Reconstructing the biogeography of a hunter-gatherer planet using machine-learning

Marcus J. Hamilton^{a,1}, Robert S. Walker^b, Briggs Buchanan^c, Damian E. Blasi^{d,e,f}, and Claire L. Bown^f

^aDepartment of Anthropology, University of Texas at San Antonio, San Antonio, Texas; ^bDepartment of Anthropology, University of Missouri, Columbia, MO; ^cDepartment of Anthropology, University of Tulsa, Tulsa, OK; ^dDepartment of Human Evolutionary Biology, Harvard University, Boston, MA; ^eMax Planck Institute for Evolutionary Anthropology, Leipzig, Germany; ^fDepartment of Linguistics, Yale University, New Haven, CT

This manuscript was compiled on May 12, 2021

1 **Estimating the total human population size (i.e., abundance) of the**
2 **preagricultural planet is important for setting the baseline expecta-**
3 **tions for human-environment interactions if all energy and material**
4 **requirements to support growth, maintenance, and well-being were**
5 **foraged from local environments. However, demographic paramet-**
6 **ters and biogeographic distributions do not preserve directly in the**
7 **archaeological record. Rather than attempting to estimate human**
8 **abundance at some specific time in the past, a principled approach**
9 **to making inferences at this scale is to ask what the human demog-**
10 **raphy and biogeography of a hypothetical planet Earth would look like**
11 **if populated by ethnographic hunter-gatherer societies. Given ethn-**
12 **ographic hunter-gatherer societies likely include the largest, dens-**
13 **est, and most complex foraging societies to have existed, we sug-**
14 **gest population inferences drawn from this sample provide an upper**
15 **bound to demographic estimates in prehistory. Our goal in this paper**
16 **is to produce principled estimates of hunter-gatherer abundance, di-**
17 **versity, and biogeography. To do this we trained an extreme gradient**
18 **boosting algorithm (XGBoost) to learn ethnographic hunter-gatherer**
19 **population densities from a large matrix of climatic, environmental,**
20 **and geographic data. We used the predictions generated by this**
21 **model to reconstruct the hunter-gatherer biogeography of the rest of**
22 **the planet. We find the human abundance of this world to be 6.1 ± 2**
23 **million with an ethnolinguistic diversity of $8,330 \pm 2,770$ populations,**
24 **most of whom would have lived near coasts and in the tropics.**

Extreme gradient boosting | Population density | Demography | Net
primary production | Biogeography

1 **I**n 2021, the global abundance of the human species is almost
2 8 billion. In 1921 it was about 2 billion, in 1821 1 billion, in
3 1621 about 500 million, and in 21 AD about 200 million (1).
4 Prior to the emergence of agriculture in the Fertile Crescent
5 about 10,000 BC, global abundance was likely no more than
6 a few million (1). If human population growth saturates at
7 11 billion around 2100 AD as predicted by the UN (2, 3)
8 human abundance will have increased 5000-fold in 13,000
9 years (Figure 1). However, reliable demographic statistics are
10 available only for the last couple of centuries (4). Even with
11 government funding and resources, national-scale population
12 estimates in the 21st century cost billions of dollars and years
13 of preparation (5). Reconstructing global scale phenomena
14 in the deeper past is even more challenging as demographic
15 parameters do not preserve in the archaeological or geological
16 records (6). Moreover, the human species has a dynamic
17 evolutionary history of adaptive radiation, range expansion,
18 founder effects, bottlenecks, fluctuating climates, and cultural
19 evolution building complex spatiotemporal dependencies into
20 any prehistoric demographic estimates.

21 A reasonable zeroth-order anthropological model of global
22 hunter-gatherer abundance is simply to multiply the average

Fig. 1. The total abundance of mammal species (blue data) and humans over time (red data) by body size on logged axes. The orange line is an OLS regression model fit to the mammal data, and the equation is top left. The dashed line is the United Nations prediction of 11 billion by AD 2100 and the solid line is the 6.1 ± 2 million estimate we develop here. Human abundance will have increased 5,000-fold over the Holocene.

23 population density of recently-recorded ethnographic hunter-gatherers, $\bar{d}_h = 0.1 \text{ km}^{-2}$, by the available terrestrial landmass
24 of the planet, $A_h = 149 \text{ million km}^2$, in which case global abundance would be $N_h = \bar{d}_h A_h = 14.9 \text{ million}$. Similarly,
25 a reasonable ecological estimate comes from the allometric
26 scaling of mammal body size, M_m , and species abundance, N_m
27 (Figure 1), where $N_m = c_m M_m^{-\beta}$, and so using the parameters
28 from the fitted model in Figure 1, the predicted total abun-
29
30

Significance Statement

Understanding the abundance of humans on planet Earth prior to the development of agriculture and the industrialized world is essential to understanding human population growth. However, the problem is that these features of human populations in the past are unknown and so must be estimated from data. We developed a machine learning approach that uses ethnographic and environmental data to reconstruct the demography and biogeography of planet Earth if populated by hunter-gatherers. Such a world would house about 6 million people divided into about 8,330 populations with a particular concentration in the tropics and along coasts.

Please provide details of author contributions here.

Please declare any competing interests here.

¹To whom correspondence should be addressed. E-mail: marcus.hamilton@utsa.com

Fig. 2. The global distribution of the reconstructed hunter-gatherer population densities (A), net primary production (B), and the biogeographic distribution of marine mammal diversity (C). Hunter-gatherer population densities tend to be highest in the tropics and in other regions of high net primary production, but also along many coastlines, particularly those with high marine mammal diversity. Outside the tropics and away from coasts population densities are much lower.

dance of a 60 kg mammal is 2.1 million. This range of 2 - 15 million bounds the range of estimates of human preagricultural abundance commonly found in the anthropological literature (1, 7–9). However, these estimates are often quite general in nature and rarely address the complex interactions that constrain hunter-gatherer biogeography and spatial ecology explicitly.

Given the complexity of human evolutionary history, a principled approach to developing more rigorous estimates of demography in the deep past is to consider a hypothetical world populated by ethnographic societies to serve as a null model. The spatial ecology of ethnographic hunter-gatherers at a coarse-grained scale is now well understood empirically and theoretically (8, 10–12). Because all energy requirements for maintenance, growth, and motility are derived from local ecosystems, environmental productivity constrains the diversity of hunter-gatherer adaptations in predictable ways (13, 14). For example, global variation in hunter-gatherer landscape use (11, 15), population density (11), and mobility (16–19) are all predicted by variation in environmental productivity. In this paper we use these insights to develop a machine-learning approach to predict the abundance, diversity, and biogeography of a hypothetical hunter-gatherer planet. In specific, we use extreme gradient boosting (XGBoost) (20), an advanced regression technique that uses a gradient boosting

decision-tree modeling framework to build complex ensemble models of independent data to predict a dependent variable (see **Supplementary Information** for details).

Model development

Our machine-learning approach begins by modeling hunter-gatherer population density as a function of a large matrix of biotic and abiotic environmental data comprised of 46 predictor variables. These data include climatic, biological, and geographic variables designed to capture environmental complexity. Given the results of this model, we then project predicted hunter-gatherer population densities across the planet based on the same matrix of environmental data. From these projections we establish estimates of the global abundance and diversity of hunter-gatherer populations. We then extract information from this reconstructed planet to examine the biogeographic distribution of population densities in terms of both the abiotic and biotic environments. For a detailed description of the data and modeling approach see the **Supplementary Information** attached to this paper.

Modeling ethnographic data. A key question in comparative anthropology is how best to use ethnographic data to reconstruct scales of demography beyond those directly observed in contemporary populations (21, 22). The only direct source of information about hunter-gatherer lifestyles comes from observations of ethnographic societies (10), all of whom have been impacted by colonialization and the expansion of the industrialized world. Simple ethnographic analogies that equate case studies in the past with case studies in the present are problematic as analogies are not homologies, and so lack mechanistic equivalence. A more principled approach is statistical: Ethnographic data can be used to make inferences about parameters of interest and to measure the errors with which those inferences are made.

Although the ethnographic record of the Anthropocene is a non-random sample, most sources of bias are likely to inflate the range of variation we observe in the ethnographic present versus the past. For example, while the spread of diseases and environmental marginalization catastrophically reduced population numbers over the colonial era, accumulating technological innovations, the warm stable climates of the Holocene, and widespread interactions with non-foraging peoples likely increased populations over the long term. Thus ecological parameters in the present are likely *less* predictable solely from environmental data than in the past because ethnographically there are additional sources of variation.

Consider population density for example. Population size per unit area is correlated with environmental productivity as the energy density of landscapes fundamentally constrains the amount of biomass that can be supported per unit area (11). However, all populations naturally cycle through periods of growth, stability and collapse and so cannot be assumed to be in equilibrium (23, 24). Moreover, in many ethnographic hunter-gatherer societies energy budgets are supplemented using technologies not available in the deeper past including metals, firearms, and combustion engines, and lifestyles incorporating horticulture, agriculture, or market economies. However, there is no *a priori* reason to believe that these interactions break the fundamental ecological symmetry between space and energy, especially as the predicted correlations are

Fig. 3. The marginal importance - SHAP values - of the predictors of hunter-gatherer population densities in the XGBoost model. Positive values to the right of center have positive marginal effects on population density and purple reflects higher impact. Negative values to the left have negative effects and yellow reflect lower impact. For example, high mean annual temperatures (ranked 4th) have a strongly positive impact on population densities and low values have a weakly negative impact. Less temperature seasonality has a high impact (ranked 11th) on population density.

observed in data (8, 11). As such, uniformitarian assumptions of ecological and demographic equivalence in the past and present can be considered uniformitarian principles (25, 26).

Results

XGBoost model of population density. The XGBoost model of hunter-gatherer population density returns a cross-validated $r^2 = 0.96$ and so variation in hunter-gatherer population densities are highly predictable from environmental, climatic, and geographic variables. Figure 3 shows the top 29 predictors of most overall importance where SHAP values (Shapley additive explanations) are the marginal impact of each predictor variable on the model conditional on all other possible interactions. Three of the top four predictors show are minimum temperature constraints during the coldest and driest parts of the year indicating that annual climatic extremes play important roles. The second highest predictor is bird diversity, which likely correlates with broader environmental variation. The fifth variable is proximity to rivers, and the sixth variable is pathogen load, consistent with Tallavaara et al. (8). This is followed by several measures of landscape complexity, such as proximity to coastlines, rivers, marine resources, and water sources. Total annual precipitation is ranked 10th in overall importance in the model. Interestingly, terrestrial mammal diversity is ranked 22nd on the list of importance, likely because much of the predictive power is captured by higher-ranked environmental predictor variables. A strength of the XGBoost model over more traditional linear models is its ability to capture the complex nonlinear responses of predictor variables and their interactions (Figure 4).

Figure 5 shows the probability distribution of modeled population densities covers the same range as the observed ethnographic sample, but the distributions change shape. The average population density of the hypothetical world is $0.03 \pm 0.01 \text{ km}^{-2}$ (Figure 5A) whereas the average population density

Fig. 4. The range of nonlinear dependencies of the marginal impact of six predictor variables on hunter-gatherer population densities captured by the XGBoost model, color-coded by pathogen load. In each panel, the y-axis is the marginal dependency of the x-axis on population density, color-coded by pathogen load (yellow to purple). A-C) Temperature and precipitation variables have negative impacts on population density at low values, and positive at high, but the responses are highly nonlinear reflecting abrupt transitions in the responses. D) Increasingly seasonal temperatures have negative impacts on population density. E) The marginal effect of distance to coast is positive at low distances and negative beyond 300 km. However, at high distances (i.e., far inland) the effect becomes positive again, probably due to the proximity to other major bodies of water. F) The marginal effect of distance to tertiary branches of rivers decreases with distance and is negative after only a few kilometers.

of the ethnographic sample is $0.11 \pm 0.07 \text{ km}^{-2}$. The highest population densities observed ethnographically are much rarer in the reconstructed world and so the mean population density is considerably less. Thus the XGBoost model is capturing - and correcting - the over-sampling of high density ethnographic societies in the training data, such as those of the Pacific Rim of North America, and mutualists of tropical regions of Southern Asia and Central Africa, such as the Chenchu of India, the Agta of the Philippines, and the Bayaka of the Congo.

Reconstructing the hunter-gatherer planet. Figure 2A shows the biogeographic distribution of predicted hunter-gatherer population densities in the reconstructed world, Figure 2B shows the global distribution of above-ground net primary production, and Figure 2C shows the global distribution of marine mammal diversity. The spatial distribution of population densities correlates with environmental productivity, with particularly high densities in the tropics ($\pm 26^\circ \text{ N/S}$ of the equator), and locally in other highly productive environments, including the southern Japanese archipelago, the east coast of Australia, and the temperate west coast of South America. Population densities are also high along coasts including the North American Pacific Rim, central America, much of coastal South America, Atlantic Europe, the Mediterranean Basin, the northern Persian Gulf and the south coast of the Caspian, coastal sub-Saharan Africa, southeast Asia, and much of coastal Australia. Population densities are often particularly high along the most diverse coasts.

Abundance and diversity estimates. To estimate the expected abundance and diversity of a hunter-gatherer planet we ran 10,000 Monte Carlo simulations of the XGBoost model, sampling with replacement from the ethnographic data. Each

Fig. 5. Population density, abundance, and diversity distributions from the XGBoost model. A) Modeled versus ethnographic population densities, where the vertical lines are the averages. B) The distribution of 10,000 Monte Carlo simulations of population abundance, with a mean of 6.1 ± 2 million. C) The distribution of 10,000 Monte Carlo estimates of ethnolinguistic population richness, with a mean of about $8,330 \pm 2,770$.

Table 1. The distribution and errors ($x \pm 1.96\sigma$) of reconstructed human population abundance, N ($\bar{d} = N/A$), ethnolinguistic diversity, S ($\bar{s} = S/A$), and landmass, A , across continents

Continent	N ($\bar{d}$)*	S ($\bar{s}$ †)	A , km ² ‡
Africa	1.59 ± 0.52 (0.07)	$2,170 \pm 720$ (70)	31.5
Australia	0.27 ± 0.09 (0.05)	370 ± 130 (44)	8.5
Eurasia	1.93 ± 0.63 (0.05)	$2,640 \pm 880$ (45)	60.2
North America	0.59 ± 0.19 (0.04)	800 ± 260 (34)	23.4
South America	1.6 ± 0.53 (0.11)	$2,200 \pm 730$ (105)	21.3
Pacific	0.11 ± 0.04 (0.04)	140 ± 50 (41)	3.6
Total	6.10 ± 2.0 (0.06)	$8,330 \pm 2,770$ (60)	149.0

*millions of people, † diversity per millions of km², ‡ millions of km²

simulation produced an abundance estimate by summing over the global population density map it produced, creating the distributions in Figure 5B and 5C. The average total human population abundance was 6.1 (4.1–8.1) million. The confidence limits around this prediction, ± 2 million, reflect the variance of the original ethnographic sample. Total abundances over 10 million were rare (9/10000 runs) and no single run exceeded 11.5 million. Similarly, total abundances of less than 3 million were rare (8/10000 runs) and the minimum abundance from a single run was 2.4 million.

To estimate ethnolinguistic diversity for each simulation we divided the total population abundance of each run by a bootstrapped geometric mean ethnolinguistic population size from the original ethnographic sample data. The bootstrapped geometric mean population size of the ethnographic societies was 732 (723–740). The average global ethnolinguistic population diversity was 8,330 (5,579–11,082). The minimum observed diversity was 3,166 and the maximum was 15,299.

Table 1 shows the predicted abundances and diversities across continents. Average population densities are predicted to have been highest in South America and Africa but considerably lower in Eurasia, Australia, and North America. The predicted number of ethnolinguistic populations in Australia, 370 ± 130 , is consistent with ethnographic estimates at European contact. While most of the Pacific was never colonized by hunter-gatherer societies, it is interesting that the low predicted diversity in Table 1 (140 ± 50) includes Papua New Guinea, ethnographically the most linguistically diverse region of the planet (27). However, the majority of societies on Papua New Guinea constituting this high diversity are horticulturalist-farmers not hunter-gatherers. Although hunter-gatherer population densities of Papua New Guinea are predicted to have been relatively high (Figure 2C), the

Fig. 6. A tree map of the distribution of hunter-gatherer cultural diversity (S and $\%S$) distributed across the 16 biomes of the reconstructed hunter-gatherer planet, for a total diversity of 8,330. 55% of hunter-gatherer diversity occurs in tropical biomes followed by grasslands, mixed forests, shrublands, hot deserts, and islands. Less than 15% of cultural diversity occurs in temperate, boreal, tundra, and polar biomes.

model predicts relatively low diversity given the size of the island, suggesting that the high levels of diversity observed ethnographically are associated with non-foraging lifestyles.

Biogeography and biodiversity estimates. Figure 6 shows the majority of hunter-gatherer populations (55%) would occur in tropical biomes, particularly evergreen forests and savannahs. Although pathogen loads are high in the tropics – and are an important predictor of population densities in the XGBoost model (see Figures 3 and 4) – tropical biomes are not only environmentally productive, but perhaps more importantly, there are many coasts in the tropics, especially in southern and southeast Asia, northern Australia, and Papua New Guinea.

Figure 7 shows the distributions of various measures of biodiversity and cultural diversity for five metrics of interest, where the blue line represents the availability of landmass along the gradient represented; data above the blue line occur at high densities (i.e., over-represented), and data below the line occur at low densities (i.e., under-represented). A striking result is that in all plots hunter-gatherer cultural diversity correlates closely with both mammal and bird species diversity. Figure 7A shows that all forms of diversity are over-represented in the tropics ($\pm 23.5^\circ$ N/S) and under-represented at more northerly latitudes ($> 23.5^\circ$ N). Figure 7B shows that hunter-gatherer diversity is over-represented in warm ($> 20^\circ\text{C}$, Figure 7C), wet ($> 1000 \text{ mm}^{-\text{yr}}$, Figure 7D), highly productive environments ($> 500 \text{ g C m}^{-2}$, Figure 7B). Correspondingly, there is an under-representation in colder ($< 5^\circ\text{C}$, Figure 7C), drier ($< 700 \text{ mm}^{-\text{yr}}$, Figure 7D), less productive environments ($< 500 \text{ g C m}^{-2}$, Figure 7B). Moreover, Figure 7E shows that diversity decreases rapidly with distance from the coast where about 25% of global hunter-gatherer abundance occurs within 25 km of the coast, and 50% occurs within 150 km. Figure 7F shows that hunter-gatherer population densities (red high, blue low) are highest near the equator, particularly along coasts.

Discussion

The hypothetical hunter-gatherer planet Earth we reconstruct here would have an abundance of 6.1 ± 2 million people with a ethnolinguistic diversity of $8,330 \pm 2,770$ populations. The majority would occur within tropical biomes (Figure 6) and

Fig. 7. Distributions of biodiversity and cultural diversity by A) latitude, B) net primary production, C) mean annual temperature, D) mean annual precipitation, E) distance from coast, and F) population density (red high, $d = 0.2$, blue low, $d = 0$) as a function of latitude and distance from coast.

one quarter of all hunter-gatherers would live less than one day's walk from a coast (25 km). Figure 2C shows that marine mammal diversity is generally high along continental coasts at all latitudes below 60°N , and regions of high hunter-gatherer population density in environments with low net primary production often coincide with particularly productive coasts, including southern South America, the south and east coasts of Africa, coastal south Asia, coastal Australia, and the west coast of North America (Figures 2A and 2C). Marine mammal diversity is high throughout the tropics, and coupled with highly productive terrestrial and marine environments (see Figures S3), these biomes support the most diverse and dense hunter-gatherer populations on the planet. Outside the tropics and inland from coasts hunter-gatherer population density drops rapidly: Figure 6 shows only 15% of cultural diversity (thus abundance) occurs in temperate, boreal, tundra, and polar environments combined, which constitute 40% of the planet's terrestrial surface area. Our results suggest that the ethnographic record of terrestrial, low density societies is biased not necessarily due to marginal environments, but because most are not coastal. Importantly, Figure 5A shows that the importance of coasts in the model is not a statistical artifact of the over-sampling of dense, complex, sedentary societies of the North American Pacific Rim, or tropical forager-farmer mutualisms in the ethnographic record, for example: these kinds of high density societies are in fact exceedingly rare in the reconstructed world, where average global population densities are substantially lower than in the ethnographic data.

While diversity decreases rapidly with the distance from coasts, Figure 7E shows there is no evidence for positive selection for coasts as the decline in diversity simply maps onto the geometric decrease in land availability with increasing distance inland. Thus hunter-gatherer population density (and biodiversity) follows an ideal free distribution with respect to landmass (28). In contrast, Figure 7B shows highly productive terrestrial environments are particularly dense with biological and cultural diversity and much lower in low productivity environments. This finding is consistent with well-established principles of biogeography where both biological and cultural diversity are higher in the tropics than predicted by environmental productivity and land mass alone (29, 30). The classic image of low density, highly mobile hunter-gatherers moving

over vast distances may be a relatively rare adaptation to specific landscapes with low environmental productivity and no direct access to coasts.

Low productivity environments far inland would have been more common in the cooler, drier climates of the late Pleistocene, as sea levels were considerably lower increasing land-mass availability and therefore the average distance to the coast. Cooler, drier climates reduced the elevation range of biomes and thus reduced the spatial extent of the tropical zone globally (31). Under these conditions low population density adaptations would have been more common than we see ethnographically, but the majority of populations would still have occurred adjacent to coasts. In fact rich, diverse coastal environments may have been even more important given the decreased availability of terrestrial resources. If 25% of hunter-gatherer populations lived within 25km of the coast in the late Pleistocene, sea level rise over the Holocene would have had a dramatic impact on the archaeological visibility of late Pleistocene hunter-gatherer lifestyles, especially in areas of shallow continental shelf (32), such as Beringia (33), Dogger Bank (34), or the Sunda shelf (35), adding to the archaeological over-representation of low density lifestyles.

Importantly, our reconstructed world of ethnographic hunter-gatherers in the Anthropocene is not the pre-agricultural world of stone age hunter-gatherers. The Holocene is a particularly warm, wet, and stable period of recent Earth history (36), suggesting average population densities would be considerably lower in the late Pleistocene. Moreover, ethnographic societies are neither a cultural nor technological analogue of past foraging societies. All ethnographic societies have been impacted by the global expansion of industrialized societies over the last 500 years. Widespread outbreaks of disease, violence, forced relocation, ecological marginalization, and habitat loss led to the decimation of indigenous societies in many parts of the world, and many ethnographic societies were first encountered over this period. All ethnographic hunter-gatherer societies have widespread interactions with non-foraging societies of the Anthropocene, incorporating innovations such as medicines, clothing, tools, metals, firearms, and combustion engines whenever useful. No ethnographic societies now use stone tools, a central feature of foraging socioeconomies throughout human evolutionary history. Many contemporary foragers practice horticulture, others participate in long-term mutualisms with farming societies, and many participate in market economies, all of which are post-Pleistocene adaptations. Similarly, the sociopolitical diversity of hunter-gatherers in the late Holocene/Anthropocene undoubtedly includes societies with institutionalized economic and political asymmetries that were likely much rarer, if not absent, in preceding time periods in many parts of the world. Particularly important here is that many of these Holocene adaptations included the increased specialization of marine coastal environments.

The relevant anthropological question here is understanding how these historical, cultural, and technological contingencies may have impacted the range of population densities we see in the ethnographic record compared to those in the past. Ethnographically there is a thousand-fold variation in population densities, from 0.003 km^{-2} in the North American Arctic and the Western Desert of Australia to 3 km^{-2} along the Californian coast, with a sample average of 0.11 km^{-2} . It is reasonable to

assume hunter-gatherer population densities have always been particularly low in warm and cold deserts, likely at the scales we observe ethnographically, but the maximum densities, such as those found in the sedentary societies of the North American Pacific Rim, are relatively recent. It is not known whether similarly dense populations occurred in the late Pleistocene, but if so they were likely much rarer than in the ethnographic record. As such, the range of population densities we observe ethnographically is probably wider than in prehistory, and if so the demographic parameters we estimate here are likely upper estimates of the pre-agricultural hunter-gatherer planet Earth.

Conclusion

Our machine-learning approach provides a detailed reconstruction of the biogeographic distribution of hunter-gatherer societies. Populations would have been particularly dense and diverse along productive coasts and in tropical biomes. Assuming our model overestimates the population density of the prehistoric world, the pre-agricultural hunter-gatherer world likely never exceeded 6.1 ± 2 million people. 6.1 ± 2 million should also be seen as a principled estimate of the scale of the human species if all energy and materials were extracted from local environments in a counterfactual Anthropocene. Human abundance is now near 8 billion, and will be near 11 billion by the end of the century, in which case the evolution of the post-foraging world will have seen a 5,000-fold increase in human abundance over the Holocene.

Materials and Methods

Our model uses extreme gradient boosting implemented through the *xgbTree* option in the *caret* package in R (37). For a complete description of the XGBoost algorithm see (20, 38). Extreme gradient boosting is an ensemble learning algorithm that uses a matrix of data to sequentially build a large number of regression tree models that maximize predictions of the dependent data by aiming to reduce model residuals in each iteration (see Figure S2). In our case the dependent variable is hunter-gatherer population density and the predictive data is a matrix of environmental data composed of 46 variables. The data include 17 climatic variables of temperature and precipitation, 3 variables of biodiversity data (mammal, bird, and marine mammal species diversity), 16 biome variables, 7 variables of physical landscape complexity, and pathogen data (see **Supplementary Information** for further discussion and sources). All data and R code are available in the **Supplementary Material** of this paper.

ACKNOWLEDGMENTS. We appreciate comments on previous versions of this manuscript from J. Robbie Burger, Miikka Tallavaara and members of a working group on Human Macroecology funded by the Center for Biological Diversity at the University of Arizona in February, 2020, where this work was first presented and discussed.

References

1. U.S. Census Bureau, Historical estimates of world population, available at: <https://www.census.gov/data/tables/time-series/demo/international-programs/historical-est-worldpop.html> (2018).
2. P Gerland, et al., World population stabilization unlikely this century. *Science* **346**, 234–237 (2014).
3. JP DeLong, O Burger, MJ Hamilton, Current demographics suggest future energy supplies will be inadequate to slow human population growth. *PLoS one* **5**, e13206 (2010).
4. KK Goldewijk, Estimating global land use change over the past 300 years: the hyde database. *Glob. biogeochemical cycles* **15**, 417–433 (2001).
5. U.S. Government Accountability Office, Decennial census, available at: <https://www.gao.gov/highrisk/decennial-census> (2021).
6. JC French, P Piris, JFL de Pablo, S Lozano, F Silva, A manifesto for palaeodemography in the twenty-first century. *Philos. transactions Royal Soc. London. Ser. B, Biol. sciences* **376**, 20190707 (year?).
7. JR Burger, TS Fristoe, Hunter-gatherer populations inform modern ecology. *Proc. Natl. Acad. Sci.* **115**, 1137–1139 (2018).
8. M Tallavaara, JT Eronen, M Luoto, Productivity, biodiversity, and pathogens influence the global hunter-gatherer population density. *Proc. Natl. Acad. Sci.* **115**, 1232–1237 (2018).
9. LR Binford, *Constructing frames of reference: an analytical method for archaeological theory building using ethnographic and environmental data sets*. (University of California Press), (2019).
10. RL Kelly, *The lifeways of hunter-gatherers: the foraging spectrum*. (Cambridge University Press), (2013).
11. MJ Hamilton, BT Milne, RS Walker, JH Brown, Nonlinear scaling of space use in human hunter-gatherers. *Proc. Natl. Acad. Sci.* **104**, 4765–4769 (2007).
12. AL Johnson, Exploring adaptive variation among hunter-gatherers with binford's frames of reference. *J. Archaeol. Res.* **22**, 1–42 (2014).
13. WR Burnside, et al., Human macroecology: Linking pattern and process in big-picture human ecology. *Biol. Rev.* **87**, 194–208 (2012).
14. MJ Hamilton, O Burger, RS Walker, *Human Ecology*. (Wiley Online Library), pp. 248–257 (2012).
15. MJ Hamilton, et al., Population stability, cooperation, and the invasibility of the human species. *Proc. Natl. Acad. Sci.* **106**, 12255–12260 (2009).
16. MJ Hamilton, J Lobo, E Rupley, H Youn, GB West, The ecological and evolutionary energetics of hunter-gatherer residential mobility. *Evol. Anthropol.* **25**, 124–132 (2016).
17. MJ Hamilton, B Buchanan, RS Walker, Scaling the size, structure, and dynamics of residentially mobile hunter-gatherer camps. *Am. Antiq.* **83**, 701–720 (2018).
18. M Grove, Hunter-gatherer movement patterns: causes and constraints. *J. Anthropol. Archaeol.* **28**, 222–233 (2009).
19. VV Venkataraman, TS Kraft, NJ Dominy, KM Endicott, Hunter-gatherer residential mobility and the marginal value of rainforest patches. *Proc. Natl. Acad. Sci.* **114**, 3097–3102 (2017).
20. T Chen, C Guestrin, XGBoost: A scalable tree boosting system in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16. (ACM, New York, NY, USA), pp. 785–794 (2016).
21. RL Kelly, *Thinking about prehistory*. (University of Utah Press), pp. 111–117 (1999).
22. FW Marlowe, Hunter-gatherers and human evolution. *Evol. Anthropol.* **14**, 54–67 (2005).
23. MJ Hamilton, RS Walker, A stochastic density-dependent model of long-term population dynamics in hunter-gatherer populations. *Evol. Ecol. Res.* **19**, 85–102 (2018).
24. MD Gurven, RJ Davison, Periodic catastrophes over human evolutionary history are necessary to explain the forager population paradox. *Proc. Natl. Acad. Sci.* **116**, 12758–12766 (2019).
25. N Howell, Toward a uniformitarian theory of human paleodemography. *J. Hum. Evol.* **5**, 25–40 (1976).
26. JC French, AT Chamberlain, Demographic uniformitarianism: the theoretical basis of prehistoric demographic research and its cross-disciplinary challenges. *Philos. Transactions Royal Soc. B* **376**, 20190720 (2021).
27. RG Gordon Jr, Ethnologue, languages of the world. <http://www.ethnologue.com/> (2005).
28. SD Fretwell, *Populations in a Seasonal Environment*. (MPB-5). (Princeton University Press) Vol. 106, (1972).
29. MJ Hamilton, RS Walker, CP Kempes, Diversity begets diversity in mammal species and human cultures. *Sci. Reports* **10**, 1–11 (2020).
30. JH Brown, Why are there so many species in the tropics? *J. Biogeogr.* **41**, 8–22 (2014).
31. M Thomas, Late quaternary environmental changes and the alluvial record in humid tropical environments. *Quat. Int.* **72**, 23–36 (2000).
32. GN Bailey, NC Flemming, Archaeology of the continental shelf: marine resources, submerged landscapes and underwater archaeology. *Quat. Sci. Rev.* **27**, 2153–2165 (2008).
33. JF Hoffecker, SA Elias, *Human ecology of Beringia*. (Columbia University Press), (2007).
34. VL Gaffney, K Thomson, S Fitch, *Mapping Doggerland: the Mesolithic landscapes of the southern North Sea*. (Archaeopress), (2007).
35. T Hanebuth, K Statterger, PM Grootes, Rapid flooding of the sunda shelf: a late-glacial sea-level record. *Science* **288**, 1033–1035 (2000).
36. H Wanner, L Mercolli, M Grosjean, S Ritz, Holocene climate variability and change; a data-based review. *J. Geol. Soc.* **172**, 254–263 (2015).
37. M Kuhn, et al., Building predictive models in r using the caret package. *J Stat Softw* **28**, 1–26 (2008).
38. XGBoost developers, Xgboost documentation, available at: <https://xgboost.readthedocs.io/en/latest/index.html> (2020).