A Stable Finite-Difference Scheme for Population Growth and Diffusion on a Map

Journal Article

Author(s):
Petersen, Wesley P.; Callegari, Simone; Lake, George R.; Tkachenko, Natalie; Weissmann, John D.; Zollikofer, Christoph P.E.

A Stable Finite-Difference Scheme for Population Growth and Diffusion on a Map

W. P. Petersen[1,2]*, S. Callegari[1], G. R. Lake[3], N. Tkachenko[1], J. D. Weissmann[1], Ch. P. E. Zollikofer[1]

1 Anthropological Institute and Museum, Univ. of Zürich, Zürich, Switzerland, **2** Seminar for Applied Mathematics, ETH, Zürich, Switzerland, **3** Institute for Computational Science, Univ. of Zürich, Zürich, Switzerland

* wpp@math.ethz.ch

Abstract

We describe a general Godunov-type splitting for numerical simulations of the Fisher–Kolmogorov–Petrovski–Piskunov growth and diffusion equation on a world map with Neumann boundary conditions. The procedure is semi-implicit, hence quite stable. Our principal application for this solver is modeling human population dispersal over geographical maps with changing paleovegetation and paleoclimate in the late Pleistocene. As a proxy for carrying capacity we use Net Primary Productivity (NPP) to predict times for human arrival in the Americas.

Introduction

There is considerable interest in modeling population dynamics at large spatial and temporal scales, for example the modern human out-of-Africa dispersal [1–4] or Neanderthal dispersal and extinction [5]. These models are required to interpret local and global patterns of genetic, phenetic and cultural variation [1, 6–9].

Fisher [10] studied the problem—via a growth-diffusion equation—of an analogous but one-dimensional situation: the propagation of an advantageous genetic mutation within an already-present population, situated along a coast line. Kolmogorov, Petrovskii and Piskunov [11] were more general; in particular, their analysis treated the two-dimensional case. Such a model (called Fisher/KPP in the following) was applied to the dispersal and growth of a population by Skellam [12], and serves as an important control for designing and validating other more complex spatiotemporal population models [5]. Coupling population dynamics with models of large-scale changes in continental topography, climate, and ecosystem productivity is necessary to understand the role of environmental constraints on patterns of genetic, phenetic, and cultural variation among human populations [5].

Here we present a stable and efficient finite-difference solver for the Fisher/KPP equation on the 2-D domains of geographical maps, and show how it can be extended to include environmental fluctuations. In a brief outline of our paper, we will: review the derivation of the Fisher/KPP equation (Section **The Fisher/KPP equation**); develop finite-difference schemes in 1 and 2 dimensions for variable environmental carrying capacity $\mathcal{K}$ (Section **Numerical**

methods and splitting) depending on both space and time $\mathcal{K}(\mathbf{x}, t)$ (Section **Space—time dependent capacity maps**); and show our application of this technique to the out-of-Africa dispersal of *Homo sapiens* by using net primal productivity (NPP) [1] as a proxy for $\mathcal{K}$ (Section **World-wide hominin dispersal**).

The Fisher/KPP equation

An intuitive way to get the Fisher/KPP equation [10–12] is as follows. A current **j** of particles (e.g., individuals) moving across an interface located at **x** is proportional to the gradient of the population density p (called "Fickian diffusion")

$$\mathbf{j} = -c \, \nabla p.$$

The rate of change of p is then given by the mass balance equation [13], which for Fickian diffusion reads

$$\frac{\partial p}{\partial t} - c \, \nabla^2 p = \rho.$$

If $\rho = 0$, this is the heat equation when $c = D/2$ and D is the diffusion coefficient. For lack of a better model, we assume c is a constant (Young & Bettinger [4]). In Subsection **Parameter optimization**, we optimize on $D = 2c$. The source term ρ is usually modeled by a logistic growth function, $\rho = \lambda p (1 - p/\mathcal{K})$, and gives the Fisher/KPP equation

$$\frac{\partial p}{\partial t} = c \, \nabla^2 p + \lambda \left(1 - \frac{p}{\mathcal{K}} \right) p, \tag{1}$$

where $\mathcal{K}$ is called *carrying capacity* and λ is the growth rate. In Eq (1) we wrote $\mathcal{K}$ as a constant. Since ρ is a local density, it can be space, **x**, and time, t, dependent. To get relative density information, all we need is that there is an *upper limit* $\bar{\mathcal{K}}$, in which case $0 \le p \le \bar{\mathcal{K}}$. If $\mathcal{K}(\mathbf{x}, t)$ is evolving, locally the solution $p \le \mathcal{K}(\mathbf{x}, t) \le \bar{\mathcal{K}}$. As in Table 1, $\bar{\mathcal{K}}$ can be scaled to unity, although obviously only the relative density $p/\bar{\mathcal{K}}$ can be computed. A scaled version for constant $\mathcal{K}$ is given in Eq (2), while space and space-time dependent $\mathcal{K}(\mathbf{x}, t)$ versions are given in the Section **Space-time dependent capacity maps**. Fluctuations in climate produce environmental changes in vegetation, sea levels, opening/closing of land bridges, waxing/waning of ice sheets, and perturbations to habitable areas in general. Thus, time-dependent environments compel us to deal with space- and time-dependent $\mathcal{K}(\mathbf{x}, t)$ (see Subsection **Time interpolation of maps**), Section **Space-time dependent capacity maps**, Eq (9). Fisher and KPP were particularly interested in the traveling wave case, $p(\mathbf{x}, t) = f(\mathbf{x}_0 + \mathbf{v}t)$. It was shown in [14, 15] that asymptotically (large t), the speed is $v = 2\sqrt{\lambda c} = \sqrt{2rD}$ in our notation. By the rescalings show in Table 1, for constant $\mathcal{K}$ the Fisher/KPP Eq (1) is written

$$\frac{\partial u}{\partial t} = \frac{1}{2} \, \nabla^2 u + (1 - u)u, \tag{2}$$

Table 1. Variables in left column are scaled versions of those in right column.

In Eq (2)	In Eq (1)
x	$\sqrt{\frac{\lambda}{2c}}\mathbf{x}$
t	λt
u	$p/\mathcal{K}$

where the only sensible solutions have $0 \leq u \leq 1$. The initial distribution $u(\mathbf{x}, 0) = u_0(\mathbf{x})$ must be defined for all $\mathbf{x}$ on the habitable regions of the map.

In Murray ([14], eq. (11.17)) our growth coefficient λ is called r and c is denoted by D, whereas in Young and Bettinger [4] the growth coefficient is R and the diffusion coefficient is K. We use r to denote radius in Section **Numerical methods and splitting**, and k for the CFL parameter, so λ and D are used here and that next section. Elsewhere, $\lambda = r$ is used interchangeably. These inputs to our code are given in units of yr^{-1} and $\mathrm{km}^2/\mathrm{yr}$, respectively.

Numerical methods and splitting

In one dimension, Eq (2) can be solved using the `MatLab` function `pdepe`, or if the system is two-dimensional but rotationally symmetric, `pdepe` can again be used with the radial part of the Laplace operator in cylindrical coordinates,

$$\nabla^2 = \frac{1}{r}\frac{\partial}{\partial r} r \frac{\partial}{\partial r} + \text{non-contributing terms},$$

requiring only that one sets a `pdepe` parameter `m=1`. What is important is that because `MatLab` function `pdepe` is robust, it is a valuable verification tool for testing our more general solver when comparisons can be made.

The finite-difference scheme

Since the map on which we will be working is a pixelized plane, an obvious method uses finite differences. First, however, let us examine the 1-D case for Eq (2). In this situation, the second order derivative becomes a differencing operator in matrix form acting on the vector $\{u_j, j = 1, n\}$, where $u_j = u(x_0 + (j - 1)\Delta x)$,

$$\frac{\partial^2 u}{\partial x^2} \to \frac{1}{(\Delta x)^2} A u,$$

where the matrix A is

$$A = \begin{pmatrix} -2 & 1 & 0 & 0 & \ldots & 0 \\ 1 & -2 & 1 & 0 & \ldots & 0 \\ 0 & 1 & -2 & 1 & \ldots & \\ 0 & 0 & 1 & \ldots & & \\ 0 & 0 & \ldots & 0 & 1 & -2 \end{pmatrix}. \tag{3}$$

If h is the time step, the Courant—Friedrichs—Lewy (CFL) parameter [16] is

$$k = \frac{h}{2(\Delta x)^2}. \tag{4}$$

An explicit integrator for Eq (2) would require $k < 1/4$ [16, 17]. In our case, because boundary conditions are so irregular on a map and there are strong spacial variations in $\mathcal{K}$, we are uninterested in a method of higher order than second because stability is more important [18].

Using this notation, the lowest order approximation is Euler's method which estimates the next step $u(t + h)$ by

$$u_E = u(t) + k A u(t) + h(1 - u(t))u(t), \tag{5}$$

which should be considered a vector equation in $u(t) = \{u_j(t), j = 1, n\}$. The logistic terms, which are diagonal, should be taken to mean $((1-u)u)_j = (1-u_j)u_j$ for $j = 1, 2, \ldots, n$. Euler's method is both low-accuracy and generally unusable if it is used alone over too many steps. This method is in principle conditionally stable if the step size is forced to be small enough. A step size which is too small, however, actually drives errors up, not down: e.g., C. Wm. Gear's Fig. 1.8 [19], page 19. Euler's method is $O(h)$ accurate so is useful as an explicit estimate (predictor) inside $O(h)$ terms, and thus yields $O(h^2)$ accuracy for such terms. An application of the trapezoidal rule yields

$$
\begin{aligned}
u(t+h) \quad = \quad & u(t) + \frac{k}{2}(Au(t+h) + Au(t)) \\
& + \frac{h}{2}((1 - u(t+h))u(t+h) + (1 - u(t))u(t))
\end{aligned}
\tag{6}
$$

and is an $O(h^2) + O((\Delta x)^2)$ accurate procedure. However, solving the quadratic vector Eq (6) for $u(t + h)$ is awkward. To the same $O(h^2)$ accuracy, we use a semi-implicit procedure which uses the Euler estimate Eq (5) to modify one of the terms, $(1-u(t+h))$, in Eq (6):

$$
\begin{aligned}
u(t+h) \quad = \quad & u(t) + \frac{k}{2}(Au(t+h) + Au(t)) \\
& + \frac{h}{2}((1 - u_E)u(t+h) + (1 - u(t))u(t)).
\end{aligned}
\tag{7}
$$

Eq (7) can be solved as a linear system,

$$
\left(1 - \frac{k}{2}A - \frac{h}{2}(1 - u_E)\right)u(t+h) = u(t) + \frac{k}{2}Au(t) + \frac{h}{2}(1 - u(t))u(t),
\tag{8}
$$

because the matrix, $1 - \frac{k}{2}A - \frac{h}{2}(1 - u_E)$, on the left hand side is explicit, as is the right hand side. That is, this matrix and the right hand side contain only old data, namely just information from the previous step, $u(t)$. Euler estimate u_E Eq (5) is an explicit one step computation using old data $u(t)$. Significant advantages are: the matrix on the left hand side is tridiagonal with constants on the sub/super-diagonals, and the diagonal terms are $O(1)$ strong. The procedure Eq (8) is only linearly stable but we find empirically that it gives good results when compared to `pdepe` when comparisons to this `MatLab` function are appropriate.

Fig 1 shows the results for $h = 1/12$, $\Delta x = 1/5$ compared to `pdepe`. Notice that at both $t = 2$ and $t = 20$ the agreement is quite satisfactory. The CFL number, $k = 1$, used to get Fig 1 is much larger than would be possible with an explicit method [17].

Space-time dependent capacity maps

Fluctuations in climate produce environmental changes in vegetation, sea levels, opening/closing of land bridges, waxing/waning of ice sheets, and perturbations to habitable areas in general. Thus, changing environments compel us to develop a procedure Eq (9) for the case when $\mathcal{K}(\mathbf{x}, t)$ depends on both space and time.
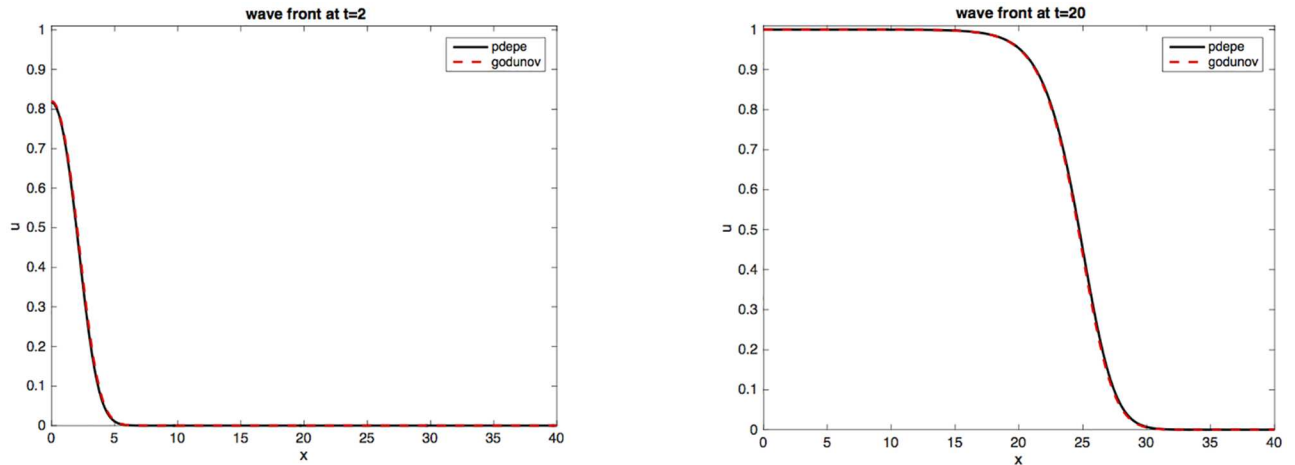
Fig 1. Godunov vs. pdepe. Left: Godunov vs. pdepe at $t = 2$. Right: same at time $t = 20$. Only $x > 0$ data are shown. The time step $h = 1/12$ and $\Delta x = 1/5$. Initial data: $u(x) = 1$ if $|x| < 1$, zero otherwise.

Since Eq (8) is basically the trapezoidal method (see section 5.3 in [17]), the modification for a space—time dependent $\mathcal{K}(\mathbf{x}, t)$ is as follows:

$$\left(1 - \frac{k}{4}A_x\right)u^\star = \left(1 + \frac{k}{4}A_x\right)u(t) \tag{9a}$$

$$u_E = u^\star + kA_y u^\star + h\left(1 - \frac{u^\star}{\mathcal{K}(t)}\right)u^\star \tag{9b}$$

$$\left(1 - \frac{k}{2}A_y - \frac{h}{2}\left(1 - \frac{u_E}{\mathcal{K}(t+h)}\right)\right)u^{\star\star} = \left(1 + \frac{k}{2}A_y + \frac{h}{2}\left(1 - \frac{u^\star}{\mathcal{K}(t)}\right)\right)u^\star \tag{9c}$$

$$\left(1 - \frac{k}{4}A_x\right)u(t+h) = \left(1 + \frac{k}{4}A_x\right)u^{\star\star}, \tag{9d}$$

where we have suppressed the $\mathbf{x}$ dependence of $\mathcal{K}(\mathbf{x}, t)$ for simplicity of notation. In Eq (9), the operators A_x and A_y are the same finite difference operators as in Eq (3) but for directions x and y, respectively. For simulations on a lattice, $u_{ij}(t) = u(x_0 + (i-1)\Delta x, y_0 + (j-1)\Delta y, t)$, where $1 \leq i \leq N_x$, $1 \leq j \leq N_y$ and $\Delta x = \Delta y$, the following gives the action of the A_x, A_y operators:

$$
\begin{aligned}
A_x u_{i,j} &= u_{i-1,j} - 2u_{i,j} + u_{i+1,j}, \\
A_y u_{i,j} &= u_{i,j-1} - 2u_{i,j} + u_{i,j+1}.
\end{aligned}
$$

A compression scheme and code outline given in Appendix **Map segmentation** show that only a maximum of one row or column (i.e., $\max(N_x, N_y)$) of storage is needed for $u^\star$ and $u^{\star\star}$.

We use the simple regularization of each version of $0 \leq u \leq \mathcal{K}$ in Eq (9) at every $\mathbf{x}$–point, namely

$$0 \leq u^\star \leq \mathcal{K}(\mathbf{x}, t),$$
$$0 \leq u_E \leq \mathcal{K}(\mathbf{x}, t+h),$$
$$0 \leq u^{\star\star} \leq \mathcal{K}(\mathbf{x}, t+h).$$

For example: use Eq (9b) to compute u_E, then set $u_E \leftarrow \max(0, u_E)$, followed by $u_E \leftarrow \min\left(\mathcal{K}(\mathbf{x}, t + h), u_E\right)$. Intermediate values $u^\star$ and $u^{\star\star}$ are constrained similarly.

Fisher/KPP on geographical maps

In order to solve Fisher/KPP on a geographical map, one segments the map into x- and y-direction pieces, imposing boundary conditions at their endpoints; the solver acts on each segment independently, changing the directions $x \to y$ Eqs (9a) to (9b), then $y \to x$ Eqs (9b) to (9d) [17] of integration as in Eq (8). This approach also lends itself to efficient parallelization. In the **Map segmentation appendix** we illustrate the scheme with an illustrative MatLab code sample.

Boundary conditions for Fisher/KPP

Two versions for boundary conditions were implemented and tested: (1) zero solution Dirichlet, and (2) zero net flux Neumann conditions. Both variations require some explanation.

Dirichlet BCs

In this case, a border of zero pixels were added to the land masses. In our Godunov method, each direction (X, respectively Y) segments acquire two zero pixel end points, and the resulting $n_{seg} + 2$ one dimensional difference equations are solved, but only the n_{seg} land mass pixel values of the solution are updated. For example, in the Appendix **Map segmentation**, an X direction segment passing through equatorial central Africa has n_{seg} pixels, each of which also has a corresponding space/time dependent carrying capacity $\mathcal{K}(x, t)$. Two zero pixels are added, one at each end, but $\mathcal{K}(x, t)$ for those additional points are not relevant, nor used. Single pixel land mass segments thus become 3 pixels, where only the middle point is updated. The differencing operator (3) becomes very simple with three elements/row (1, -2, 1). Simulations with high resolution, say $N_x = 720$ and $N_y = 360$, show that the coastal region populations look too small. This is because of smoothing between zero (water) and the second adjacent pixel away. Our ancestors fished, so we concluded Dirichlet BCs seem inappropriate.

Neumann BCs

In consequence, we use zero net flux, $\partial u/\partial x = 0$ or $\partial u/\partial y = 0$ into the sea conditions. This is more awkward and two issues must be dealt with:

- Derivative information (u_x, u_y to $O(\Delta x^2)$) must be available, or approximated. From LeVeque [17], we know that when $n_{seg} > 3$, we may use $\partial u/\partial x \approx \left(\frac{3}{2}u(x) - 2u(x + \Delta x) + \frac{1}{2}u(x + 2\Delta x)\right)/\Delta x$ as a one-sided (to the right, here) difference approximation. If $n_{seg} \leq 3$, however, things are clumsier.

- Furthermore, because our Godunov splitting advances two one half steps of the heat equation per time step, there is the question of whether the solver simulates these half-steps in a properly posed way. It is well known (again [17]) that the heat equation with Neumann boundary conditions can be an ill-posed problem. This is easy to see: $u_t = \frac{1}{2}u_{xx}$ with $u_x = 0$ at the end points remains invariant to the shift $u \to u + C$, for any constant C. Thus, in that case, the solution cannot be unique. Uniqueness must be imposed by the logistics term in (2). That term is not invariant under the shift, so uniqueness is assured. Only single half time—steps for the heat equation steps, Eqs (9a) and (9d), are taken.

If $n_{seg} > 3$, the resulting linear equations to be solved are of the form Eq (9c).

$$\begin{pmatrix} 1, & -4/3, & 1/3 & \cdots & 0 \\ -\dfrac{k}{2}, & 1+k-\dfrac{h}{2}(1-u_{E,2}), & -\dfrac{k}{2} & \cdots & 0 \\ 0 & -\dfrac{k}{2}, & 1+k-\dfrac{h}{2}(1-u_{E,3}), & -\dfrac{k}{2} & \cdots \\ & \cdots & \cdots & \cdots & \\ 0 & \cdots & 1/3, & -4/3, & 1 \end{pmatrix} u^{\star\star} \tag{10}$$
$$= RHS,$$

where $RHS_1 = RHS_{n_{seg}} = 0$. Elements $u_{E,j}$ denote the j-th elements of the Euler estimate vectors u_E. Note that this system has bandwidth 5, not 3. However, the upper and lower 2nd sub/super-diagonals contain only one element (1/3). As LeVeque [17] points out, this only *slightly disturbs the tridiagonal structure*. To be able to still use a tridiagonal solver, we modify Eq (10), by using explicit Heun rule estimates for u_3 and $u_{n_{seg}-2}$ and move those to the right hand side: that is,

$$RHS_1 \rightarrow -\frac{1}{3} u_3^{est}, \quad \text{and} \quad RHS_{n_{seg}} \rightarrow -\frac{1}{3} u_{n_{seg}-2}^{est},$$

and remove the first and n_{seg}-th row elements 1/3 in Eq (10). The u_1 and u_{nseg} elements can be updated after the tridiagonal system solution to correct for the explicit estimates, if desired. We find no discernible differences doing so.

Three special cases remain: $n_{seg} = 1, 2, 3$.

- If $n_{seg} = 1$, only the logistics term can be used: set $A_y = 0$ in (11b) and (11c) and update the single element $u_1^{\star\star}$. This must be kept non-negative, however.

- Surprisingly, the $n_{seg} = 2$ is the most awkward. Eq (11) use the 2nd order differencing operator on the old (previous time step) solutions,

$$A_x = \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix},$$

and likewise A_y. The resulting solutions of Eqs (9a) and (9d) contain limited derivative information, hence the Neumann boundary conditions require $u_1^\star = u_2^\star$, $u_1^{\star\star} = u_2^{\star\star}$, and $u_1(t + h) = u_2(t + h)$ and all these must be non-negative. Similar restrictions are also made on the Euler estimate in (11b).

- When $n_{seg} = 3$ the resulting 3-vector equations for $u^\star$ $u^{\star\star}$, and $u(t + h)$ all take a form

$$\begin{pmatrix} 1 & -4/3 & 1/3 \\ S_1 & S_2 & S_1 \\ 1/3 & -4/3 & 1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} 0 \\ RHS_2 \\ 0 \end{pmatrix}.$$

A couple of simple manipulations show $u_1 = u_2 = u_3$, which says that there is no derivative information if Neumann boundary conditions are imposed on this $n_{seg} = 3$ case. The solution is $u_2 = RHS_2/(2S_1 + S_2)$, and thus we get the other two. They must also be kept non-negative.

World-wide hominin dispersal

We now focus on our principal application using the methods presented above: the world-wide dispersal of *Homo sapiens* out of Africa.

Capacity maps

Our construction of a time-dependent $\mathcal{K}$ uses Net Primary Productivity (NPP) as a proxy [1]. The Miami model [20] was originally formulated in 1972 to estimate NPP (in (grams of carbon in dry organic matter)/m^2/day) from annual temperature and rainfall [21]. In order to compute our NPP maps, we obtained the temperature and precipitation data from simulations by the BRIDGE program [22] organized at the University of Bristol [23–26]. The simulation data that we use are computed on a 96×73 grid, which we interpolate to size 720×360 and convert to NPP maps [27] by applying the formulas given in [20].

World—wide NPP data are difficult to obtain, so our Miami model-like maps are somewhat rough. Via testing, we found our Godunov solver to be quite robust with respect to abrupt $\mathbf{x}$ −steps in the carrying capacity $\mathcal{K}(\mathbf{x}, t)$.

Time interpolation of maps

We assembled 61 NPP maps, from 120 kya to 1 kya. These are in 4 kilo—year steps for the first 10; 2 ky steps for the next 21; then 1 ky steps for the remainder. The time stepper in our Godunov scheme has no information about continuity between these discrete NPP maps, so an interpolation scheme needs to be used.

If we have available some estimates for the carrying capacities $\mathcal{K}_L$, $\mathcal{K}_H$ at times t_L, t_H, one possible estimate for $\mathcal{K}(t)$ at times $t_L \leq t \leq t_H$ in between is a homotopy,

$$\mathcal{K}(\mathbf{x}, t) = \mathcal{K}_L(\mathbf{x})(1 - S(t)) + \mathcal{K}_H(\mathbf{x})S(t),$$

where a sigmoid function $0 = S(t_L) \leq S(t) \leq S(t_H) = 1$ smoothly interpolates between lower and higher time frames. There are many choices available, such as that used in [1]. We use the following variant.

Start with the classical sigmoid

$$S(z) = \left(1 + e^{-z}\right)^{-1} \tag{11}$$

which is zero at $z = -\infty$ and unity at $z = +\infty$. The $-\infty < z < +\infty$ interval is not what we want, but the following $t \mapsto z$ transformation permits several variants:

$$z = \frac{2\Delta T(t - t_L) - (\Delta T)^2}{((t - t_L)(\Delta T - (t - t_L)))^\nu}. \tag{12}$$

where $\Delta T = t_H - t_L$. Notice that $S(z(t_L)) = 0$ and $S(z(t_H)) = 1$. The exponent ν in Eq (12) gives some freedom in choosing a particular form for z for almost any $\nu > 0$. If $\nu < 1/2$, d^{2}S/dt^2 has more than two sign changes, so $\nu \geq 1/2$ is preferable. With choice $\nu = 1/2$, the interpolant is nearly a straight line: see Fig 2. However, its turn-up at $t = t_L$ and turn-down at $t = t_H$ numerically resemble very quick derivative changes. So, we choose $\nu = 1$ which also makes $z(t)$ time scale independent. At both ends, all derivatives of $S(z(t))$ in t smoothly vanish. We also have the forward/backward symmetry $S(z(t_H - t)) = 1 - S(z(t))$ for $t_L \leq t \leq t_H$, as does [1].

Parameter optimization

Other tasks remain: optimize the $r = \lambda$ and $D = 2c$ parameters in Eq (1); choose our starting value T_s, the start time for our simulation; and choose a threshold value for arrival. For
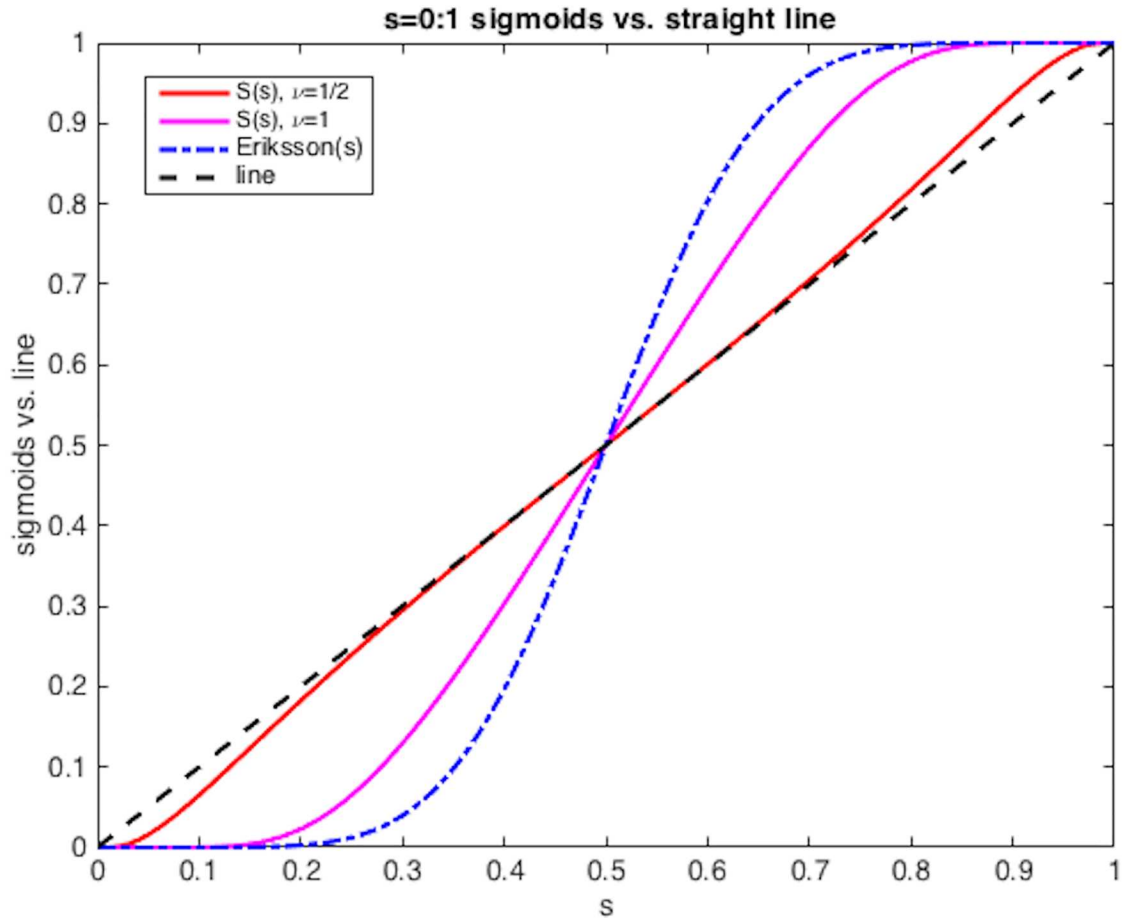
Fig 2. Time frame interpolation methods. The interpolation maps $\mathcal{K}$ at different times: sigmoids Eq (12) for $v = 1/2$ and $v = 1$, straight line, and Eriksson's $f(f(f(t)))$ model [1]. Time scaling is $s = (t - t_L)/(t_H - t_L)$.

doi:10.1371/journal.pone.0167514.g002

example, a migration wave reaching some fraction of the local carrying capacity at a studied site would constitute *arrival*. Roughly, the distance from Eritrea/Djibouti to any site, arriving at time t_A, looks like

$$\text{distance} = a\sqrt{rD}\, t_A + b\sqrt{Dt_A}$$

where a, b are unknown fractions of the distance. The terms consist of a nearly constant speed ($\sim \sqrt{rD}$) traveling wave, and a diffusion $\sqrt{Dt_A}$, respectively. This is a scaling argument using the units $D \sim$ km^2/yr and $r \sim$ yr^{-1}. A fraction ccFRAC $\approx 1/2$ of the carrying capacity would mostly probe the first term, $a\sqrt{rD}\, t_A$. Conversely, a tiny value of ccFRAC weighs too much of the diffusion tail $b\sqrt{Dt_A}$. The right-hand frame of Fig 1 shows that the wave consists of a steep traveling wave with extended diffusion tails. We found that arrival times are relatively insensitive to values of ccFRAC ~ 0.2. Larger values can produce much later arrival times, as in Table 2 and its missing error bars. Hence, our choice is ccFRAC = 0.1: When the incoming population reaches 10 percent of the local carrying capacity, it has *arrived*. For this ccFRAC choice, the differences between NPP and FEP (*Food Extraction Potential* [28]) arrivals did not seem significant. The opening of the Bering Straight "bridge" occurs before any of our simulation start-times, T_s, so arrival times in the Americas become almost linearly increasing for

Table 2. Predicted arrival times for 10 sites [29–31]. The furthest apart American sites†, Meadowcroft and Fell's Cave, were used for optimization. Error estimates give approximate sensitivities to the coarse r steps in the left-hand plot of Fig 3. Entry **NA** means the half-filling wave did not arrive. All values are in kilo—years before present.

Early Archaeological sites			
site	arch. range	arrival at 10% CC	arrival at 50% CC
Meadowcroft†	16–14	19.0 ± 0.7	17.5
Taima-Taima	14 ?	13.4 ± 0.9	11.8
El Inga	9	13.2 ± 0.9	11.8
Pachamachay	12–10.5	17.4 ± 0.7	NA
Pikimachay	22.2–12	12.5 ± 0.9	11.4
Pedra Furada	12–5	11.4 ± 0.9	10.3
Tagua-Tagua	11.5	13.4 ± 0.9	11.8
Monte Verde	14.6–12.5	9.9 ± 0.9	8.6
Page-Ladson	14.6	19.4 ± 0.7	18.3
Fell's Cave†	11	9.3 ± 1.0	7.6

doi:10.1371/journal.pone.0167514.t002

$T_s \geq$ 45kya. Hence, T_s = 45kya value is a reasonable lower limit. Our NPP data and masks are later than 120 kya, after the beginning of the Tarantian period. A Bering Strait crossing was possible until the Holocene period began around 10 kya. Thus migration data would permit a transit until closure. Arrival in northeastern India is known to have happened before 40kya, so T_s cannot be that small. Thus to limit our optimization space, we choose T_s = 45kya. For this choice of T_s, we still have to choose r, D. Some estimates were known [4, 5] to be approximately $D \approx 200 \text{km}^2/\text{yr}$ and $r \approx 2.0 \times 10^{-3} \text{ yr}^{-1}$. With the hope that we might be more precise, we optimize on the following RMS parameter,

$$\text{RMS} = \sqrt{\frac{1}{2}\left((\text{sim}_{MC} - \text{arch}_{MC})^2 + (\text{sim}_{FC} - \text{arch}_{FC})^2\right)},$$

where arch_{MC} = earliest estimated population of Meadowcroft, PA; similarly, arch_{FC} = earliest known population of Fell's Cave, near the tip of Cape Horn, Isla Grande, Tierra del Fuego
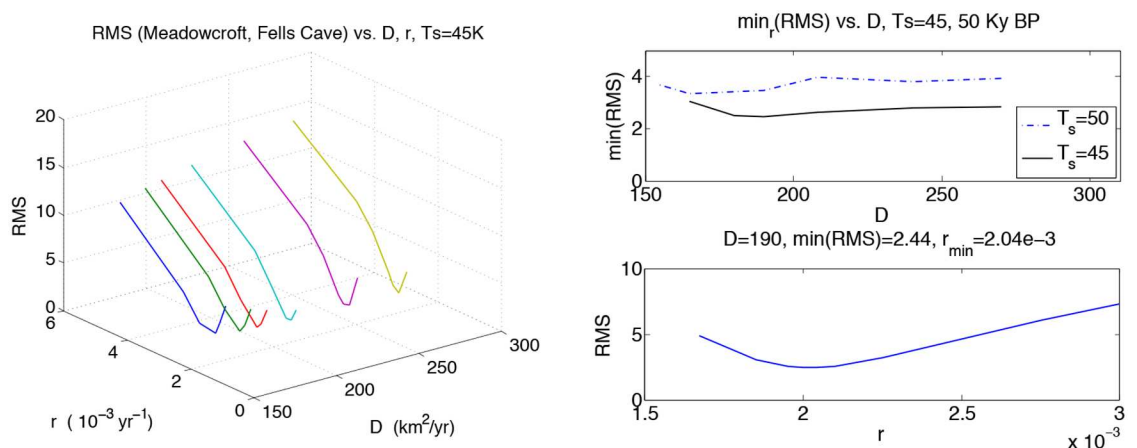

Fig 3. Diffusion and growth parameters optimization. Left: the RMS value of simulation compared to earliest known populations of Meadowcroft and Fell's Cave. The approximate optimal value of D is 190km²/yr. The minimum RMS value is shallow in D but sensitive to r = λ, showing we are probing both the traveling wave and diffusion tail. Right: a detailed min$_r$(RMS) optimization in D and in r at D = 190: $\lambda_{min} = r_{min} = 2.04 \times 10^{-3} \text{ yr}^{-1}$.

doi:10.1371/journal.pone.0167514.g003

[29, 30]. Our simulation values sim_{MC} and sim_{FC} depend on the diffusion/growth parameters r, D, the *arrival* threshold CCFRAC, and starting time T_s.

Population dispersal

An initial population at $t = T_s = 45$ kya at an Eritrean is shown in Fig 4. The integration units are scaled following Table 1 such that $\lambda = 2.04 \cdot 10^{-3} \, y^{-1}$ and $D = 2c = 190 \, km^2/y$ (consistent with those used in [4]). The solver is then run using time frame $\mathcal{K}$ maps described above, down to 1 kya before present. The remaining plots (Figs 5, 6, 7, 8 and 9) display the resulting population dispersal simulation on space-time dependent $\mathcal{K}$ maps. Using population parameters optimized in Subsection **Parameter optimization** gives results consistent with the literature. The gross features of the late (45-50 kya) out-of-Africa dispersal of *Homo sapiens* are reproduced [32], e.g. the colonization of Western Europe before 40 kya and that of South America before 14 kya [31]. As the upper plot in the right-hand panel of Fig 3 shows, for these (r, D) parameters, arrival times increase as $T_s \geq 45$ increases. Figs 4, 5, 6, 7, 8 and 9 show our results graphically. Furthermore, the predictions of arrival times shown in the table of Table 2 are also quite reasonable considering the uncertainties involved in our crude NPP maps. The 2nd column of Table 2 also indicates archaeological uncertainties.
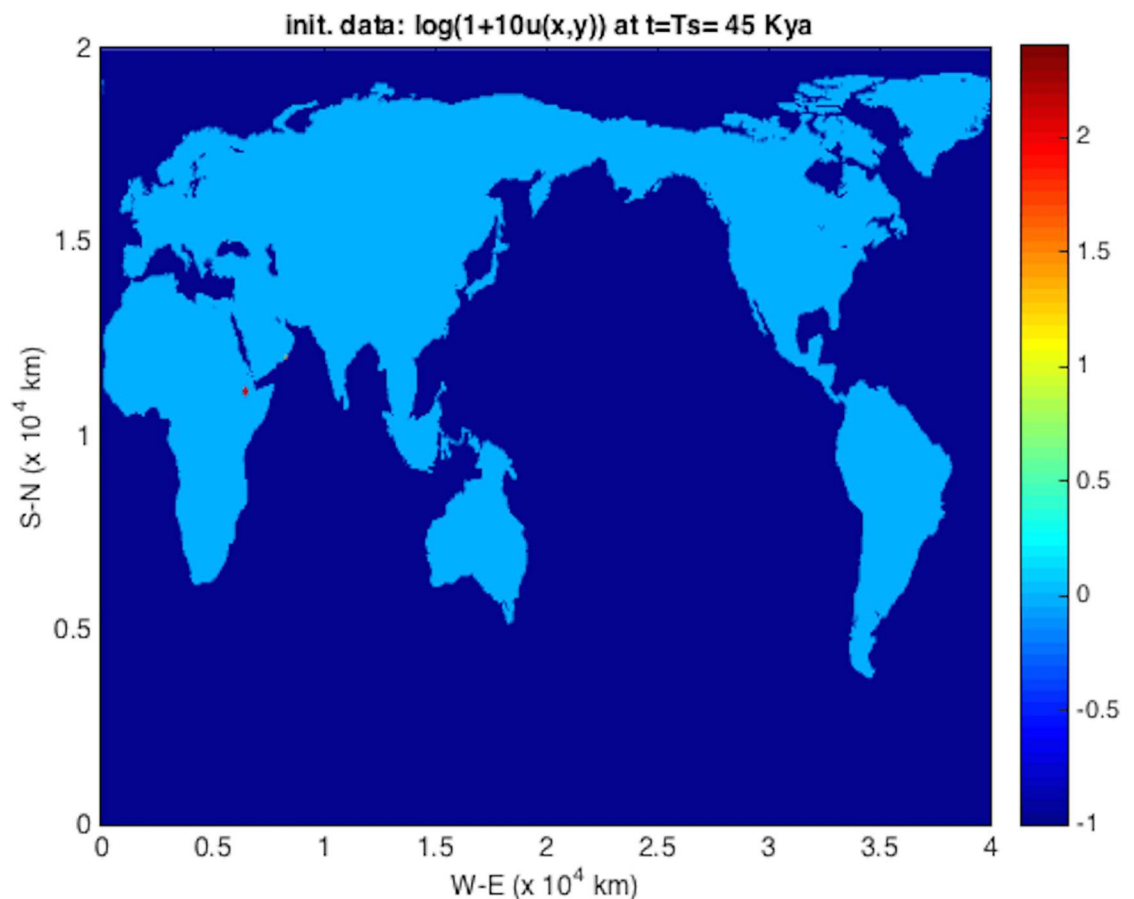
Fig 4. Color-coding and labeling for all of out-of-Africa dispersal plots. Dark blue shows *water*, with $\mathcal{K} = 0$. Growth rate is $\lambda = 2.04 \cdot 10^{-3} \, y^{-1}$, and diffusion coefficient $D = 190 km^2/y$. The initial distribution at $T_s = 45$ kya is denoted by a small dark red spot in Eritrea, west of the Bab al Mandab straight: a $\sigma = 3$ pixel Gaussian with peak u = local $\mathcal{K} = 0.72$, color scale 2.1.
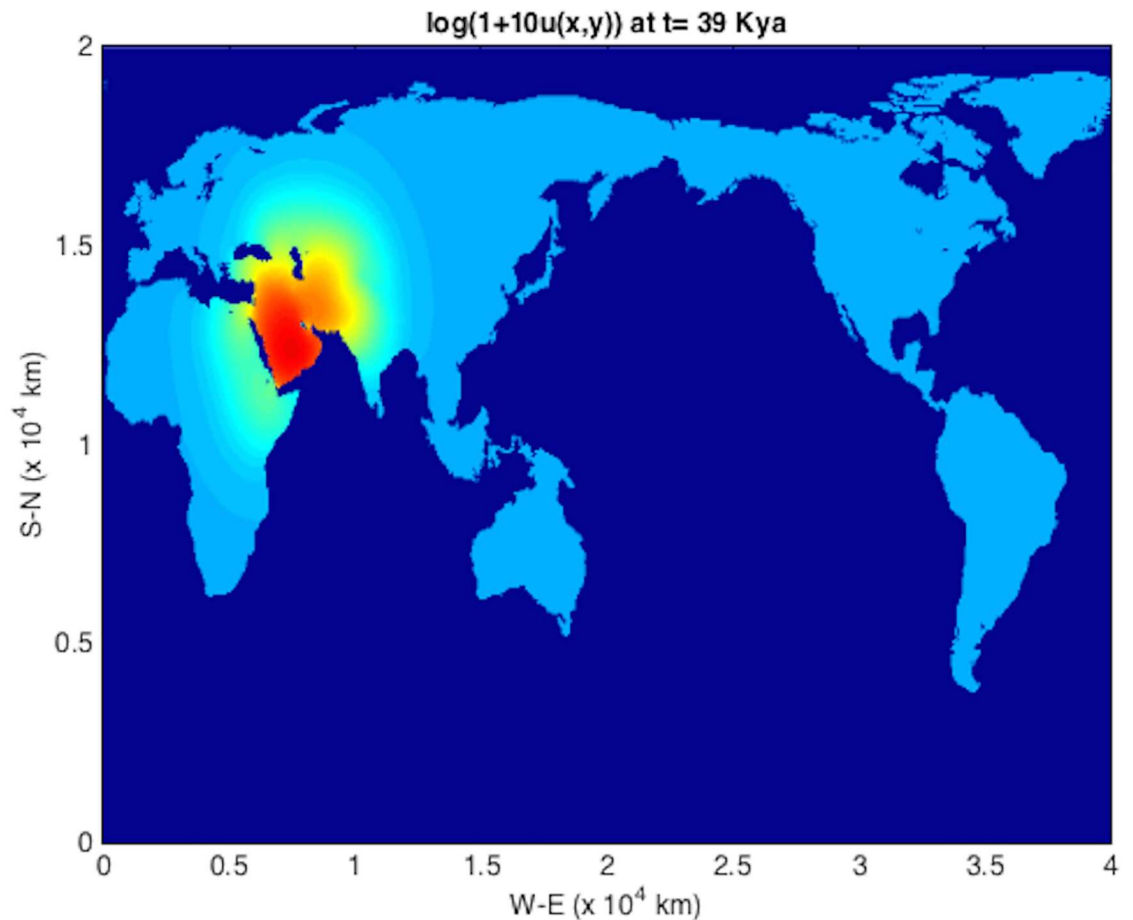
doi:10.1371/journal.pone.0167514.g004

Fig 5. log(1 + 10*u*) plot for out-of-Africa dispersal population distribution at *t* = 39 kya.

Conclusions

In this paper, we present a semi-implicit Godunov scheme for the Fisher/KPP equation with a variable carrying capacity $\mathcal{K}$. In one dimension, the expected traveling wave [10, 11] develops as shown in Fig 1. As in the one-dimensional situation, a traveling wave also develops as expected [11] in 2-D. Plots in Subsection **Population dispersal** clearly show these evolving waves.

In the Section **Space-time dependent capacity maps** we describe our procedure for handling **x** and time dependent $\mathcal{K}(\mathbf{x}, t)$, Eq (9).

Finally, we apply our scheme to our principal objective of population dynamics: the out-of-Africa dispersal of *Homo sapiens*. On the Mercator projected world map, by using vegetation net primal productivity (NPP) as a proxy for carrying capacity, we get Table 2 which shows that interpolating in time between discrete $\mathcal{K}-$ maps yields stable and reasonable results. These results show ancestor arrival in NE India before 40 kya, then a crossing of the Bering Strait before 10 kya. Multiple routes into South Asia [33] are also evident. Honesty requires that we admit our size $(56 \text{km})^2$ pixels do not resolve the two crossing points at Bab al Mandab and Sinai adequately. Additionally, we make no attempt to model coastlines effectively. Neither simple boats nor easier passages along beaches are included in our model: only Neumann
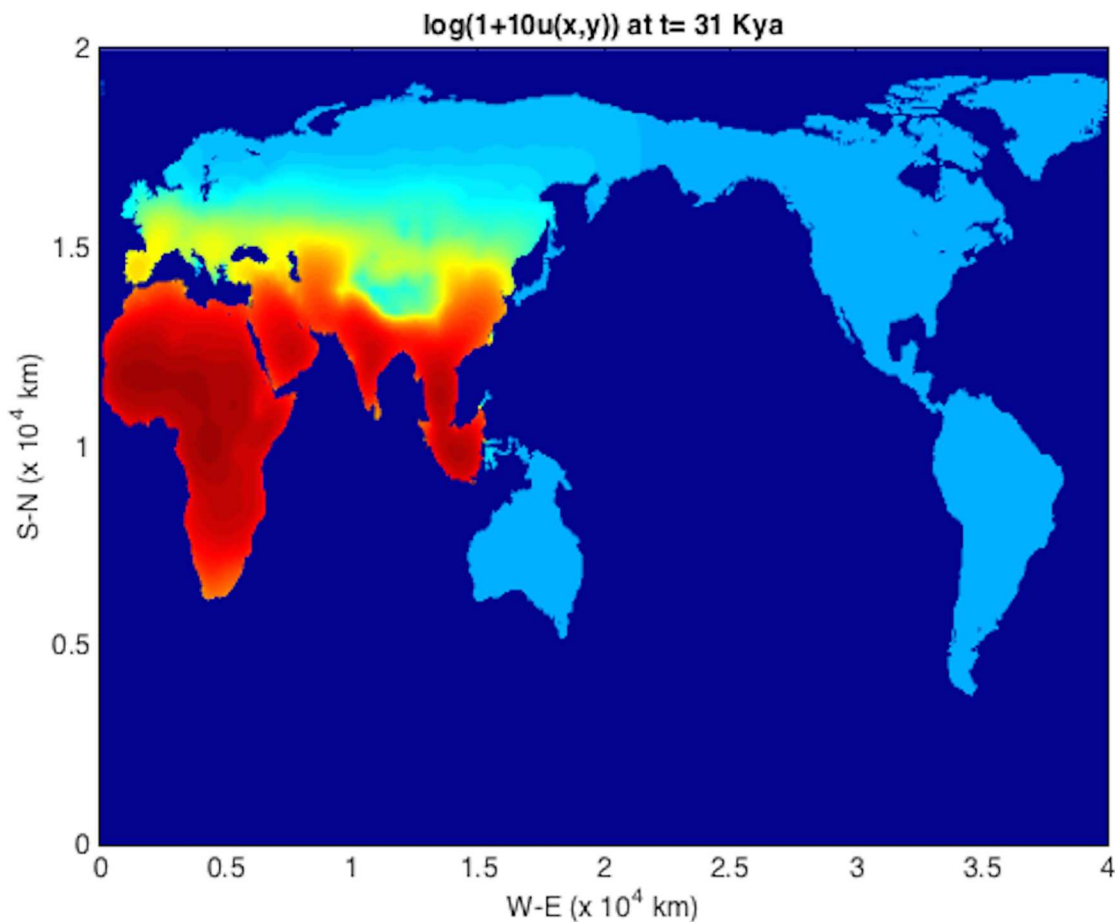
Fig 6. log(1 + 10*u*) plot for out-of-Africa dispersal population distribution at *t* = 31 kya.

boundary conditions determine such movements here. Better modeling of coastlines and more detailed NPP maps would improve our results.

The core computations performed by our solver are *independent* tridiagonal solutions, which can be easily parallelized to deal with larger grids (e.g. [34], p. 116). In order to improve numerical performance, in the Appendix **Map segmentation**, we discuss a compressed storage scheme to integrate the Fisher/KPP equation on a projected world map. About 71% of the earth's surface is water, so this compressed storage reduces computational work by the same amount.

Provenance

For this paper, the simulations were run on either a Mac Mini, 2.3 GHz Intel Core i7, or a 2.8 GHz i7 `MacBook` (both OS 10.9.5). On the Mini, `MatLab` R2015b was used, R2015a on the MacBook laptop. Our codes and maps are available [27].

Map segmentation appendix

Our solver on a geographical map uses a map outline, i.e., a rectangular grid with 1's in habitable regions, and 0's in the water, see Fig 10. Each Godunov direction step only involves
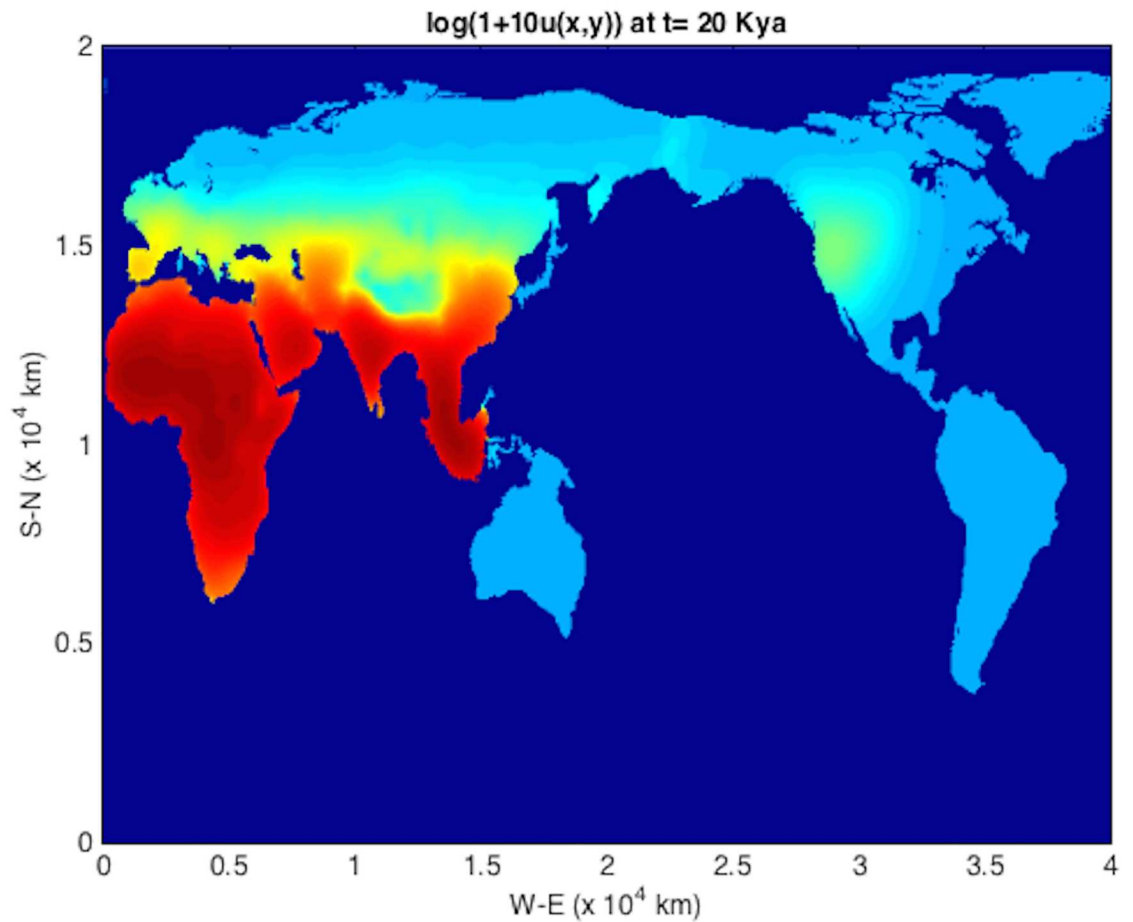
Fig 7. log(1 + 10u) plot for out-of-Africa dispersal population distribution at t = 20 kya. Notice that the bridge across the Bering Straight is not yet closed.

independent rows, resp. columns, indexed by starts/ends of habitable segments. Independent columns, $i = 1 \ldots NX$ in Fig 10 will have `nysegs(i)` of habitable segments, whose start and end positions are `ystart_seg(k)` and `yend_seg(k)`, respectively, where $k = 1 \ldots$ `nysegs(i)`. Likewise, for $j = 1 \ldots NY$ rows, each have `nxsegs(j)` also with start/end positions. A 100×50 example, Fig 10, shows row 26 has 4 segments of varying size. Column 87 has 5 segments.

Each time a new carrying capacity map is loaded (up to 61 total) [1, 22] a suitable mask is computed from $\mathcal{K}_L$ and $\mathcal{K}_H$, then resegmentation is performed: see Subsection **Time interpolation of maps**. Resegmentation facilitates time changing coastlines.

We include here an illustrative MatLab code sample of our Godunov-Strang-Yoshida scheme. Notation: `cfl` = Eq (4), carrying capacities `kL` at t, `kH` at $t + h$. Over multiple timesteps, `godunovstep1` which is used in the two half-steps, $h/2$, of Eq (9) can be doubled up: at step t_{n+1} Eq (9a) can be combined with Eq (9d) used at t_n for one full step, h. This is common for all trapezoidal rule integrations: for example, equation (2.1.4) in [35], or in a stochastic setting [36].
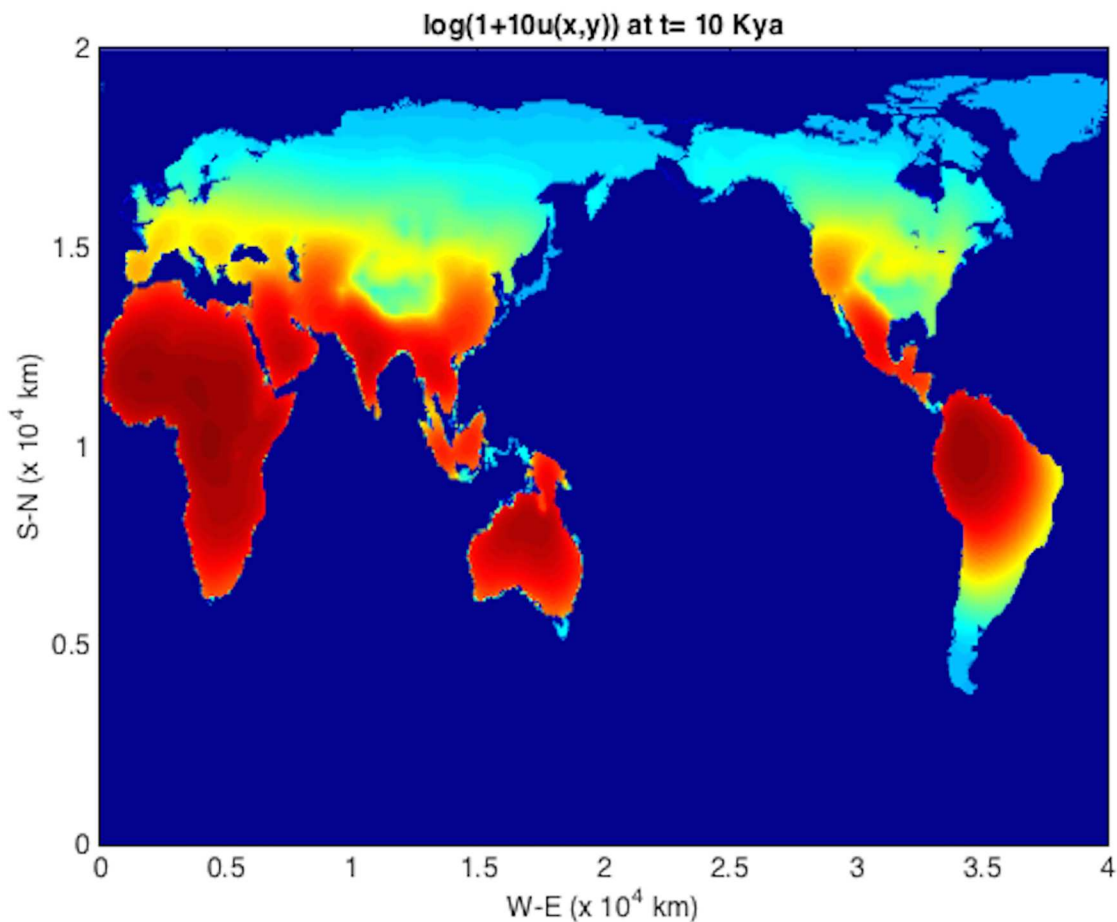
Fig 8. log(1 + 10*u*) plot for out-of-Africa dispersal population distribution at *t* = 10 kya. The bridge across the Bering Straight is now closed.

```
% NY X direction updates for 1/2-step1, Y direction for step2
  locx = 0;
  for j = 1:NY
      nsegs = nxsegs(j);
      for k = 1:nsegs
          istart = xstart_seg(locx+k); iend = xend_seg(locx+k);
          ninseg = iend-istart+1;
          u0(1:ninseg) = u(istart:iend,j);
% Eq (9a) solution, sc1, sc2 are static workspaces
          ut = godunovstep1(ninseg,h,cfl,u0,sc1,sc2);
          u(istart:iend,j) = ut(1:ninseg);
      end
      locx = locx + nsegs;
  end
```
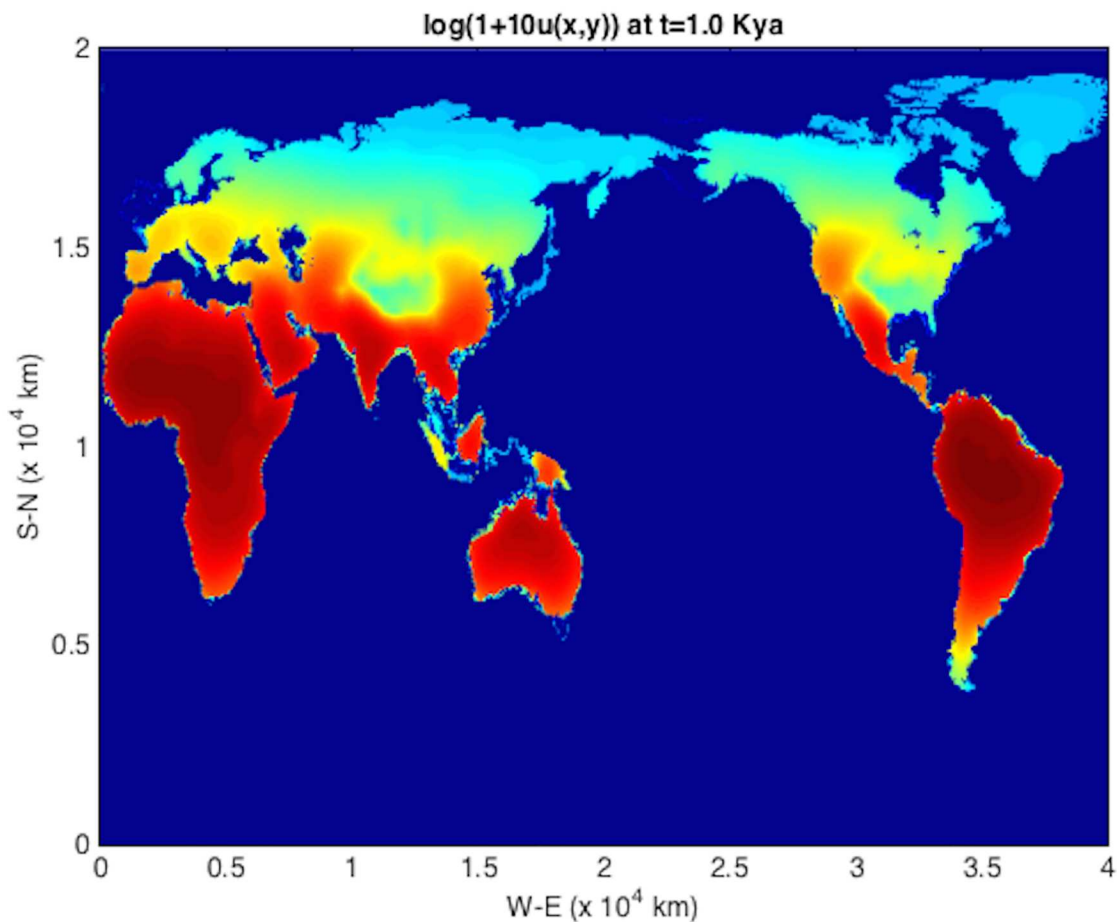
Fig 9. log(1 + 10u) plot for out-of-Africa dispersal population distribution at t = 1 kya.

```
% do NX Y dir. updates step2, wt0, wt1 = time interpolation weights
  locy = 0;
  for i = 1:NX
      nsegs = nysegs(i);
      for k = 1:nsegs
          jstart = ystart_seg(locy+k); jend = yend_seg(locy+k);
          ninseg = jend-jstart+1;
          u1(1:ninseg) = u(i,jstart:jend)';
          kL(1:ninseg) = w_L(i,jstart:jend)*(1-wt0) ...
                        + w_H(i,jstart:jend)*wt0;
          kH(1:ninseg) = w_L(i,jstart:jend)*(1-wth) ...
                        + w_H(i,jstart:jend)*wth;
% Eq (9c) solution:
          ut = godunovstep2(ninseg,h,cfl,u1,sc1,sc2,kL,kH);
          u(i,jstart:jend) = ut(1:ninseg);
      end
      locy = locy + nsegs;
  end
```
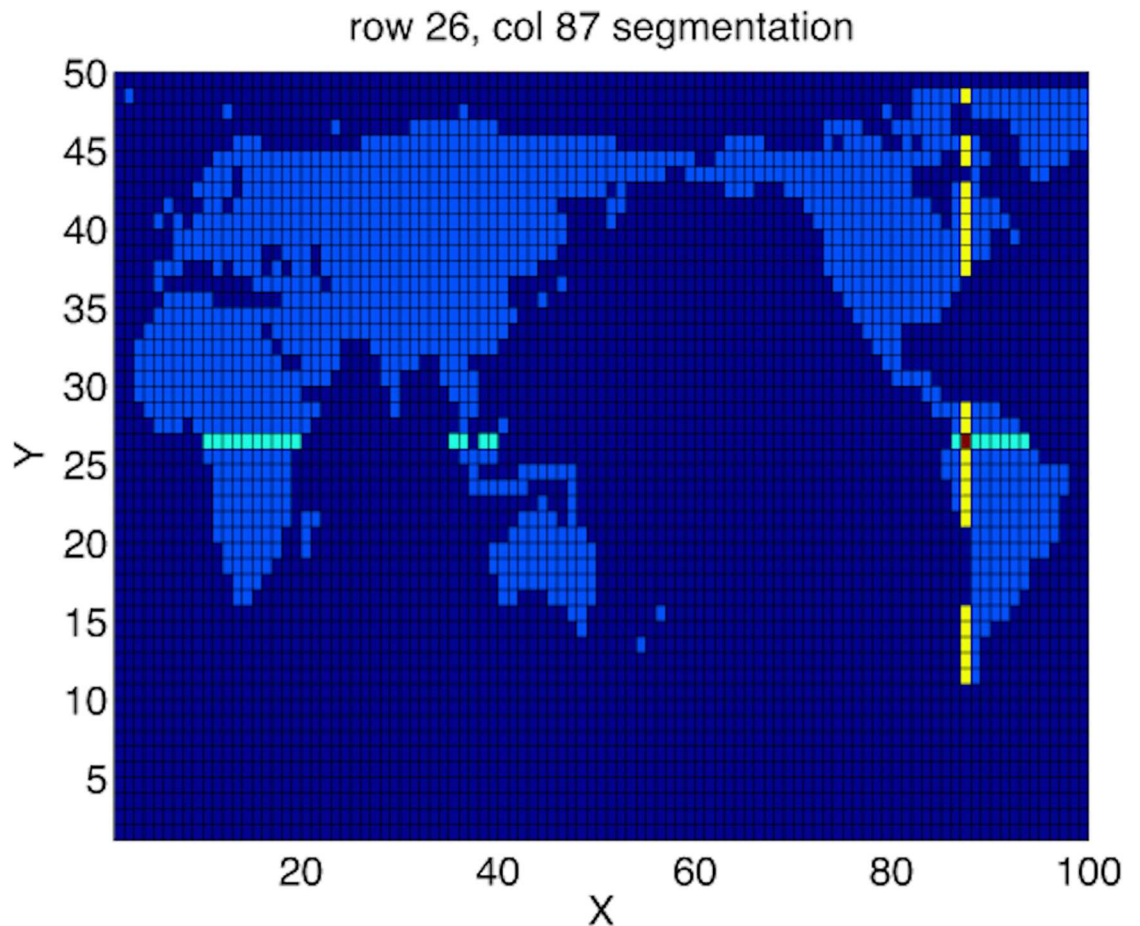
Fig 10. Habitable map regions. Land is set to 1 (light blue) and water 0 (dark blue) on a low-resolution $N_x = 100$, $N_y = 50$ grid. Row 26 has 4 habitable segments (cyan), column 87 has 5 segments (yellow).

doi:10.1371/journal.pone.0167514.g010

```
% Repeat godunovstep1, as above for Eq (9d)
% locx = 0;
% for j = 1:NY
%       ETC
% end
```

Acknowledgments

Author Contributions

Conceptualization: WPP CPEZ.

Data curation: WPP.

Formal analysis: WPP.

References

1. Eriksson A, Betti L, Friend AD, Lycett SJ, Singarayer JS, von Cramon-Taubadel N, et al. Late Pleistocene climate change and the global expansion of anatomically modern humans. Proceedings of the National Academy of Sciences. 2012; 109(40):16089–16094. doi: 10.1073/pnas.1209494109

2. Henn BM, Cavalli-Sforza L, Feldman MW. The great human expansion. Proceedings of the National Academy of Sciences. 2012; 109(44):17758–17764. doi: 10.1073/pnas.1212380109

3. Nikitas P, Nikita E. A study of hominin dispersal out-of-Africa using computer simulations. J of human evolution. 2005; 49(5):602–617. doi: 10.1016/j.jhevol.2005.07.001

4. Young DA, Bettinger RL. Simulating the global human expansion in the late Pleistocene. J of Archaeological Science. 1995; 22(1):89–92. doi: 10.1016/S0305-4403(95)80165-0

5. Callegari S, Weissmann JD, Tkachenko N, Petersen WP, Lake GR, Ponce de León M, et al. An agent-based model of human dispersals at a global scale. Advances in Complex Systems. 2013; 16(4 and 5) 1350023–21 pages. URL: https://doi.org/10.1142/S0219525913500239. doi: 10.1142/S0219525913500239

6. Bouckaert R, Lemey P, Dunn M, Greenhill SJ, Alekseyenko AV, Drummond AJ, et al. Mapping the origins and expansion of the Indo-European language family. Science. 2012; 337(6097):957–960. doi: 10.1126/science.1219669 PMID: 22923579

7. Manica A, Amos W, Balloux F, Hanihara T. The effect of ancient population bottlenecks on human phenotypic variation. Nature. 2007; 448(7151):346–348. doi: 10.1038/nature05951 PMID: 17637668

8. Pinhasi R, Fort J, Ammerman AJ. Tracing the origin and spread of agriculture in Europe. PLoS Biology. 2005; 3(12):e410. doi: 10.1371/journal.pbio.0030410 PMID: 16292981

9. Relethford J. Global patterns of isolation by distance based on genetic and morphological data. Human Biology. 2004; 76(4):499–513. doi: 10.1353/hub.2004.0060 PMID: 15754968

10. Fisher RA. The wave of advance of advantageous genes. Annals of Eugenics. 1937; 7(4):355–369. doi: 10.1111/j.1469-1809.1937.tb02153.x

11. Kolmogorov A, Petrovsky I, Piskunov N. Etude de l'équation de la diffusion avec croissance de la quantité de matiere et son applicationa un probleme biologique. Moscow Univ Bull Math. 1937; 1:1–25.

12. Skellam JG. Random dispersal in theoretical populations. Biometrika. 1951; 38(1/2):196–218. doi: 10.2307/2332328 PMID: 14848123

13. Reichl LE. A modern course in statistical physics. Wiley-VHC Verlag Gmbh, KGaH, Weinheim; 2004.

14. Murray JD. Mathematical Biology. Springer-Verlag, New York; 2002.

15. Bramson M. Convergence of solutions of the Kolmogorov equation to traveling waves. vol. 44. American Mathematical Society; 1983. Number 285 (end of volume).

16. Strang G, Aarikka K. Introduction to applied mathematics. vol. 16. Wellesley-Cambridge Press, Wellesley, MA; 1986.

17. LeVeque RJ. Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems. SIAM, Philadelphia, PA; 2007.

18. Godunov SK, Ryabenkii VS. Difference schemes: an introduction to the underlying theory. Elsevier, New York; 1987.

19. Gear CW. Numerical initial value problems in ordinary differential equations. Prentice Hall Inc., Englewood Cliffs, NJ; 1971.

20. Grieser J, Gommes R, Cofield S, Bernardi M. World maps of climatological net primary production of biomass, NPP. Viale delle Terme di Caracalla, 00153, Rome, Italy: The Agromet Group, SDRN Food and Agriculture Organization of the UN; 2006. URL: http://www.fao.org.

21. Lieth HH, Whittaker RR, et al. Primary productivity of the biosphere. Springer-Verlag, New York; 1975.

22. Bridge Project, Univ. of Bristol; 2003. URL: http://www.bridge.bris.ac.uk.

23. Bigelow NH, Brubaker LB, Edwards ME, Harrison SP, Prentice IC, Anderson PM, et al. Climate change and Arctic ecosystems: 1. Vegetation changes north of 55 N between the last glacial maximum, mid-Holocene, and present. J of Geophysical Research. 2003; 108(D19):8170. doi: 10.1029/2002JD002558

24. Harrison S, Yu G, Takahara H, Prentice I. Palaeovegetation (Communications arising): diversity of temperate plants in east Asia. Nature. 2001; 413(6852):129–130. doi: 10.1038/35093166 PMID: 11557970

25. Pickett EJ, Harrison SP, Hope G, Harle K, Dodson JR, Peter Kershaw A, et al. Pollen-based reconstructions of biome distributions for Australia, Southeast Asia and the Pacific (SEAPAC region) at 0, 6000 and 18,000 14C yr BP. J of Biogeography. 2004; 31(9):1381–1444. doi: 10.1111/j.1365-2699.2004.01001.x

26. Prentice IC, Jolly D. Mid-Holocene and glacial-maximum vegetation geography of the northern continents and Africa. J of Biogeography. 2000; 27(3):507–519. doi: 10.1046/j.1365-2699.2000.00425.x

27. Provenance: maps and codes; 2016. URL: http://www.zora.uzh.ch/128296/ or https://doi.org/10.5176/uzh-128296.

28. Wirtz K, Lemmen C. A global dynamic model for the neolithic transition. Climate Change. 2003; 59 (3):333–367. doi: 10.1023/A:1024858532005

29. Burenhult G. Die ersten Menschen die Ursprünge des Menschen bis 10000 vor Christus. Bechtermünz, Augsburg; 2000.

30. Lavallée D. The first South Americans: the peopling of a continent from the earliest evidence to high culture. University of Utah Press; 2000.

31. Halligan JJ, Waters MR, Perrotti A, Owens IJ, Feinberg JM, Bourner MD, et al. Pre-Clovis occupation 14,550 years ago at the Page-Ladson site, Florida, and the peopling of the Americas. Sci Advances. 2016;13 May, electronic: E1600375:1–8.

32. Forster P. Ice Ages and the mitochondrial DNA chronology of human dispersals: a review. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences. 2004; 359(1442):255–264. doi: 10.1098/rstb.2003.1394 PMID: 15101581

33. Reyes-Centento H, Ghirotto S, Detroit F, Grimaud-Herve D, Barbujani G. Genomic and cranial phenotype data support multiple modern human dispersals from Africa and a southern route into Asia. Proceedings of the National Academy of Sciences. 2014; 111(20):7248–7253. URL: https://doi.org/10.1073/pnas.1323666111. doi: 10.1073/pnas.1323666111

34. Petersen WP, Arbenz P. Introduction to parallel computing. Oxford Univ. Press; 2004.

35. Davis PJ, Rabinowitz P. Methods of Numerical Integration. Academic Press, Orlando, Florida; 1984.

36. Perret C, Petersen WP. A weak second-order split-step method for numerical simulations of stochastic differential equations. BIT, Numerical Mathematics. 2014;(54):801–821. doi: 10.1007/s10543-014-0482-4