was done by multiplying the weighting maps. Then this map was compared with the maps that came out of the integration.

4. Results

In this chapter, the key findings of this research are presented from the literature review and the GIS analysis.

4.1 Literature review

During the literature review articles were analysed that compared regional historical landcover data to the HYDE database. The goal of the literature review was 1) to answer the first research question '*Is the HYDE allocation method for historical cropland cover correct?*'; and 2) to find specific proxies to focus on during the GIS analysis by answering the second research question and its sub-questions '*How can the allocation rules be improved to generate credible historical cropland cover data?*'.

From the literature review it becomes apparent that allocating historical landcover data is a very complex matter. All of the studies that were analysed indicated that the total amount of cropland areas that was accounted for by HYDE was generally in correspondence with regional document-based reconstructions. However, the differences in the spatial patterns, e.g. the location of the cropland cover, were substantial. The main point of critic that was found in the reviewed articles was that the HYDE 3.2.1 dataset is predominantly based on algorithms that rely on present-day environmental variables. Moreover, there is a lack of human indicators in the allocation methods.

4.1.1 Human indicator

The study of Wu, Fang & Ye (2022) suggests the introduction of a new proxy, which provides a possible answer for the first sub-question '*Are there alternative proxies that should be considered in the allocation of historical cropland?*'. They designed a new allocation method that incorporates human factors which could reflect both the location and the chronological order of human cultivation activities. The approach is a settlement density based allocation method for historical cropland. Firstly, settlements were selected on the basis of various human factors associated with the reclamation of cropland. Then, settlement density was used to conduct a method for gridded allocation of historical cropland cover. The method was tested in a case study in the Jilin Province of China.

The allocation method worked well for the case study as it provided a better regional cropland pattern. The study also advocated the potential of the method to be applied for global historical cropland cover reconstruction. Wu et al. (2022) state that there are abundant records of historical settlements, which can mainly be sourced from gazetteers, ancient maps and archaeological data. However, it will be very time consuming in order to gather worldwide settlement data, that is also publicly available. Therefore, it is questionable whether it is feasible to integrate this allocation method into the HYDE model. Still, it is an interesting theory to take into account when redesigning the allocation methods for HYDE.

This approach relates to the study of Zhang, Fang, Liang & Yang (2021) that concluded that a good strategy for cropland allocation would be to select representative human indicators on a feasible spatial scale according to the differences and similarities between regions. Then, with these indicators, a wider scope LUCC reconstruction could be realized by combining the expertise of the humanities and proper modelling techniques. Furthermore, Widgren (2018) also suggests to incorporate a human factor in the allocation process. This study attempted the mapping of global agrarian systems instead of precise land use cover. Thus, multiple articles suggest including an extra weighting proxy that somehow reflects the human influence.

4.1.2 Land suitability

Multiple studies suggest that when the natural factors are used as the basis of the allocation, it is best to use the dominant physiogeographic factors per region (Wu et al., 2022; Zhang et al., 2022). This relates to the sub-question '*Should the proxies be weighted differently in comparison to each other?*'. For example, the study of Li, He & Zhang (2016) showed that the weight of the distance to water is overestimated by HYDE for China. Therefore, a suggestion would be to give less weight to this proxy in the GIS analysis and in the future for a new HYDE model. Next to this, Wei, Widgren & Li et al. (2021) concluded that for Scandinavia specifically, some factors can be excluded since these do not show much deviation. They state that elevation and slope should be the leading physiogeographic factors in this region, since climate is relatively the same and soil properties are not a limitation. Furthermore, for Germany and China the cropland patterns along rivers from HYDE do not correspond with reconstructions based on region specific historical records (Wei et al., 2019; Yuan et al., 2017; Zhang, He, & Li, 2013; Zhang, Fang, Liang, & Yang, 2021). In conclusion, the dominant physiogeographic factors vary from region to region and should thus be weighted distinctively per region according to the differentiation of the physiogeographic environment.

4.2 GIS analysis

With the GIS analysis some of the proxies used by HYDE were analysed. The GIS analysis was done to answer the second research question '*How can the allocation rules be improved to generate credible historical cropland cover data?*'. The sub-questions helped to structure the research. Multiple changes to the allocation method were done. Firstly, the weighting and normalisation of the climate proxy was changed in two ways: the temperature map was weighted differently and a rainfall map was added. Secondly, the integration of the different proxies into one historical weighting map was changed. Here the results of the GIS analysis are presented.

4.2.1 Weighting and normalisation of the climate proxy

The second sub-question is '*Should there be other weighting and/or normalizing methods for the proxies*'. To answer this sub-question, a GIS analysis was done with the climate proxy. HYDE weighted the land suitability for cultivation of the climate proxy by only taking the temperature map and then normalizing it. The normalization by HYDE was done with the highest temperature representing higher suitability for cultivation and areas with a temperature below 0 were excluded. For this research, the suitability of the climate proxy was weighted by combining the temperature map together with the rainfall map. Furthermore, the temperature map was weighted differently because higher temperatures are not necessarily better for cultivation.

The weighting of the temperature map was done by firstly researching the optimal temperature for cultivation. There was no clear answer for this coming from the literature. However, the Climate Indicators for Agriculture report said that nightly temperatures above 68F (20°C) result in lower crop yields (Walsch et al., 2020). Next to this, Van Velthuizen (2007) stated that areas with temperatures lower than 5°C are not suitable for agriculture. For this research, it is not of high importance what the actual optimal temperature for cropland cultivation is. Moreover, the aim of this part of the research is to show the effect of a different reclassification method on the outcome of the final land suitability map. It is evident that the highest temperature is not the best for agriculture, however, it is unclear what average annual temperature is the most optimal for cropland cultivation. Therefore, 15°C was taken as an example for this research. Subsequently, the lower and higher temperatures were weighted to this optimal temperature. The reclass table that was used to weigh the temperature map is represented in

Table 1. The weighted map for temperature as well as the original map can be found in Figure A1 and A2 in the appendix.

Table 1: Weighting of temperature proxy

Old value (°C)	**New value (°C)**
-28 – 0	0
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	**15**
16	14
17	13
18	12
19	11
20	10
21	9
22	8
23	7
24	6
25	5
26	4
27	3
28	2
29	1
30	0

For the climate proxy, the normalised temperature map was merged with the rainfall map. Figure 2 and 3 represent the weighted and normalized suitability maps for the climate proxy, resulting from the mergence. Two different weighting methods were used to integrate the climatic factors into one map. The first method that was used for Figure 2 is the multiplying method. For Figure 3 the square root method was used. The formulas for these methods are respectively formula 2 and formula 3 for Figure 2 and Figure 3.

$$W_{cl} = r_{temp} \times r_{rain} \qquad (2)$$

$$W_{cl} = \sqrt{r_{temp} \times r_{rain}} \qquad (3)$$

Where r_{temp} is the normalised temperature raster map; r_{rain} is the normalised rainfall raster map; and W_{cl} is the weighted and normalized suitability map for the climate proxy, represented by *n_climate_mult* and *n_climate_sqr* in Figure 2 and 3 respectively.

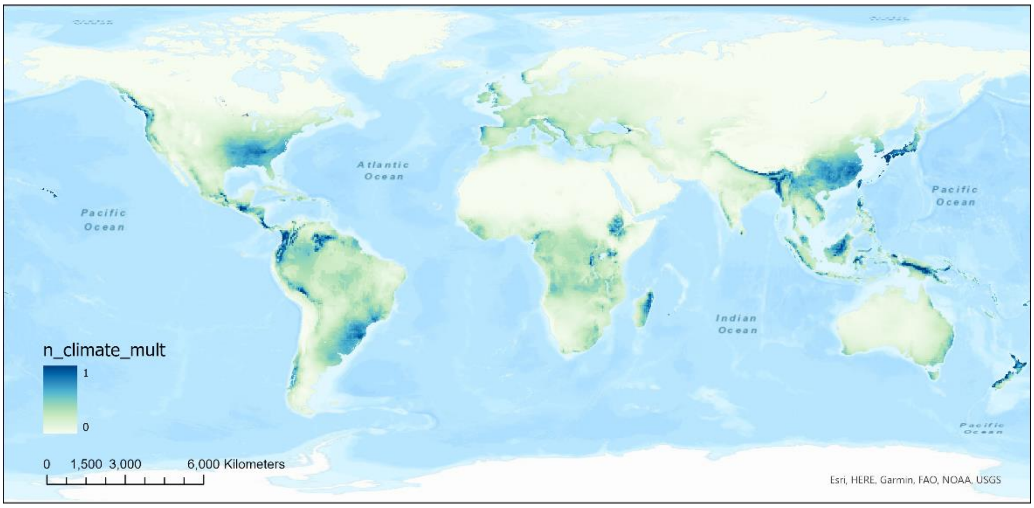

Figure 2: Normalized suitability map for the climate proxy, weighted with the multiplying method.

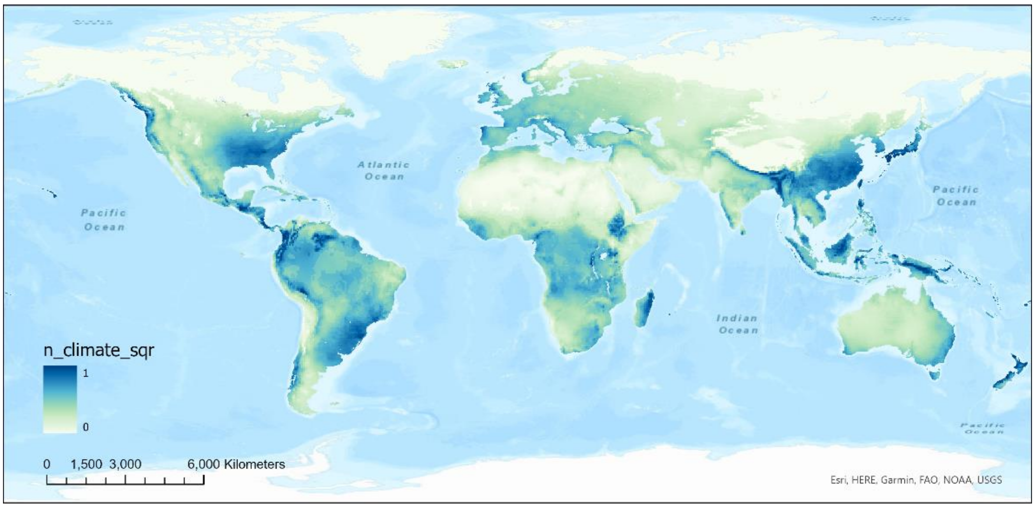

Figure 3: Normalized suitability map for the climate proxy, weighted with the square root method.

There is a clear difference visible between the two maps. The map in Figure 3 has more (dark) blue areas than the map in Figure 2. This means that the square root method tends to cause more allocation of cropland; that is, more cropland area will be allocated to areas with high climatic land suitability. The square root method makes it more dispersed, while with the multiplying method the values are lower and closer together. In comparison to the original climate proxy from HYDE (Figure A1), where only the temperature map was taken into account and the highest temperature meant the highest land suitability, the climate proxy from this study shows a different distribution of the land suitability.

4.2.2 Integration of different physiogeographic factors into one weighting map

To answer the third sub-question, '*Should the proxies be weighted differently in comparison to each other*', two proxies were taken as research examples. The weighted and normalized climate map and distance to water map were merged together into one map. This is done to demonstrate the different ways in which the proxies can be integrated into one weighting map reflecting the total land suitability, which is used to allocate the historical cropland cover. In this research, three methods of integrating the different physiogeographic factors into one ultimate weighting map for cropland cover were tested. The methods are reflected in formula 4, 5 and 6 and the resulting maps are presented respectively in Figure 4, 5 and 6.

$$r_1 \times r_2 \times \ldots \times r_n \qquad (4)\ \text{multiplying}$$

$$\sqrt[n]{r_1 \times r_2 \times \ldots \times r_n} \qquad (5)\ n\text{th root}$$

$$\sqrt[(a+b+\cdots+z)]{{r_1}^a \times {r_2}^b \times \ldots \times {r_n}^z} \qquad (6)\ n\text{th root with varied weighting}$$

Where r_1 reflects one of the raster maps of a physiogeographic factor; n reflects the number of physiogeographic factors involved in calculating the land suitability; and *a*-*z* reflect the weighting of a certain factor.

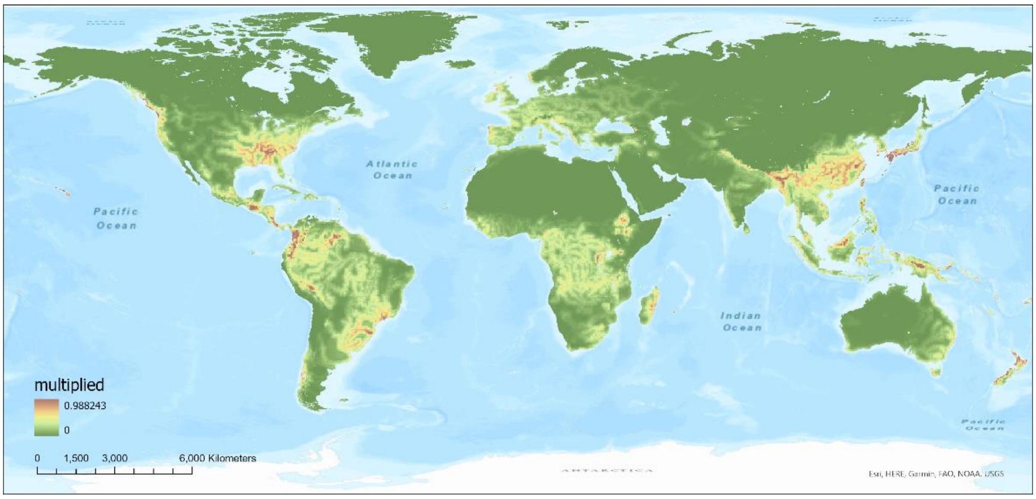

Figure 4: Integrated weighting map (climate and distance to water) using the multiplying method.

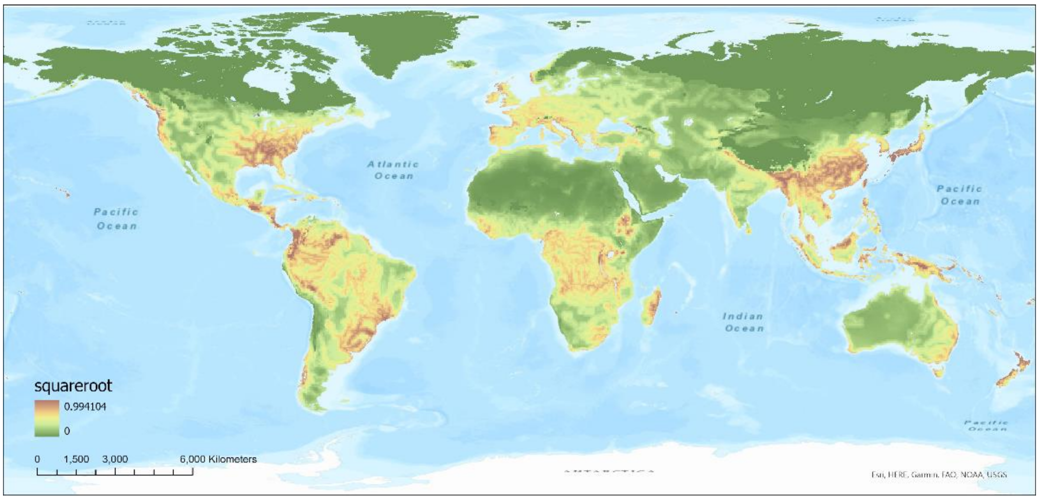

Figure 5: Integrated weighting map (climate and distance to water) using the nth *root method.*

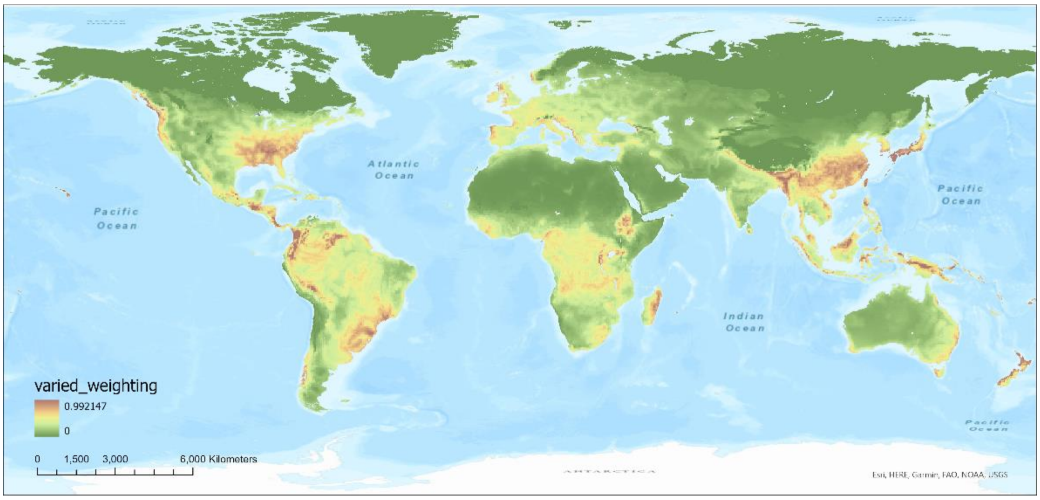

Figure 6: Integrated weighting map (climate and distance to water) using the nth *root method with varied weighting.*

It is visible that there is a difference in output between the integration methods. The multiplying method results in less allocation of cropland. This is already the case when two proxies are included and will be more so when there are multiple proxies in the calculation. The *n*th root method assures that it is more evenly spread out. However, in Figure 5 it is visible that there is mainly allocation near the river patterns, although studies suggest that this factor should not be given this much weight. Therefore, with the third method, it is possible to give a different weighting to the various proxies. In the literature it was also found that this is important in the allocation of historical cropland (Li et al., 2016; Wu et al., 2022; Zhang et al., 2022).

4.2.3 Comparison to HYDE

The map in Figure 7 shows how the two proxies would be integrated using the methods from HYDE. What stands out is that in this map the cropland pattern is almost exclusively located along the rivers and coastlines. The areas around rivers with higher temperatures have a darker coloured pattern, which means that more cropland would be allocated to these areas in comparison with the lighter coloured patterns. It is interesting to see that there is such a clear difference between Figure 7 and Figure 4, where the same integration method (multiplying) is used. Here it is visible what the influence is of changing the weighting of the temperature and adding a rainfall map.

Figure 7: Reconstruction of the integrated weighting map from HYDE.

5. Discussion

In this section the key findings are discussed and a reflection of the changes to the allocation methods is provided. Next to this, the limitations of this research are mentioned as well as suggestions for future research.

5.1 Key findings

The aim of the research was to critically reflect on the allocation method for historical cropland cover used for HYDE. The research question can be answered as follows: the HYDE allocation method for historical cropland cover could be improved to be completely correct, but for now it is the best approach to globally estimate historical LUCC. The literature research has shown that there are several suggestions for alternative methods of cropland allocation. However, these methods are very difficult and time consuming when they are to be implemented on a global scale. Therefore, the feasibility of implementing a new factor such as settlement density has to be further analysed.

The GIS analysis answers the second research question. The allocation rules can be improved by changing the climate proxy. In the analysis the weighting of the temperature map is changed. For the sake of this research, 15°C is taken as the optimal temperature to be able to clearly show the difference. For the actual implementation of this change in allocation rule, it should be further investigated what the real average annual temperature is for optimal cultivation. This research showed how this optimal temperature can be used to weigh the temperature map to and, with the addition of the rainfall map, how this influences the climate proxy to accurately represent the land suitability for cultivation. The GIS analysis as well as studies from the literature review suggest to add the rainfall component to the climate proxy to improve the allocation method.

Another finding of this research is that the allocation proxies should be weighted differently to one another. The literature review concluded that the dominant proxies vary from region to region. Thus, it would be best to have an allocation algorithm based on the regional land suitability of cultivation. In this GIS analysis it was tested how this differentiation of physiogeographic factors could be realised. Moreover, three different ways of integrating these factors were tested. The *n*th root with varied weighting method allows the proxies to be weighted differently in comparison with each other. The next step is to enhance this method by differentiating the weighting per region.

5.2 Implications and suggestions

The different integration methods evidently result in different weighting maps for land suitability. In this research, climate and distance to water were taken as examples to test the different methods. In order to find out which methods works best in accurately reconstructing historical cropland cover, the methods should be tested with all of the proxies in the calculation. Then, the weighting map that comes out of this has to be run through the HYDE model to be able to compare the output, historical cropland maps, with regional reconstructions and with older versions of HYDE. Since it was not possible to execute this extensive research in the scope of this study, it is a suggestion for future research. Next to this, future research aims at discovering which proxies should be weighted more heavily.

In general, the HYDE allocation methods do not result in the most spatially accurate historical cropland reconstructions. However, this is not the goal of HYDE. It is used for 'the bigger picture', i.e. global climate models, meaning that the overall trend should be accurate, but the specific allocation does not have to be perfect. Regardless, the uncertainties related to the dataset in terms of its proxies and input data lead to a higher risk of faulty use of the data as input for further assessment. Users often misinterpret the modelling results as actual reconstructions, which poses a challenge for the research community.

This research has had added value because it is necessary to critically reflect on existing research and methods. Next to this, HYDE is used for influential climate research; therefore, it is important that the allocation is as accurate as possible. Thus, it remains essential to look into better ways of globally allocating historical LUCC. This can only be achieved by more interdisciplinary collaboration between different specialties and fields of science. For example, archaeological evidence is very useful as empirical data for the reconstruction of cropland. Disciplines such as historical ecology can contribute making the many existing local case studies more accessible to other researchers.

6. Concluding remarks

This thesis aimed to critically reflect on the HYDE allocation rules for historical cropland cover. Based on the literature review it can be concluded that the current allocation methods could be improved to be completely correct. The results of the GIS analysis established that this can be achieved by changing the climate proxy and using varied weighting for the integration of the proxies into the final historical weighting map. The climate proxy can best be indicated by rainfall and temperature and weighting the temperature proxy to the optimal temperature for cultivation instead of the highest temperature. From the literature review the suggestion of a human indicator came forward; however, the feasibility of implementing this on a global scale is yet to be researched. Next to this, the literature review as well as the GIS analysis concluded that the allocation proxies should be weighted differently to one another, even varying from region to region. The conducted research provides suggestions for improvement of the allocation rules which will hopefully be taken into consideration when the next version of HYDE will be realised.

7. Acknowledgements

I would like to thank my supervisor Kees Klein Goldewijk for helping me with this research. Without your input, provision of the according data and explicit guidance in the GIS analysis I could not have carried out this research. I also appreciate the steering feedback and the freedom you gave in shaping my own research. I would also like to thank my girlfriend Kira for supporting me mentally throughout the process and for giving advice when I got stuck.

8. References

Díaz, S., Settele, J., Brondízio, E. S., Ngo, H. T., Agard, J., Arneth, A., Balvanera, P., Brauman, K. A., Butchart, S. H. M., Chan, K. M. A., Lucas, A. G., Ichii, K., Liu, J., Subramanian, S. M., Midgley, G. F., Miloslavich, P., Molnár, Z., Obura, D., Pfaff, A., … Zayas, C. N. (2019). Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science*, *366*(6471). https://doi.org/10.1126/SCIENCE.AAX3100/SUPPL_FILE/AAX3100-DIAZ-SM.PDF

Ellis, E. C., Kaplan, J. O., Fuller, D. Q., Vavrus, S., Klein Goldewijk, K., & Verburg, P. H. (2013). Used planet: A global history. *Proceedings of the National Academy of Sciences*, *110*(20), 7978–7985. https://doi.org/10.1073/PNAS.1217241110

Ellis, E. C., & Ramankutty, N. (2008). Putting people in the map: anthropogenic biomes of the world. *Frontiers in Ecology and the Environment*, *6*(8), 439–447. https://doi.org/10.1890/070062

ESA, 2016. Three global LC maps for the 2000, 2005 and 2010 epochs, European Space Agency, available at: https://maps.elie.ucl.ac.be/CCI/viewer/index.php

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., Stuart Chapin, F., Coe, M. T., Daily, G. C., Gibbs, H. K., Helkowski, J. H., Holloway, T., Howard, E. A., Kucharik, C. J., Monfreda, C., Patz, J. A., Colin Prentice, I., Ramankutty, N., & Snyder, P. K. (2005). Global

Consequences of Land Use. *Science*, *309*(5734), 570–574. https://doi.org/10.1126/science.1111772

Hollmann, R., Merchant, C. J., Saunders, R., Downy, C., Buchwitz, M., Cazenave, A., Chuvieco, E., Defourny, P., De Leeuw, G., Forsberg, R., Holzer-Popp, T., Paul, F., Sandven, S., Sathyendranath, S., Van Roozendael, M., & Wagner, W. (2013). The ESA Climate Change Initiative: Satellite Data Records for Essential Climate Variables. *Bulletin of the American Meteorological Society*, *94*(10), 1541–1552. https://doi.org/10.1175/BAMS-D-11-00254.1

Hua, L., Li, S., Gao, D., & Li, W. (2022). Uncertainties of Global Historical Land Use Datasets in Pasture Reconstruction for the Tibetan Plateau. *Remote Sensing*, *14*(15), 3777. https://doi.org/10.3390/rs14153777

Hurtt, G. C., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., Doelman, J. C., Fisk, J., Fujimori, S., Klein Goldewijk, K., Hasegawa, T., Havlik, P., Heinimann, A., Humpenöder, F., Jungclaus, J., Kaplan, J. O., Kennedy, J., Krisztin, T., Lawrence, D., … Zhang, X. (2020). Harmonization of global land use change and management for the period 850-2100 (LUH2) for CMIP6. *Geoscientific Model Development*, *13*(11), 5425–5464. https://doi.org/10.5194/GMD-13-5425-2020

Kaplan, J. O., Krumhardt, K. M., Ellis, E. C., Ruddiman, W. F., Lemmen, C., & Klein Goldewijk, K. (2011). Holocene carbon emissions as a result of anthropogenic land cover change. *Https://Doi.Org/10.1177/0959683610386983*, *21*(5), 775–791. https://doi.org/10.1177/0959683610386983

Klein Goldewijk, K. (2001). Estimating global land use change over the past 300 years: The HYDE Database. *Global Biogeochemical Cycles*, *15*(2), 417–433. https://doi.org/10.1029/1999GB001232

Klein Goldewijk, K., Beusen, A., Doelman, J., & Stehfest, E. (2017). Anthropogenic land use estimates for the Holocene - HYDE 3.2. *Earth System Science Data*, *9*(2), 927–953. https://doi.org/10.5194/essd-9-927-2017

Klein Goldewijk, K., Beusen, A., Van Drecht, G., & De Vos, M. (2011). The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Global Ecology and Biogeography*, *20*(1), 73–86. https://doi.org/10.1111/j.1466-8238.2010.00587.x

Klein Goldewijk, K., & Verburg, P. H. (2013). Uncertainties in global-scale reconstructions of historical land use: An illustration using the HYDE data set. *Landscape Ecology*, *28*(5), 861–877. https://doi.org/10.1007/S10980-013-9877-X/FIGURES/10

Li, S., He, F., & Zhang, X. (2016). A spatially explicit reconstruction of cropland cover in China from 1661 to 1996. *Regional Environmental Change*, *16*, 417–428. https://doi.org/10.1007/s10113-014-0751-4

Li, S., He, F., Zhang, X., & Zhou, T. (2019). Evaluation of global historical land use scenarios based on regional datasets on the Qinghai–Tibet Area. *Science of the Total Environment*, *657*, 1615–1628. https://doi.org/10.1016/J.SCITOTENV.2018.12.136

Pongratz, J., Reick, C., Raddatz, T., & Claussen, M. (2008). A reconstruction of global agricultural areas and land cover for the last millennium. *Global Biogeochemical Cycles*, *22*(3). https://doi.org/10.1029/2007GB003153

Ramankutty, N., & Foley, J. A. (1998). Characterizing patterns of global land use: An analysis of global croplands data. *Global Biogeochemical Cycles*, *12*(4), 667–685. https://doi.org/10.1029/98GB02512

Taherdoost, H. (2016). Validity and reliability of the research instruments: How to test the validation of a questionnaire/survey in a research. *International Journal of Academic Research in Management*, *5*(3), 28-36. https://doi.org/10.2139/ssrn.3205040

Van Velthuizen, H. (2007). *Mapping biophysical factors that influence agricultural production and rural vulnerability*. Food & Agriculture Org.

Walsh, M., Backlund, P., Buja, L., DeGaetano, A., Melnick, R., Prokopy, L., Takle, E., Todey, D., & Ziska, L. (2020). Climate Indicators for Agriculture. In *USDA Technical Bulletin* (Vol. 1953). https://doi.org/10.22004/AG.ECON.320800

Wei, X., Widgren, M., Li, B., Ye, Y., Fang, X., Zhang, C., & Chen, T. (2021). Dataset of 1 km cropland cover from 1690 to 1999 in Scandinavia. *Earth System Science Data*, *13*(6), 3035–3056. https://doi.org/10.5194/essd-13-3035-2021

Wei, X., Ye, Y., Zhang, Q., Li, B., & Wei, Z. (2019). Reconstruction of cropland change in North China Plain Area over the past 300 years. *Global and Planetary Change*, *176*, 60–70. https://doi.org/10.1016/J.GLOPLACHA.2019.01.010

Widgren, M. (2018). Mapping Global Agricultural History: A Map and Gazetteer for Sub-Saharan Africa, c. 1800 AD BT. In A. M. Mercuri, A. C. D'Andrea, R. Fornaciari, & A. Höhn (Eds.), *Plants and People in the African Past* (pp. 303–327). Springer International Publishing. https://doi.org/10.1007/978-3-319-89839-1_15

Wu, Z., Fang, X., & Ye, Y. (2022). A Settlement Density Based Allocation Method for Historical Cropland Cover: A Case Study of Jilin Province, China. *Land*, *11*(8), 1374. https://doi.org/10.3390/LAND11081374

Yang, X., Jin, X., Guo, B., Long, Y., & Zhou, Y. (2015). Research on reconstructing spatial distribution of historical cropland over 300years in traditional cultivated regions of China. *Global and Planetary Change*, *128*, 90–102. https://doi.org/10.1016/J.GLOPLACHA.2015.02.007

Yuan, C., Ye, Y., Tang, C., & Fang, X. (2017). Accuracy comparison of gridded historical cultivated land data in Jiangsu and Anhui provinces. *Chinese Geographical Science*, *27*(2), 273–285. https://doi.org/10.1007/S11769-017-0862-1/METRICS

Zhang, C., Fang, X., Ye, Y., Tang, C., Wu, Z., Zheng, X., Zhang, D., Jiang, C., Li, J., Li, Y., & Zhao, Z. (2022). A spatially explicit reconstruction of cropland cover in China around 1850 C.E. employing new land suitability based gridded allocation algorithm. *Quaternary International*, *641*, 62–73. https://doi.org/https://doi.org/10.1016/j.quaint.2022.06.001

Zhang, D., Fang, X., Liang, &, & Yang, E. (2021). Comparison of the HYDE cropland data over the past millennium with regional historical evidence from Germany. *Regional Environmental Change*, *21*(15). https://doi.org/10.1007/s10113-020-01735-1/Published

Zhang, X.-Z., He, F.-N., & Li, S.-C. (2013). Reconstructed cropland in the mid-eleventh century in the traditional agricultural area of China: implications of comparisons among datasets. *Regional Environmental Change*, *13*(5), 969–977. https://doi.org/10.1007/s10113-012-0390-6

Appendix A

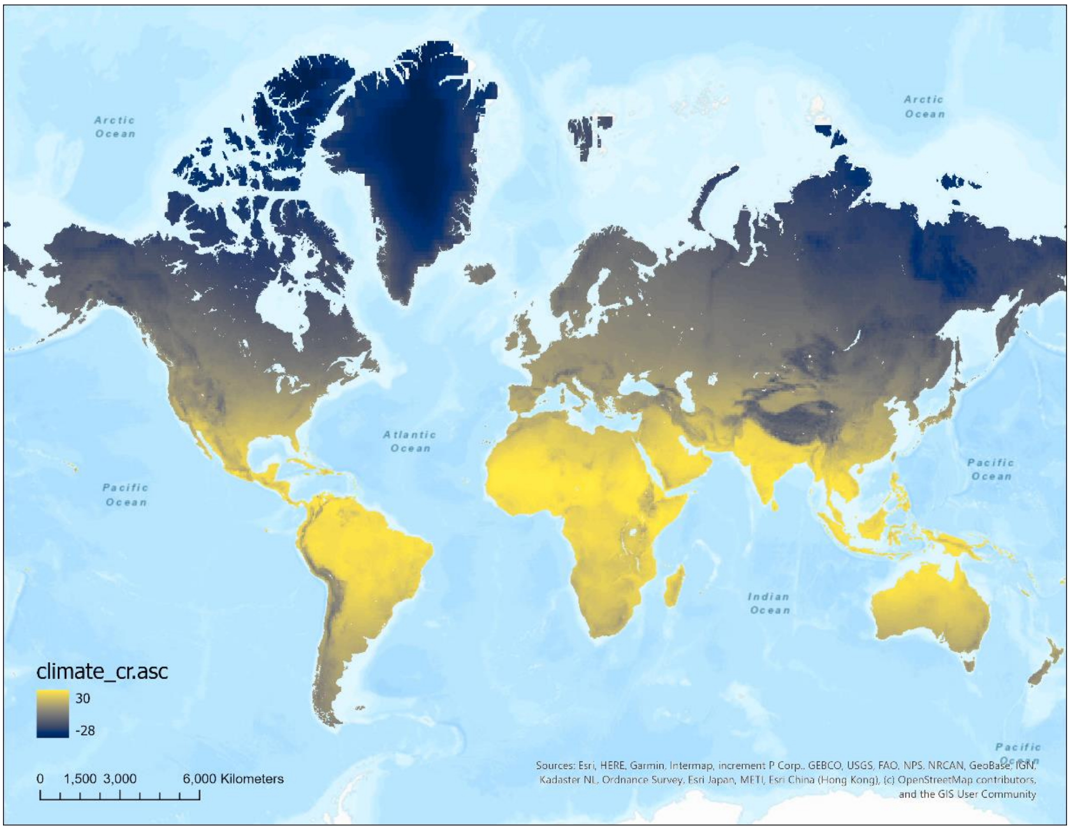

Figure A1: Original temperature map

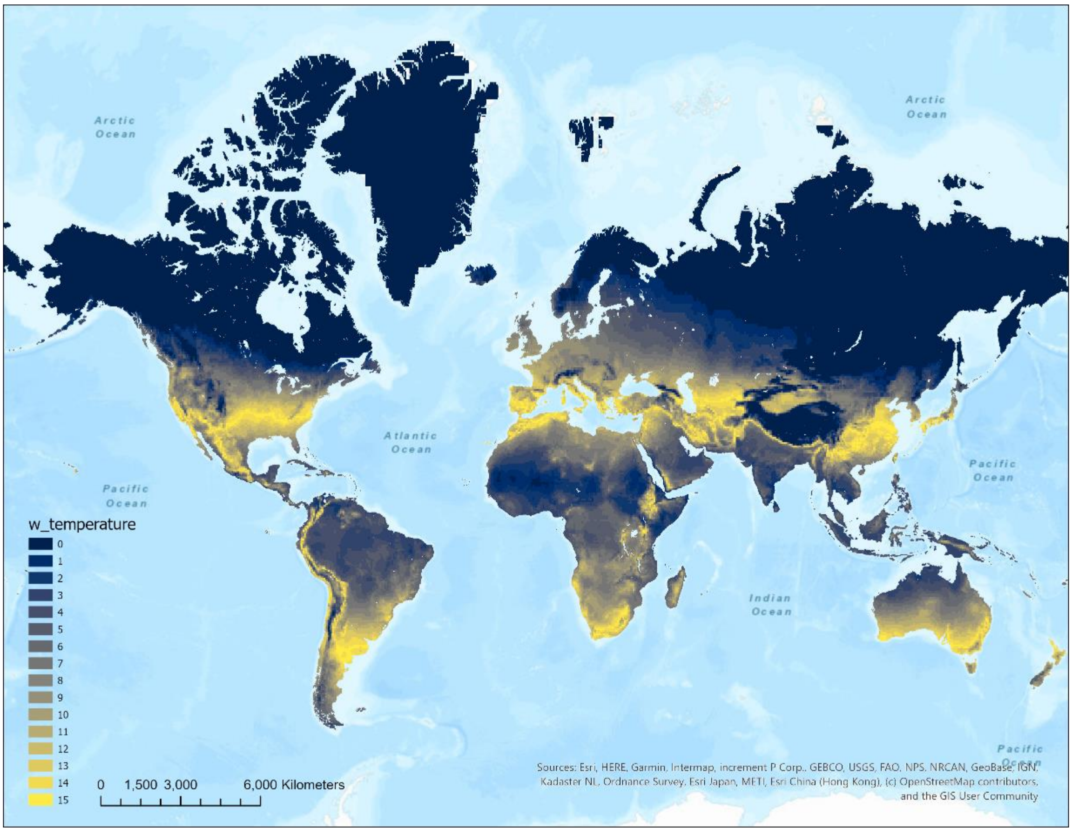

Figure A2: Weighted temperature map