

¹ Bayesian inference of prehistoric population dynamics from multiple
² proxies: a case study from the North of the Swiss Alps

³ Martin Hinz^{1,2,*} Joe Roe¹ Julian Laabs³ Caroline Heitz³ Jan Kolář^{4,5}

⁴ 08 December, 2022

⁵ **Abstract**

Robust estimates of population are essential to the study of human–environment relations and socio-ecological dynamics in the past. Population size and density can directly inform reconstructions of prehistoric group size, social organisation, economic constraints, exchange, and political and social institutions. In this pilot study, we present an approach that we believe can be usefully transferred to other regions, as well as refined and extended to greatly advance our understanding of prehistoric demography. Here, we present a Bayesian hierarchical model suite that uses Negative Binomial linear model and state-space representation to produce absolute estimates of past population size and density. At its core, the statistical model is as follows: if the proxies as a whole have a positive delta, a Negative Binomial draw for the new number of settlements is biased by a positive amount (and similarly for negative deltas). Using the area North of the main ridge of the Swiss Alps in prehistoric times (6000–1000 BCE) as a case study, we show that combining multiple proxies (site counts, radiocarbon dates, dendrochronological dates, and landscape openness) produces a more robust reconstruction of population dynamics than any single proxy alone. The model’s estimates of the credibility of its prediction, and the relative weight it affords to individual proxies through time, give further insights into the relative reliability of the evidence currently available for paleodemographic research. Our prediction of population development of the case study area accords well with the current understanding in the wider literature, but provides a more precise and higher-resolution estimate that is less sensitive to spurious fluctuations in the proxy data than existing approaches, especially the popular summed probability distribution of radiocarbon dates. The archaeological record provides several potential proxies of human population dynamics, but individually they are inaccurate, biased, and sparse in their spatial and temporal coverage. Similarly, current methods for estimating past population dynamics are often simplistic: they work on limited spatial scales, tend to rely on a single proxy, and are rarely able to infer population size or density in absolute terms. In contemporary demography, it is becoming increasingly common to use Bayesian statistics to estimate population trends and project them into the future. Bayesian approach is the natural and principled one for fusing data while appropriately propagating uncertainty. This makes it possible to qualify the uncertainty and credibility attached to forecasts. These same characteristics make it well-suited to applications to archaeological data in paleodemographic studies.

³³ ¹ Institute of Archaeological Sciences, University of Bern

³⁴ ² Oeschger Centre for Climate Change Research, University of Bern

³⁵ ³ CRC 1266 - Scales of Transformation, University of Kiel

³⁶ ⁴ Department of Vegetation Ecology, Institute of Botany of the Czech Academy of Sciences

³⁷ ⁵ Institute of Archaeology and Museology, Faculty of Arts, Masaryk University

³⁸ * Correspondence: Martin Hinz <martin.hinz@iaw.unibe.ch>

³⁹ Keywords: Prehistoric demography; Bayesian modelling; Multi-proxy; Settlement dynamics

⁴⁰ Highlights: - Bayesian modelling can integrate multiple, heterogeneous population proxies from the ar-

⁴¹ chaeological record - Our initial model produces more robust, high-resolution estimates of past population

⁴² dynamics than previous, single-proxy approaches - We provide absolute estimates of population size and

⁴³ density on the area north of the Swiss Alpes in prehistoric times (6000–1000 BCE) based on an initial value

44 1. Introduction

45 Prehistorians have long recognised demography as a fundamental force in human cultural evolution (Childe,
46 1936). Despite decades of interest in the population dynamics of prehistoric societies, concrete estimates
47 of population size and density before written records remain elusive. Though the archaeological record
48 provides multiple possible demographic proxies (Müller and Diachenko, 2019), a lack of access to this data
49 and methodological tools for turning it into quantitative estimates has left the conclusions drawn from it
50 vague and superficial (Hassan, 1981). As a result, ‘expert estimates’ transferred from ethnographic parallels
51 have often taken the place of direct inference from archaeological evidence (Morris, 2013; Turchin et al.,
52 2015).

53 Prehistoric demography has experienced a resurgence in interest in recent years (Riede, 2009 and others in
54 same issue; Shennan, 2000), partly explained by a renewed interest in human–environment relations and
55 human impact, necessarily requiring an assessment of population size. Kintigh et al. (2014) list human influence,
56 dominance, population size, and population growth amongst their ‘grand challenges’ for archaeology
57 in the 21st century.

58 In particular, the ‘dates as data’ technique (Rick, 1987), using the frequency of radiocarbon dates as a proxy
59 for population dynamics, has been significantly developed in the last decade (e.g. Shennan et al., 2013) and
60 widely applied to archaeological contexts worldwide (Crema, 2022). This approach has contributed greatly
61 to our understanding of prehistoric demography, but is not without its critics (Attenbrow and Hiscock,
62 2015; Carleton and Groucutt, 2021; Price et al., 2021). While the methodology continues to evolve and
63 address these critiques (Crema, 2022), it remains subject to fundamental problems. We believe that these
64 problems cannot be overcome by methodological refinements in this area alone. One of those problems is,
65 that most analysis of this kind rely on a single proxy (French et al., 2021; Schmidt et al., 2021). Other
66 data for demographic processes, such as age composition of grave groups (Bocquet-Appel, 2008; Kohler and
67 Reese, 2014) exist, as does the use of environmental data (Lechterbeck et al., 2014). The combination of
68 a large number of indicators can compensate for the weaknesses of the individual proxies used and open
69 up completely new possibilities for insights (Wolpert et al., 2021, p. 10). A Bayesian approach offers
70 a robust, quantitative methodology for inferring prehistoric population dynamics from multiple proxies,
71 including summed radiocarbon dates. Problems such as statistical dispersion of sum calibration and effects
72 of the calibration curve will be mitigated respectively detected by this combination with indicators that are
73 not subject to these effects. Moreover, in our approach, the mutual evaluation of the proxies in terms of
74 concordance penalises divergent behaviour of an indicator (due to intrinsic biases) by reducing its relevance
75 for the resulting estimate. While this does not eliminate these problems entirely, it does reduce them in a
76 numerically manageable way for large data sets.

77 In the following article we present our methodology and demonstrate its application with a case study in
78 the area north of the Swiss Alps. We use a type of Bayesian Generalised Linear Model (Kruschke, 2015) to
79 combine the different proxies. Their assessment (weight and credibility intervals) results from the modelling
80 and the combination of the data itself and is not defined a priori. Furthermore, we pursue the concept of a
81 state space model (Auger-Méthé et al., 2021), a time series model in which a time series is interpreted as the
82 result of a noisy observation of a stochastic process. We evaluate the number of settlements at time t as a
83 function of the number of settlements at time t-1, modified according to the changes in the proxies. For this
84 purpose, we also use these as the difference between these points in time and transform them accordingly.
85 In this sense, it is not a deductive model for testing a hypothesis, but a reconstructive (abductive) one,
86 such it is also used in environmental reconstructions (Inkpen and Wilson, 2009). In this respect, we hope to
87 convince the readers that such an approach is far better suited to realising reliable reconstructions of past
88 population developments than any of the previous approaches.

⁸⁹ **2. Background**

⁹⁰ **2.1. Population estimation in prehistory**

⁹¹ Proxies currently used for the estimation of population size in prehistory (following Müller and Diachenko,
⁹² 2019) can roughly be divided into three groups: ethnographic analogies; deductive estimates from eco-
⁹³ logical/economic factors; and the interpolation of frequencies of archaeological features (e.g. settlements,
⁹⁴ structures, individual finds). Three basic problems are common to all these approaches:

- ⁹⁵ **1. Reliance on single proxy:** Most investigations use only one source of evidence. Although multi-
proxy approaches exist, the individual proxies only serve to support each other or the main estimator,
without explicitly combining them.
- ⁹⁸ **2. Uncertainty in measurements:** All archaeological evidence is inherently uncertain which is carried
through to derived measurements. However, in most studies, single curves are presented as estimates,
and no uncertainty measurement exists.
- ¹⁰¹ **3. Lack of a linking function:** By ‘linking function’, we mean something that allows for the proxy data
to be interpreted in terms of actual population size or density. This could be absolute, i.e. a numerical
estimate of population, or relative, i.e. a means of scaling changes in the proxy value to changes in
population. Lack of suitable frameworks and ‘calibration’ data means that this is rarely presented
alongside proxy estimates. In the best cases, there is a qualitative assessment of the informative value
of the proxy, not sufficiently accounting for the complex nature of archaeological data.

¹⁰⁷ Our current implementation, with an abductive understanding of the data and the need to set a determinant
initial value for the reconstruction, represents a first step towards a fully developed end-to-end hierarchical
Bayesian model capable of solving these problems.

¹¹⁰ Furthermore, the types of archaeological data commonly used as population proxies share a number of
problematic characteristics, being:

- ¹¹² • **Limited:** We have only incomplete data, and it is usually not very informative.
- ¹¹³ • **Unevenly distributed:** For example, although there is a good data on settlement frequencies for
some regions, these regions are very unevenly distributed over time and space.
- ¹¹⁵ • **Noisy:** Frequently individual proxies are strongly influenced by factors unrelated to population, for
example taphonomic conditions or depositional biases.
- ¹¹⁷ • **Unreliable:** Research strategies, research history and varying levels of resources available to re-
searchers strongly affect the nature of compiled datasets. Systematic distortions are the rule rather
than the exception.
- ¹²⁰ • **Heterogeneous:** All potential proxies have different spatio-temporal scales, granularity, information
value, scales, and data formats.
- ¹²² • **Indirect:** We will never have direct data on prehistoric population; only proxy data that is thought to
be a reasonable substitute. The functions linking the proxy data with the desired quantity (population)
are unknown.
- ¹²⁵ • **Contradictory:** When considering several proxies, differences in linking functions, data quality and
noisiness inevitably lead to different results.

¹²⁷ Many, if not all, of these problems can be ameliorated through a) the explicit, quantitative integration of
¹²⁸ multiple proxies; and b) the use of a Bayesian approach to take account of and estimate uncertainty.

¹²⁹ **2.2. Hierarchical Bayesian demographic models**

¹³⁰ Many of the problems with archaeological population proxies are shared with contemporary demography. In
¹³¹ response, demographers have increasingly turned to Bayesian methods to estimate and forecast contemporary

132 population dynamics. For example, Bryant and Zhang (2018) consider Bayesian data modelling a solution
133 to exactly the kind of problems that affect archaeological data. Bayesian approaches are well suited for
134 limited, unreliable and noisy data. Various data sources, even contradictory data, can be brought into a
135 common framework and used to support one another. These methods also provide a quantitative estimate of
136 the likelihood and uncertainty of the model's resulting predictions (or in our case retro-dictions). Bayesian
137 approaches are also capable of accounting for spatially and temporally incomplete data: where this data
138 is missing, the uncertainty increases, but this does not prevent general modelling and estimation. Finally,
139 hierarchically-structured model suites, with sub-models for each individual proxy, can be used to estimate
140 linking functions between them and the value to be modelled, thanks to the interaction of a large number
141 of evidence.

142 This modeling technique can thus be used to join different lines of evidence horizontally and vertically and
143 combine their results into a overall estimate, including an assessment of their reliability: contradicting data
144 lead to a lower overall reliability, while a mutual support to smaller confidence intervals. If there is no
145 systematic bias that affects all data sources to the same extent, this results in the most reliable estimate
146 possible through the most heterogeneous set of data sources. Even though data fusion can be achieved with
147 different techniques, the Bayesian approach seems to us to be excellently suited for our purposes due to its
148 flexibility and robustness.

149 Bayesian radiocarbon calibration is a similar, well-established application in archaeology, where radiometric
150 uncertainty is modelled based on prior stratigraphic information. More recently, archaeologists have also
151 used Bayesian modelling techniques for testing hypotheses relating to demographic models based on ^{14}C data
152 (e.g. Crema and Shoda, 2021). This approach differs from the one presented here in that, in these analyses,
153 deductive models are generated and their plausibility is tested on the basis of ^{14}C data only. This is a clear
154 step forward to a model-based, scientific analysis. However, the use of only one proxy, exclusively for testing
155 hypotheses developed independently, creates problems comparable to those of the inductive approaches used
156 so far: without a combination with other indicators, one is largely limited to the problems and conditions
157 of summation calibration. However, recent work (Carleton and Groucutt, 2021; Price et al., 2021) shows a
158 possible way out here, and might be a promising alternative to abductive approaches like ours.

159 We attempt to make Bayesian hierarchical techniques usable for archaeological reconstructions. We want to
160 show, in a reproducible and practical form using a case study, how Bayesian methods can make a decisive
161 contribution to a better assessment of population development, crucial for the reconstruction of the human
162 past, even in for periods for which we only have very patchy, noisy and unreliable data.

163 2.3. The Bayesian approach

164 Bayesian statistics relies on the premise that there is always some prior assumption, even if very rough,
165 about the probability of an event. This assumption is adjusted by observing data, by checking how credible
166 these priors are (likelihood, see also Bryant and Zhang, 2018, p. 66). This is Bayesian updating (cf. also
167 Kruschke, 2015, especially 15–25), resulting in the posterior probability distribution, which represents not a
168 point prediction. Small amounts of data lead to a broad distribution not strongly localised and restricted.
169 Thus, we simultaneously obtained a result and an estimate the credibility interval, given the data.

170 This Bayesian learning is iterative and sequential, so that the result of one Bayesian inference can form the
171 prior of another (Kruschke, 2015, p. 17). This allows different information to be combined (Bryant and
172 Zhang, 2018, pp. 219–224), as it has long been exploited by archaeology in using stratigraphic information
173 to make radiometric dating more accurate (Ramsey, 1995).

174 This also makes a hierarchical formulation of problem domains possible. Parameters that are necessary for
175 an estimation, such as the relationship of population density to the deforestation signal in pollen data, need
176 not be specified explicitly, but can be given by probability distributions and then estimated in the model
177 itself (Bryant and Zhang, 2018, p. 186). The more data available, the more degrees of freedom can be
178 estimated with a reasonable width of credibility intervals (Kruschke, 2015, p. 112). For the estimation of
179 these parameters, submodels have to be created describing the relationship of the data to the characteristics
180 of the parameter (Kruschke, 2015, pp. 221–222).

¹⁸¹ 3. Materials: population proxy data

¹⁸² Our case study area north of the Swiss Alps (Figure 1) covers about one third of Switzerland's territory and
¹⁸³ comprises the partly flat, but largely hilly area between the Jura Mountains and the Alps. It is favourable
¹⁸⁴ for settlement and agriculture; the Swiss Plateau between Lake Zurich and Lake Geneva is by far the most
¹⁸⁵ densely populated region of the Switzerland today. This serves as our core region of interest because it
¹⁸⁶ is here that archaeological data is most abundant and accessible. The region has a very diverse natural
¹⁸⁷ landscape: shaped by glaciers during the ice ages, the many lakes and bogs provide excellent preservation
¹⁸⁸ conditions for the numerous Neolithic and Bronze Age lakeside settlements, and a rich source for vegetation
¹⁸⁹ reconstructions by means of pollen analyses. Thanks to very active and efficient archaeological research and
¹⁹⁰ heritage management there is an abundance of archaeological information, including known sites as well as
¹⁹¹ dendrochronological and ¹⁴C data.

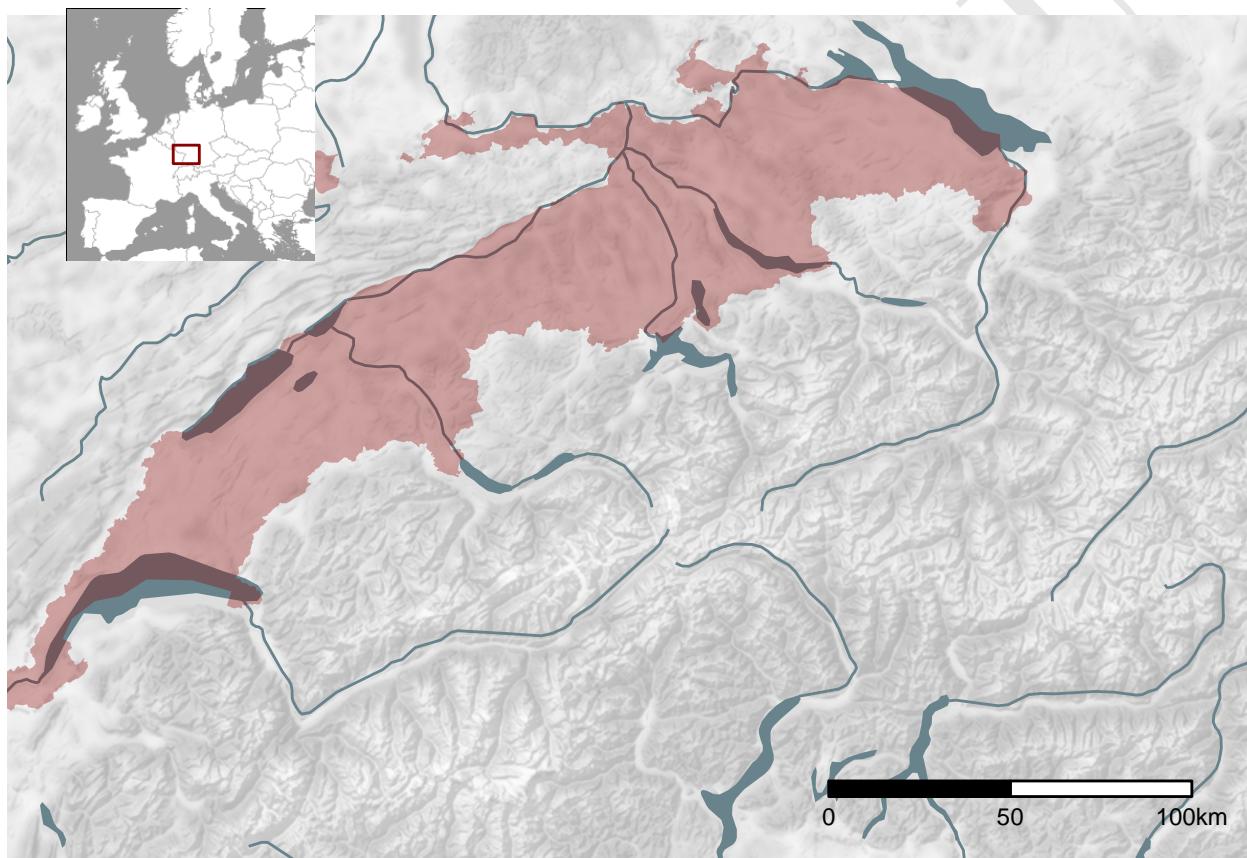


Figure 1: Location and extent of the Swiss Plateau as biogeographical region (based on swisstopo) including additional low altitude areas in the north of Switzerland (regions along the High Rhine between Schaffhausen and Basel).

¹⁹² Our case study targets the period between 6000–1000 BCE. The lower limit of this time window was chosen
¹⁹³ to avoid the so-called 'Hallstatt plateau' in the Northern Hemisphere radiocarbon calibration curve, which
¹⁹⁴ causes difficulties for the using sum calibration and similar approaches. The upper limit coincides with post-
¹⁹⁵ glacial changes in pollen spectra, before which the openness indicator is highly unlikely to reflect human
¹⁹⁶ influence.

¹⁹⁷ A large number of different proxies can be integrated into a model of this type, provided that these ob-
¹⁹⁸ servations a) can be understood as dependent on the population density in the past, and b) a model-like
¹⁹⁹ description of this dependence can be created. Table 1 provides a non-exhaustive list. Furthermore, it is

200 possible to combine analytical deductive approaches with the inductive proxies, such as ethnographic analogies
201 and economic modelling. For our case study, we used a landscape openness indicator; an aoristic sum
202 of typological dated sites; a radiocarbon based settlement count (sum calibration); and frequency data for
203 dendro-dated lakeshore settlements in the Three Lakes region (western Swiss Plateau).

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

Possible Proxies
Expert estimates
Extrapolation of buried individuals
Burial anthropology
Settlement data, number of houses
Settlement data, settlement size
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
...

204 For all proxies used, it is a common assumption that they are positively correlated with past population levels.
205 In some cases we have reprocessed them in this respect in order to emphasise this relationship even more
206 strongly in our opinion (example: pollen data). This is of course a strong assumption, but we believe it is
207 justified, as this has been widely accepted in the scientific literature, and these proxies are used accordingly
208 (each in isolation). Our ambition with this article is to present an abductive model to help improve the
209 current practice of proxy use. Our ambition is not to justify or verify the use of these proxies itself. This
210 can only be done through empirical, hypothesis-testing approaches with known response observations. What
211 our method can do, however, is to infer, through a weighting resulting from the modelling, which indicators
212 seem to be better suited in terms of their variability and fit with other indicators to trace a population
213 development in the past.

214 3.1. Dendro-dated lakeshore settlements

215 From the Neolithic onwards, known settlement areas in Switzerland concentrate along its rivers and lakes
216 (Christian Lüthi, 2009). Thus, our working region offers on the one hand excellent data for population
217 estimation, but on the other hand poses very specific problems for such an undertaking. If we have high-
218 resolution information on the temporal sequence of individual settlements at the lakeside settlements by
219 means of dendro data, this also might cause a research problem with regard to the ^{14}C data often used as a
220 proxy.

221 The dataset we use for the number of dendro-dated wetland settlements in the Three Lakes region was
222 collected by Julian Laabs for his PhD thesis (Laabs, 2019). The time series used here runs from 3900 to
223 800 BCE, and contains the number of chronologically registered fell phases at individual settlements. This
224 results in a time series that reflects the settlement of the lakeshores in the Neolithic and Bronze Age periods.

225 **3.2. Radiocarbon Based Settlement Count**

226 Summing up the probabilities of radiocarbon dates can undoubtedly be considered the standard approach
227 in demographic reconstruction in archaeology today. In our variant of the summation calibration, we sum
228 up the individual data in each site so that each site receives a weight of 1. These individual probabilities are
229 then summed up over all sites, resulting in a radiocarbon based settlement count (RBSC). In many respects,
230 this follows the same logic as generally applied summation calibration, where binning achieves much the
231 same outcome.

232 The logic for RBSC can be derived from the Poisson binomial distribution: such a distribution can be
233 described as the sum of independent Bernoulli trials as it represents the probability estimate for the existence
234 or use of a site per unit time. The expected value per time unit for the number of simultaneously existing
235 sites/settlements is the sum of the individual probabilities for the sites, i.e. the area under the curve of the
236 probabilities of ^{14}C data. Incidentally, the same logic applies to the aoristic sum (shown below). The main
237 problem with this proxy is therefore not so much its statistical nature, but mainly the systematic biases
238 that arise from the conservation and finding practicalities in the research process. To counter these, and to
239 detect and mitigate the systematic biases, hierarchical modelling such as that carried out in this study is
240 appropriate.

241 The dataset for the RBSC primarily consists of data from the XRONOS database (<https://xronos.ch>),
242 supplemented by dates from the unpublished PhD thesis of Julian Laabs (Laabs, 2019) and the data collection
243 of Martínez-Grau et al. (2021). It contains a total of 1135 single ^{14}C data from 246 sites (see Figure 2). The
244 dates in the dataset range in ^{14}C years from 10730 to 235 uncal BP. This time window extends beyond the
245 study horizon in order to minimise boundary effects.

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

249 Dendrodated lakeshore settlements and RBSC are two fairly direct proxies for settlement numbers over time.
250 However, both are subject to certain constraints, especially on the Swiss Plateau, but also beyond. They
251 require the discovery of corresponding settlement sites. This is subject to preservation and discovery filters,
252 which distort the results of the derived estimates. These biases are not connected to past population trends,
253 but result from research activity, preservation and publication status, as well as dating strategies. Therefore,
254 in our model we supplement these two direct proxies with such proxies that are more independent of these
255 biases.

256 **3.3. Aoristic sum**

257 To add another archaeological indicator of occupation, we include relative dating information obtained from
258 the Swiss cantonal archaeology/heritage management authorities (Figure 3). These are primarily derived
259 from scattered surface finds, which often have a low dating accuracy. We incorporate this data into our model
260 as a typologically dated aorist time series. The dating accuracy is only in the range of archaeological periods,
261 but the advantage is that we are not bound to the conditions and problems of radiocarbon dating and thus
262 methodological issues of sum calibration can be avoided (scatter, calibration curve effects). Furthermore,
263 this data provides an independent indicator with regard to the methodology of the ^{14}C data, even if they are
264 influenced by similar transmission filters and archaeological conditions as the evaluation of ^{14}C data. Data
265 from 4321 sites were included in the aoristic sum, which is a very rough indicator due to the low dating
266 accuracy offered by archaeological phases, but which nevertheless has an important role in the normalisation
267 of the data due to its independence from calibration effects.

268 **3.4. Landscape openness**

269 Natural conditions in the Swiss lakes enable not only highly precise dating of archaeological sites, but also
270 a very dense network of pollen analysis. We make use of this by generating a supra-regional openness

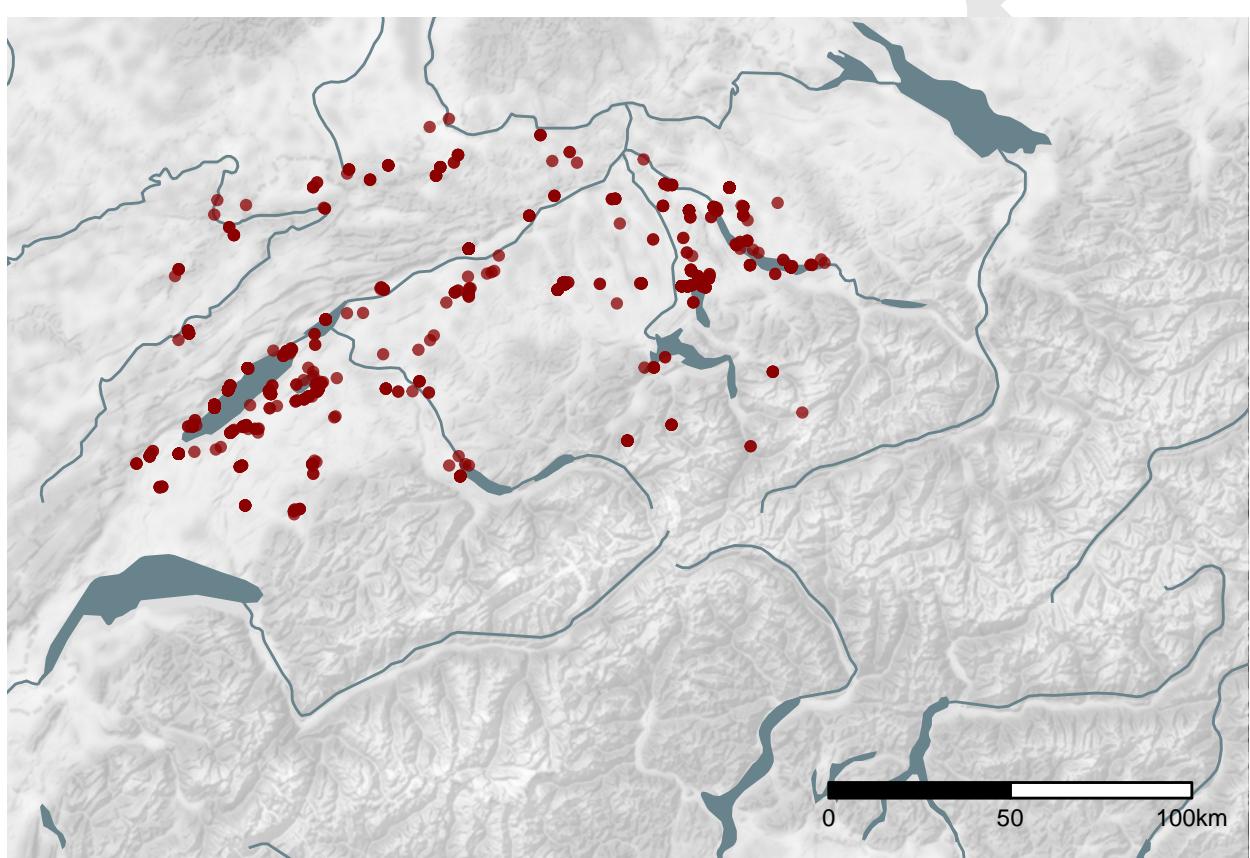


Figure 2: The location of the ^{14}C dated sites in the dataset.

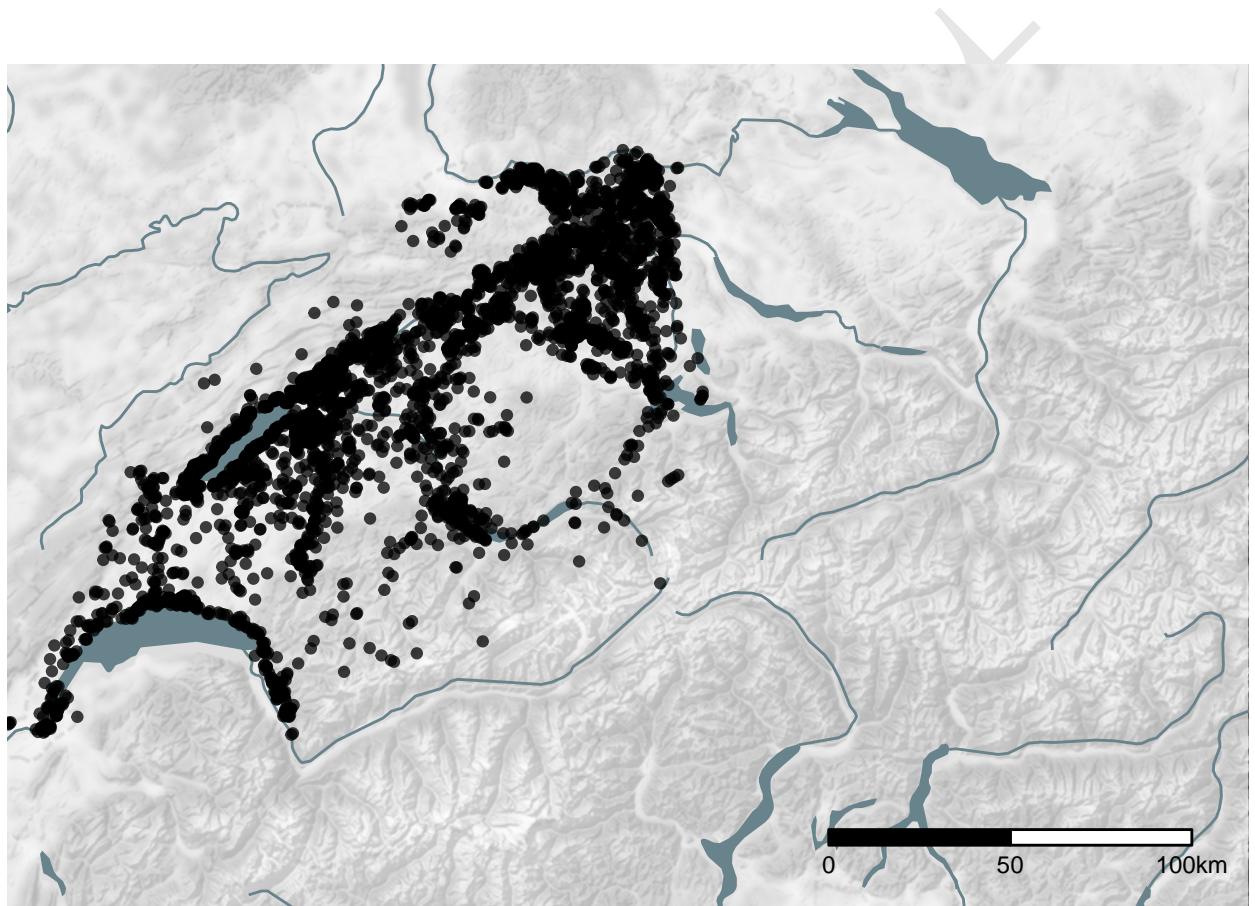


Figure 3: Location of the sites from the find reports of cantonal archaeology (heritage management) authorities. Locations are ‘fuzzed’ by approximately 1 km.

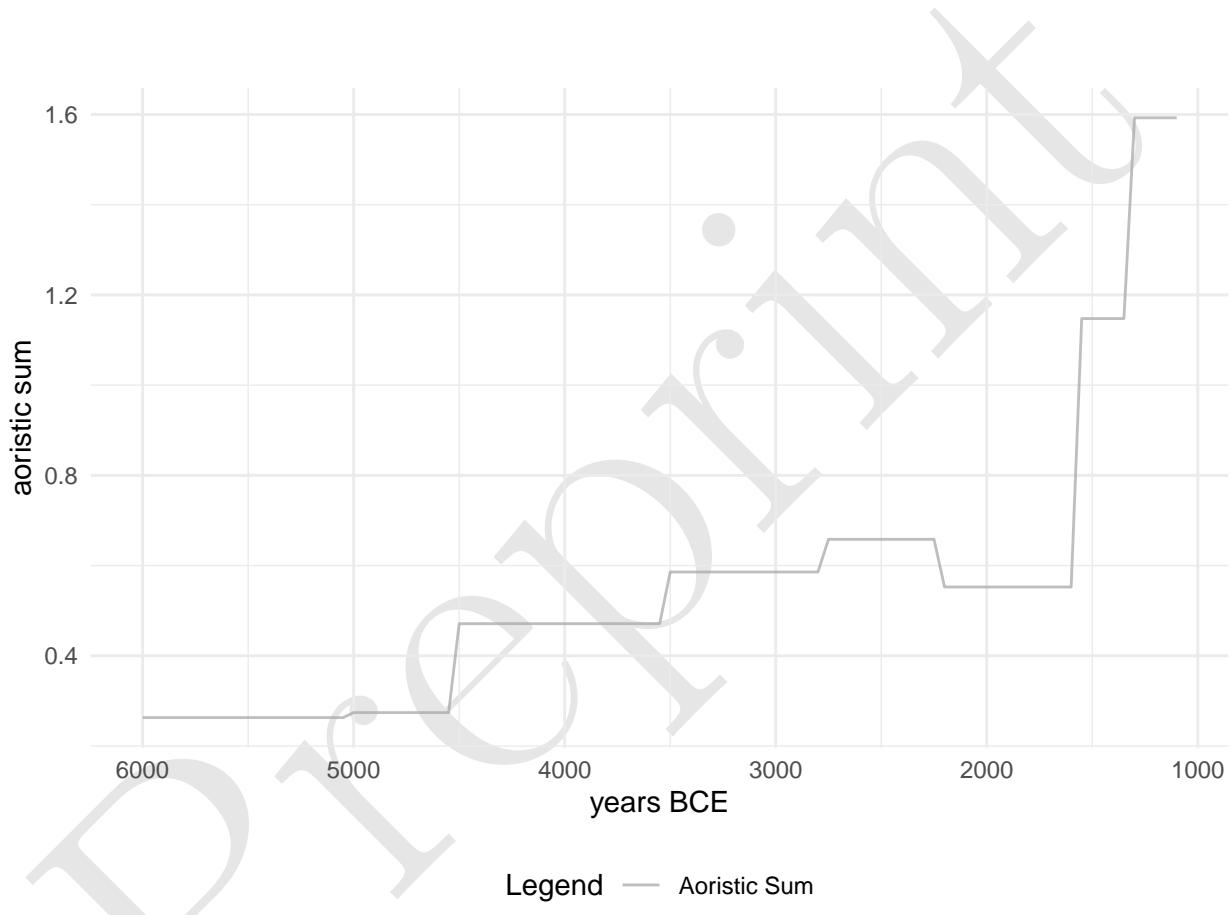


Figure 4: Aoristic sum of archaeological sites used in the analysis.

271 indicator for the vegetation from the pollen data (Figure 5). This proxy has the specific advantage that it is
272 not dependent on archaeological preservation conditions, making it particularly valuable for compensating
273 systematic distortions that result from archaeological taphonomy and period-specific settlement patterns.

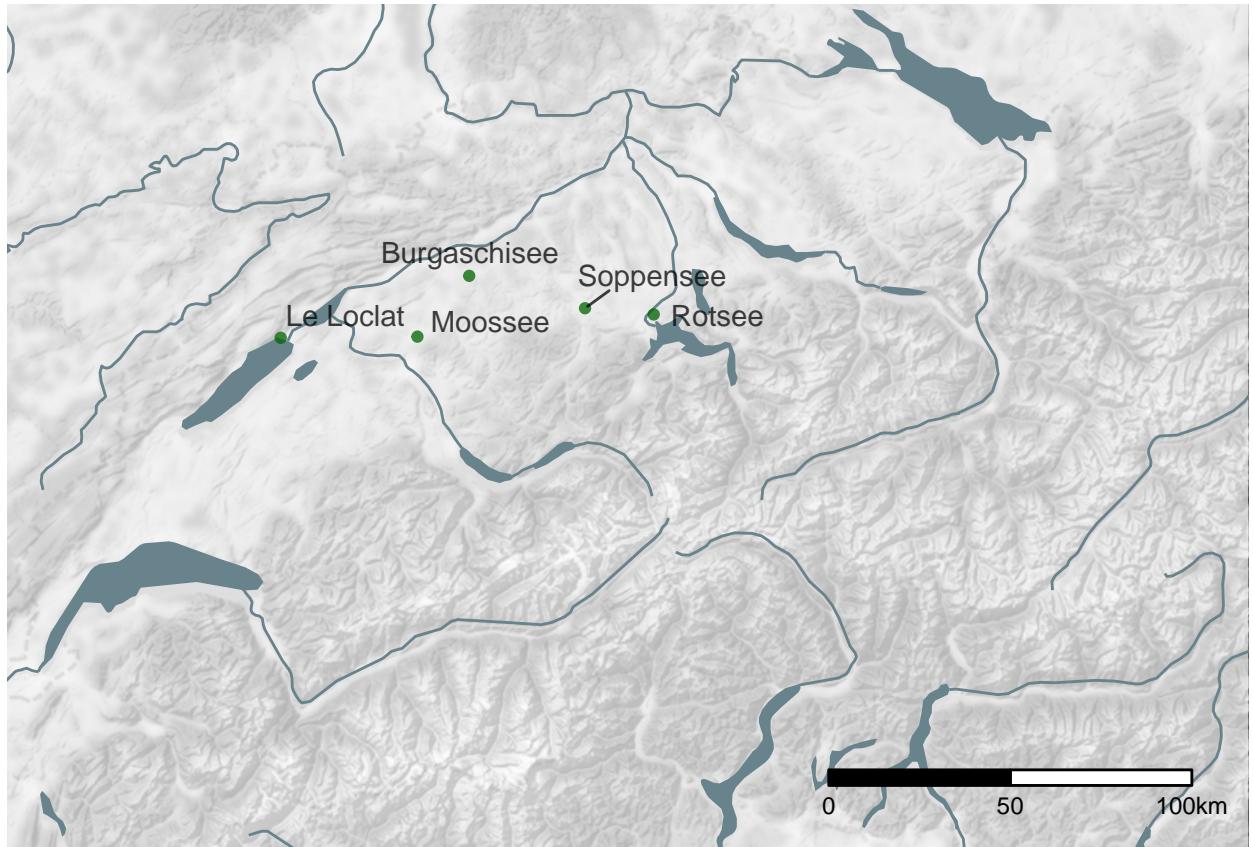


Figure 5: Location of the pollen profiles used for the openness indicator.

274 The utilisation of this proxy is based on the assumption that the higher the population density in an area, the
275 greater the human influence on the natural environment (Lechterbeck et al., 2014), and that the extensiveness
276 of agricultural activity in an area is related to human population density (Zimmermann, 2004). Evidence
277 of land clearance in pollen diagrams can therefore provide further indications of population dynamics where
278 humans can be assumed to be the main driver of this process, which is the case in much of Europe. The full
279 procedure for deriving this proxy from several different pollen diagrams is detailed in a previous publication
280 (Heitz et al., 2021). In this study, five pollen diagrams from sites mainly in the hinterland of the large
281 Alpine lakes were used. The technical steps are also documented in the accompanying R compendium. The
282 percentage pollen data, based on a pollen sum of all terrestrial taxa of the individual sites, was combined
283 into one data set by means of a principal component analysis (Figure 6). Only terrestrial pollen taxa
284 with a frequency of more than one third and, if present, with an average frequency of at least 0.1% were
285 selected to reduce potential disturbance by rare species. Cereal pollen was explicitly retained as an important
286 anthropogenic indicator. As each sample is absolutely dated, the data on the x-axis can be plotted against
287 the openness value on the y-axis to obtain a time series time series for land clearance.

288 4. Methods: Bayesian model

289 RBSC, openness and the dendro-dated settlement data was smoothed by a moving average with a 50 years
290 window, corresponding to the unified sampling interval for all proxies. The choice of interval reflects a

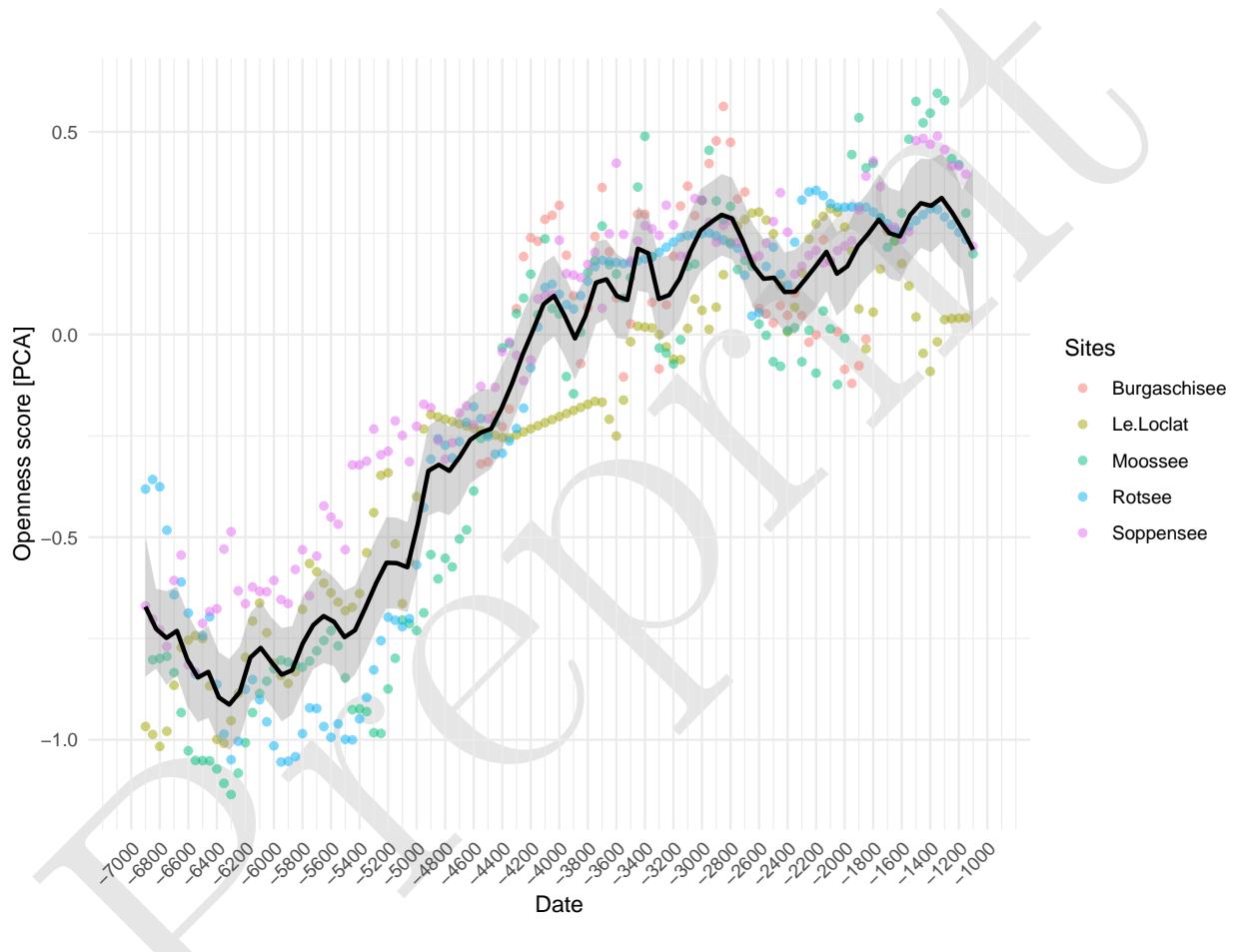


Figure 6: Value on the first dimension of the PCA against dating of the samples for the individual pollen profiles and their combined average value as the openness indicator.

291 trade-off between the data quality of the proxies and the complexity of the model. The aoristic sum was not
292 smoothed, because it already has a very coarse temporal resolution. In the construction of our ‘observational
293 model’, we considered all these proxies as informative of the number of settlements located in the north of
294 the Swiss Alps. Population development is simulated in a ‘process model’ using a Negative Binomial process.

295 4.1. Process model

296 A special class of Bayesian hierarchical models are so-called ‘state space models’, specifically designed for
297 time series. They follow two principles. First, a hidden or latent process is assumed, representing the state
298 of the (vector) variable of interest z through the entire time series. Every state of variable z as z_t at a specific
299 time is bound by a Markov process to the state of variable z at time before, at $t - 1$. Second, it is assumed
300 that certain observations, represented in variable y , are dependent on the state of variable z at time t . This
301 implies that a relationship between the individual manifestation of variable y is generated over time via the
302 hidden variable z , which is not directly observable.

303 In the following I reproduce the mathematical representation according to Auger-Méthé et al. (2021):

$$z_t = \beta z_{t-1} + \epsilon_t \quad (1)$$

$$x_t = \alpha z_t + \eta_t \quad (2)$$

304 “The autocorrelation in the states is captured by the parameter β . The observations are a function of the
305 states only and the parameter α allows the observation at time t to be a biased estimate of the state at time
306 t .” (Auger-Méthé et al., 2021, pp. 4–5). Here ϵ is the process variation, and η the observation error, which
307 can be modelled with a suitable distribution. In our current implementation we ignore α and thus assume
308 that x_t is an unbiased observation. We compensate for this (see below) by defining β as a simplex with sum
309 1. This leads to the intuition that each of the proxies is seen as more or less good compared to the others,
310 but in their sum they represent the best possible description of z_t (given the available data).

311 To clarify the structure of the dependencies, the model can also be represented in terms of probability
312 distributions:

$$f(z_t|z_{t-1}, \theta_p), \quad t = 1, \dots, T \quad (3)$$

$$f(x_t|z_t, \theta_o), \quad t = 1, \dots, T \quad (4)$$

313 Here f and g are two probability density functions, and θ_p and θ_o are vectors of parameters associated with
314 each equation. Eq. 3 describes the autocorrelation in state values as a first-order Markov process, and Eq.
315 4 describes how observations depend simply on the states. This definition also demonstrates that states are
316 random variables and thus that SSMs are a type of hierarchical mode (Auger-Méthé et al., 2021, p. 5).

317 This structure makes these models particularly suitable for population reconstruction using archaeological
318 and other data. Population density itself is not directly measurable: all we have at our disposal are
319 observations derived by unknown linking functions.

320 Our overall model is broken down into several hierarchically-connected individual elements. The process
321 model represents the population development itself, without already being explicitly parameterised with
322 data. Here we assume that the latent variable ‘number of sites’ (z) is strongly autocorrelated across different
323 time periods (z_t). The number of sites in 3000 BCE is strongly conditioned by the number of sites in 3050
324 BCE, and so on. The population at time t results from the population at time $t - 1$ times a parameter β ,
325 which represents the population change at this time (see Equation 1).

326 A univariate discrete Poisson distribution is particularly suitable for modelling frequencies, numbers of events
 327 that occur independently of each other at a constant mean rate in a fixed time interval or spatial area. It is
 328 determined by a real parameter $\lambda > 0$, describing the expected value and the variance. Thus, the relationship
 329 shown above can be rearranged as follows:

$$z_t \sim Poisson(\lambda_t) \quad (5)$$

330 where

$$\log(\lambda_t) = \log(z_{t-1}) + \sum_{i=1}^n \beta_i x_{t,i} \quad (6)$$

331 which can be rearranged to

$$\lambda_t = z_{t-1} e^{\beta^T \delta x_t} \quad (7)$$

332 β is a parameter vector of weightings that sums to 1, a vector transpose is indicated by T , and x_t is a vector
 333 of proxy difference values.

334 Using a link function ensures that λ , which must always be positive, can also be described by variables that
 335 may also be negative. β serves as slope factor, as in a normal linear regression. Here, it functions as a scaling
 336 factor for the individual proxies. z_t is to be understood as an intercept, representing a baseline when there
 337 were no change due to the variables. This is the desired behaviour: λ is equal to the value of the population
 338 in the previous time period, plus or minus the changes resulting from the variables.

339 A Poisson distribution has the constraint that the variance is equal to the mean. This distribution may have
 340 been empirically very useful for modelling respective processes. However, we were advised in the peer review
 341 of the article that this limitation may not be justified. Therefore, we have rebuilt and run the model using
 342 a Negative Binomial Distribution, and present here the results of this process model, contrasting them with
 343 those of the Poisson model.

344 A negative binomial distribution can be seen as a generalisation of a Poisson distribution. In this case, the
 345 expected value and variance are parameterised separately, which gives the model an additional degree of
 346 freedom. This affects the credibility intervals, which become correspondingly wider. A negative binomial
 347 distribution is described by a parameter of the probability of success p , and a parameter r , which indicates
 348 the rate. Therefore, the implementation of the process model in this variant looks like this:

$$z_t \sim NB(r, p_t) \quad p_t = \frac{r}{r + \lambda_t} \quad r \sim halfCauchy(0, 25) \quad (8)$$

349 The negative binomial distribution has a variance $\frac{r(1-p)}{p^2}$, whereby the distribution becomes identical to
 350 the Poisson distribution in the limit $r \rightarrow \infty$. In practise, values above 50 are considered to produce a
 351 distribution that approaches a Poisson Distribution. In the model, r is estimated from the data using a
 352 half-Cauchy distribution with a location parameter of 25 as prior. This represents a weakly informative
 353 prior according to the standard specifications for a scale hyperparameter (Gelman, 2006).

354 Population size z_t as well as population change λ_t are time-dependent. At each individual point in time,
 355 these variables will take on different values. But we can assume that the population change will not exceed
 356 certain limits (*max_change_rate*), though it is not possible to specify this at this point.

$$\begin{aligned} max_growth_rate &\sim Gamma(shape = 5, scale = 0.05) \\ z_t/z_{t-1} &< (max_growth_rate + 1) \\ z_{t-1}/z_t &< (max_growth_rate + 1) \end{aligned} \quad (9)$$

357 A gamma distribution is defined $[0, \infty)$. This parameterisation used here centres the probability mass in
 358 the range 0-1; adding 1 makes this range 1-2. This prevents the number of sites from explosively increasing

359 between two time periods, which would lead to problems for the convergence of the model. The estimation
 360 of this parameter for the entire model, as well as the estimation of the respective population change per time
 361 section, results from the modelling and the interaction with the data.

362 4.2. Observational model

363 In this initial implementation, the observational model is essentially a Negative Binomial regression, where
 364 the proxies are used to inform the change in the number of settlements between time steps. The individual
 365 proxies were z-normalised and absolute differences between time steps were then computed. If the value of
 366 the proxy increases, this results in a positive difference from the previous time step, and vice versa.

$$x'_t = \frac{x_t - \bar{x}}{\sigma_x^2} \mid \sigma_x^2 := \text{Standard Deviation} \quad (10)$$

$$\delta x'_t = x'_t - x'_{t-1}$$

367 The fact that we use the difference here, and not the actual value, may need some explanation.
 368 The sum of the resulting differences between the time steps, together with the settlement number of the
 369 previous step as the expected value, then forms λ_t : the expected value for the settlement number of the
 370 current time step (see Equation 6).
 371 This means that we consider the development of population size and thus the number of settlements as
 372 autocorrelated per se. There are two reasons for this: First, the proxies we use are obviously autocorrelated.
 373 This can be seen as empirical confirmation that such an autocorrelation can also be regarded as probable
 374 for the latent variable of number of sites behind it. On the other hand, a model that does not assume
 375 such an autocorrelation would mean that population development is primarily controlled by migration, since
 376 the population already present in an area would play no role in this. We think it more likely that it is
 377 primarily births and death events, together with migration, that are responsible for number of sites and
 378 their change. Therefore, we view the population at time t (and therefore the number of sites) as a function
 379 of the population at time t-1 plus a change that can be estimated from the change in the proxies. In this
 380 way, we make all time steps interdependent, which constrains the model in a comprehensible and justifiable
 381 way, and thus narrows the probability range for the prediction by providing additional information in the
 382 model. Here, β_i is a scaling factor that represents the influence of the changes in the individual proxies. It
 383 is a confidence value of the model for the respective proxy, so that the sum of all β_i results in 1.

$$\sum_{i=1}^n \beta_i = 1 \quad (11)$$

384 A Dirichlet distribution—a multivariate generalization of the beta distribution—is commonly used for this
 385 purpose in hierarchical Bayesian modelling. Its density function gives the probabilities of i different exclusive
 386 events. It has a parameter vector $\alpha = (\alpha_1, \dots, \alpha_i) \mid (\alpha_1, \dots, \alpha_i) > 0$, for which we have chosen a weakly infor-
 387 mative log-normal prior. The priors for the log-normal distribution in turn come from a weakly informative
 388 exponential distribution for the mean and a log-normal distribution with μ of 1 and σ_{log} of 0.1:

$$\begin{aligned} \beta_i &\sim \text{Dirichlet}(\alpha_{1-i}) \\ \alpha_i &\sim \text{Lognormal}(\mu_{\alpha_{i-1}}, \sigma_{\alpha_{i-1}}^2) \\ \mu_{\alpha_{i-1}} &\sim \text{Exp}(1) \\ \sigma_{\alpha_{i-1}}^2 &\sim \text{Lognormal}(1, 0.1) \end{aligned} \quad (12)$$

389 Intuitively, we consider the sum of the changes of the proxies as best possible representation of the change in
 390 number of settlements, given the data that we have. That is, the share of each individual proxy is variable

391 and is estimated within the model. This is represented by β in the mathematical model. In the computational
392 model, we named this parameter p (for percentage influence)

393 Another way of approaching this would be to take an end-to-end Bayesian approach (eg. Price et al., 2021).
394 Here we would model the relationships between the proxies and the population size directly, and let the
395 model estimate the parameters of this process. This was our original approach. We had to abandon it for
396 the time being due to the underdetermined nature of such a model with our data. However, we are convinced
397 that such an approach would lead to a better model that propagates uncertainty more appropriately and
398 avoids certain biases associated with the conversion to a fixed time series and the associated biases due to
399 preprocessing. We plan to revert to this more principled approach when applying our model on a larger scale
400 with a regionalisation and an extension of the proxies.

401 The error value is represented by the Negative Binomial process in the process model. In this implementa-
402 tion, the model finds the best possible combination between the individual proxies to describe a settlement
403 dynamic. The number of sites is converted into population density using (certainly debatable) parameters
404 defined by us, but which are only scaling factors for the intermediate value of number of settlements. We
405 assume that each site represents a number of people that is poisson distributed around the value 50, a com-
406 promise, as both Mesolithic and Neolithic and Bronze Age settlement communities need to be represented.
407 An evidence-based estimate data series of the temporal development of settlement sizes could enhance this
408 specification. From the number of sites and the mean number of individuals a population density can be
409 calculated using the case study area (12649 km^2), making the models estimate comparable with estimates
410 from other sources or the literature.

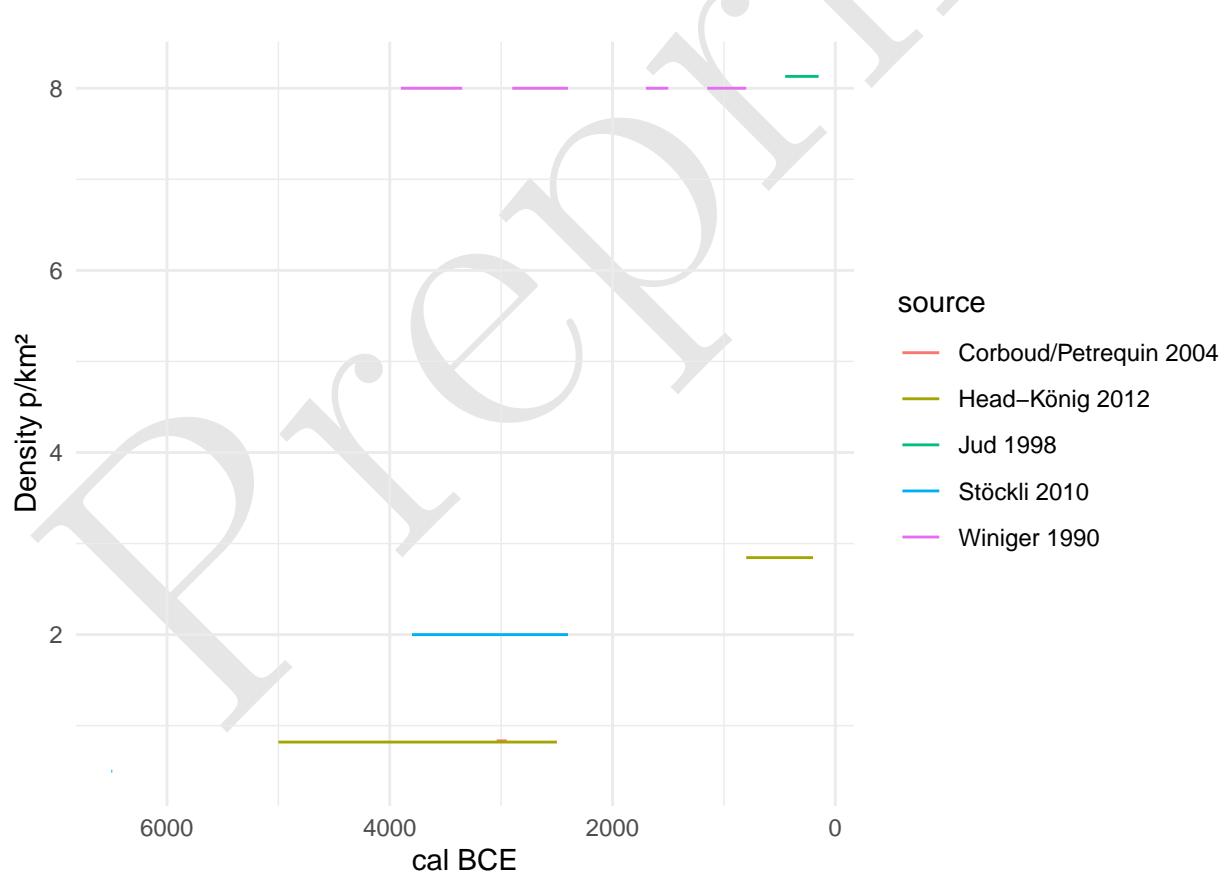


Figure 7: Expert estimate of population density on the Swiss Plateau.

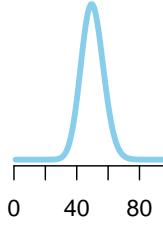
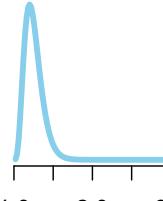
411 **4.3. Model fitting**

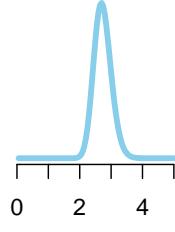
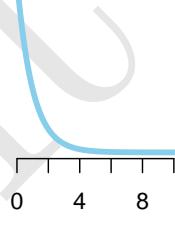
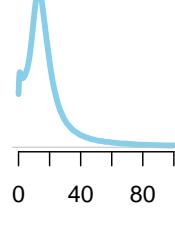
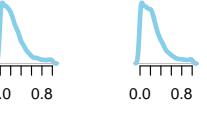
412 The model was fitted using the R package *nimble* (version 0.11.1, R version 4.1.3), using 4 parallel chains.
413 Achieving and ensuring convergence and sufficient effective samples (10000) for a reliable assessment of the
414 highest posterior density interval was carried out in steps.

- 415 1) the model was initialised for each chain and run for 100000 iterations (with a thinning of 10). On
416 a reasonably capable computer (Linux, Intel(R) Xeon(R) CPU E3-1240 v5 @ 3.50GHz, 4 cores, 8
417 threads), this takes approximately a minute.
418 2) the run was extended until convergence could be determined using Gelman and Rubin's convergence
419 diagnostic, the criterion being that a potential scale reduction factor of less than 1.1 was achieved for
420 all monitored variables. Convergence occurred after about thirty seconds.
421 3) Due to the high correlation of the parameters and thus a low sampling efficiency, the collection of at
422 least 10,000 effective samples for all parameters took about five hours.

423 A starting value of 5 p/km² for the population density of the Late Bronze Age (1000 BCE) was taken from
424 the literature, which may represent a general average value for all prehistoric population estimates (Nikulka,
425 2016, p. 258). For the model, this was set as the mean of a normal distribution with a standard deviation of
426 0.5, which should give enough leeway for deviations resulting from the data. Nevertheless, it should be noted
427 that our resulting estimate is strongly conditioned by this predefined value, especially in the later sections.
428 For traceplots and the prior-posterior overlap, as well as density functions of the posterior samples of the
429 individual parameters, please refer to the supplementary material.

Table 2: Priors and fixed parameters used in the model.

Priors	Value	Plot/Comment
MeanSiteSize	dpois(50)	
max_growth_rate	dgamma(shape = 5, scale=0.05) + 1	

Priors	Value	Plot/Comment
mu_alpha	dlnorm(1,sdlog=0.1)	
a_alpha	dexp(1)	
alpha	dlnorm(mu_alpha[j],sdlog=a_alpha[j])	
p	ddirch(alpha[1:4])	
Parameters		
nEnd	5	
AreaSwissPlateau	12649 km ²	
Initial Values		
lambda _{1:nYears}	$\log(1 - 10^{\frac{1}{nYears-1}})$	exponential increase of the factor 10
PopDens _{1:nYears}	nEnd (=5)	
nSites _{1:nYears}	50	

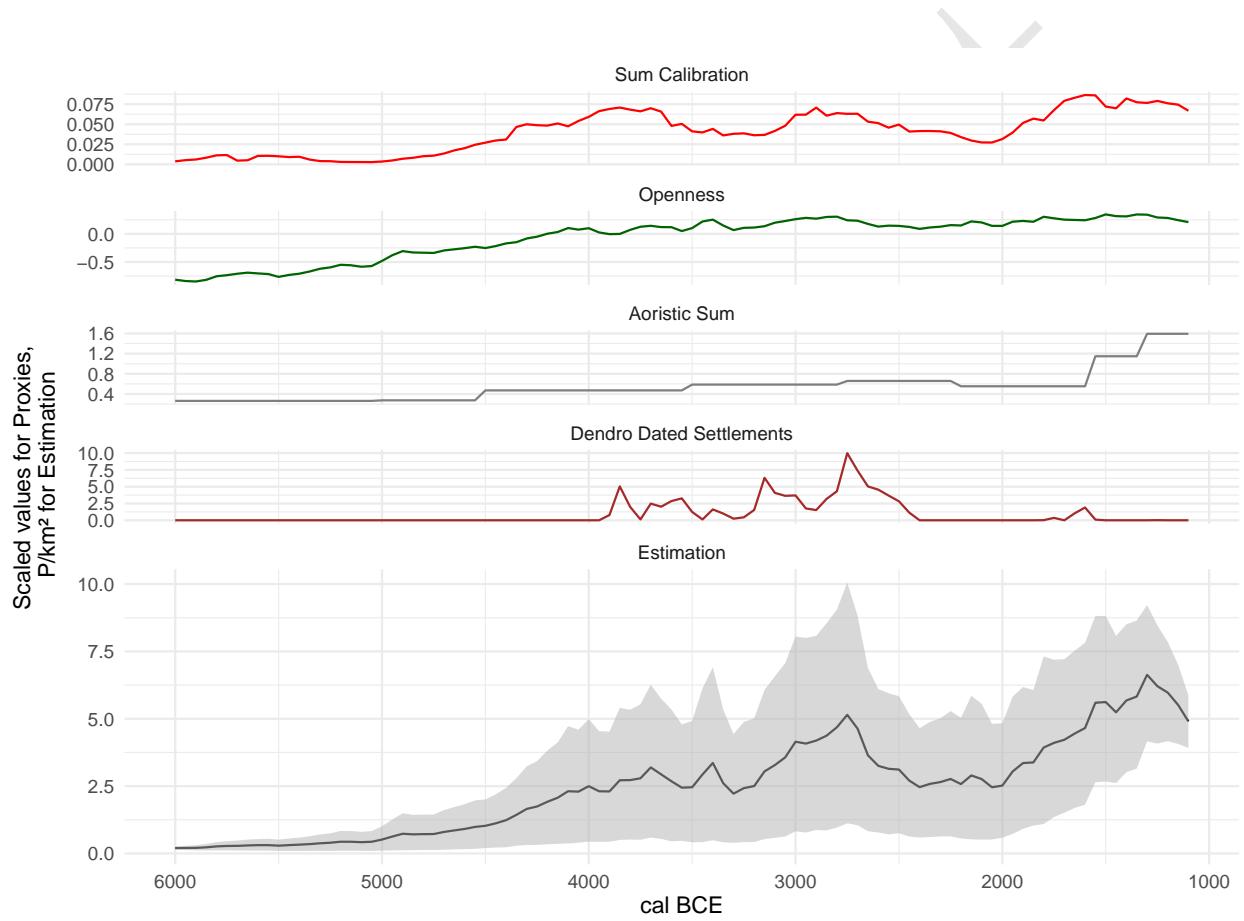


Figure 8: Estimate of population density predicted by the model. The four input proxies are also plotted (scaled) for comparison.

430 **5. Results**

431 **5.1. Negative Binomial Model**

432 The population density estimated by the model (Figure 8) ranges between 0.2 p/km² for the beginning (6000
 433 BCE) and 4.8 p/km² for the end of the estimate (1000 BCE), reaching a maximum of 6.5 p/km² for around
 434 1250 BCE. This remains within the range considered plausible according to expert estimates. There are
 435 clear peaks around 1250 BCE and around 2750 BCE, which corresponds to the beginning of the influence of
 436 Corded Ware ceramic styles (Hafner, 2004).

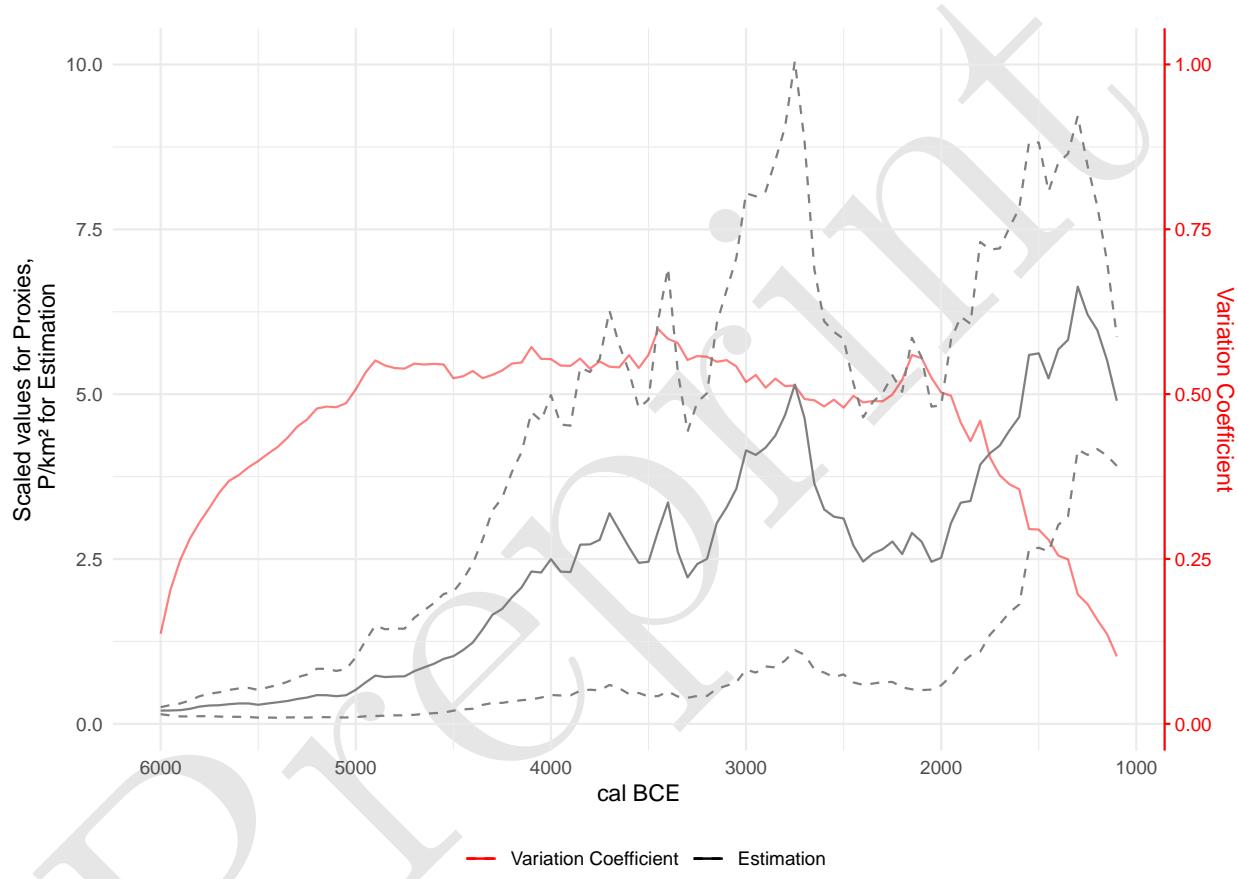


Figure 9: Variability of the model estimate of population density over time, with the estimate itself for reference.

437 The temporal distribution of variability in the estimate (Figure 9) allows us assess at which time steps the
 438 uncertainty is greater due to e.g. contradictions in the proxies. The coefficient of variation is 0.13 for the
 439 beginning and 0.1 for the end of the estimate, with the greatest variability (0.59) seen around 2150 BCE.
 440 This is not surprising as there are fewer archaeological contexts recorded from the earlier phase of the Early
 441 Bronze Age, c. 2200-1800 BCE. This picture changes from c. 1800 BCE onwards (David-Elbali, 2000; Hafner,
 442 1995). The beginning and end of the time series are relatively clearly determined, resulting from the *a priori*
 443 setting of final population density, but also from the uniformity of the proxies during these periods. Overall,
 444 the variability is relatively stable over the entire estimation and averages 35% of the respective mean.

445 The parameter p reflects the relative weight given to the individual proxies. Its posterior distribution (Figure
 446 10) shows that the model weights the openness indicator the highest, averaging slightly above 60%, followed
 447 by the RBSC, with an average of about 20%. The aoristic sum is slightly above 10%, whereas the importance

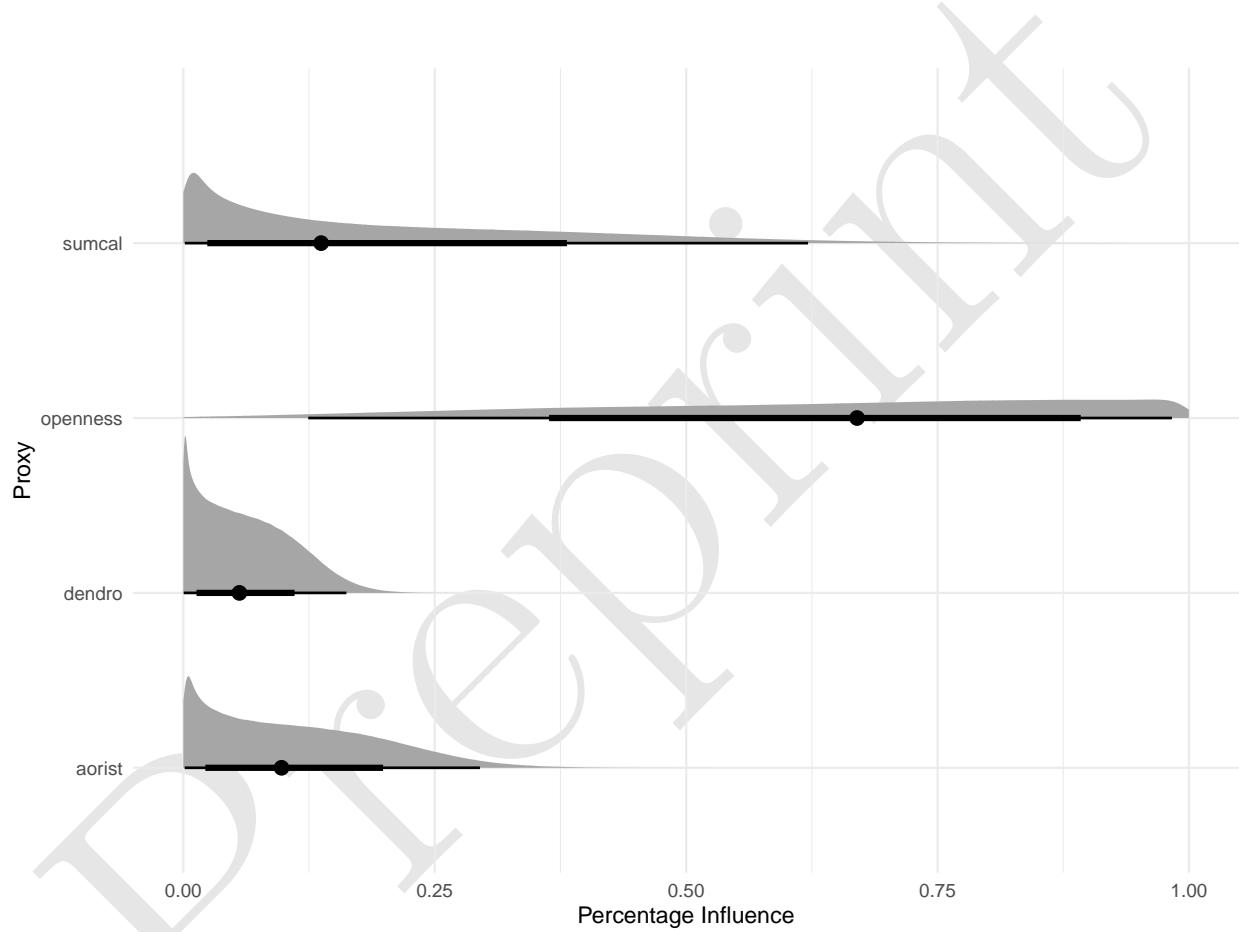


Figure 10: Distribution of influence ratios of proxies on model's final estimation of number of sites.

448 of the dendro-dated settlements is below 10%. The reason for the latter is certainly that there are no lakeshore
449 settlements over large areas of the time window, and therefore the overall confidence in the proxy is low.
450 The aoristic sum is flat for long periods, making it difficult to integrate with other proxies. The RBSC
451 shows very strong short-term fluctuations, at least partly due to the calibration curve, which suggests that
452 it does not reliably represent a continuous population trend. Its fluctuations have an impact on the model's
453 estimate, albeit to a lesser extent than the general trend.

454 5.2. Poisson Model

455 In the model, the parameter r was also estimated for the negative binomial model, which indicates the
456 convergence of the distribution to a Poisson distribution. If this value exceeds 50, then a Poisson model is
457 essentially given. The posterior distribution of this parameter looks as follows:

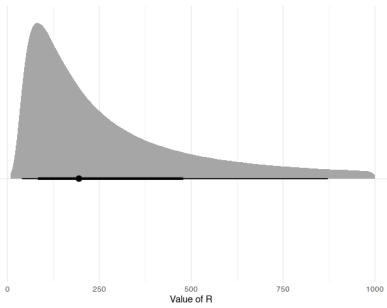


Figure 11: Posterior distribution of the rate parameter r .

458 Clearly, the central tendency of this distribution is above 50. The long tail is caused by the fact that above
459 a certain threshold the exact parameter r no longer makes a difference, since the distribution approaches a
460 Poisson distribution. We can therefore assume that the result essentially speaks for a support of a Poisson
461 distribution from the data. If we take this as a basis, then the mean values of the estimate are essentially
462 identical, but the credibility intervals are smaller (Fig. 12)

463 5.3. Sensitivity Analysis

464 The model has essentially two predefined parameters. These are the mean site size and the end value for
465 the population. The mean site size should not have a significant influence on the analysis, as it represents
466 a pure scaling factor. In contrast, the given start value defines the course of the reconstruction and should
467 therefore have a dominant influence on the model estimates.

468 In the sensitivity analysis (see supplementary material or GitHub repository), the model was run with 20
469 parameter settings each, with the one used in the model forming the middle of all cases.

470 The analyses carried out (Fig. 13 resp. 14) show a clear pattern and confirm the assumptions. The model
471 is not sensitive to mean site size, but is determined by final population density.

472 6. Discussion

473 6.1. Reliability of the overall estimation

474 We consider each of the proxies used to be intrinsically flawed and biased. Therefore, we assume that none of
475 them alone is suitable to generate a population estimate with sufficient credibility. Especially since the use

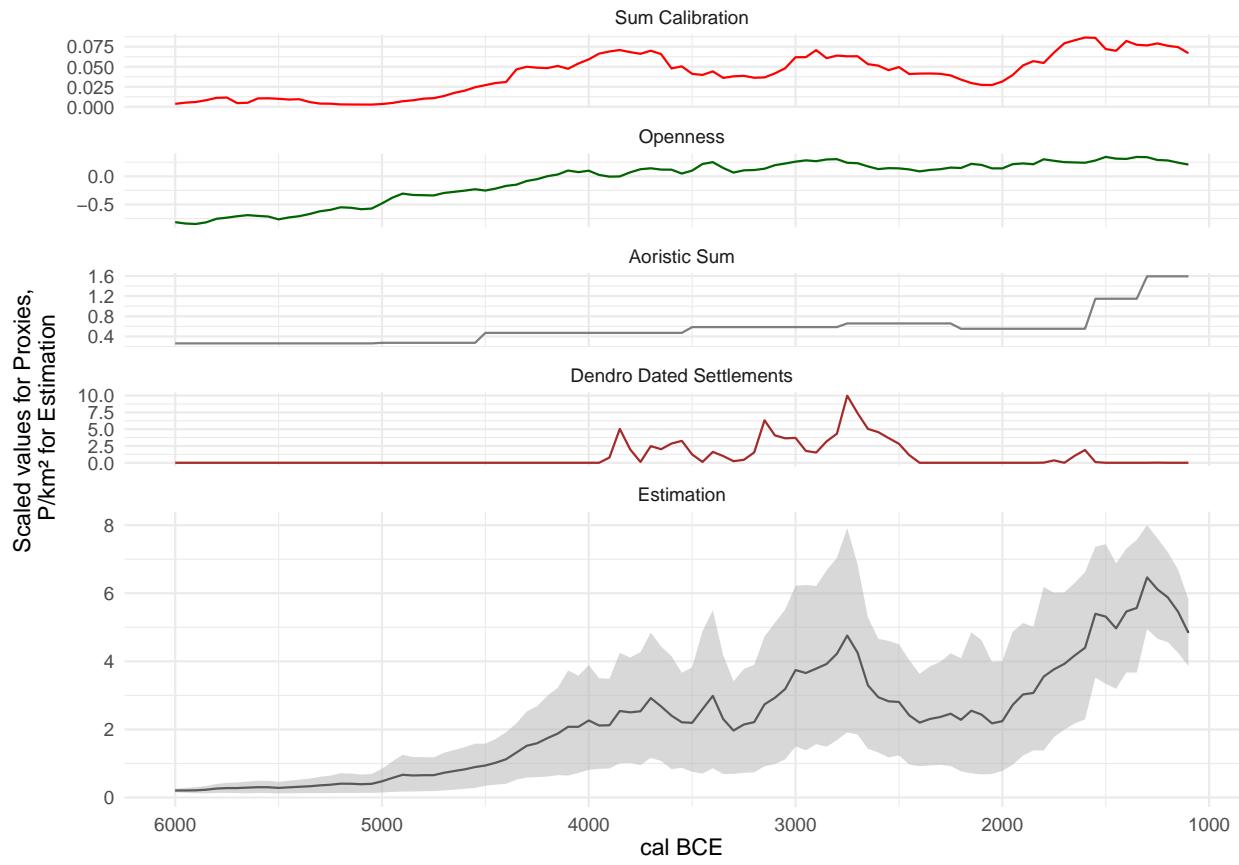


Figure 12: Estimate of population density predicted by the model, here with the Poisson distribution in the Process model. The four input proxies are also plotted (scaled) for comparison.

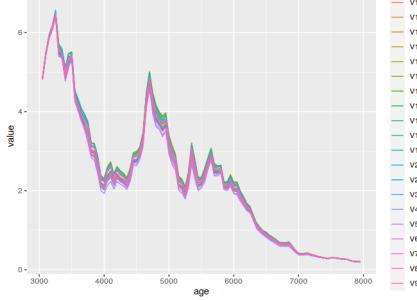


Figure 13: Estimate of population density predicted by the model in a sensibility analysis sweep over the parameter mean site size.

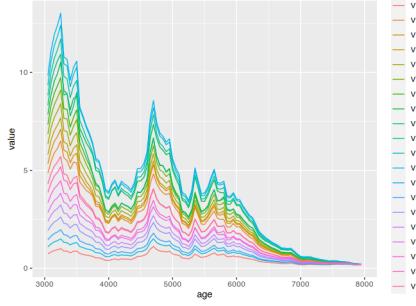


Figure 14: Estimate of population density predicted by the model in a sensibility analysis sweep over the parameter final population density.

476 of a single proxy makes it impossible to specify a confidence interval, as its value must be used as a direct
477 indication of a change in population size.

478 Rather, we assume that all proxies used are influenced, to a greater or lesser extent, by a proportion of
479 random noise as well as by one of the actual signal. As long as there is no ground truth that can be used
480 for calibration, this cannot be fully differentiated. What we assume, however, is that the noise between the
481 individual proxies is uncorrelated, but the signal must essentially produce a coherent pattern.

482 However, other reasons are also conceivable. For example, a correlation between two parameters can also
483 be caused by the fact that they are affected by the same confounding influence. All archaeologically derived
484 proxies are subject to preservation conditions. If no signs of human activity are preserved for a certain space
485 at a certain time based on the surviving archaeological evidence, then all these indicators will also show a
486 settlement gap. It is therefore all the more important to use independent proxies that are not affected by the
487 same distorting influences for a robust reconstruction. This is the reason why the openness indicator is an
488 essential part of this study, as it is not confounded by the archaeological record. Other similarly independent
489 indicators are to be integrated in future applications of our approach to make the result even more robust.

490 Moreover, our modelling approach in terms of an abductive reconstructive model is also such that in its
491 present form the model provides the best possible summary of the proxies used and thus the best possible
492 reconstruction given the available information. It must be emphasised that such a model does not attempt
493 to directly represent the process of the relationship between proxy and latent variable. We treat all proxies
494 as equal (false), and leave it to coherence with respect to the underlying process (the latent population
495 development) to provide a weighting. In our view, this corresponds to a (pragmatic) implementation of
496 the general scientific process. However, this also means that we prefer a useful model to a less useful one
497 without wanting to (or being able to) make a statement about the concept of truth of a model. Especially in
498 comparison to the uncritical application of a purely radiocarbon data-based proxy, as it is currently used in
499 large scale, we see the significance of such a model and the philosophy behind it as more than justified.

500 6.2. Reliability of individual proxies

501 Comparing the model's overall estimate with the individual proxies provides several insights into the quality
502 of these records. The sum calibration (here: RBSC), currently the most frequently used proxy for (relative)
503 population change in prehistory, has its large fluctuations damped when considered alongside other proxies.
504 This is especially true of the first fluctuation shortly after 4000 BCE. The expected increase in archaeological
505 remains with the onset of Neolithisation is still clearly visible, but the overall curve is much flatter than the
506 RBSC itself. The period between 3950 and 3700 BCE, contemporaneous with the first major settlement of
507 the Three Lakes regions' lakeshores, coincides with a noticeable plateau in the calibration curve, producing
508 an overestimation of the ^{14}C density. A second maximum, after 3000 BCE, is supported by the other
509 proxies, and is consequently much more reflected in the overall estimate, coinciding with a smaller and
510 shorter plateau. The rise towards the Middle and Late Bronze Age is also supported by the other proxies,

511 without a significant pattern in the calibration curve. We may conclude that the model is successful in using
512 information from other proxies to sift ‘real’ fluctuations in the summed radiocarbon record from artefacts of
513 the calibration curve.

514 On average, the model weights the RBSC at about 20%, significantly less than the 60% afforded to the
515 openness indicator. After an initial increase, which is easily explained by spread of agriculture, the openness
516 indicator tends to fluctuate less and thus has a dampening effect on the overall estimate. In general, this
517 trend in the RBSC is well reflected in land openness, while changes within the Neolithic and Bronze Age are
518 more gradual. However, it should be noted that temporal changes in land use strategies are not yet modelled
519 in the current implementation. This is clearly a potential for improvement for future model generations.

520 The aoristic sum remains flat over long spans of time. It is not until the Middle and Late Bronze Age that
521 we see a significant rise, which is also apparent in the model’s overall estimate. It remains to be seen to
522 what extent modelling of the taphonomic loss (Surovell et al., 2009) could be integrated in this approach.

523 The number of simultaneously existing lakeshore settlements is a temporally and spatially limited estimator,
524 but extremely reliable. Its limitations are reflected in the low overall confidence of the model, since its value
525 is zero over long stretches. However, where it has information potential, such as around and shortly after
526 3800 BCE, 3200 BCE, 1600 BCE or especially around 2750 BCE, its fluctuations have a noticeable influence
527 on the overall estimate. This highlights another potential of our approach: where a proxy has little structure
528 and thus little significance, or where its trends cannot be linked to other indicators, it consequently has little
529 influence. For periods in which it can provide information, however, this will also feed into the overall model,
530 despite a low overall confidence in the estimator.

531 6.3. Prehistoric population dynamics north of the Swiss Alps

532 In order to review the reconstruction against the background of established archaeological knowledge, it is
533 useful to overlay conventionally-defined archaeological phase boundaries (Hafner, 2005) on the results of our
534 model (Figure 15).

535 The Early and Middle Neolithic are hardly documented in Switzerland. We must assume a low level of
536 settlement, probably mainly by mobile groups. Isolated Neolithic sites of the LBK and later groups are known
537 in the periphery of Switzerland, but they play a subordinate role (Ebersbach et al., 2012). The evidence
538 of the Neolithic is dense from the so-called Upper Neolithic onwards, connected with the typochronological
539 pottery phases of Egolzwil (late 5th millennium BCE) and Cortaillod respectively Pfyn (first half of the 4th
540 millennium BCE). The first lake shore settlements north of the Alps date to this time too. Here we see a clear
541 increase in the estimated population in the model. In the transition to the Late Neolithic, we know from
542 the lakeshore settlements the so-called Horgen Gap (Hafner, 2005). This is also visible as a slight decrease
543 in the model. In another study (Heitz et al., 2021) we demonstrated that this is in fact probably not a
544 decline in population. Rather communities relocated their settlements to the hinterland of the large lakes in
545 times of stronger lake level rises due to climatic changes. In the Late Neolithic, associated with the Horgen
546 pottery, we then see a clear increase in the settlement intensity, which peaks and breaks off at the transition
547 to the Final Neolithic (Hafner, 2004). In the second half of the Early Bronze Age, during which lakeshores
548 were resettled to a smaller extend, there is again a clear increase in population size according to the model,
549 continuing until the Late Bronze Age. The general trends fit very well with the previous reconstructions of
550 population development for Switzerland (see eg. Lechterbeck et al., 2014), while offering higher precision
551 and higher resolution.

552 7. Conclusions

553 The model we present is a step towards a fully integrated hierarchical Bayesian model for estimating ab-
554 solute population sizes. With this model we can already combine different data (qualities) and evaluate
555 the reliability of the estimates at different points in time. Currently, absolute estimation is still done using
556 an arbitrary starting value. These estimates can be a basis for further studies where relative measures of

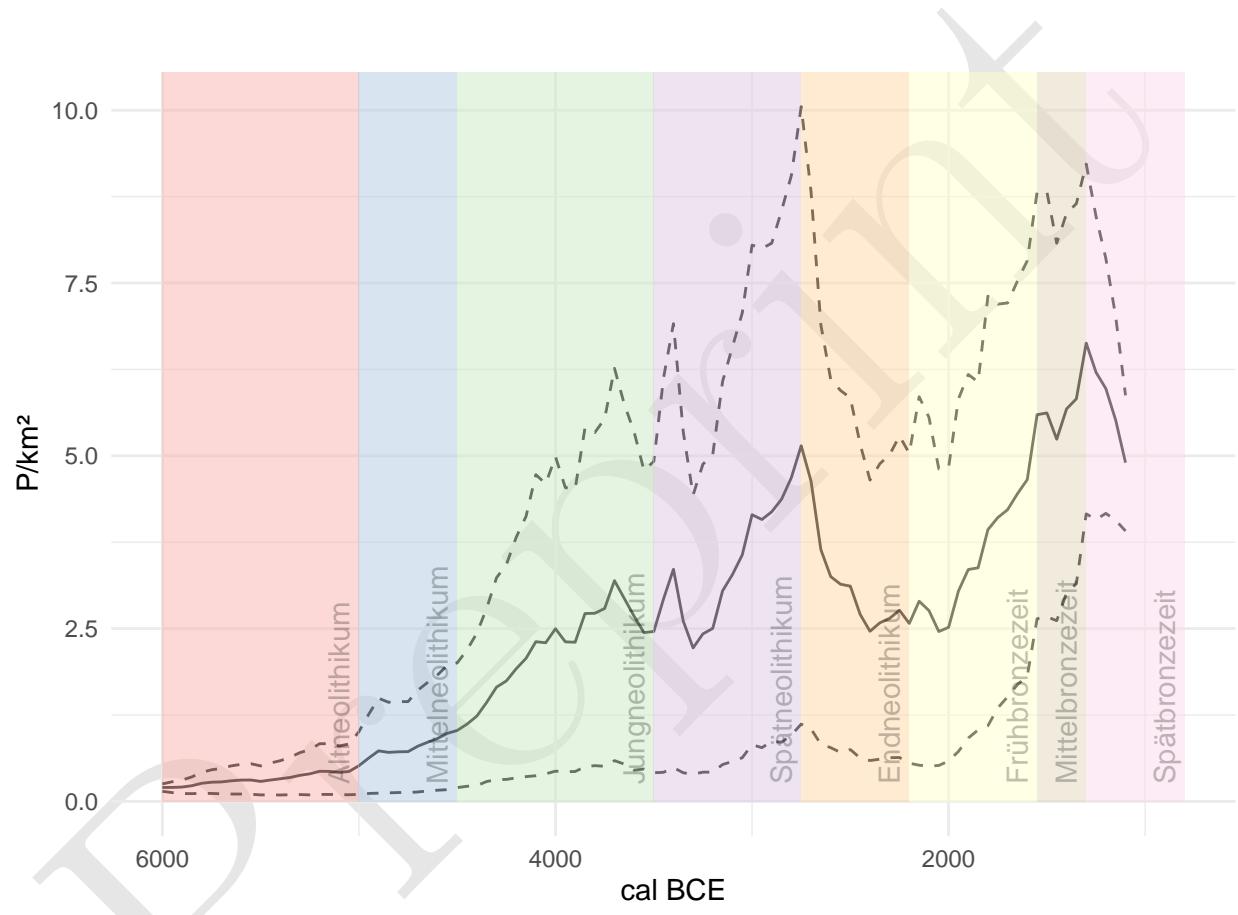


Figure 15: Estimate of population density in relation to the established chronology of the case study area north of the Swiss Alps.

557 population development are not helpful, such as long-term land use studies. Modelling of large-scale socio-
558 ecological systems based on archaeological data does not have to rely deductive, asynchronous population
559 models (e.g. carrying capacity or ethnographic analogues).

560 We have also demonstrated that, with Bayesian hierarchical modelling, it is possible to achieve a true
561 multi-proxy analysis – as opposed to a juxtaposition of different indicators. This opens up the possibility
562 of quantitatively linking different records and assessing their credibility. We are also able to specifying
563 a confidence interval for the overall estimate. The result is a firmer basis for reconstructing population
564 dynamics and settlement patterns in prehistory.

565 Nevertheless, we consider this model as only the first step towards a more sophisticated Bayesian approach.
566 We have trusted the individual proxies in aggregate, without individualised measurement error. Our esti-
567 mates are based on a limited number of sources, almost all of which are subject to taphonomic biases in the
568 archaeological record. Consequently, we can only transform the model's prediction into an absolute estimate
569 of population density with predefined parameters: settlement size and the initial value of the reconstruction.
570 Overcoming this limitation would represent a major refinement of our approach.

571 Incorporating additional proxies independent of the immediate, time-dependent conditions of the archaeo-
572 logical record could be one way to achieve this. These could be data on settlement sizes, parameters for
573 economic-ecological carrying capacity, demographic data from burial groups or archaeogenetic data on popu-
574 lation sizes. This data is available to varying degrees in different regions. On the Swiss Plateau, for example,
575 we have little data on human remains over large spans of prehistory, in contrast to the abundance of wetland
576 settlements.

577 Another direction of development is to reconsider and improve some decisions currently made for practical
578 reasons regarding the modelling approach. For example, one proposed alternative is to use the values of
579 the proxies directly instead of using them via normalisation and differencing as an indicator of the change
580 in a population from time $t-1$ to time t . This would require that explicit data models and observation
581 models for the individual proxies become part of the hierarchical model in order to show in detail their
582 dependence on the respective population numbers and to have their parameters estimated as probability
583 distributions in the model. Consequently, we would have to compensate for the resulting degrees of freedom
584 with additional information (more proxies, regionalisation), and we would have to cope with the increased
585 numerical complexity and the resulting higher computational effort. This can be done, for example, by using
586 a high performance computing environment, and is intended for the extension of the model.

587 To apply this approach to other regions, the proxies we use here would have to be adapted to fit local
588 conditions and research histories. By means of large-scale modelling, however, it would be possible to
589 supplement gaps in the data in one region with data from another by regionalisation and a partial transfer
590 of information (partial pooling). Such an extension would be the next logical step in the improvement of
591 the model, to which end we hope to be able to contribute a further study in the near future.

592 8. Acknowledgements

593 Data collection were conducted as part of the project 'Beyond lake settlements' in the doctoral thesis of Julian
594 Laabs, funded by the SNF (project number 152862, PI Albert Hafner) and as part of the XRONOS project,
595 also funded by the SNSF (project number 198153, PI Martin Hinz). The development of the openness index
596 took place within the framework of the project Time and Temporality in Archaeology (project number 194326,
597 PI Caroline Heitz), inspired by the cooperation within the project QuantHum (project 169371, PI Marco
598 Conedera), both also funded by the SNSF. Jan Kolář was supported by a long-term research development
599 project (RV67985939) and by a grant from the Czech Science Foundation (19-20970Y). Julian Laabs was
600 also funded by the German Research Foundation (DFG), Collaborative Research Centre 1266 'Scales of
601 Transformation' (project number 2901391021). We also thank the Institute of Archaeological Sciences of
602 the University of Bern for its support and faith in the outcome of our modelling project. Finally, we thank
603 Michael Holton Price and W. Christopher Carleton for their really thorough review, helpful comments and
604 valuable suggestions, which will certainly improve this manuscript significantly.

605 **9. Code availability**

606 The computer code used to generate the Bayesian Population model is provided in full in the Supplementary
607 Information, together with information about the program and version used. The R code and Data are avail-
608 able online at <https://github.com/MartinHinz/bayesian.demographic.reconstruction.2022> and is archived at
609 <https://doi.org/10.5281/zenodo.6594498>.

610 10. References

- 611 Attenbrow, V., Hiscock, P., 2015. Dates and demography: Are radiometric dates a robust proxy for long-term
612 prehistoric demographic change? *Archaeology in Oceania* 50, 30–36. <https://doi.org/10.1002/arco.5052>
- 613 Auger-Méthé, M., Newman, K., Cole, D., Empacher, F., Gryba, R., King, A.A., Leos-Barajas, V., Mills
614 Flemming, J., Nielsen, A., Petris, G., Thomas, L., 2021. A guide to state-space modeling of ecological
615 time series. *Ecological Monographs* 91, e01470. <https://doi.org/10.1002/ecm.1470>
- 616 Bocquet-Appel, J.-P., 2008. Explaining the Neolithic Demographic Transition, in: Bocquet-Appel, J.-P., Bar-
617 Yosef, O. (Eds.), *The Neolithic Demographic Transition and Its Consequences*. Springer Netherlands,
618 Dordrecht, pp. 35–55. https://doi.org/10.1007/978-1-4020-8539-0_3
- 619 Bryant, J., Zhang, J.L., 2018. *Bayesian Demographic Estimation and Forecasting*. Chapman; Hall/CRC.
620 <https://doi.org/10.1201/9780429452987>
- 621 Carleton, W.C., Groucutt, H.S., 2021. Sum things are not what they seem: Problems with point-wise
622 interpretations and quantitative analyses of proxies based on aggregated radiocarbon dates. *The Holocene*
623 31, 630–643. <https://doi.org/10.1177/0959683620981700>
- 624 Childe, V.G. (Ed.), 1936. *Man Makes Himself*. Watts & Co., London.
- 625 Christian Lüthi, 2009. Mittelland (region).
- 626 Crema, E.R., 2022. Statistical Inference of Prehistoric Demography from Frequency Distributions of Radio-
627 carbon Dates: A Review and a Guide for the Perplexed. *Journal of Archaeological Method and Theory*.
628 <https://doi.org/10.1007/s10816-022-09559-5>
- 629 Crema, E.R., Bevan, A., 2021. Inference from large sets of radiocarbon dates: software and methods.
630 Radiocarbon 63, 23–39. <https://doi.org/10.1017/RDC.2020.95>
- 631 Crema, E.R., Shoda, S., 2021. A Bayesian approach for fitting and comparing demographic growth models
632 of radiocarbon dates: A case study on the Jomon-Yayoi transition in Kyushu (Japan). *PloS One* 16,
633 e0251695. <https://doi.org/10.1371/journal.pone.0251695>
- 634 David-Elbiali, M., 2000. La Suisse occidentale au IIe millénaire av. J.-C. : Chronologie, culture, intégration
635 européenne. *Cahiers d'archéologie romande*, Lausanne.
- 636 Ebersbach, R., Kühn, M., Stopp, B., Schibler, J., 2012. Die Nutzung neuer Lebensräume in der Schweiz
637 und angrenzenden Gebieten im 5. Jtsd. V. Chr. : Siedlungs- und wirtschaftsarchäologische Aspekte.
638 *Jahrbuch Archäologie Schweiz* 95, 7–34. <https://doi.org/10.5169/SEALS-392483>
- 639 French, J.C., Riris, P., Fernández-López de Pablo, J., Lozano, S., Silva, F., 2021. A manifesto for palaeodem-
640 ography in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological
641 Sciences* 376, 20190707. <https://doi.org/10.1098/rstb.2019.0707>
- 642 Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models (comment on article
643 by Browne and Draper). *Bayesian Analysis* 1. <https://doi.org/10.1214/06-BA117A>
- 644 Hafner, A., 2005. Neolithikum: Raum/Zeit-Ordnung und neue Denkmodelle. *Archäologie im Kanton Bern*
645 6B, 431–498.
- 646 Hafner, A., 2004. Vom Spät- zum Endneolithikum. Wandel und Kontinuität um 2700 v. Chr. In Mit-
647 teleuropa. - Vom Endneolithikum zur Frühbronzezeit: Wandel und Kontinuität zwischen 2400 und 1500
648 v.Chr. in: Beier, H.J., Einicke, R. (Eds.), *Varia Neolithica*. 3. Beiträge Zur Ur- Und Frühgeschichte
649 Mitteleuropas. Beier&Beran, Weissbach, pp. 213–249.
- 650 Hafner, A., 1995. Die Frühe Bronzezeit in der Westschweiz. Funde und Befunde aus Siedlungen, Gräbern
651 und Horten der entwickelten Frühbronzezeit. *Staatlicher Lehrmittelverlag Bern* 1995, Bern.
- 652 Hassan, F.A., 1981. *Demographic archaeology, Studies in archaeology*. Academic Press, New York.
- 653 Heitz, C., Hinz, M., Laabs, J., Hafner, A., 2021. Mobility as resilience capacity in northern Alpine Neolithic
654 settlement communities. <https://doi.org/10.17863/CAM.79042>
- 655 Inkpen, R., Wilson, G.P., 2009. Explaining the past: Abductive and Bayesian reasoning. *The Holocene* 19,
656 329–334. <https://doi.org/10.1177/0959683608100577>
- 657 Kintigh, K.W., Altschul, J.H., Beaudry, M.C., Drennan, R.D., Kinzig, A.P., Kohler, T.A., Limp, W.F.,
658 Maschner, H.D.G., Michener, W.K., Pauketat, T.R., Peregrine, P., Sabloff, J.A., Wilkinson, T.J., Wright,
659 H.T., Zeder, M.A., 2014. Grand challenges for archaeology. *Proceedings of the National Academy of
660 Sciences* 111, 879–880. <https://doi.org/10.1073/pnas.1324000111>
- 661 Kohler, T.A., Reese, K.M., 2014. Long and spatially variable Neolithic Demographic Transition in the
662 North American Southwest. *Proceedings of the National Academy of Sciences* 111, 10101–10106. <https://doi.org/10.1073/pnas.1324000111>

- 663 //doi.org/10.1073/pnas.1404367111
- 664 Kruschke, J.K., 2015. Doing Bayesian Data Analysis. Elsevier. <https://doi.org/10.1016/B978-0-12-405888-0.09999-2>
- 665 Laabs, J., 2019. Populations- und landnutzungsmodellierung der neolithischen und bronzezeitlichen
666 westschweiz (PhD thesis). Bern.
- 667 Lechterbeck, J., Edinborough, K., Kerig, T., Fyfe, R., Roberts, N., Shennan, S., 2014. Is Neolithic land
668 use correlated with demography? An evaluation of pollen-derived land cover and radiocarbon-inferred
669 demographic change from Central Europe. *The Holocene* 24, 1297–1307. <https://doi.org/10.1177/0959683614540952>
- 670 Martínez-Grau, H., Morell-Rovira, B., Antolín, F., 2021. Radiocarbon Dates Associated to Neolithic Con-
671 texts (Ca. 5900 – 2000 Cal BC) from the Northwestern Mediterranean Arch to the High Rhine Area.
672 *Journal of Open Archaeology Data* 9, 1. <https://doi.org/10.5334/joad.72>
- 673 Morris, I., 2013. The measure of civilization: How social development decides the fate of nations. Princeton
674 University Press, Princeton.
- 675 Müller, J., Diachenko, A., 2019. Tracing long-term demographic changes: The issue of spatial scales. *PLOS
676 ONE* 14, e0208739. <https://doi.org/10.1371/journal.pone.0208739>
- 677 Nikulka, F., 2016. Archäologische Demographie. Methoden, Daten und Bevölkerung der europäischen
678 Bronze- und Eisenzeiten. Sidestone Press, Leiden.
- 679 Price, M.H., Capriles, J.M., Hoggarth, J.A., Bocinsky, R.K., Ebert, C.E., Jones, J.H., 2021. End-to-end
680 Bayesian analysis for summarizing sets of radiocarbon dates. *Journal of Archaeological Science* 135,
681 105473. <https://doi.org/10.1016/j.jas.2021.105473>
- 682 Ramsey, C.B., 1995. Radiocarbon calibration and analysis of stratigraphy; the OxCal program. *Radiocarbon*
683 37, 425430.
- 684 Rick, J.W., 1987. Dates as Data: An Examination of the Peruvian Preceramic Radiocarbon Record. *American
685 Antiquity* 52, 55–73. <https://doi.org/10.2307/281060>
- 686 Riede, F., 2009. Climate and demography in early prehistory: Using calibrated (14)C dates as population
687 proxies. *Human Biology* 81, 309–337. <https://doi.org/10.3378/027.081.0311>
- 688 Schmidt, I., Hilpert, J., Kretschmer, I., Peters, R., Broich, M., Schiesberg, S., Vogels, O., Wendt, K.P.,
689 Zimmermann, A., Maier, A., 2021. Approaching prehistoric demography: Proxies, scales and scope of
690 the Cologne Protocol in European contexts. *Philosophical Transactions of the Royal Society B: Biological
691 Sciences* 376, 20190714. <https://doi.org/10.1098/rstb.2019.0714>
- 692 Shennan, S., 2000. Population, Culture History, and the Dynamics of Culture Change. *Current Anthropology*
693 41, 811–835.
- 694 Shennan, S., Downey, S.S., Timpton, A., Edinborough, K., Colledge, S., Kerig, T., Manning, K., Thomas,
695 M.G., 2013. Regional population collapse followed initial agriculture booms in mid-Holocene Europe.
696 *Nature Communications* 4, 2486. <https://doi.org/10.1038/ncomms3486>
- 697 Surovell, T.A., Byrd Finley, J., Smith, G.M., Brantingham, P.J., Kelly, R., 2009. Correcting temporal
698 frequency distributions for taphonomic bias. *Journal of Archaeological Science* 36, 1715–1724. <https://doi.org/10.1016/j.jas.2009.03.029>
- 699 Turchin, P., Brennan, R., Currie, T., Feeney, K., Francois, P., Hoyer, D., Manning, J., Marciniak, A.,
700 Mullins, D., Palmisano, A., Peregrine, P., Turner, E.A.L., Whitehouse, H., 2015. Seshat: The Global
701 History Databank. *Cliodynamics* 6. <https://doi.org/10.21237/C7clio6127917>
- 702 Wolpert, D.H., Price, M.H., Crabtree, S.A., Kohler, T.A., Jost, J., Evans, J., Stadler, P.F., Shimao, H.,
703 Laubichler, M.D., 2021. The Past as a Stochastic Process. <https://doi.org/10.48550/arXiv.2112.05876>
- 704 Zimmermann, A., 2004. Landschaftsarchäologie II. Überlegungen zu prinzipien einer landschaftsarchäologie.

708 11. Author contributions

- 709 • *Martin Hinz*: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation,
710 Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization
711 • *Joe Roe*: Software, Validation, Writing - Review & Editing
712 • *Julian Laabs*: Investigation, Data Curation, Writing - Review & Editing
713 • *Caroline Heitz*: Conceptualization, Investigation, Writing - Review & Editing
714 • *Jan Kolář*: Conceptualization, Writing - Review & Editing

715 12. Colophon

716 This report was generated on 2022-12-08 19:06:28 using the following computational environment and de-
717 pendencies:

```
718 #> - Session info -----
719 #>   setting  value
720 #>   version R version 4.2.1 (2022-06-23)
721 #>   os        macOS Big Sur ... 10.16
722 #>   system    x86_64, darwin17.0
723 #>   ui        X11
724 #>   language (EN)
725 #>   collate  en_US.UTF-8
726 #>   ctype    en_US.UTF-8
727 #>   tz       Europe/Berlin
728 #>   date     2022-12-08
729 #>   pandoc   2.19.2 @ /Applications/RStudio.app/Contents/MacOS/quarto/bin/tools/ (via rmarkdown)
730 #>
731 #> - Packages -----
732 #>   package      * version date (UTC) lib source
733 #>   assertthat    0.2.1   2019-03-21 [1] CRAN (R 4.2.0)
734 #>   bitops        1.0-7   2021-04-24 [1] CRAN (R 4.2.0)
735 #>   bookdown      0.30    2022-11-09 [1] CRAN (R 4.2.0)
736 #>   cachem        1.0.6   2021-08-19 [1] CRAN (R 4.2.0)
737 #>   callr         3.7.2   2022-08-22 [1] CRAN (R 4.2.0)
738 #>   class         7.3-20  2022-01-16 [1] CRAN (R 4.2.1)
739 #>   classInt      0.4-8   2022-09-29 [1] CRAN (R 4.2.0)
740 #>   cli            3.4.1   2022-09-23 [1] CRAN (R 4.2.0)
741 #>   colorspace    2.0-3   2022-02-21 [1] CRAN (R 4.2.0)
742 #>   cowplot        * 1.1.1  2020-12-30 [1] CRAN (R 4.2.0)
743 #>   crayon        1.5.2   2022-09-29 [1] CRAN (R 4.2.0)
744 #>   curl           4.3.3   2022-10-06 [1] CRAN (R 4.2.0)
745 #>   DBI            1.1.3   2022-06-18 [1] CRAN (R 4.2.0)
746 #>   devtools       2.4.5   2022-10-11 [1] CRAN (R 4.2.0)
747 #>   digest          0.6.29  2021-12-01 [1] CRAN (R 4.2.0)
748 #>   dplyr          1.0.10  2022-09-01 [1] CRAN (R 4.2.0)
749 #>   e1071          1.7-11  2022-06-07 [1] CRAN (R 4.2.0)
750 #>   ellipsis        0.3.2   2021-04-29 [1] CRAN (R 4.2.0)
751 #>   evaluate        0.17    2022-10-07 [1] CRAN (R 4.2.0)
752 #>   fansi           1.0.3   2022-03-24 [1] CRAN (R 4.2.0)
753 #>   farver          2.1.1   2022-07-06 [1] CRAN (R 4.2.0)
754 #>   fastmap         1.1.0   2021-01-25 [1] CRAN (R 4.2.0)
755 #>   foreign         0.8-82  2022-01-16 [1] CRAN (R 4.2.1)
```

```

756 #> fs           1.5.2   2021-12-08 [1] CRAN (R 4.2.0)
757 #> generics     0.1.3   2022-07-05 [1] CRAN (R 4.2.0)
758 #> ggmap        * 3.0.0   2019-02-05 [1] CRAN (R 4.2.0)
759 #> ggplot2      * 3.3.6   2022-05-03 [1] CRAN (R 4.2.0)
760 #> ggrepel       * 0.9.1   2021-01-15 [1] CRAN (R 4.2.0)
761 #> ggsn         * 0.5.0   2019-02-18 [1] CRAN (R 4.2.0)
762 #> glue          1.6.2   2022-02-24 [1] CRAN (R 4.2.0)
763 #> gtable        0.3.1   2022-09-01 [1] CRAN (R 4.2.0)
764 #> here          * 1.0.1   2020-12-13 [1] CRAN (R 4.2.0)
765 #> highr         0.9     2021-04-16 [1] CRAN (R 4.2.0)
766 #> htmltools     0.5.3   2022-07-18 [1] CRAN (R 4.2.0)
767 #> htmlwidgets   1.5.4   2021-09-08 [1] CRAN (R 4.2.0)
768 #> httpuv        1.6.6   2022-09-08 [1] CRAN (R 4.2.0)
769 #> httr          1.4.4   2022-08-17 [1] CRAN (R 4.2.0)
770 #> jpeg          0.1-9   2021-07-24 [1] CRAN (R 4.2.0)
771 #> KernSmooth    2.23-20  2021-05-03 [1] CRAN (R 4.2.1)
772 #> knitr         1.40    2022-08-24 [1] CRAN (R 4.2.0)
773 #> labeling       0.4.2   2020-10-20 [1] CRAN (R 4.2.0)
774 #> later          1.3.0   2021-08-18 [1] CRAN (R 4.2.0)
775 #> lattice        0.20-45  2021-09-22 [1] CRAN (R 4.2.1)
776 #> lifecycle      1.0.3   2022-10-07 [1] CRAN (R 4.2.0)
777 #> magrittr       2.0.3   2022-03-30 [1] CRAN (R 4.2.0)
778 #> maptools       1.1-5   2022-10-21 [1] CRAN (R 4.2.0)
779 #> memoise        2.0.1   2021-11-26 [1] CRAN (R 4.2.0)
780 #> mime            0.12    2021-09-28 [1] CRAN (R 4.2.0)
781 #> miniUI         0.1.1.1  2018-05-18 [1] CRAN (R 4.2.0)
782 #> munsell        0.5.0   2018-06-12 [1] CRAN (R 4.2.0)
783 #> pillar          1.8.1   2022-08-19 [1] CRAN (R 4.2.0)
784 #> pkgbuild       1.3.1   2021-12-20 [1] CRAN (R 4.2.0)
785 #> pkgconfig      2.0.3   2019-09-22 [1] CRAN (R 4.2.0)
786 #> pkgload         1.3.0   2022-06-27 [1] CRAN (R 4.2.0)
787 #> plyr            1.8.7   2022-03-24 [1] CRAN (R 4.2.0)
788 #> png              0.1-7   2013-12-03 [1] CRAN (R 4.2.0)
789 #> prettyunits    1.1.1   2020-01-24 [1] CRAN (R 4.2.0)
790 #> processx       3.7.0   2022-07-07 [1] CRAN (R 4.2.0)
791 #> profvis        0.3.7   2020-11-02 [1] CRAN (R 4.2.0)
792 #> promises        1.2.0.1  2021-02-11 [1] CRAN (R 4.2.0)
793 #> proxy           0.4-27  2022-06-09 [1] CRAN (R 4.2.0)
794 #> ps               1.7.1   2022-06-18 [1] CRAN (R 4.2.0)
795 #> purrr           0.3.5   2022-10-06 [1] CRAN (R 4.2.0)
796 #> R6               2.5.1   2021-08-19 [1] CRAN (R 4.2.0)
797 #> RColorBrewer   1.1-3   2022-04-03 [1] CRAN (R 4.2.0)
798 #> Rcpp             1.0.9   2022-07-08 [1] CRAN (R 4.2.0)
799 #> remotes          2.4.2   2021-11-30 [1] CRAN (R 4.2.0)
800 #> rgdal           1.6-2   2022-11-09 [1] CRAN (R 4.2.0)
801 #> RgoogleMaps    1.4.5.3  2020-02-12 [1] CRAN (R 4.2.0)
802 #> rjson            0.2.21  2022-01-09 [1] CRAN (R 4.2.0)
803 #> rlang            1.0.6   2022-09-24 [1] CRAN (R 4.2.0)
804 #> rmarkdown        2.17    2022-10-07 [1] CRAN (R 4.2.0)
805 #> rnaturalearth  * 0.1.0   2017-03-21 [1] CRAN (R 4.2.0)
806 #> rprojroot        2.0.3   2022-04-02 [1] CRAN (R 4.2.0)
807 #> rstudioapi       0.14    2022-08-22 [1] CRAN (R 4.2.0)
808 #> s2               1.1.0   2022-07-18 [1] CRAN (R 4.2.0)
809 #> scales           1.2.1   2022-08-20 [1] CRAN (R 4.2.0)

```

```
810 #> sessioninfo      1.2.2  2021-12-06 [1] CRAN (R 4.2.0)
811 #> sf                * 1.0-8  2022-07-14 [1] CRAN (R 4.2.0)
812 #> shiny              1.7.3  2022-10-25 [1] CRAN (R 4.2.0)
813 #> sp                * 1.5-0  2022-06-05 [1] CRAN (R 4.2.0)
814 #> stringi             1.7.8  2022-07-11 [1] CRAN (R 4.2.0)
815 #> stringr             1.4.1  2022-08-20 [1] CRAN (R 4.2.0)
816 #> tibble              3.1.8  2022-07-22 [1] CRAN (R 4.2.0)
817 #> tidyverse            1.2.1  2022-09-08 [1] CRAN (R 4.2.0)
818 #> tidyselect           1.2.0  2022-10-10 [1] CRAN (R 4.2.0)
819 #> units               0.8-0  2022-02-05 [1] CRAN (R 4.2.0)
820 #> urlchecker          1.0.1  2021-11-30 [1] CRAN (R 4.2.0)
821 #> usethis              2.1.6  2022-05-25 [1] CRAN (R 4.2.0)
822 #> utf8                1.2.2  2021-07-24 [1] CRAN (R 4.2.0)
823 #> vctrs               0.4.2  2022-09-29 [1] CRAN (R 4.2.0)
824 #> withr               2.5.0  2022-03-03 [1] CRAN (R 4.2.0)
825 #> wk                  0.7.0  2022-10-13 [1] CRAN (R 4.2.0)
826 #> xfun                 0.33   2022-09-12 [1] CRAN (R 4.2.0)
827 #> xtable              1.8-4  2019-04-21 [1] CRAN (R 4.2.0)
828 #> yaml                 2.3.5  2022-02-21 [1] CRAN (R 4.2.0)
829 #>
830 #> [1] /Library/Frameworks/R.framework/Versions/4.2/Resources/library
831 #>
832 #> -----
```