Bayesian Hierarchical Model for estimating holocene prehistoric population development on the Swiss Plateau

Martin Hinz[1,2,*] Joe Roe[1,2] Julian Laabs[3] Caroline Heitz[1,2,4] Jan Kolář[5,6]

09 April, 2022

Abstract

In order to understand human-environment relationships in the past and the dynamics of socio-ecological systems, it is essential to have a robust estimate of population sizes. Size and density of population are crucial factors for the reconstruction of group sizes, the character and range of political institutions for social organization, the constraints and possibilities of economic practice, the number of people available for collective activities, for economic and social exchange systems, and the creation and maintenance of collective identities. And, of course, the human impact on the environment. Most estimates of population sizes currently work on limited spatial scales or use proxies that hardly allow for absolute numbers (e.g. sum calibration). The figures in the archaeological discourses on population density in general currently come from vague sources and are often decades old. In demography in general, the last decade has triggered an upswing in the application of Bayesian methods, so that a Bayesian demography has been announced. Bayesian demography is currently at the forefront of methodological developments in this area, but has reached such maturity that the United Nations has been using it since 2015 for population forecasts and projections. The Bayesian approach offers the possibility of combining heterogeneous data and at the same time qualifying them in terms of uncertainty and credibility. This is precisely where it becomes very interesting for archaeological data, since they are mostly inaccurate, biased, sparse, simplified and often based on simplistic assumptions. In this contribution, issues and possibilities of the application of Bayesian demography in the field of archaeological (prehistoric) population size and density estimation will be discussed and the results of a pilot study will be presented, which should lead to a more comprehensive project for the reconstruction of demographic developments in prehistory on a large scale.

[1] Institute of Archaeological Sciences, University of Bern
[2] Oeschger Centre for Climate Change Research, University of Bern
[3] CRC 1266 - Scales of Transformation, University of Kiel
[4] Associated Researcher, School of Archaeology, University of Oxford
[5] Department of Vegetation Ecology, Institute of Botany of the Czech Academy of Sciences
[6] Institute of Archaeology and Museology, Faculty of Arts, Masaryk University

[*] Correspondence: Martin Hinz <martin.hinz@iaw.unibe.ch>

Keywords: keyword 1; keyword 2; keyword 3

Highlights: These are the highlights.

1 Introduction

The study of population size and population density as important demographic factors has a long history in archaeology. Although questions of population extent and distribution have been dominant fields since the beginning of archaeological research, the lack of access to this information, data on it, or methodological tools for the use of proxies has long made the statements made on this topic for prehistoric times vague and superficial (Hassan, 1981), and they were often based on uncritical transfer of ethnographic parallels.

Gordon Childe (1936) was one of the first to stress the importance of estimating the density and characteristics of prehistoric populations and sought to examine the role of human populations in the course of cultural evolution (Hassan, 1981).

Attempts to estimate the size and density of prehistoric populations from archaeological remains were made by Hack (1942), Colton (1949), Frankfort (1950) and Cook (1946). Cook and Heizer (1965, 1968) and Naroll (1962) provided rules derived from ethnographic contexts for estimating the size of prehistoric populations. Cook and Heizer (1966) made the first attempt to estimate the population growth rate during the Neolithic (Hassan, 1981).

At the end of the 1960s, however, with the emergence of the processual archaeology, such investigations became increasingly popular. It was the rise of the ecological paradigm that gave even greater weight to population studies, especially those dealing with population ecology (Hassan, 1981).

After research interest turned increasingly to other aspects with the emergence of post-process trends, a strong boom in the study of population sizes can be observed especially since the 2010s (or even earlier, (Shennan, 2000)). This is probably due, on the one hand, to the new shift towards the study of human-environment relationships, for which an assessment of the size of human populations is indispensable. This is true especially for an evaluation of the human impact on the natural environment. Another reason is certainly the emergence of the 'dates as data' approach (Rick, 1987), which only really took off with the publications around the London UCL group. Moreover, Kintigh et al. (2014) listed human influence, dominance and population size and growth as one of the key elements where methodological improvements are necessary in the article on the Grand Challenges for archaeology, in the area of 'human-environment interactions'.

Müller and Diachenko (Müller and Diachenko, 2019) has compiled a list of currently used approaches and proxies for estimating prehistoric population sizes. These can be roughly divided into ethnographic parallels, ecologic-economic deductive estimates, and the counting and spatial interpolation of archaeological features (settlements, houses, individual find types). From this short literature review three basic problems are common to all these approaches and the studies based on them:

1. Proxy is based on a single line of evidence. Although multiproxy approaches exist, in such cases the individual proxies only serve to support each other or the proxy used in the main case. There is no combination of results to increase the accuracy of each of the inaccurate and biased proxies.
2. The uncertainty of evidence is not adequately reflected. In most cases, individual curves are presented as estimates, but the size of the error bars for these estimates is almost never apparent.
3. Nature of the archaeological data is taken into account only insufficiently. What is meant here is that they take little account of the inherent uncertainties of the proxies themselves, for lack of an evaluative framework, and that transfer functions from the proxy to the actual size to be observed, i.e. population size or density, are hardly ever evaluated more intensively. In the best case, a critical appraisal of the informative value is provided, but a quantification of the relationship between, for example, a doubling of a proxy value and the change in population density does not take place and cannot take place for lack of a suitable framework or external data for calibration.

Beispielebringen?Bringsomeexampleshere...

Archaeological data used to estimate population trends have the following characteristics:

- Limited: We have only incomplete data that can be used for these purposes, and they are usually very low in information value.
- Unevenly distributed: Although there is a good data on settlement frequencies for some regions, and these are sometimes in very high temporal resolution, such favorable situations are very unevenly distributed over time and space.
- Noisy: Often other influences, not related to population development, affect the characteristics of individual proxies. Preservation conditions, depositional behavior or social aspects should be mentioned here as an example.
- Unreliable: Comprehensive information and derived data are influenced by research strategies. Systematic distortions are possible or rather the rule.

- Highly heterogeneous: The number of artefacts at a site, the number of houses, the demographic conditions in cemeteries, the estimates of human impact based on pollen analyses: All these are data that may serve our purposes, but which are available in completely different spatio-temporal scales, granularities and information values, and simply represent technically very different data formats and scales.
- Indirect: It is actually always proxy data that are collected as substitute values for the information that is actually needed (i.e. population density or size). The transfer functions that link the collected data with the data to be obtained are unknown and are at best established by observing the (also ethnological) presence. However, since living conditions, circumstances and economic parameters differ greatly between the present time and prehistory, these transfer functions can only be evaluated as a first approximation.
- Contradicting: Particularly when considering several proxies, differences in transfer functions and data quality and noisiness inevitably lead to contrasts. However, this usually only leads to the rejection of a proxy for a certain time period, without mutually supporting or contradictory information leading to a quantitative evaluation of the proxy and its uncertainties.

In recent years, a new method has been established in the field of demographic estimations. Bayesian demographic forecasting has been designed to solve all these problems, which also exist for demographic estimations using recent data. In their book on the subject, Bryant and Zhang (Bryant and Zhang, 2018) state that Bayesian data modeling can be used to deal with exactly those problems that also affect archaeological data. Especially for limited, unreliable and noisy data, Bayesian approaches are optimally suited. Various even contradictory data can be brought into a common framework and support each other. Similarly, unlikelihood and uncertainty of a model approach can be quantified. This is well-accepted in archaeological applications through the Bayesian calibration and modelling of stratigraphic conditions and radiometric data. Furthermore, spatially and temporally missing data can be taken into account: Where these data are missing, the uncertainty automatically increases, but this does not prevent general modelling and estimation. Finally, hierarchically structured model suites, in which sub-models are created for the utilisation of individual proxies, can also be used to estimate transfer functions between individual proxies and the value to be modelled, thanks to the interaction of a large number of data sources and evidence.

Bayesian modelling techniques have also been used recently as a tool for hypothesis testing of demographic trends or underlying models based on 14C data.Most notably are the recently published papers by Crema et al. However, the approach taken by Crema et al. is clearly different from the one presented in this paper.In these analyses, deductive models are generated and their plausibility is tested on the basis of 14C data only. This is a clear step in the direction of a model-based, and thus scientific, analysis.However, on the one hand, the use of only one proxy, and its use exclusively for testing hypotheses developed independently, creates problems comparable to those of inductive approaches used so far: due to the lack of a combination with other indicators, one is limited to the problems and conditions of sum calibration as a tool. Furthermore, this approach loses significant potential information that would be gained by a direct evaluation of the time series.Thus, the credibility of a model can only be checked as a whole, without the dynamic developments that can arise in the course of demographic processes being represented. Therefore, we would like to better exploit the capabilities of Bayesian hierarchical models through a combination of inductive data analysis and model-passed data integration of different proxies.

This contribution now represents an attempt to make these techniques usable for archaeological reconstructions. In addition to a presentation of the basics and possible procedures, we want to show in the following, in a reproducible and practical form using a case study, how Bayesian methods can also make a decisive contribution to a better assessment of population development. These assessments are crucial for the reconstruction of the human past, even in for periods for which we only have very patchy, noisy and unreliable data.

2 Background

At the heart of all Bayesian statistics is the concept of updating a given prior assumption with new data and expressing this in probabilities (refs.). Our assumptions about the demographic development of the

past must naturally be very conservative. In the logic of the basic Bayesian equation, these assumptions represent the prior (probability). In conjunction with the data that are included as the likelihood of the prior, a posterior (probability) results, which represents the Bayesian learning from data. It is also in the nature of the approach that in real applications there is no point prediction, but in most cases a probability distribution for the prediction. Thus we simultaneously obtained a result and an estimate of the confidence intervals, or better, the credibility interval given the data.

This Bayesian learning is iterative and sequential, so that the result of one Bayesian inference can form the prior of another, i.e. it is an additive process. Moreover, at the conceptual level, this allows different sources of information to be combined. This enables that they can be mapped to the same set and the same real-world domain, e.g. probabilities. This fact has long been exploited by archaeology in using stratigraphic information to make radiometric dating more accurate. ^{14}C dates and stratigraphy are something completely different, but both can be mapped to the probability of older and younger and combined in this way. The same is also conceivable (and feasible) when it comes to the probability of population sizes or population densities or their derived dynamics (increase, decrease).

A further characteristic of Bayesian modeling is that, due to the fact that results of a Bayesian inference can be regarded as a priority for the next one, a hierarchical formulation of problem domains is possible. Parameters that are necessary for an estimation, such as the relationship of population density to the clearing signal in pollen data, need not be specified explicitly, but can be given by probability distributions and then estimated in the model itself. The more data available, the more degrees of freedom can be estimated with a reasonable loss of confidence or a reasonable width of credibility intervals. For the estimation of these parameters, in turn, submodels have to be created which describe the relationship of the data to the characteristics of the parameter. This can be carried out over several levels, depending on necessity.

This modeling technique can thus be used to combine different lines of evidence horizontally and vertically and in this way combine their results into a common conclusion and estimate, which at the same time includes an estimation of their reliability: if the data contradict each other, the overall reliability will be lower. If they support each other, the confidence interval will be smaller. And if there is no systematic bias that affects all data sources to the same extent, it should be possible to arrive at the most reliable estimate possible through the most heterogeneous set of data sources.

A special class of Bayesian hierarchical models are so-called State Space Models (also known as Hidden Markov Models). These are specifically designed for time series and follow two basic principles. First, a hidden or latent process is assumed, which represents the state of the variable of interest x_t at all times. For the course of the states of this variable over time, it is assumed that every state of variable x in the future, as well as in the past, is bound by a Markov process to the state of variable x at time t. At the same time, it is assumed that certain observations, represented in variable y, are dependent on the state of variable x at time t. This means that there is a relationship between the variable x and the state of variable y over time. This implies that a relationship between the individual states of variable y is generated over time via the hidden variable x, which itself is not observable.

This basic structure of the model makes it particularly useful and suitable for the purpose of demographic reconstruction using archaeological, but also other data, which depend on population density in the past. This population density itself is not accessible or measurable by our means. All we have at our disposal are observations derived by unknown transfer functions. These can be of very different natures, such as number of archaeologically observable settlements, or other effects that can be observed through time series, and which are influenced by population in a given area. In our example, these are the openness indicators from pollen data, which we can interpret primarily in terms of human influence and its intensity. On a more abstract level, we can also include expert estimates, as these also happen on (often unspecified) bases that are at least indirectly influenced by past population density. The specifics of the individual data we used in this modelling are discussed in the section 3.2.

3 Materials and Methods

3.1 Process Model

The overall model for the estimation of demographic developments is broken down into several hierarchically interconnected individual elements in accordance with the basic structure of a state space model. The basis is a process model that represents the demographic development itself in terms of a structure model, without this model already being explicitly configured with data.

In this process model we assume that the latent variable number of sites is strongly autocorrelated across different time periods. In principle, one can represent a population development in such a way that the population at time t results from the population at time $t-1$ times a parameter λ, which represents the population change at this time. This results in the following very simple formula:

$$N_t = N_{t-1} * \lambda_t$$

A distribution that is particularly suitable for modelling frequencies is the Poisson distribution. It is a univariate discrete probability distribution that can be used to model the number of events that occur independently of each other at a constant mean rate in a fixed time interval or spatial area. It is determined by a real parameter $\lambda > 0$, which describes the expected value and simultaneously the variance of the distribution. Thus, the relationship shown above can also be rearranged as follows:

$$N_t \sim dpois(\lambda_t)$$
$$\lambda_t = N_t$$

If we now have information about the change in population development, we can use this to enter it into the model via a change in λ. This is done in the form of a regression: If $x \in R^n$ is a vector of independent variables, then the model takes the form

$$log(E(Y \mid x)) = \alpha + \beta' x$$

In the case of population trends, the expected value would be equal to the value of the population in the previous time period, plus or minus the changes resulting from the variables:

$$log(\lambda_t) = log(N_{t-1}) + \sum_{i=1}^{n} \beta_i x_{t,i}$$

The values for population size N_t as well as for population change λ_t are time-dependent. At each individual point in time in the time series, these variables can or will also take on different values. However, we can narrow down the structure of population change even further. We can assume that, considered overall over time, the population change the population change will not exceed certain limits over the entire time span, without it being possible here to specify this already.

Thus, we can define the limits, the max_change_rate as an time independent variable, again without already specifying them with fixed values at this stage. The estimation of these parameters for the entire model, as well as the estimation of the respective population change per time section, results from the modelling and the interaction with the data, respectively. Overall, this represents a hierarchical model that can be noted as follows:

$$max\_growth\_rate \sim dgamma(shape = 5, scale = 0.05)$$
$$N_t / N_{t-1} < (max\_growth\_rate + 1)$$
$$N_{t-1} / N_t < (max\_growth\_rate + 1)$$

The gamma distribution used centres the probability in a range $[0, 1[$, adding 1 makes this range $[1 - 2[$ (See also Figure 1).

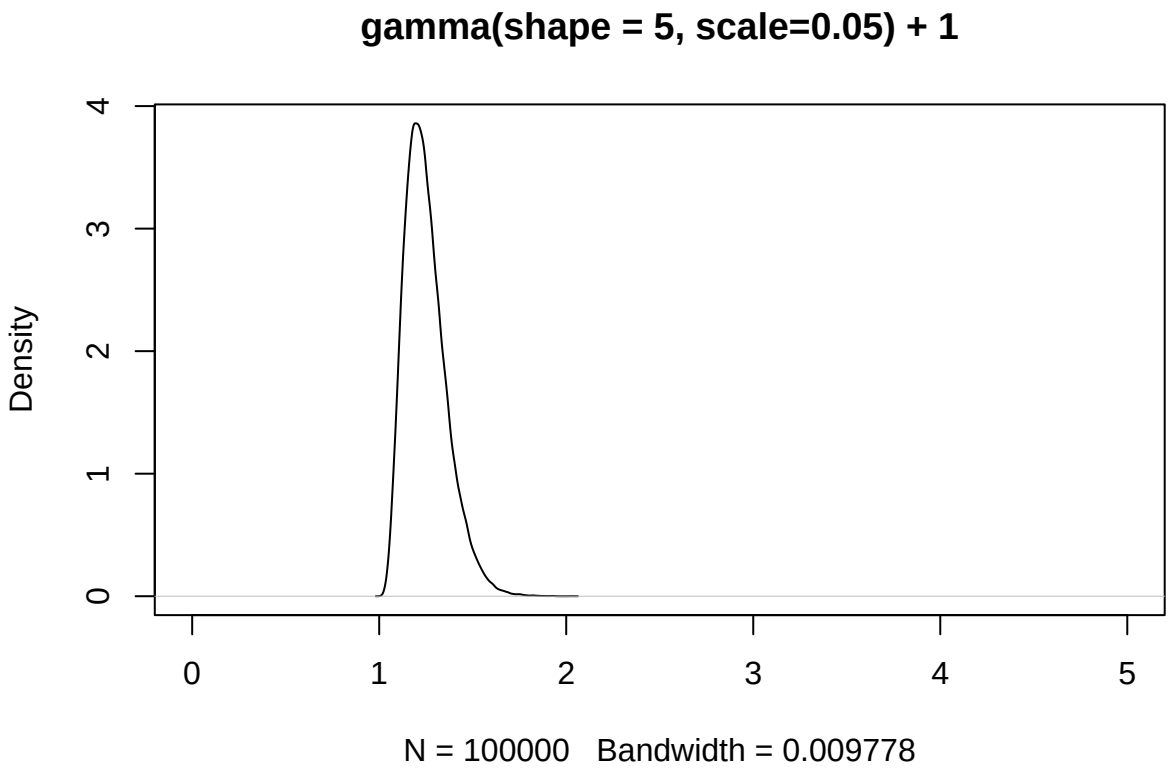

Figure 1: The gamma-distributed maximum growth rate prior of the model.

This prevents the number of sites from explosively increasing between two time periods, which would lead to problems for the convergence of the model. The interaction of these parameters results in the following prior probability distribution for λ and thus the growth (or change) rate of the population:

3.2 Data/Proxies

In principle, a large number of different data sources can be integrated into the overall model as observations, provided that these observations a) can be understood as dependent on the population density in the past, and b) a model-like description of this dependence can be created. A non-exhaustive list can be found in the following table 1:

Table 1: A incomplete list of possible observation that can be linked to population developments in the past. Proxies used in this study are highlighted.

Proxies
Expert estimates
Ethnographic Analogies
Carrying Capacity
Economic modelling
Extrapolation of buried individuals
Burial anthropology
Settlement data, number of houses
Settlement data, settlement size

Proxies
Aoristic analysis
Dendro dates
Amount of archaeological objects
Radiocarbon sum calibration
Estimates based on specific object types
Human impact from pollen or colluvial data
aDNA based estimates
...

Our working region consists of the Swiss Plateau, which on the one hand offers excellent data for demographic estimation, but on the other hand poses very specific problems for such an undertaking. If we have high-resolution information on the temporal sequence of individual settlements at the lakeside settlements by means of dendro data, this also might cause a research problem with regard to the ^{14}C data often used as a proxy.

The dataset we use for the number of dendrodated wetland settlements in the Three Lakes region was collected by Julian Laabs for his PhD thesis (Laabs, 2019). Details on the creation of this data series will be published at the referenced location. The time series used here runs from 3900 to 800 BCE, and contains the number of chronologically registered fell phases at individual settlements. This results in a time series that reflects the settlement of the lakeshores in the Neolithic and Bronze Age periods.

Since many periods of the Swiss prehistoric dry land settlements are unknown, and archaeology is based primarily on the wet land settlements, the dendro data are used primarily here, since they are much more precise. As a result, ^{14}C data could be largely absent in this time period. If the amount of ^{14}C data were now evaluated directly and naively over time, it would show that in the phases of particularly active settlement of the lakeshore, the ^{14}C sum calibration shows minima.

The dataset for the ^{14}C sum calibration primarily consists of data from the XRONOS database, supplemented by dates from the unpublished PhD thesis of Julian Laabs (Laabs, 2019) and the data collection of (Martínez-Grau et al., 2021). It contains a total of 1135 single ^{14}C data from 246 sites. The dates were selected so that their distribution area coincides with the catchment area of the pollen proxy explained below (see also Figure 2). The dates in the dataset range in ^{14}C years from 10730 to 235 BP uncal. This time window extends beyond the study horizon in order to minimise boundary effects, which, however, cannot be completely avoided due to the Hallstatt plateau. But it was precisely to reduce its influence that the time window of the investigation was chosen to be 7000 – 1000 BCE.

We binned the data at site levels to obtain a temporally dispersed count and thus an expected value of contemporaneous ^{14}C dated sites. For the creation of the cumulative calibration, the corresponding functions of the R package rcarbon (Crema and Bevan, 2021) were used in the standard settings.

We can now compare these two data sets (Figure 3). In fact, there is a not uninteresting fit between the two data series. However, it must be assumed that the two dating methods, even if they would contradict each other, actually complement each other, and thus allow a better overall unified picture of the actual settlement density than each of the individual proxies would allow on their own.

In order to add another indicator of archaeological evidence of occupation, we have included the data of the Cantonal Archaeology (Figure 4), and thus the Heritage Management, which are primarily derived from scattered surface finds, and which often have a low depth of information and thus dating accuracy. This information is incorporated into our model as a typologically dated aorist time series. Although the dating accuracy is very low, the advantage here lies in the fact that we are not bound to the conditions and problems of radiocarbon dates, and thus involved in the issue of sum calibration. Much more, these data provide an independent indicator with regard to the methodology of the ^{14}C data, even if they are influenced by the same transmission filters and archaeological conditions as the evaluation of ^{14}C data. Data from x sites were included in the aoristic sum, which is a very rough indicator due to the low dating accuracy in archaeological

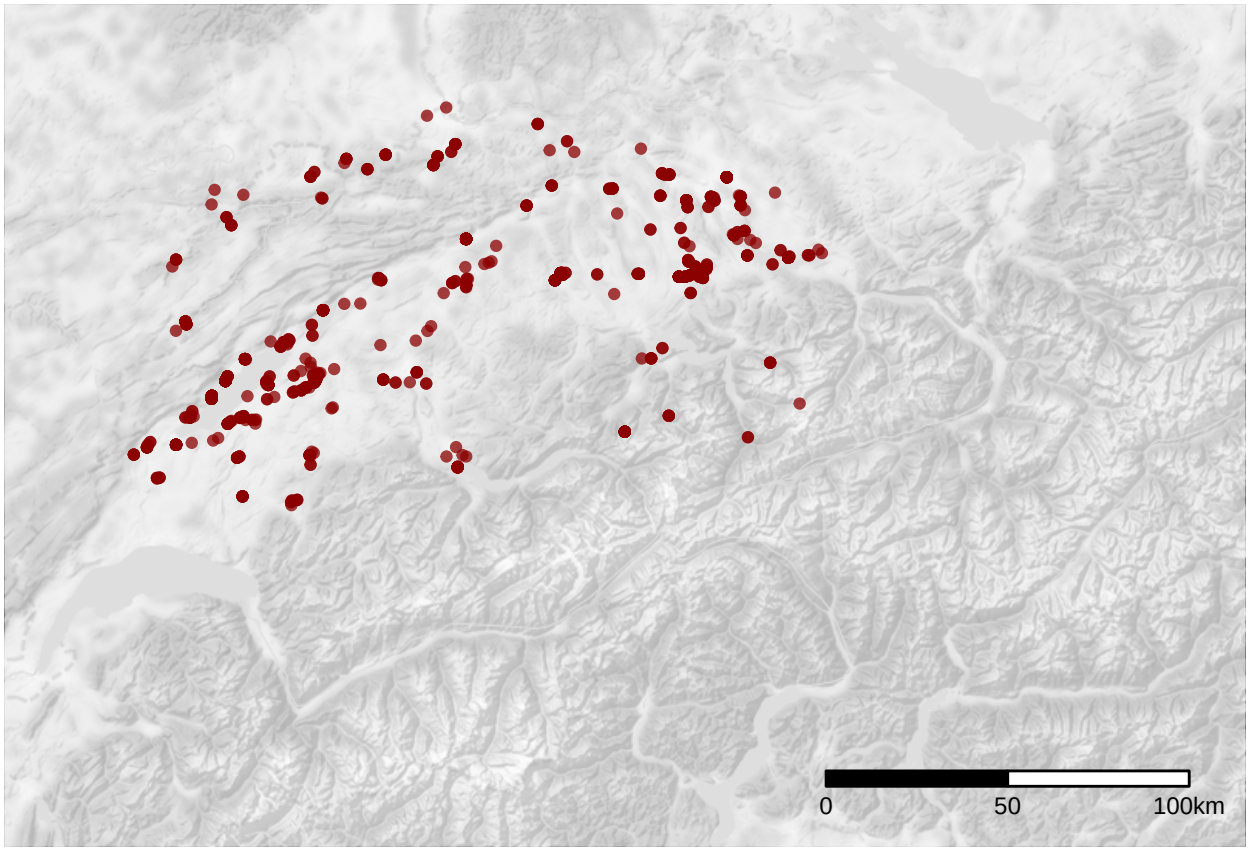

Figure 2: The location of the ^{14}C dated sites in the dataset.

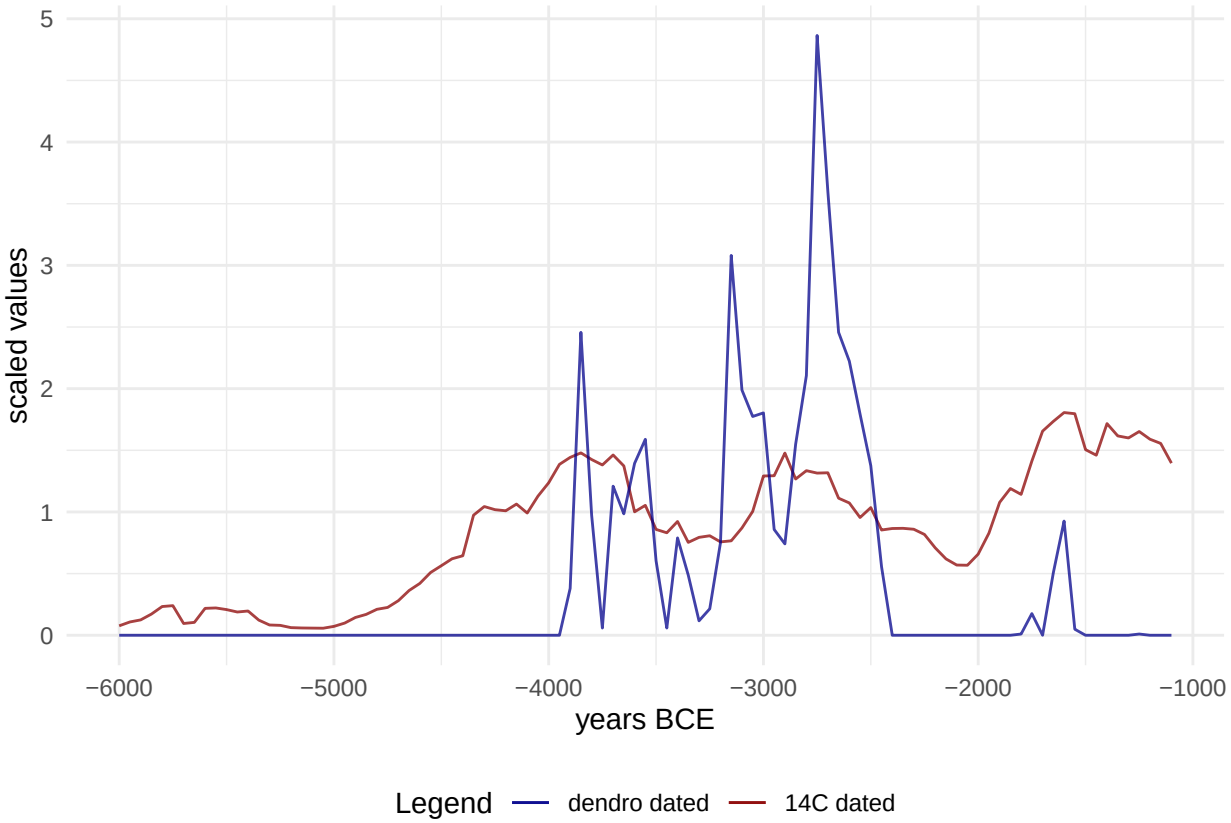

Figure 3: Comparison of the scaled number of ^{14}C and dendrodated sites over time in the dataset used.

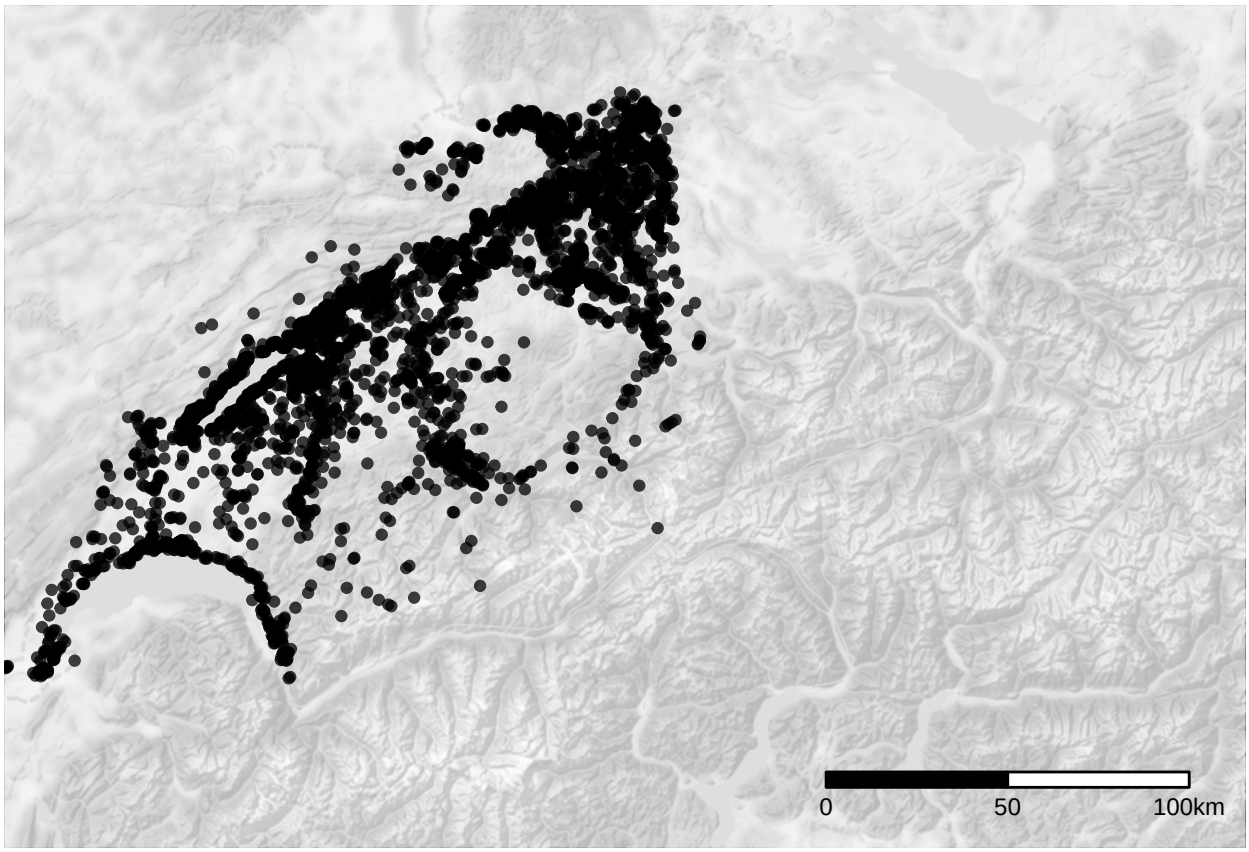

Figure 4: The location of the sites from the find reports of the cantonal archaeology (heritage management). The locations are fuzzified by ~1km.

phases, but which nevertheless has an important role in the normalisation of the data due to its independence from calibration effects.

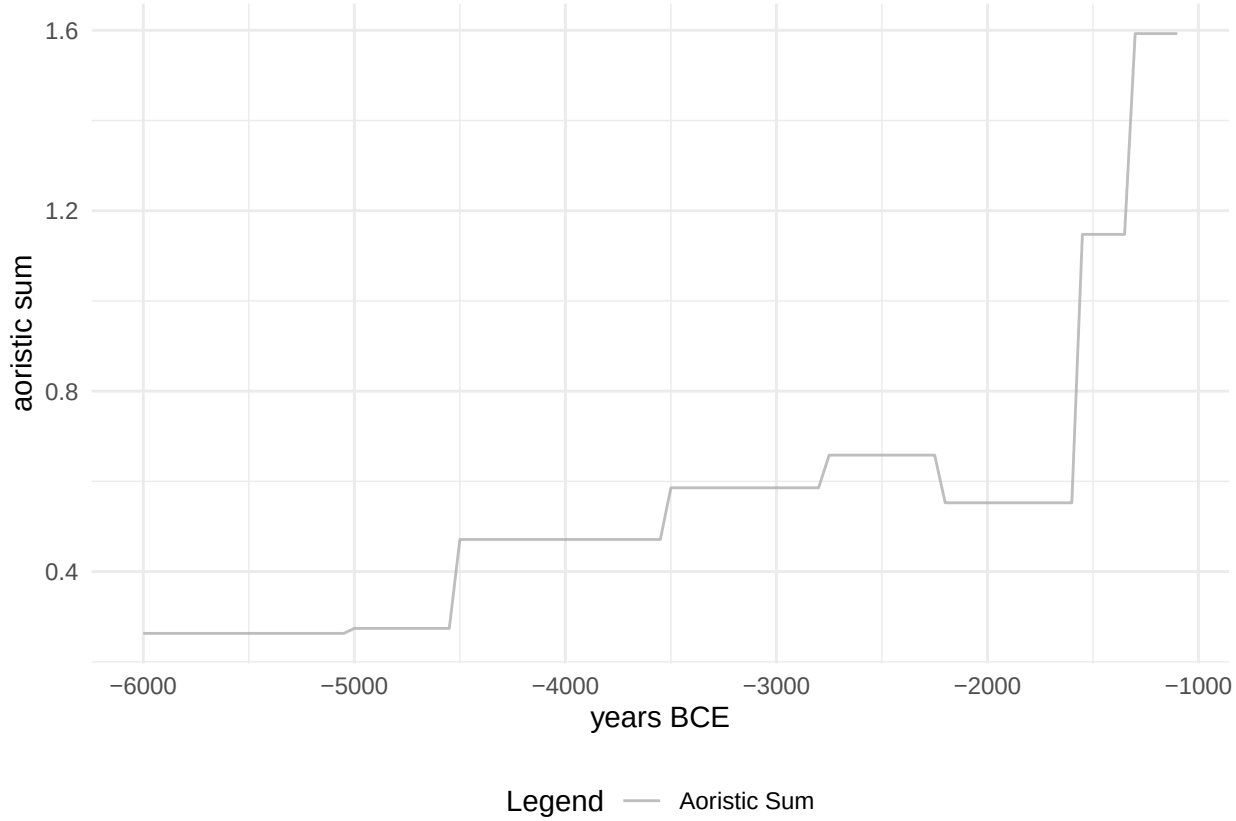

Figure 5: Comparison of the scaled number of ^{14}C and dendrodated sites over time in the dataset used.

The natural conditions provided by the many lakes enable not only highly precise dating of archaeological sites, but also a very dense network of pollen analysis. We make use of this fact by generating a supra-regional openness indicator for the vegetation from the pollen data (Figure 6). This indicator has the specific advantage that it is not dependent on preservation conditions, as archaeological indicators are. This makes it particularly valuable for indicating or compensating for systematic distortions that result from temporally specific settlement patterns and archaeological preservation conditions.

The utilisation of this proxy is based on the assumption that the higher the population density in an area, the greater the human influence on the natural environment (Lechterbeck et al., 2014), and that the agricultural space of an area is closely related to the population density (Zimmermann, 2004). Evidence of land clearing in pollen diagrams can therefore give further indications regarding population dynamics. The exact methodological procedure for obtaining this proxy from several, in our case 5, different pollen diagrams of sites mainly in the hinterland of the large Alpine lakes can be found in a previous publication (Heitz et al., 2021). The technical procedure is also documented in the accompanying R-script. The percentage pollen data based on a pollen sum of all terrestrial taxa of the individual sites were combined into one data set by means of a principal component analysis (Figure 7). Only terrestrial pollen taxa with a frequency of more than 1/3 and, if present, with an average frequency of at least more than 0.1% were selected to reduce potential disturbance by rare species. Cereal pollen was explicitly retained as an important anthropogenic indicator. As each sample is absolutely dated, the data on the x-axis can be plotted against the openness value on the y-axis to obtain a time series time series for land clearing.

Unfortunately, in the Swiss Plateau there is no data from burials that could be usefully incorporated into such this model. Preservation conditions again simply prevent their use in this study. Nevertheless, we see a

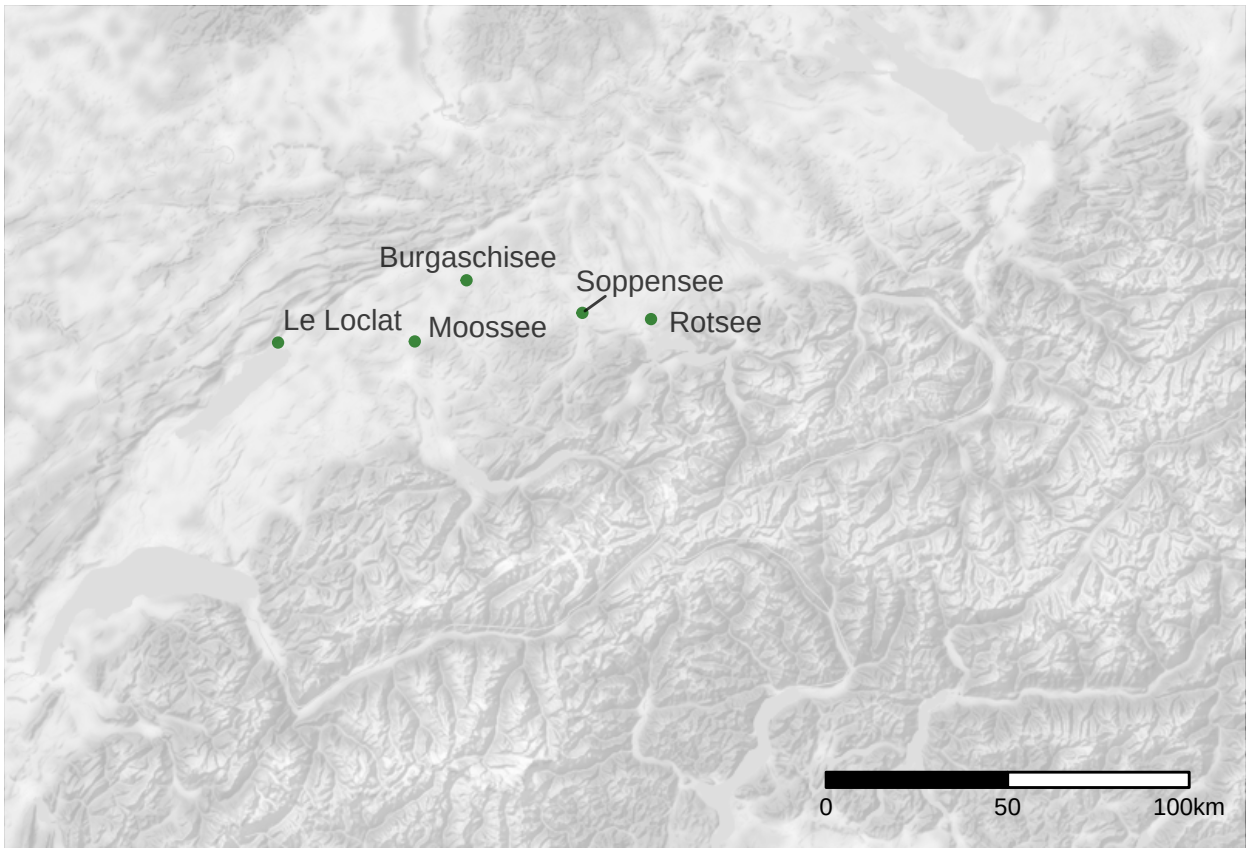

Figure 6: Location of the pollen profiles used for the openness indicator.

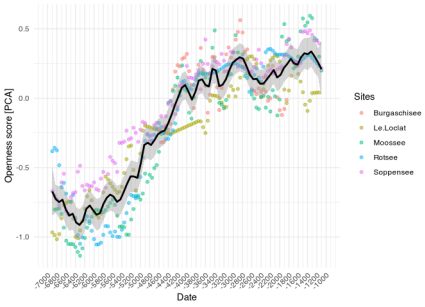

Figure 7: Value on the first dimension of the PCA against dating of the samples for the individual pollen profiles and their combined average value as the openness indicator.

very high potential for other regions in the integration of demographic indicators from burial data in order to enlarge the canon of methods and the range of proxies.

The values of the sum calibration, the openness index and the dendrodated settlements were smoothed by means of a moving average with a window of 50 years. Since the aoristic sum already had a very coarse temporal resolution, this was not applied for this measure. The range of the smoothing window corresponded to the sample interval, with which a unified resolution of 50 years was obtained for all proxies as time slices for the model. In addition, all data were restricted to a window of observation of 7000-1000 BCE. The upper limit is set by the post-glacial changes in the pollen spectra, which are hardly associated with human influence before 7000 in the working area, and which would consequently distort the openness indicator. The lower limit is defined by the Hallstatt Plateau (with buffer), which would have direct and indirect research influences on the ^{14}C proxy.

3.3 Observational Model

During the development of the overall model, we abandoned the implementation of dedicated observation models adapted to the conditions of the individual proxies and their generating processes. In previous implementations, which we have also presented at various conferences, the underdetermination by the currently usable data of the model with many degrees of freedom led to equifinality of the solutions and, thus, to a high path dependency of the individual model runs. Therefore, it was almost impossible to achieve convergence of the overall model. Nevertheless, we believe that for a future application of the model with more data, a larger geographical coverage and especially a regionalised approach with information transfer by means of partial pooling, this more specific approach will be feasible and a very useful approach.

The implementation we introduce in this paper represents a Poisson regression in which the indicators used inform the change in the number of settlements from time step to time step. For this purpose, the individual proxies were z-normalised. The absolute differences from one time step to another were then computed from the resulting time series. Thus, if the value of the proxy increases, this results in a positive difference from the previous time step, and vice versa.

$$z_t = \frac{x_t - \bar{x}}{\sigma_x} \mid \sigma_x := \textit{Standard Deviation}$$
$$\delta z_t = z_t - z_{t-1}$$

The sum of the resulting differences between the time steps, together with the settlement number of the previous step as the expected value, then forms λ_t as the expected value for the settlement number of the current time step.

$$log(\lambda_t) = log(N_{t-1}) + \sum_{i=1}^{n} \beta_i \delta z_{i,t}$$

Here, β_i is a scaling factor that represents the influence of the respective proxy. It is a confidence value of the model for the respective proxy, so that the sum of all β_i results in 1.

$$\sum_{i=1}^{n} \beta_i = 1$$

A probability distribution that can be used for this purpose in a hierarchical Bayesian model is the Dirichlet distribution, which is a multivariate generalization of the beta distribution, commonly used as prior distributions in Bayesian statistics. Its density function gives the probabilities of i different exclusive events. It has a parameter vector $\alpha = (\alpha_1, ..., \alpha_i) \mid (\alpha_1, ..., \alpha_i) > 0$, for which we have chosen a weakly informative log-normal prior. The priors for the log-normal distribution in turn come from a weakly informative exponential distribution for the mean and a log-nomal distribution with μ of 1 and σ_{log} of 0.1:

$$
\begin{aligned}
\beta_i &\sim Dir(\alpha_{1-i}) \\
\alpha_i &\sim LogNormal(\mu_{alpha_i}, \sigma_{alpha_i}) \\
\mu_{alpha_i} &\sim Exp(1) \\
\sigma_{alpha_i} &\sim LogNormal(1, 0.1)
\end{aligned}
$$

As an intuition, this means that we consider the sum of the proxies as determinant for the number of settlements. The estimation therefore assumes that all proxies together give the best possible estimation result, whereby the share of each individual proxy is considered variable and is estimated within the model. The error value is represented by the Poisson process in the process model, rather than directly as an estimation error for the individual proxies. Thus, our model does not correspond to a classical state space model, where the measured values are each considered to be error-prone. In the implementation, the model finds the best possible combination or compromise between the individual indicators to describe a settlement dynamic that is given by them.

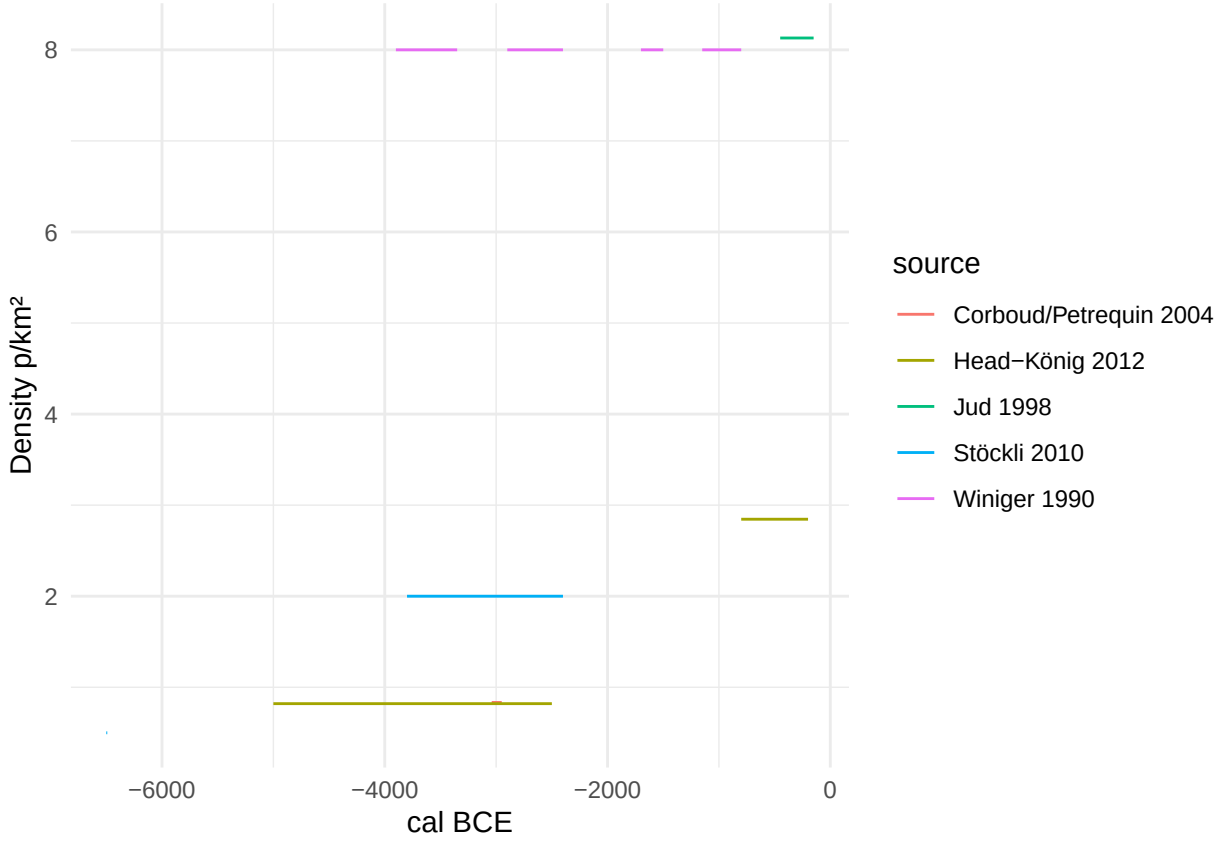

Figure 8: Expert estimations for the population density on the Swiss Plateau from different authors.

In earlier implementations, expert estimates were also integrated into the model. However, since these are highly contradictory for the working area (Figure 8, therefore have little influence on the model, and also lead to a significantly longer runtime, we have refrained from doing so in the current implementation. For future applications of the model with a larger geographical range and thus a higher information density, however, this would again be a factor that could be integrated in a useful way.

3.4 Model fitting

The model was fitted using the R package nimble (version 0.11.1, R version 4.1.3). For this purpose, 4 chains were run in parallel. Achieving and ensuring convergence and sufficient effective samples (10000) for a reliable

assessment of the highest posterior density interval was carried out in steps.

In a first run, the model was initialised for each chain and run for 100000 iterations (with a thinning of 10). The computer used (Linux, Intel(R) Xeon(R) CPU E3-1240 v5 @ 3.50GHz, 4 cores, 8 threads) needed approx. 1 min for this.

In a second step, the run was extended until convergence could be determined using Gelman and Rubin's convergence diagnostic, the criterion being that a potential scale reduction factor of less than 1.1 was achieved for all monitored variables. Convergence occurred after about 30 seconds.

Due to the high correlation of the parameters and thus a low sampling efficiency, the collection of at least 10,000 effective samples for all parameters in the third step took about 5 hours.

For the fitting process, a starting value for a population density of the Late Bronze Age (1000 BCE) was taken from the literature, which may represent a general average value for all prehistoric population estimates (Nikulka, 2016, p. 258). For the model, this was set as the mean of a normal distribution with a standard deviation of 0.5, which should give enough leeway for deviations resulting from the data. Nevertheless, especially the late part of the reconstruction is of course clearly influenced by this predefined value.

For the traceplots and the prior-posterior overlap as well as the density functions of the posterior samples of the individual parameters, please refer to the supplementary material.

4 Results

Within the model, the parameter p was estimated, which reflects the proportion of the individual proxies used in the estimation of the number of settlements, as well as the number of settlements or transforms the resulting population density under the assumption that per settlement a number of inhabitants lived poisson-distributed around a λ of 50 persons. This parameter is variable, but has only a scaling influence on the final estimate of population density.

By looking at the distribution of posterior samples for the share of each proxy (Figure 9), it is clear that the model weights the openness indicator the highest. The average is slightly above 60%. The next most important indicator is the sum calibration value, which has an average of about 20%. The aoristic sum is slightly above 10%, whereas the importance of the dendrodated settlements is below 10%.

The reason for the latter is certainly that there are no lakeshore settlements over large areas of the time window, and therefore the proxy achieves a low confidence value in comparison with the other estimators. In the case of the aoristic sum, it is certainly the fact that it is flat over large sections and has little structure, making it difficult to relate to the other estimators. The sum calibration shows very strong short-term fluctuations, which are presumably at least partly due to the effects of the calibration curve, and which also make this proxy seem ill-suited to reliably represent a continuous population trend. Nevertheless, its fluctuations do have an impact on the resulting overall estimate of the development of the number of settlements, albeit to a lesser extent.

The population density estimate (Figure 10) ranges on average between 0.2 p/km² for the beginning of the estimate (6000 BCE) and 4.8 p/km² for the end of the estimate (1000 BCE), reaching a maximum of 6.5 p/km² for the time slice 1250 BCE. Thus, the estimate remains within the values that are also considered plausible by the expert estimates. Clear peaks are reached around 1250 BCE (also the maximum value), as well as around 2750 BCE, which corresponds to the onset of the Corded Ware in the Swiss Plateau. However, we will go into further detail in the discussion on evaluations concerning the connections to cultural phenomena.

Furthermore, it may be relevant to look at the temporal distribution of the variability in the estimate (Figure 11). The coefficient of variation is 0.13 for the beginning and 0.1 for the end of the estimate, the greatest variability is reached around 2150 BCE with 0.4679342. The beginning and end of the time series are relatively clearly determined. The end results from the a priori setting of the parameter, but also here as at the beginning of the series the proxies are very uniform, which explains the low variability. Overall, the

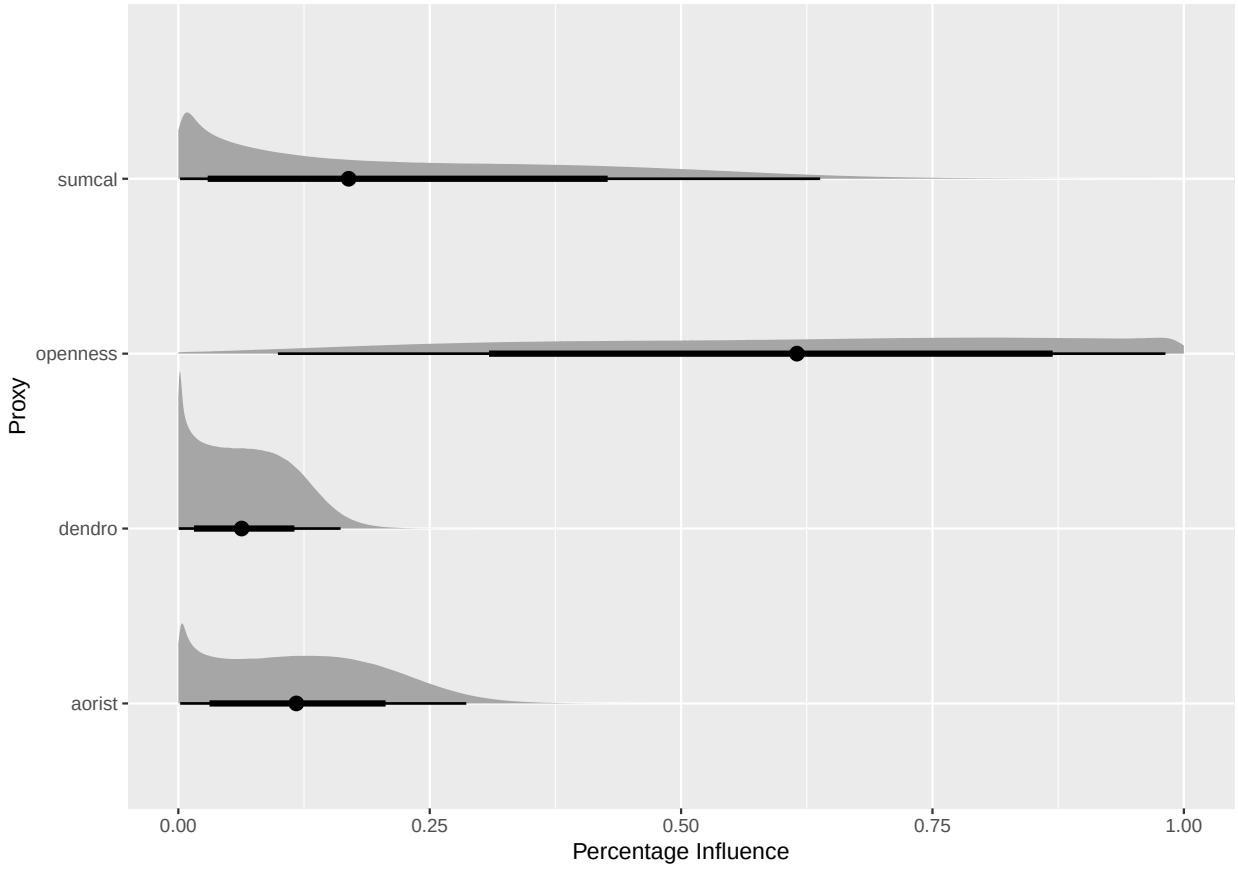

Figure 9: Distribution of the influence ratio of the different proxies on the final estimation of the number of sites.

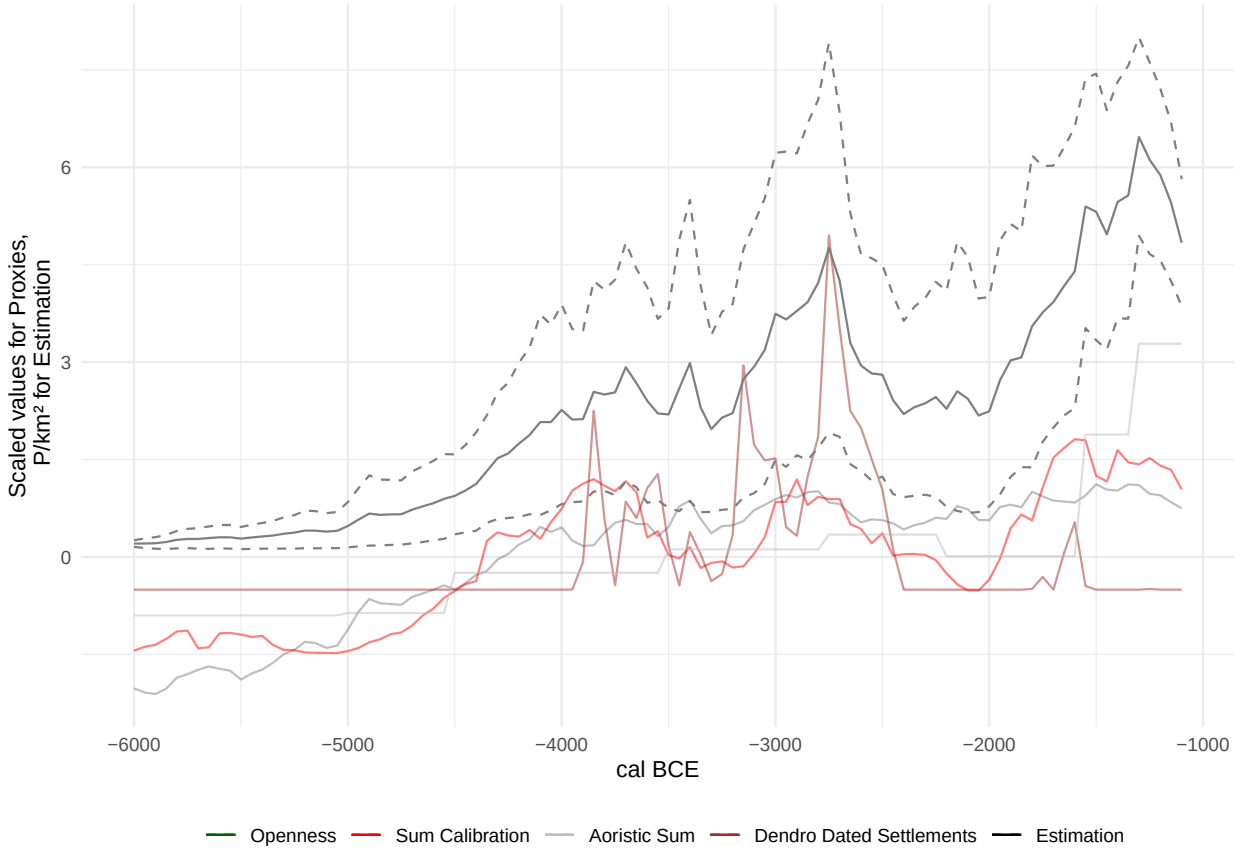

Figure 10: The estimate of the population density resulting from the model and the three proxies, which are also plotted (scaled) for comparison.

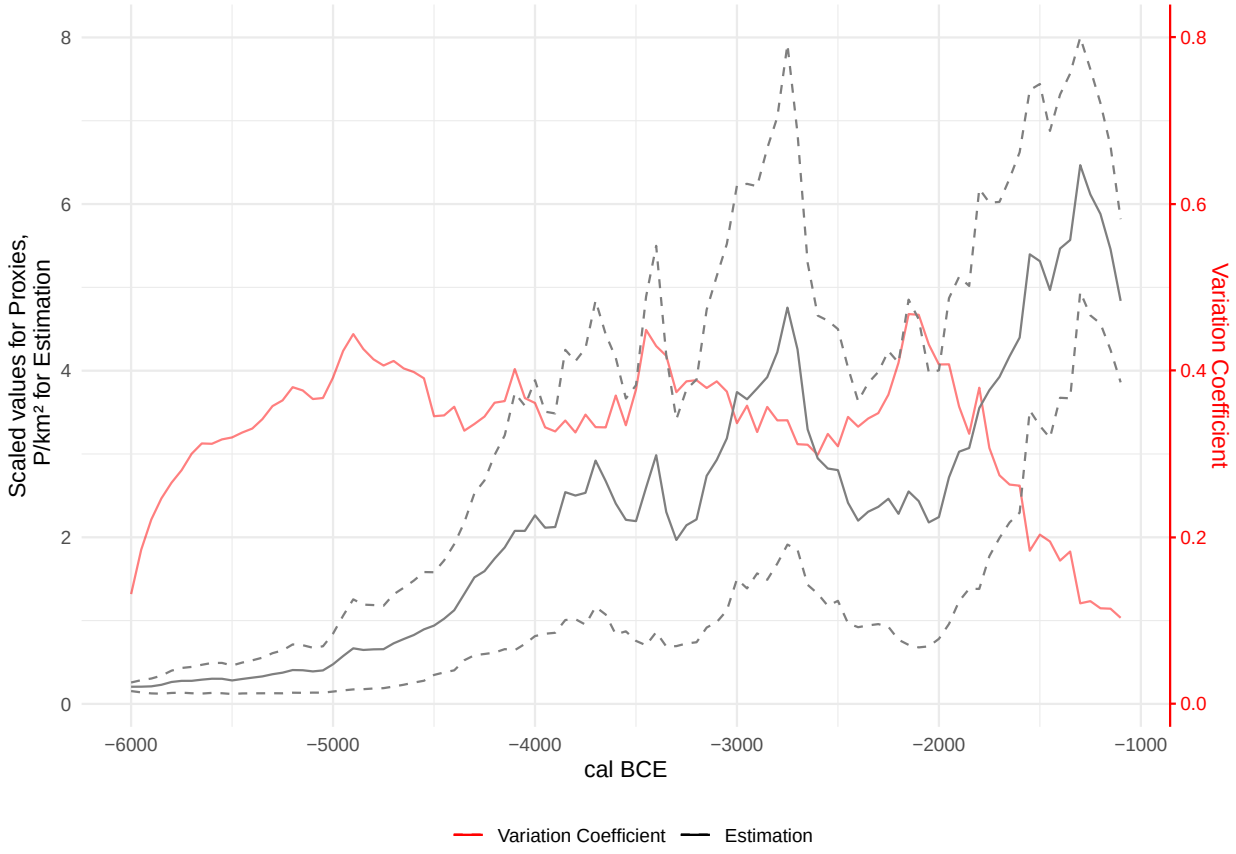

Figure 11: The estimate of the population density resulting from the model and the three proxies, which are also plotted (scaled) for comparison.

variability is relatively uniform over the entire course of the estimation and averages over all time slices at 33% of the respective mean.

5 Discussion

6 Conclusion

7 Acknowledgements

8 References

Bryant, J., Zhang, J.L., 2018. Bayesian Demographic Estimation and Forecasting. Chapman; Hall/CRC. https://doi.org/10.1201/9780429452987

Childe, V.G. (Ed.), 1936. Man Makes Himself. Watts & Co., London.

Colton, H.S., 1949. The prehistoric population of the Flagstaff area. Plateau 22, 21–25.

Cook, S.F., 1946. A reconsideration of shell mounds with respect to population and nutrition. American Antiquity 12, 51–53.

Crema, E.R., Bevan, A., 2021. Inference from large sets of radiocarbon dates: software and methods. Radiocarbon 63, 23–39. https://doi.org/10.1017/RDC.2020.95

Frankfort, H., 1950. Town planning in ancient Mesopotamia. Town Planning Review 21, 98–115.

Hack, J.F., 1942. The changing physical environment of the Hopi Indians of Arizona, Papers of the Peabody Mu- seum of American Archaeology and Ethnolog. Harvard University, Cambridge.

Hassan, F.A., 1981. Demographic archaeology, Studies in archaeology. Academic Press, New York.

Heitz, C., Hinz, M., Laabs, J., Hafner, A., 2021. Mobility as resilience capacity in northern Alpine Neolithic settlement communities. https://doi.org/10.17863/CAM.79042

Kintigh, K.W., Altschul, J.H., Beaudry, M.C., Drennan, R.D., Kinzig, A.P., Kohler, T.A., Limp, W.F., Maschner, H.D.G., Michener, W.K., Pauketat, T.R., Peregrine, P., Sabloff, J.A., Wilkinson, T.J., Wright, H.T., Zeder, M.A., 2014. Grand challenges for archaeology. Proceedings of the National Academy of Sciences 111, 879–880. https://doi.org/10.1073/pnas.1324000111

Laabs, J., 2019. Populations- und landnutzungsmodellierung der neolithischen und bronzezeitlichen westschweiz (PhD thesis). Bern.

Lechterbeck, J., Edinborough, K., Kerig, T., Fyfe, R., Roberts, N., Shennan, S., 2014. Is Neolithic land use correlated with demography? An evaluation of pollen-derived land cover and radiocarbon-inferred demographic change from Central Europe. The Holocene 24, 1297–1307. https://doi.org/10.1177/0959683614540952

Martínez-Grau, H., Morell-Rovira, B., Antolín, F., 2021. Radiocarbon Dates Associated to Neolithic Contexts (Ca. 5900 – 2000 Cal BC) from the Northwestern Mediterranean Arch to the High Rhine Area. Journal of Open Archaeology Data 9, 1. https://doi.org/10.5334/joad.72

Müller, J., Diachenko, A., 2019. Tracing long-term demographic changes: The issue of spatial scales. PLOS ONE 14, e0208739. https://doi.org/10.1371/journal.pone.0208739

Nikulka, F., 2016. Archäologische Demographie. Methoden, Daten und Bevölkerung der europäischen Bronze- und Eisenzeiten. Sidestone Press, Leiden.

Rick, J.W., 1987. Dates as Data: An Examination of the Peruvian Preceramic Radiocarbon Record. American Antiquity 52, 55–73. https://doi.org/10.2307/281060

Shennan, S., 2000. Population, Culture History, and the Dynamics of Culture Change. Current Anthropology 41, 811–835.

Zimmermann, A., 2004. Landschaftsarchäologie II. Überlegungen zu prinzipien einer landschaftsarchäologie.

8.0.1 Colophon

This report was generated on 2022-04-09 17:45:31 using the following computational environment and dependencies:

```
#> - Session info ---------------------------------------------------------------
#>  setting  value
#>  version  R version 4.1.3 (2022-03-10)
#>  os       Manjaro Linux
#>  system   x86_64, linux-gnu
#>  ui       X11
#>  language (EN)
#>  collate  C
#>  ctype    de_DE.UTF-8
#>  tz       Europe/Zurich
#>  date     2022-04-09
#>
#> - Packages -------------------------------------------------------------------
#>  package       * version date       lib source
#>  assertthat      0.2.1   2019-03-21 [1] CRAN (R 4.1.0)
#>  bitops          1.0-7   2021-04-24 [1] CRAN (R 4.1.0)
#>  bookdown        0.25    2022-03-16 [1] CRAN (R 4.1.3)
#>  cachem          1.0.5   2021-05-15 [1] CRAN (R 4.1.0)
#>  callr           3.7.0   2021-04-20 [1] CRAN (R 4.1.0)
#>  class           7.3-20  2022-01-16 [2] CRAN (R 4.1.3)
#>  classInt        0.4-3   2020-04-07 [1] CRAN (R 4.1.0)
#>  cli             3.1.0   2021-10-27 [1] CRAN (R 4.1.1)
#>  colorspace      2.0-1   2021-05-04 [1] CRAN (R 4.1.0)
#>  crayon          1.4.1   2021-02-08 [1] CRAN (R 4.1.0)
#>  curl            4.3.2   2021-06-23 [1] CRAN (R 4.1.1)
#>  DBI             1.1.1   2021-01-15 [1] CRAN (R 4.1.0)
#>  desc            1.4.0   2021-09-28 [1] CRAN (R 4.1.1)
#>  devtools        2.4.2   2021-06-07 [1] CRAN (R 4.1.2)
#>  digest          0.6.27  2020-10-24 [1] CRAN (R 4.1.0)
#>  dplyr           1.0.7   2021-06-18 [1] CRAN (R 4.1.0)
#>  e1071           1.7-7   2021-05-23 [1] CRAN (R 4.1.0)
#>  ellipsis        0.3.2   2021-04-29 [1] CRAN (R 4.1.0)
#>  evaluate        0.14    2019-05-28 [1] CRAN (R 4.1.0)
#>  fansi           0.5.0   2021-05-25 [1] CRAN (R 4.1.0)
#>  farver          2.1.0   2021-02-28 [1] CRAN (R 4.1.0)
#>  fastmap         1.1.0   2021-01-25 [1] CRAN (R 4.1.0)
#>  foreign         0.8-82  2022-01-16 [2] CRAN (R 4.1.3)
#>  fs              1.5.0   2020-07-31 [1] CRAN (R 4.1.0)
#>  generics        0.1.0   2020-10-31 [1] CRAN (R 4.1.0)
#>  ggmap         * 3.0.0   2019-02-05 [1] CRAN (R 4.1.3)
#>  ggplot2       * 3.3.5   2021-06-25 [1] CRAN (R 4.1.2)
#>  ggrepel       * 0.9.1   2021-01-15 [1] CRAN (R 4.1.0)
#>  ggsn          * 0.5.0   2019-02-18 [1] CRAN (R 4.1.3)
#>  glue            1.4.2   2020-08-27 [1] CRAN (R 4.1.0)
#>  gtable          0.3.0   2019-03-25 [1] CRAN (R 4.1.0)
#>  highr           0.9     2021-04-16 [1] CRAN (R 4.1.0)
#>  htmltools       0.5.2   2021-08-25 [1] CRAN (R 4.1.2)
#>  httr            1.4.2   2020-07-20 [1] CRAN (R 4.1.0)
#>  jpeg            0.1-9   2021-07-24 [1] CRAN (R 4.1.0)
#>  KernSmooth      2.23-20 2021-05-03 [2] CRAN (R 4.1.3)
```

```
#>  knitr           1.33    2021-04-24 [1] CRAN (R 4.1.0)
#>  labeling        0.4.2   2020-10-20 [1] CRAN (R 4.1.0)
#>  lattice         0.20-45 2021-09-22 [2] CRAN (R 4.1.3)
#>  lifecycle       1.0.0   2021-02-15 [1] CRAN (R 4.1.0)
#>  magrittr        2.0.1   2020-11-17 [1] CRAN (R 4.1.0)
#>  maptools        1.1-1   2021-03-15 [1] CRAN (R 4.1.0)
#>  memoise         2.0.0   2021-01-26 [1] CRAN (R 4.1.0)
#>  munsell         0.5.0   2018-06-12 [1] CRAN (R 4.1.0)
#>  pillar          1.6.1   2021-05-16 [1] CRAN (R 4.1.0)
#>  pkgbuild        1.2.0   2020-12-15 [1] CRAN (R 4.1.0)
#>  pkgconfig       2.0.3   2019-09-22 [1] CRAN (R 4.1.0)
#>  pkgload         1.2.1   2021-04-06 [1] CRAN (R 4.1.0)
#>  plyr            1.8.6   2020-03-03 [1] CRAN (R 4.1.1)
#>  png             0.1-7   2013-12-03 [1] CRAN (R 4.1.0)
#>  prettyunits     1.1.1   2020-01-24 [1] CRAN (R 4.1.0)
#>  processx        3.5.2   2021-04-30 [1] CRAN (R 4.1.0)
#>  proxy           0.4-26  2021-06-07 [1] CRAN (R 4.1.0)
#>  ps              1.6.0   2021-02-28 [1] CRAN (R 4.1.0)
#>  purrr           0.3.4   2020-04-17 [1] CRAN (R 4.1.0)
#>  R6              2.5.0   2020-10-28 [1] CRAN (R 4.1.0)
#>  Rcpp            1.0.7   2021-07-07 [1] CRAN (R 4.1.2)
#>  remotes         2.4.0   2021-06-02 [1] CRAN (R 4.1.0)
#>  RgoogleMaps     1.4.5.3 2020-02-12 [1] CRAN (R 4.1.3)
#>  rjson           0.2.21  2022-01-09 [1] CRAN (R 4.1.3)
#>  rlang           0.4.11  2021-04-30 [1] CRAN (R 4.1.0)
#>  rmarkdown       2.13    2022-03-10 [1] CRAN (R 4.1.3)
#>  rnaturalearth * 0.1.0   2017-03-21 [1] CRAN (R 4.1.1)
#>  rprojroot       2.0.2   2020-11-15 [1] CRAN (R 4.1.0)
#>  rstudioapi      0.13    2020-11-12 [1] CRAN (R 4.1.0)
#>  scales          1.1.1   2020-05-11 [1] CRAN (R 4.1.0)
#>  sessioninfo     1.1.1   2018-11-05 [1] CRAN (R 4.1.0)
#>  sf            * 1.0-6   2022-02-04 [1] CRAN (R 4.1.2)
#>  sp            * 1.4-5   2021-01-10 [1] CRAN (R 4.1.0)
#>  stringi         1.7.6   2021-11-29 [1] CRAN (R 4.1.2)
#>  stringr         1.4.0   2019-02-10 [1] CRAN (R 4.1.0)
#>  testthat        3.0.2   2021-02-14 [1] CRAN (R 4.1.0)
#>  tibble          3.1.2   2021-05-16 [1] CRAN (R 4.1.0)
#>  tidyr           1.1.3   2021-03-03 [1] CRAN (R 4.1.0)
#>  tidyselect      1.1.1   2021-04-30 [1] CRAN (R 4.1.0)
#>  units           0.7-2   2021-06-08 [1] CRAN (R 4.1.0)
#>  usethis         2.1.3   2021-10-27 [1] CRAN (R 4.1.1)
#>  utf8            1.2.1   2021-03-12 [1] CRAN (R 4.1.0)
#>  vctrs           0.3.8   2021-04-29 [1] CRAN (R 4.1.0)
#>  withr           2.4.2   2021-04-18 [1] CRAN (R 4.1.0)
#>  xfun            0.30    2022-03-02 [1] CRAN (R 4.1.3)
#>  yaml            2.2.1   2020-02-01 [1] CRAN (R 4.1.0)
#>
#> [1] /home/martin/R/x86_64-pc-linux-gnu-library/4.1
#> [2] /usr/lib/R/library
```