Short communication

Simulations of human migration into North America are more sensitive to demography than choice of palaeoclimate model

Julia A. Pilowsky a,b,*, Andrea Manica c, Stuart Brown a,d, Carsten Rahbek b,e,f, Damien A. Fordham b,g

a The Environment Institute and School of Biological Sciences, University of Adelaide, SA 5005, Australia
b Center for Macroecology, Evolution, and Climate, GLOBE Institute, University of Copenhagen, Copenhagen Ø 2100, Denmark
c Department of Zoology, University of Cambridge, Cambridge, England United Kingdom
d Section for Evolutionary Genomics, GLOBE Institute, University of Copenhagen, Copenhagen K 1350, Denmark
e Danish Institute for Advanced Study, University of Southern Denmark, Odense 5230, Denmark
f Department of Ecology, Peking University, Beijing 100871, China
g Center for Global Mountain Biodiversity, GLOBE Institute, University of Copenhagen, Copenhagen Ø 2100, Denmark

ARTICLE INFO

Keywords:
Human migration
Sensitivity analysis
Process-explicit model
Paleoecology
Macroecology

ABSTRACT

Reconstructions of the spatiotemporal dynamics of human dispersal away from evolutionary origins in Africa are important for determining the ecological consequences of the arrival of anatomically modern humans in naïve landscapes and interpreting inferences from ancient genomes on indigenous population history. While efforts have been made to independently validate these projections against the archaeological record and contemporary measures of genetic diversity, there has been no comprehensive assessment of how parameter values and choice of palaeoclimate model affect projections of early human migration. We simulated human migration into North America with a process-explicit migration model using simulated palaeoclimate data from two different atmosphere-ocean general circulation models and did a sensitivity analysis on the outputs using a machine learning algorithm. We found that simulated human migration into North America was more sensitive to uncertainty in demographic parameters than choice of atmosphere-ocean general circulation model used for simulating climate-human interactions. Our findings indicate that the accuracy of process-explicit human migration models will be improved with further research on the population dynamics of ancient humans, and that uncertainties in model parameters must be considered in estimates of the timing and rate of human colonisation and their consequence on biodiversity.

1. Introduction

Early human migration has been reconstructed indirectly (Beyer et al., 2021), correlatively (Giampoudakis et al., 2017) and process-explicitly (Timmermann and Friedrich, 2016), allowing pathways for the expansion of modern humans to be identified by inferring or modelling relationships between climatic conditions, occupancy and population growth (Erikkson et al., 2012; Steele et al., 1998). Process-explicit models have advantages over correlation reconstructions and inferences based on climate metrics because they explicitly capture demographic responses to changing climatic and environmental conditions in model simulations (Pilowsky et al., 2022). However, they are generally data intensive, with complex model structures, often resulting in high variability amongst simulations of early human migration owing to large uncertainties in underlying demographic parameters (Timmermann and Friedrich, 2016). Furthermore, most models are fitted to a single set of simulated climatic reconstructions. It is unclear how different assumptions and biases in palaeoclimate simulations (Solomon et al., 2007) affect model projections of human migration, and how important these effects are relative to uncertainties in demographic parameters. Sensitivity analyses can help improve projections of human expansion from process-explicit macroecology models by identifying parameters that contribute the most to model output, those that are insignificant and can be potentially omitted from the model, and those that need refining to improve model accuracy (Hamby, 1994).

* Corresponding author at: University of Adelaide, Adelaide, SA 5005, Australia.
E-mail addresses: julia.pilowsky@adelaide.edu.au (J.A. Pilowsky), damien.fordham@adelaide.edu.au (D.A. Fordham).

https://doi.org/10.1016/j.ecolmodel.2022.110115
Received 20 June 2022; Received in revised form 22 August 2022; Accepted 24 August 2022
Available online 12 September 2022
0304-3800/© 2022 Elsevier B.V. All rights reserved.
The Climate-Informed Spatial Genetic Model (CISGeM) is one example of a process-explicit model of human dispersal out of Africa, which has been validated using genetic distances between contemporary human populations (Eriksson et al., 2012). Its outputs include human arrival times on non-African continents and islands, as well as spatial maps of effective population size (a proxy for relative abundance (Fordham et al., 2014)) from 120 kg-years before present (ka BP). The simulated outputs of CISGeM have been used to parametrise and inform other models of phenomena including megafaunal extinctions (Fordham et al., 2022) and species range dynamics (Canteri et al., 2022). However, CISGeM has never been subjected to a sensitivity analysis, meaning there is no knowledge of the importance of demographic parameters and climatic conditions on model projections. Here we simulate human migration into North America in the Paleocene using CISGeM parametrised with two widely used atmosphere-ocean general circulation models (AOGCMs): the Hadley Centre Coupled Model, version 3 (HadCM3) (Singarayer and Valdes, 2010) and the Community Climate System Model version 3 (CCSM3) (Yeager et al., 2009). Transient Climate Evolution (TraCE-21ka) simulation (Z. Liu et al., 2009), and predicts present-day climate patterns with verified hindcast skill (Fordham et al., 2017). Importantly, both HadCM3 and TraCE-21ka model ice sheet dynamics using the ICE-5 G reconstruction (Peltier, 2004), meaning that ice sheet barriers to human dispersal in CISGeM models were identical in simulations regardless of palaeoclimatic model (Movie S1). We spatially downscaled data from both models to the equal-area resolution of CISGeM (100 km width). See Supplementary Information for details.

2. Material and methods

2.1. Human expansion model

We modelled the peopling of North America using CISGeM (Climate-Informed Spatial Genetic Model), which is a process- and spatially-explicit population model of global human migration during the late Pleistocene and Holocene (Eriksson et al., 2012). The model is driven by demographic processes responding to glacial-interglacial ice-land-sea dynamics, and spatiotemporal variation in net primary productivity that affects carrying capacities. The latter has been shown to be an important driver of population density for hunter-gatherers (Tallavaara et al., 2018; but see Zhu et al., 2021). Previous model testing has shown that CISGeM accurately reconstructs global genetic diversity and human arrival times on the non-African continents (Eriksson et al., 2012; Raghavan et al., 2015). See Supplementary Information for more details on the model structure of CISGeM.

Model parameters in CISGeM have been optimised using pattern-oriented Bayesian Computation (Grimm et al., 2005) and Approximate Bayesian Computation (Csilléry et al., 2010). In this study, we used the posterior ranges of optimised model parameters to generate 4950 plausible CISGeM models, each with different parameter values (Table S1). These posterior ranges have been used elsewhere to reconstruct human migration rates in North America and Eurasia using CISGeM (Canteri et al., 2022; Fordham et al., 2022). We used Latin hypercube sampling to generate a stratified random subset of parameter input values for simulations by specifying the posterior range for each parameter and sampling all portions of the distributions (Stein, 1987). We then ran each of these models using palaeoclimatic data from two AOGCMs, and did a global sensitivity analysis (Antoniadis et al., 2021) to determine the influence of demographic parameters and climate model parametrisation on CISGeM projections of human colonisation of North America (Figure S2).

2.2. Climate data

Plausible models (n = 4950) were simulated using palaeoclimatic AOGCM data from HadCM3 (Singarayer and Valdes, 2010) and the CCSM3 TraCE-21ka simulation (Z. Liu et al., 2009). These two palaeoclimatic models were chosen because their climatic outputs are most frequently used in macroecological models (Blos et al., 2013; Theodoridis et al., 2020), including approaches that simulate colonisation and extinction processes (Canteri et al., 2022; Fordham et al., 2022; He et al., 2013). Their high usage in ecological models reflects their temporal coverage, which tends to be more continuous than many other widely accessible palaeoclimatic datasets (Armstrong et al., 2019; S. C. Brown et al., 2020; Fordham et al., 2017), many of which are limited to widely spaced snapshots of key climatic periods (J. L. Brown et al., 2018; Lima-Ribeiro et al., 2015). While projections from the HadCM3 have been shown to be congruent with those from the CCSM3 TraCE-21ka simulation for some climatic parameters in some regions and time points (Armstrong et al., 2019), there are important local-to-regional differences between projections from these AOGCMs (Burke et al., 2018; Kageyama et al., 2018), including in North America (Fig. 1).

Unlike the TraCE-21ka simulation, the HadCM3 is not a fully transient climate model, meaning that outputs from HadCM3 are climate snapshots rather than continuous projections. Climate snapshots from the HadCM3 outputs (separated by ≥ 1 ka) were temporally downscaled to 25 year time-steps to match the timestep of CISGeM simulations using a stochastic weather generator, which draws random values from empirical distributions adjusted to fit the temperature and precipitation intervals found in the climate data (Semenov and Barrow, 2002). The grid cell resolution of HadCM3 data is 3.75° longitude × 2.5° latitude. Forcings include orbital forced insolation changes, changes in long-lived greenhouse gases, and meltwater from evolving ice sheets. These are the same forcings used in TraCE-21ka, with a key difference that HadCM3 does not account for vegetation-air-ocean interactions (Collins et al., 2006).

The TraCE-21ka simulation (Z. Liu et al., 2009) uses the CCSM3 (Yeager et al., 2006) to reconstruct daily global climate conditions at a spatial resolution of 3.75° longitude × 3.75 latitude (over land and sea) for the last 21,000 years. It accurately reproduces major climatic features associated with the most recent deglaciation event (Z. Liu et al., 2009), and predicts present-day climate patterns with verified hindcast skill (Fordham et al., 2017). Importantly, both HadCM3 and TraCE-21ka model ice sheet dynamics using the ICE-5 G reconstruction (Peltier, 2004), meaning that ice sheet barriers to human dispersal in CISGeM models were identical in simulations regardless of palaeoclimatic model (Movie S1). We spatially downscaled data from both models to the equal-area resolution of CISGeM (100 km width). See Supplementary Information for details.

2.3. Simulations

We ran a single replicate of CISGeM for each combination of plausible parameters and recorded the simulated effective population size at each hex cell and time point. Previously, it has been shown that running a single simulation iteration per parameter sample is optimal for sensitivity analysis if the parameter space is extensively sampled (Prowse et al., 2016). All simulations were global, began at the same starting location in East Africa at 120 ka BP, and proceeded until present (0 BP, 1950 C.E.) at 25-year time steps (Eriksson et al., 2012).

We identified, a priori, time of movement out of Alaska and rate of expansion through North America as two important metrics of regional human migration that are likely to be sensitive to changes in demographic parameters and variation in climate model projections. This is because climatic change facilitated the initial movement of people into North America (Becerra-Valdivia and Higham, 2020), and the speed of this movement was constrained by demographic processes and their interaction with climate and environmental conditions (Timmermann and Friederich, 2016). We calculated time of movement out of Alaska (after 19 ka BP) and rate of expansion through North America (14.7 to 11 ka BP) for each projection. Movement out of Alaska was calculated as the time when the population-weighted centroid of the leading edge of the human range (Watts et al., 2013) crossed 130 W or 51 N. Rate of expansion through North America was calculated as the rate of movement, in kilometres per year, of the population-weighted centroid of the
leading edge of the human range. See Supplementary Information for more details on how these variables were calculated. CISGeM projections of time of movement out of Alaska were independently validated using inferences of the timing of arrival of Clovis culture in North America (13,250 to 12,800 years BP; Waters and Stafford 2007).

2.4. Sensitivity analysis

To determine which parameters contribute most to model projections of human expansion in North America, we did a global sensitivity analysis using our summary metrics of time of movement out of Alaska and rate of expansion through North America (Antoniadis et al., 2021). Sensitivity analyses were done in two ways: (i) using only CISGeM models simulated using HadCM3 climate data (demographic-only sensitivity analysis); (ii) using models simulated with climate and precipitation data from HadCM3 and CCSM3 TRaCE-21ka palaeoclimate models (demographic + climate sensitivity analysis). This two-step approach was done because CISGeM were originally optimised using HadCM3 climate data (Eriksson et al., 2012). The sensitivity analysis did not account for potentially important structural uncertainties in CISGeM, including human generation length and the simulated sequence of modelled demographic processes.

We determined the sensitivity of timing of movement out of Alaska and expansion rate using random forest learning methods (Antoniadis et al., 2021).
et al., 2021) following techniques established for process-explicit macroecology models (Pearson et al., 2014). We tuned the hyperparameters using k-fold cross-validation, choosing number of variables sampled per split and minimum node size by minimizing RMSE (RMSE = 478 ± 13.6 for exit from Alaska and RMSE = 479 ± 15.3 for expansion rate). We assessed variable importance using unscaled permutation importance (Strobl et al., 2007). See Supplementary Information for details.

3. Results

While range size of humans in North America varied according to AOGCM (Fig. 1), time of movement out of Alaska and rate of human migration were most sensitive to uncertainty in key demographic parameters (Fig. 2). The demographic-only sensitivity analysis, done using HadCM3 model-based simulations only, revealed: i) time of movement out of Alaska was most sensitive to colonisation rate, upper net primary productivity threshold for carrying capacity and population growth rate; while ii) population-weighted rate of expansion was most sensitive to population growth rate and colonisation rate (Fig. 2). This order of relative importance remained unchanged when the sensitivity analysis was done on simulations with varying temperature and precipitation inputs from the two AOGCMs (demographic + climate sensitivity analysis) (Fig. 2). This indicates a relatively low sensitivity of CISGeM projections to pronounced differences in palaeoclimate conditions in North America according to AOGCM (Fig. 1) when compared to uncertainties in demographic model parameters.

Independent tests of CISGeM projections of time of movement out of Alaska showed that simulations of land migration from CISGeM parametrised with TraCE21-ka climate data gave a median exit date from Alaska that was closer to the estimated Clovis arrival (median: 14,375 years BP, MAD: 482) compared to simulations parametrised with HadCM3 data (median: 15,000 years BP, MAD: 111). The difference for TrnCe-21ka and HadCM3 was 1144 years (95% confidence interval (CI) = 1138–1150 years) and 1682 years (CI = 1663–1700 years), respectively. Model projections of migration patterns into North America and relative Ne for both models can be accessed on Figshare (Pilowsky et al., 2022).

4. Conclusions

While projections of the peopling of North America from process-explicit models vary in response to two choices of AOGCM, uncertainties in key demographic parameters have a disproportionately larger influence on simulations of time of movement out of Alaska and rate of expansion through North America. This shows the likely importance of considering uncertainties in the demographic parameters of process-explicit model projections of timing, rate and mechanisms of initial human expansion across continents (Raghavan et al., 2015), and the broader ecological consequences of human colonisation on biodiversity (Canteri et al., 2022; Fordham et al., 2022).

While arrival times of humans in different regions have been established archaeologically with reasonable certainty (Goebel et al., 2008; Groucutt et al., 2015), and dispersal rates have been inferred from genomic analysis of aDNA (Rasmussen et al., 2011), the pattern of human growth and expansion has been more difficult to reconstruct at fine spatiotemporal scales. Consequently, projections of early human migration across continents are still uncertain (H. Liu et al., 2006). This is partly because of overly simplistic parametrisation of the relationship between net primary productivity and population growth (Zhu et al., 2021) and large uncertainties in other demographic parameters, including dispersal (French et al., 2021).

Resolving these issues should be a priority, given how sensitive the rate of human migration in North America is to rates of population growth and colonisation. Promising avenues of research that could reduce uncertainty in early human demography include Bayesian analysis of spatiotemporal distributions of radiocarbon dates (Price et al., 2020); phylogenetic analysis of the human palaeoproteome, which is more resistant to degradation over long timescales compared to the palaeogenome (Welker, 2018); and sampling of environmental DNA, which can detect arrival and movement of small populations better than the archaeological or fossil record (Wang et al., 2021).

Our finding that uncertainty in projections of human migration from process-explicit models is only weakly sensitive to the choice of underlying palaeoclimate model is in stark contrast to findings for correlative models of species distributions (Beaumont et al., 2007; Tuck et al., 2006), which model demographic processes implicitly, not explicitly (Pilowsky et al., 2022). When interpreting the generality of this result, it is important to recognise that CISGeM simulates pathways for the global expansion of modern humans. Therefore, in other regions and time periods, the parametrisation of palaeoclimate could have a larger effect on human migration, especially since migration occurred at different rates in different regions. While we tested the sensitivity of CISGeM to palaeoclimate uncertainty using two AOGCMs with very different climate sensitivities (Mason-Delmote et al., 2013), spatiotemporal uncertainty could potentially be greater in North America if more models were considered. Nevertheless, our results highlight the importance of realistically capturing demographic mechanisms in process-explicit human migration models.

CRediT authorship contribution statement

Julia A. Pilowsky: Conceptualization, Validation, Formal analysis, Writing – original draft, Visualization. Andrea Manica: Methodology, Software, Writing – review & editing. Stuart Brown: Formal analysis, Writing – review & editing. Carsten Rahbek: Conceptualization, Supervision. Damien A. Fordham: Funding acquisition, Conceptualization, Supervision, Methodology, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data are available on Figshare and have been cited in the text.

Acknowledgments

DAF acknowledges funding from the Australian Research Council (FT140101192, DP180102392), and a residency fellowship from Danmarks Nationalbank. CR received funding from DNRF-CMEC (DNRF96) and from Villum Fonden (grant no. 25925).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ecolmodel.2022.110115.

References
