

Spatial scales in human movement between reservoirs of infection

Robert J. Hardwick^{a,b,c,*}, Carolin Vegvari^{a,c}, Benjamin Collyer^{a,c}, James E. Truscott^{a,b,c}, Roy M. Anderson^{a,b,c}

^aLondon Centre for Neglected Tropical Disease Research (LCNTDR), Department of Infectious Disease Epidemiology, St. Marys Campus, Imperial College London, London WC2 1PG, UK

^bThe DeWorm3 Project, The Natural History Museum of London, London, UK

^cMRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London, UK

Abstract

The life cycle of parasitic organisms that are the cause of much morbidity in humans often depend on reservoirs of infection for transmission into their hosts. Understanding the daily, monthly and yearly movement patterns of individuals between reservoirs is therefore of great importance to implementers of control policies seeking to eliminate various parasitic diseases as a public health problem. This is due to the fact that the underlying spatial extent of the reservoir of infection, which drives transmission, can be strongly affected by inputs from external sources, i.e., individuals who are not spatially attributed to the region defined by the reservoir itself can still migrate and contribute to it. In order to study the importance of these effects, we build and examine a novel theoretical model of human movement between spatially-distributed focal points for infection clustered into regions defined as ‘reservoirs of infection’. Using our model, we vary the spatial scale of human movement defined around focal points and explicitly calculate how varying this definition can influence the temporal stability of the effective transmission dynamics — an effect which should strongly influence how control measures, e.g., mass drug administration (MDA), define evaluation units (EUs). Considering the helminth parasites as our main example, by varying the spatial scale of human movement, we demonstrate that a critical scale exists around infectious focal points at which the migration rate into their associated reservoir can be neglected for practical purposes. This scale varies by species and geographic region, but is generalisable as a concept to infectious reservoirs of varying spatial extents and shapes. Our model is designed to be applicable to a very general pattern of infectious disease transmission modified by the migration of infected individuals between clustered communities. In particular, it may be readily used to study the spatial structure of hosts for macroparasites with temporally stationary distributions of infectious focal point locations over the timescales of interest, which is viable for the soil-transmitted helminths and schistosomes. Additional developments will be necessary to consider diseases with moving reservoirs, such as vector-borne filarial worm diseases.

Keywords:

Mathematical models, Spatial infection model, Control policies, Monitoring and Evaluation

1. Introduction and background

Defining the spatial scales over which transmission should be considered is a relatively recent research area in the context of neglected tropical diseases (NTDs), where much of the effort has been focused on the development of geostatistical methods for each disease in turn [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. In particular, mathematical models of helminth transmission have only recently begun to incorporate the dynamical effects of human movement between reservoirs of infection [9, 12, 13, 14].

Helminth infections (or helminthiasis) are a class of macroparasitic diseases which include, among others, the soil-transmitted helminth (STH) infections, schistosomiasis and lymphatic filariasis (LF). Following World Health Organisation (WHO) guidelines, free drugs have been donated by pharmaceutical companies since 2010 [15, 16, 17, 18, 19] to countries significantly affected by helminth-associated morbidity. Control initiatives have been developed specifically to investigate

the prospects for elimination as a public health problem in the long term [20, 21].

In Fig. 1 we have illustrated the essential role that infectious focal points have on driving the transmission of human helminth infections. In the case of STHs, these focal points consist of eggs, or larval stages in the soil, that are either ingested or enter the body via skin penetration. Similarly, in the case of schistosomiasis transmission, these focal points exist in water sources where larvae, that are released by freshwater snails (intermediate hosts), penetrate the skin to infect human hosts.

It may be possible to minimise the effect human movement has on evaluation units (EUs) of, e.g., mass drug administration (MDA) control programmes, by specifying a critical spatial scale around the focal points within a reservoir (and corresponding to a human community) over which one may sample and treat infections. At this scale (or larger), the number of migrants contributing to the reservoir per unit time is potentially low enough such that their effect may essentially be neglected. If such a scale exists it would provide a natural level of res-

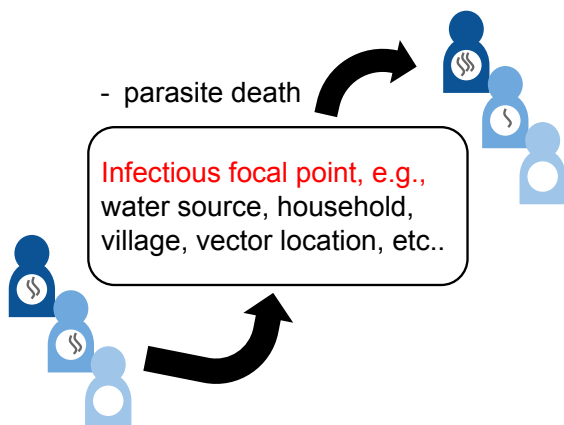


Figure 1: Illustration of the transmission dynamics of human helminth infections via a defined focal point.

olution for those generating spatial maps of the prevalence of infections in a given region. Note that for some NTDs, implementation units (IUs) and evaluation units (EUs) differ. In this paper, we shall look at the spatial scale of EUs because there is currently no evidence-based framework for choosing an appropriate spatial scale at which programme managers can evaluate the impact of NTD programmes. We will demonstrate the existence of this scale, which is expected to vary between helminths and regions, by developing a spatial model of human movement between defined locations which is consistent with the current literature.

The distances travelled by human movement patterns are not purely random, but exhibit certain known characteristics similar to those of known random walks, e.g., the same heavy-tailed distributions as those of Lévy flights [22, 23]. Furthermore, in a variety of countries and levels of urbanisation, Ref. [24] uses mobile phone data in many different countries to conclude that there is evidence for some universality in the distribution shape of daily work-home commute distances, which appear to fit a broken power-law. Note that the distributions are similar to, e.g., those observed for radial distance in Refs. [25, 26], where these works make use of geolocated tweets and GPS data, respectively. The findings in all cases above, and other examples of broken power-laws for human mobility in the NTD disease modelling literature (see, e.g., Refs. [27, 28]), will motivate us to use a similar description to develop a simple model of human movement from households to focal points of infection in this work.

In Sec. 2, we derive a broken power-law model with one jump per individual, which both replicates the distribution behaviour and also suggests a possible mechanism for its origin. The model we introduce, however, will remain flexible (by varying parameters) to other possible power-law descriptions for human mobility and so we will not be wholly reliant on one possible description. In Sec. 3 we briefly investigate a computationally-efficient extension to this model which uses an approximation to describe multiple jumps per individual for movement patterns with longer timescales — both of which might be important for an accurate representation of the rele-

vant human behaviour.

Having developed our model and illustrated its possible extensions, in Sec. 4 we apply it to obtain a critical spatial scale at which the migration of individuals to focal points clustered as infectious reservoirs declines to a negligible level. This result is shown to vary with helminth species, human movement patterns and regional geometry of settlement patterns. The analysis also provides insight into the most important pieces of (often missing) information which are necessary to build an accurate model of human migration patterns between reservoirs of infection. Collecting such information will be key to the success of future helminth control programmes in reaching their targets, e.g., achieving STH elimination as a public health problem [29]. Lastly, in Sec. 5, we discuss how the spatial extent of reservoirs of infection considered here should influence the scale of EUs for control programmes and conclude with a summary of our findings and prospects for future work.

2. A one-jump model

2.1. Locations of infectious focal points and households

There are important ethical implications to the public availability of data on both the spatial locations and movement patterns of individuals over various scales in time, especially in low- and middle-income countries [30]. Among these are data privacy and the potential identifiability of individuals through the use of ‘big data’ by governments, companies and universities, and the potential use of this information for non-research purposes. Therefore, access to such high-quality data for study is scarce. This represents a challenge for those concerned with the various consequences of such movements and their methods of model validation. In this paper, we present mathematical models that can be used in the absence of access to high-quality movement data. Model dynamics are implemented on the spatial patterns of, e.g., building locations, obtained from real-world datasets, e.g., the high resolution settlement layer dataset generated by the Facebook Connectivity Lab [31]. Our methodology should allow researchers to draw useful conclusions on the impact of human spatial movement on NTD programmes from readily available geospatial data (coordinates of households, settlements and environmental features, such as water bodies). Other data sources, e.g., call data records (from mobile phones) and data from migration questionnaires in control studies, may be available under certain circumstances and can easily be integrated in our modelling framework. The models we introduce here should also be able to take this specific case-study information into account by appropriate parameter inference.

Let us first consider the embedding of spatial locations of households as points in a 2-dimensional flat space. In addition to these household points there will also exist focal points of infectious contact that depend on the helminth species, e.g., near buildings, water sources, parts of farmland with large concentrations of infectious material and households themselves. Choosing a particular focal point, we will use the observed spatial distributions of buildings in the high resolution settlement layer dataset [31] as a proxy for the spatial distribution

of household locations around it. Therefore, the expectation of having $n_{<r}$ nearby households or buildings within a radial distance of r from the focal point, i.e., $E(n_{<r})$, scales according to the following radially-dependent power law

$$E(n_{<r}) \propto r^{\alpha(r)}. \quad (1)$$

It is important to note that, since at first we consider only a single focal point from which household distances is measured, the distribution of focal points themselves is not relevant to the calculation. The spatial distribution of focal points will, however, become relevant later on when we cluster them into reservoirs of infection and investigate their spatial extent. Note, however, that households themselves may be a reasonable tracer of the underlying distribution for infectious focal points of STH [32, 33], though our calculations throughout will not depend on this possibility.

In the cases of STH and schistosomiasis transmission, although it is clear that the location and importance of their infectious focal points can vary over time — e.g., a chance decline in snail population for a particular location at the edge of a fresh water source leads to a drop in infectious spread — throughout this work, we shall assume that their spatial distribution is effectively temporally stationary over the timescales of interest. Note that for some NTDs, the infectious contact event locations can also be mobile, e.g., in the case of lymphatic filariasis transmission, where the larval stage of microfilariae enter the body by bites from infected mosquitoes — hence, the contact event locations will migrate with the human and mosquito populations.

By averaging over the observed point spread patterns in the high resolution settlement layer dataset, we obtain the following cumulative neighbouring point number distributions as a function of radial distance. This can be compared to the radial dependence of the power-law index $\alpha(r)$ we have quoted in Eq. (1). In order to give an indication of the effect of spatial heterogeneity in the distances between buildings for this averaged quantity, we have plotted some example lines (left column of plots) and the sample mean lines with root-mean-square deviation (RMSD) shaded regions of $\alpha(r)$ (right column of plots) from individual randomly sampled initial building locations for different countries in Fig. 2. In Central Malawi and Northern Benin, up to some substantial variance, below a transitioning radial scale $r < r_{tr}$, the index seems to have a mean of $\alpha \simeq 1$ and above this scale $r > r_{tr}$ the mean of the index begins a transition to $\alpha \rightarrow 2$. In the Central Ivory Coast, we see that the mean trend rapidly varies between $\alpha \simeq 1.5$ and 0.5 below the transition scale $r < r_{tr}$, and then at radial distances much greater than this scale $r \gg r_{tr}$ most of the power-laws appear to begin a transition to $\alpha \rightarrow 2$. The substantial variation below the transition scale in the Ivory Coast is likely due to a greater degree of clustering on particularly small scales (note the drop towards $\alpha \simeq 0$ before the transition scale indicates this as very few new points are added to the cumulative total for this range).

For all countries and initial locations in Fig. 2, we can confirm that the majority of the power-laws begin to converge to a mean of $\alpha \rightarrow 2$ above a particular transition scale r_{tr} which

varies substantially between location. This convergence reflects the emergence of statistical spatial homogeneity only on very large spatial scales, but this is clearly not present below r_{tr} in all cases. The power-law signature of statistical homogeneity being $\alpha = 2$ should become clear by noting that another way to generate it is to draw a Poisson point process with an intensity $rdr \propto r^2$ (the flat Euclidean measure in 2 dimensions). Note that statistical isotropy is also implicitly confirmed by choosing different $r = 0$ locations for each line and observing the same behaviour up to some noise that is captured by the RMSD in the right column of plots. Deviations from this trend towards $\alpha = 2$ (particularly in the Ivory Coast) are also present in many other locations in the same countries, due to the presence of substantial global heterogeneity, e.g., through proximity to large voids in the distribution of points, which exist for geographic reasons.

2.2. Obtaining a model using maximum entropy arguments

In this section we provide a simple argument to generate the broken power-law frequency distribution of work-home commute distances (as observed and described in, e.g., Kung et al. [24] and Zhao et al. [26]) using maximum entropy arguments. The distributions of movement distances we derive will be used to model the migration of individuals from their household locations to a focal point of infection. In the absence of access high-quality site-specific data that we discussed in the previous section, we are using the work-home commuter distance distributions as a guide for the likely patterns of movement. Work-home commuter distances should be a reasonable proxy for the overall movement of individuals in many areas. However, the model that we develop in this section is adaptive and flexible to modification, even in circumstances where this proxy is inappropriate, we will be able to vary distribution parameters to quantify their impact on our conclusions.

For greater analytic insight in the expressions we derive in this section, and throughout most of the subsequent arguments made in this article, we shall now assume that Eq. (1) has a constant index $\alpha(r) = \alpha$. When considering the heterogeneity in the real power-law behaviour below r_{tr} in Fig. 2, this assumption will not be appropriate in all locations/situations. Note, however, the results we obtain in all cases may be easily numerically generalised to take into account complex transitions between different power-law behaviours, as are observed in Fig. 2. In addition, choosing several key values for constant α (which fit within the range of observed power law behaviours) in what follows will aid with quantifying its impact on our conclusions.

Assuming that each individual jumps to a single new location and then returns back to their starting position over the course of each day, we may follow a simple argument that incidentally generates a distribution over travel distances which is consistent with those observed and described in, e.g., Kung et al. [24] and Zhao et al. [26]. Let us take the distribution of neighbouring points to be isotropic. Defining $p(r)$ as the probability density function (PDF) of an individual jumping to a given location as a function of its radial distance, one may then transform the flat radial probability measure $p(r)dr$ by a Jaco-

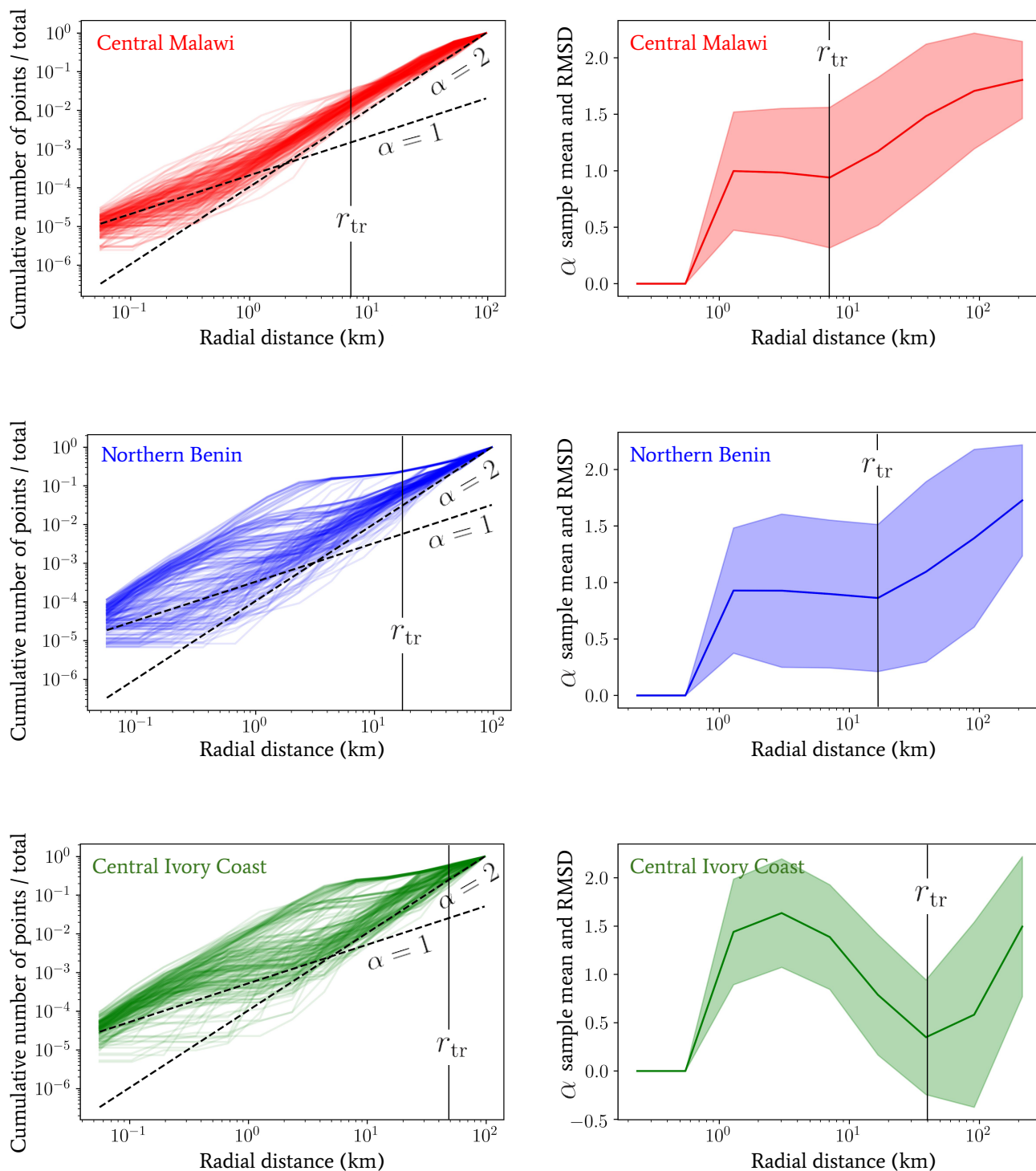


Figure 2: *Left column*: The cumulative number of buildings (vertical axis) below a given radial distance (horizontal axis) from sample locations in Malawi, Benin and the Ivory Coast obtained from the high resolution settlement layer dataset generated by the Facebook Connectivity Lab [31]. Lines for different constant α and r_{tr} values are provided only as rough indicators for comparison. *Right column*: The sample mean lines and root-mean-square deviation (RMSD) shaded regions of the power law index α plotted against radial distance in each case. The RMSD was calculated by computing discrete derivatives between log-frequency bins as a function of log-radial distance. In each country, the collection of sample locations was randomly drawn from: the central region of Malawi, including Lilongwe; the wide northern region of Benin, spanning between Kandi and Djougou; and a wide central region of the Ivory Coast, including Yamoussoukro.

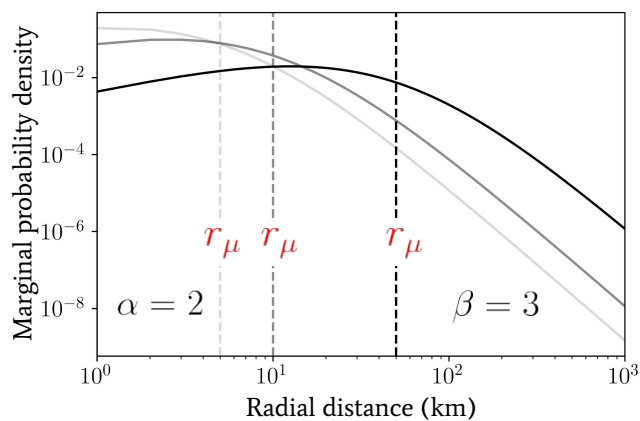
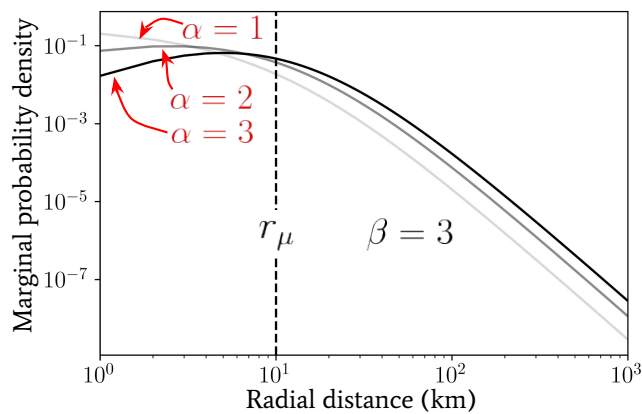
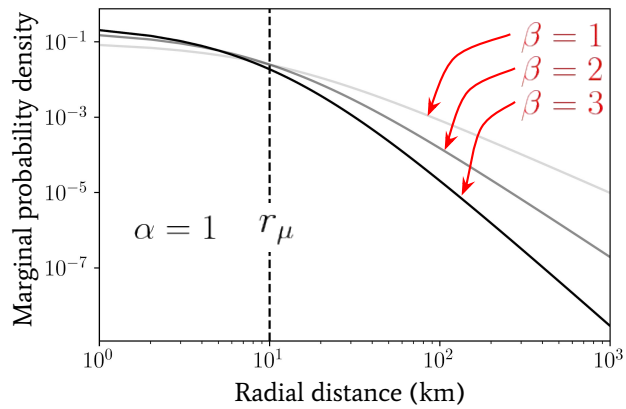


Figure 3: Plots of the marginal probability density $p(r; \alpha, \beta, r_\mu)$ as a function of radial distance (in units of km) generated using Eq. (5) for a range of α, β and r_μ values for comparison.

bian $v(r) = dE(n_{<r})/dr$ like so

$$p(r)dr \rightarrow p(r)v(r)dr \propto p(r)r^{\alpha-1}dr, \quad (2)$$

which accounts for the power-law density of points that may depart from linear scaling according to Eq. (1). Another way to view this rescaling is to consider the expectation of the probability measure for a radially-generated, statistically isotropic Poisson point process.

Note that if we additionally assume that the PDF of the jump distribution for each individual has a mean σ then the maximum entropy PDF is that of an exponential distribution $p(r) = p(r; \sigma) = \text{Exp}(r; 1/\sigma)$, then by using Eq. (2) we can trivially demonstrate that the resulting normalised PDF is a Gamma distribution

$$\begin{aligned} p(r; \alpha, \sigma) &= \frac{v(r)\text{Exp}(r; 1/\sigma)}{\int_0^\infty v(r)\text{Exp}(r; 1/\sigma) dr} \\ &= \text{Gamma}(r; \alpha, 1/\sigma). \end{aligned} \quad (3)$$

One possible interpretation for σ here is a number, with physical units of distance, which quantifies the access of that individual to varying degrees of transportation methods that are themselves capable of travelling varying distances. To clarify, without access to any motorised transportation, an individual may receive a low value for σ , and in contrast, if an individual has access to a vehicle then they may receive a larger σ value. Despite this simple suggested explanation, the presence of variable scales for human travel distance distributions is not contingent on access to vehicular transportation at all. Indeed, human walking patterns themselves are known to exhibit multi-distance-scale characteristics in many contexts — see, e.g., Ref. [34].

Assuming that $1/\sigma$ values may vary across the population of individuals with a distribution of known mean $E(1/\sigma)$ and expected scaling $E(\ln \sigma)$, the maximum entropy distribution for values of $1/\sigma$ is another gamma distribution. Choosing specific parameter values, we therefore have

$$p(1/\sigma) = \text{Gamma}(1/\sigma; \beta, r_\mu), \quad (4)$$

where the resulting marginal jump PDF is obtained by integration in the following way

$$\begin{aligned} p(r; \alpha, \beta, r_\mu) &= \int_0^\infty d(1/\sigma) \text{Gamma}(r; \alpha, 1/\sigma) \text{Gamma}(1/\sigma; \beta, r_\mu) \\ &= \int_0^\infty d(1/\sigma) \frac{r^{\alpha-1} e^{-r/\sigma}}{\Gamma(\alpha) \sigma^\alpha} \frac{r_\mu^\beta (1/\sigma)^{\beta-1} e^{-r_\mu(1/\sigma)}}{\Gamma(\beta)} \\ &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{r_\mu^\beta r^{\alpha-1}}{(r_\mu + r)^{\alpha+\beta}} \equiv \text{Beta}'(r/r_\mu; \alpha, \beta, 1, r_\mu), \end{aligned} \quad (5)$$

where $\text{Beta}'(\cdot; \cdot, \cdot, \cdot, \cdot)$ denotes the generalised beta prime distribution (or beta distribution of the second kind), which takes a form similar to that of a beta distribution. Note that, to avoid a divergent integral, one must specify that $\alpha, \beta, r_\mu \in \mathbb{R}^+$.

239 The main motivations for using maximum entropy proba-
 240 bility distributions in the case above (and in general) include
 241 the beneficial properties that such distributions have when lim-
 242 ited *a priori* information is available. In the present instance,
 243 these properties translate into distributions which account for
 244 the maximum amount of potential variability in outcomes while
 245 still satisfying the basic pre-defined classifying properties, e.g.,
 246 the gamma distribution we have used above is maximally entrop-
 247 ic under the condition that the mean and expected scaling of
 248 the distribution are known. Furthermore, it is often reasonable
 249 to expect that the steady states of many systems observed across
 250 the sciences are well-described by maximum entropy configura-
 251 tions.

252 We introduced another radial spatial scale r_μ (distinct from r_{tr}
 253 in the previous section) in Eq. (4) whose effect, in conjunction
 254 with α and β , we shall explore. In Fig. 3 we have plotted the
 255 marginal jump PDF given by Eq. (5) for a range of parameter
 256 values α, β and r_μ . By inspection of Eq. (5), in the limit where
 257 $r \ll r_\mu$ the distribution scales as $\sim r^{\alpha-1}$, whereas in the oppo-
 258 site limit $r \gg r_\mu$ the distribution exhibits a scaling $\sim r^{-\beta-1}$. The
 259 scale of r_μ therefore plays the role of separating two regimes
 260 in the distribution of jump lengths for a given individual. For
 261 radial distances below r_μ , the geometric distribution of avail-
 262 able spatial locations to jump to (which is encoded in the α pa-
 263 rameter) dominates the behaviour, whereas, for radial distances
 264 above r_μ , the predisposition of individuals to travel a given
 265 distance combines with the geometry of points to give the distri-
 266 bution behaviour and this is encoded in the β parameter of the
 267 second gamma distribution in Eq. (4).

268 So far we have *not* assumed that the parameter r_μ in the distri-
 269 bution over inverse-jump scales $1/\sigma$ given by Eq. (4) is related
 270 to any purely geometrically-driven break in the power laws of
 271 Fig. 2. By comparing our observed power laws for the Ivory
 272 Coast with the observed work-home commuter distance distri-
 273 butions in Ref. [24], for instance, we see that $r_\mu \ll r_{tr}$ in general.
 274 This is perhaps unsurprising since the scale r_μ may arise due to
 275 the availability of easy access to vehicles that allow for longer-
 276 distance travel for either work or school, or water sources near
 277 a given settlement. Despite this fact, it is intriguing to note that
 278 the radial scale associated to the peak in clustering ($\alpha \approx 1.5$
 279 well below r_{tr}) for the Ivory Coast dataset in Fig. 2 occurs at
 280 the same rough scale that is observed for r_μ in Ref. [24]. We
 281 leave the investigation into the potential link between these two
 282 scales for future work.

283 Despite statistical isotropy and homogeneity on large spatial
 284 scales, we have observed that the variation between the power-
 285 laws for cumulative building numbers observed in Fig. 2, and
 286 hence the α index of Eq. (1), exhibits a large degree of location
 287 dependence between regions. In Fig. 4, we see the effect that
 288 this small-scale heterogeneity in the distances between build-
 289 ings has on the probability density for our one-jump model. We
 290 plot a comparison between the binned frequency of 10^4 indi-
 291 viduals (with exponential jump distributions each with a scale
 292 drawn from Eq. (4)) simulated to travel (solid lines) between
 293 two buildings on real-world map data. The data for building
 294 locations corresponds to the same regions of each country as in
 295 Fig. 2, and the marginalised jump probability densities derived

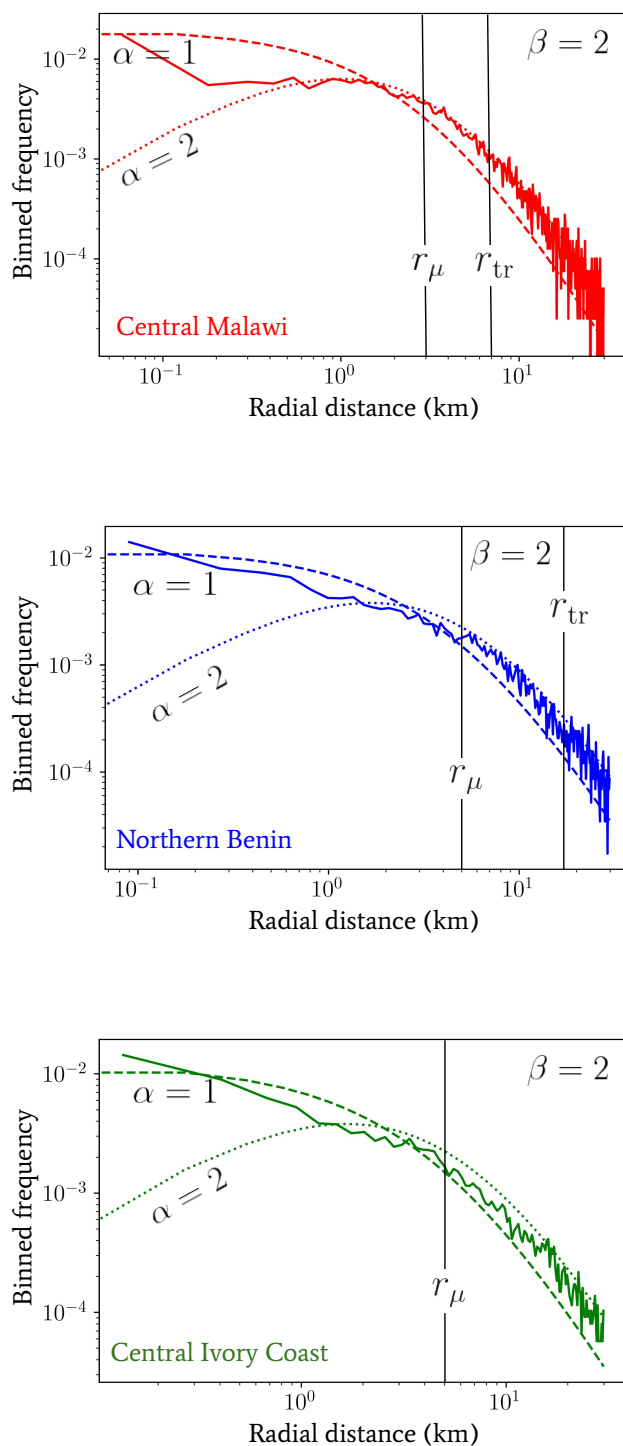


Figure 4: The binned frequency of 10^4 individuals simulated to travel a between two buildings (solid lines) whose locations correspond to data from the same regions of each country as in Fig. 2 (where some globally-applied uniform random point density thinning has been used to reduce computational load of exploring the distribution tails). This corresponds to the same marginalised jump probability density as Eq. (5) but instead of using a geometric power-law given by Eq. (1) we have implemented our single jumps on real-world map data (with exponential jump distributions each with a scale drawn from Eq. (4)). For comparison, we have also plotted some jump probabilities calculated using Eq. (5) with $\alpha = 1$ (dashed lines) and $\alpha = 2$ (dotted lines). In all cases we have fixed $\beta = 2$.

296 from Eq. (5) for choices of $\alpha = 1, 2$. In all cases we have fixed
297 $\beta = 2$ to a single value for easier comparison of different lines;
298 this value is approximately the same as the tails observed in the
299 distance distributions of Kung et al. [24].

300 The agreement between the $\alpha = 1, 2$ curves and the distances
301 generated from single simulated jumps between buildings on
302 the real-world map data appears to be variable between coun-
303 tries. Overall, it is unsurprising that the shapes of the curves all
304 seem to agree for scales $r > r_\mu$, because the asymptotic $\sim r^{\beta-1}$
305 behaviour dominates in all cases, which is consistent with our
306 model. In the opposing limit, however, the small-scale variabil-
307 ity in power laws for each country dominates and the departure
308 from our constant- α models appears more pronounced — in
309 favour of more transitory distance-dependent behaviour. Note
310 also that the range in variability that the region between our
311 $\alpha = 1$ and $\alpha = 2$ lines can account for is quite substantial. Such
312 a range in behaviour motivates the use of both $\alpha = 1$ and $\alpha = 2$
313 in quoting the most important results that follow so as to cor-
314 rectly assess the impact of small-scale region-specific building
315 distributions.

316 3. Multi-jump processes

317 In the previous section, a single jump model from a house-
318 hold to an infectious focal point location was developed in order
319 to describe the movement of individuals between locations be-
320 fore contributing infectious material. An important caveat to
321 this treatment of movement patterns is that one must also con-
322 sider the possibility that multiple movements are performed ei-
323 ther: between buildings before making a contribution to a focal
324 point, or between infectious focal points themselves, supplying
325 each some portion of infectious material in turn. The latter of
326 these will be very difficult to quantify without some form of
327 data, but for the former, we can explore some potential modifi-
328 cations to our model.

329 In this section, we will briefly explore an extension to the
330 basic model of human movement presented in Sec. 2 to in-
331 clude successive jumps by an individual to multiple buildings
332 before reaching an infectious focal point. The type of human
333 movement we aim to capture is not just that of daily commut-
334 ing, but also potentially longer distance travel across multiple
335 days/weeks. The latter form of movement can contribute to the
336 variability in migrant seasonal labour or family visits which can
337 change the population numbers in a given region on a yearly
338 timescale.

339 3.1. Homogeneous randomly-directed jumps

Following along similar lines to Sec. 2.2, let us now con-
sider an individual who completes multiple successive jumps
with exponentially-distributed lengths r_i (where we have in-
dexed each successive jump event with an i) with a fixed mean
scale σ within a day. For a fixed jump rate \mathcal{J} in time, consider
the following compound Poisson process for a two-dimensional
vector $\mathbf{x}(t)$ which encodes the 2-dimensional Euclidean coordi-

nate position of an individual over time

$$\mathbf{x}(t) = \sum_{i=1}^{\infty} \begin{bmatrix} r_i \cos(\theta_i) \\ r_i \sin(\theta_i) \end{bmatrix} \mathbb{1}_{[t_i, \infty)}(t) \quad (6)$$

$$t_i - t_{i-1} = \Delta t_i \sim \text{Exp}(\Delta t_i; \mathcal{J}) = \mathcal{J} e^{-\mathcal{J} \Delta t_i} \quad (7)$$

$$\theta_i - \theta_{i-1} = \Delta \theta_i \sim \text{U}(\Delta \theta_i; -\pi, \pi) = \frac{1}{2\pi} \Theta(\pi - \theta_i) \Theta(\pi + \theta_i), \quad (8)$$

where $\text{U}(\Delta \theta_i; -\pi, \pi)$ is the uniform distribution PDF (assum-
ing statistical isotropy) over the change in angle and $\Theta(\cdot)$ is a
Heaviside function. In Eq. (6) above, note that we are using an
indicator function $\mathbb{1}_A(t)$ which takes value unity if $t \in A$, else
zero.

The process specified by Eq. (6) assumes isotropy of both
the geometric distribution of the points and a uniform-random
direction choice of the individual. The distribution $p[\mathbf{x}(t)]$ is
not known as a closed-form expression, however, it clearly has
both a vanishing first moment

$$\begin{aligned} \mathbb{E}[\mathbf{x}(t)] &= \\ & \int_{-\pi}^{\pi} d\theta_i \int_0^{\infty} dr_i \sum_{j=1}^{\infty} \text{Pois}[j; \mathcal{J}(t - t_0)] \sum_{i=1}^j \begin{bmatrix} r_i \cos(\theta_i) \\ r_i \sin(\theta_i) \end{bmatrix} p(r_i; \alpha, \sigma) \\ &= \begin{bmatrix} \mathcal{J}(t - t_0) \mathbb{E}(r_i) \mathbb{E}[\cos(\theta_i)] \\ \mathcal{J}(t - t_0) \mathbb{E}(r_i) \mathbb{E}[\sin(\theta_i)] \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \end{aligned} \quad (9)$$

and a second moment which scales according to σ^2

$$\begin{aligned} \mathbb{E}[\mathbf{x}(t) \cdot \mathbf{x}(t)] &= 2\pi \mathcal{J}(t - t_0) \mathbb{E}(r_i^2) \\ &= 2\pi \mathcal{J}(t - t_0) (\alpha + \alpha^2) \sigma^2. \end{aligned} \quad (10)$$

Furthermore, with a scale distribution over jump lengths ap-
plied — as in Eq. (5) — the second moment given in Eq. (10)
becomes

$$\mathbb{E}[\mathbf{x}(t) \cdot \mathbf{x}(t)] = 2\pi \mathcal{J}(t - t_0) \frac{(\alpha + \alpha^2) r_\mu^2}{(\beta - 1)(\beta - 2)}. \quad (11)$$

In the left panel of Fig. 5 we have plotted the binned fre-
quency of total daily distances travelled by a population of
 5×10^4 individuals following the process defined by Eq. (6) and
drawing each individual's inverse-jump scale $1/\sigma$ from Eq. (4).
We have also fixed $\mathcal{J} = 1, 20$ per day (in red and black lines, re-
spectively), $\alpha = 1$ and $\beta = 1, 2, 3$ as indicated by the increasing
opacity within each triplet of lines. Contrasting these distribu-
tions, it is immediately clear that by increasing the jump rate
 \mathcal{J} , the effect on the distribution of daily distances travelled is
similar to increasing the value of α — see the middle panel of
Fig. 3 for comparison.

342 3.2. Homogeneous unidirectional jumps

Human movement patterns are not truly random, and instead,
one might anticipate a strong directional dependence, e.g., to
long distance travel of individuals over the course of multiple
days to an intended destination. Due to this fact, let us try to
quantify the effect of this behaviour by assuming the logical

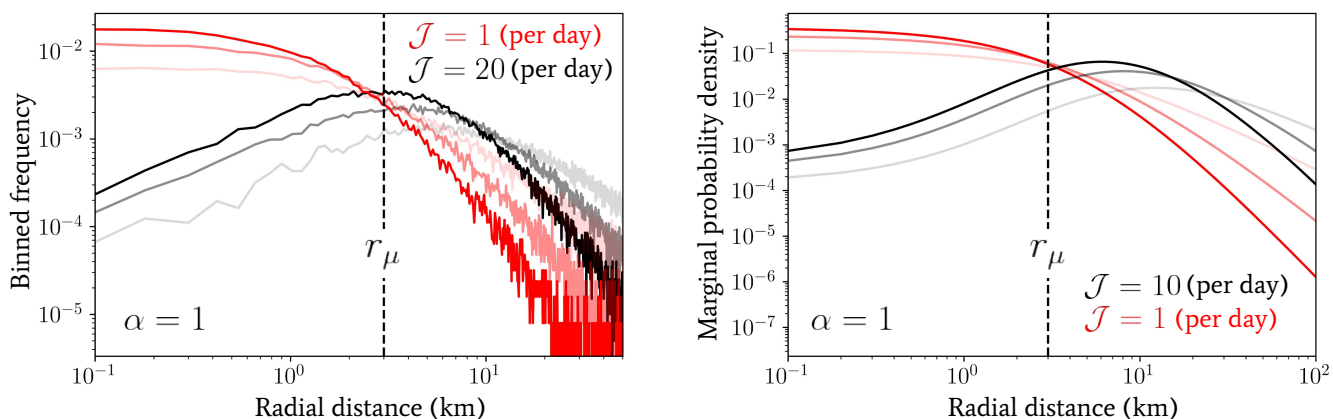


Figure 5: Numerical plots of the distance distributions generated by the multi-jump processes introduced in Sec. 3.1 with randomly-directed (left panel) and in Sec. 3.2 with unidirectional jumps (right panel). In the left panel, a population of 5×10^4 individuals following the process defined by Eq. (6) have been drawn (drawing from Eq. (4) for their jump scale predispositions) and the binned frequency of the total distance evaluated at the end of the day with jump rates of $\mathcal{J} = 1, 20$ per day (in red and black lines, respectively), $\alpha = 1$ and $\beta = 1, 2, 3$ as indicated by the increasing opacity within each triplet of lines. In the right panel, the marginal probability density given by Eq. (15) is depicted with jump rates of $\mathcal{J} = 1, 10$ per day (in red and black lines, respectively), $\alpha = 1$ and $\beta = 1, 2, 3$ as indicated by the increasing opacity within each triplet of lines.

extreme, i.e., that the angular direction of an individual's long-distance travel is fixed to a particular value. Following multiple successive jumps while assuming no variation in angle, Eq. (6) then becomes a compound Poisson process for the total jump distance $x(t) \equiv |x(t)|$ in a single dimension. This process has a known characteristic function

$$\varphi_x(s) = \exp \{ \mathcal{J}(t - t_0) [\varphi_r(s) - 1] \}, \quad (12)$$

where $\varphi_r(s)$ denotes the characteristic function of the stationary increments given by Eq. (3). Hence, we may write

$$\varphi_x(s) = \exp \{ \mathcal{J}(t - t_0) [(1 - \sigma i s)^{-\alpha} - 1] \}. \quad (13)$$

which has no closed-form inverse transformation, but is useful as an expression to calculate the moments. Alternatively, in order to find the distribution over $x(t)$, we may marginalise over the number of jumps performed (which is Poisson-distributed by construction), which are themselves drawn from the jump distance distribution in the following way

$$p[r = x(t) | \theta_i = \theta_{i-1}] = \sum_{j=0}^{\infty} \text{Pois}[j; \mathcal{J}(t - t_0)] \text{Gamma}[r; j\alpha, 1/\sigma]. \quad (14)$$

Therefore, if $\alpha = 1$, Eq. (14) yields

$$p[r = x(t) | \theta_i = \theta_{i-1}, \alpha = 1] = \sqrt{\frac{\mathcal{J}(t - t_0)}{\sigma r}} e^{-\mathcal{J}(t - t_0) - \frac{r}{\sigma}} \mathcal{I}_1 \left[2 \sqrt{\frac{\mathcal{J}(t - t_0) r}{\sigma}} \right], \quad (15)$$

where $\mathcal{I}_n(\cdot)$ is the modified Bessel function of the first kind.

In the right panel of Fig. 5 we have plotted the marginal probability density given by integrating Eqs. (15) and (4) over $1/\sigma$

for direct comparison with the one-jump model of Eq. (5). We have used values of $\mathcal{J} = 1, 10$ per day (in red and black lines, respectively), $\alpha = 1$ and $\beta = 1, 2, 3$ as indicated by the increasing opacity within each triplet of lines. Comparing this distribution with the one observed for randomly-directed multiple jumps (see Sec. 3.1), a similar, but even more pronounced, effect by varying \mathcal{J} is observed — since only a value of $\mathcal{J} = 10$ is required to achieve roughly the same change. This is to be expected, as the unidirectional jumpers will always travel at least the same (or more often) a greater distance in total in comparison with the randomly-directed ones, which will likely induce a more severe deformation of the overall distribution when using the former.

The parameter degeneracy between α of the one-jump model and \mathcal{J} in the multi-jump processes described above suggests a test can be performed, to either validate a model, or discard it in favour of a modified version for a given situation with real data. By combining collected data on the total distance travelled per day and the local geometry of distances between buildings travelled to in a given region, the jump rate for each individual (if the same) could be statistically inferred. Comparing this inferred value to real data would provide a test of the movement models we have suggested in this work or potentially provide insight into where they may be improved to better reflect real human daily movement in a given setting.

The simple multi-jump models we have considered in this section have assumed that the distribution for the cumulative number of nearby buildings is homogeneous for each new jump. This approximation is not likely to work well on real-world map data as we have shown there is substantial small-scale heterogeneity in the distances between buildings exhibited in, e.g., Fig. 2. Furthermore, if the multi-jump movement model is purely diffusive over the building locations themselves (itself

393 an assumption which must be investigated further), then there
394 will be correlations in the location distributions between suc-
395 cessive jumps that can lead to the distribution of jump distances
396 exhibiting localising effects akin to those exhibited by power-
397 law random banded matrices [35]. In other words, when an
398 individual jumps from a cluster of points to, e.g., a point on the
399 edge of their cluster, then they may experience a stronger ‘pull’
400 back towards the centre of this cluster on the next jump due to
401 the anisotropy in their apparent local point density. This effect
402 should be investigated in future work to evaluate its relevance
403 to movement patterns.

404 A more sophisticated method to modify the model and deal
405 with (uncorrelated) small-scale heterogeneity could be to sam-
406 ple from the numerically-obtained power laws from Fig. 2 to
407 emulate the spatial heterogeneity in building distances on the
408 real-world map data directly in the processes of Sec. 3. This
409 method would have significant advantages, in terms of compu-
410 tational complexity, to direct simulation methods over the real
411 world map data.

412 4. The effect on reservoirs of infection

413 4.1. A coarse-grained stochastic reservoir network

414 Up until this point, the movement patterns we have dis-
415 cussed have not been contextualised in an NTD disease model.
416 Since our current (static focal point) framework is best-suited to
417 schistosomiasis and STH transmission, we shall use their well-
418 studied transmission models as a reference throughout this sec-
419 tion [36].

420 Let us now consider a group of focal points that are clus-
421 tered into a ‘reservoir of infection’. Such a cluster is a source
422 of infectious material from which disease-free individuals are
423 exposed to new infections and, for those already with infec-
424 tions, to potentially increase them in intensity. With each of
425 these clusters, let us now also associate a population of indi-
426 viduals with household locations that are sufficiently close such
427 that the majority of their infectious inputs must be into one of
428 their associated reservoir’s focal points. Note that the vague-
429 ness in definition of ‘sufficiently close’ above immediately ex-
430 poses some of the difficulties associated to defining an EU for
431 the purpose of assessing local prevalence or intensity. Our work
432 in the previous sections of this paper will allow us to proceed
433 with a new approach to quantifying the spatial scales associ-
434 ated with the proximity of households to these focal points of
435 infection. We will then use this approach to understand how
436 spatial scales may be incorporated into the effective dynamical
437 description of STH and schistosomiasis transmission.

Let us now denote the set of indices which identify all in-
dividuals who live in households that are less than a radial dis-
tance r_Λ away from a focal point that is included in the reservoir
cluster as S_Λ . Additionally, to make a reservoir cluster, let us
use r_Λ as a radial separation distance threshold below which any
two focal points are clustered together to be part of the same
infectious reservoir. For reservoirs of infection of either schis-
tosomiasis or STH, assuming that the associated human pop-
ulation number N_Λ (the number of elements in S_Λ) does not

change, the force of infection (FOI) $\Lambda(t)$ at time t is updated
according to

$$d\Lambda(t) = -d_{\text{res}}\Lambda(t)dt + \frac{d_{\text{res}}(d + d_w)R_0}{N_\Lambda} \sum_{i \in S_\Lambda} d\ell_i(t), \quad (16)$$

where d_{res} is the death rate of the infectious material, d_w is
the worm death rate, d is the human death rate, R_0 is the basic re-
production number and $\ell_i(t)$ is a compound Poisson process asso-
ciated to the time-dependent infectious material input of the
 i -th individual. Since helminths are dioecious, the definition
for R_0 , differs from its standard interpretation in standard mi-
croparasite diseases since the life cycle of these macroparasites
depends on fertilisation of female worms (and hence the pres-
ence of both sexes) within a host [36]. Note also that in the
equation above, and throughout, we shall neglect age structure
in our description — though our formalism can be adapted to
include this. This choice is for simplicity in presentation and
will not affect our main conclusions.

By clustering focal points into a reservoir of infection by us-
ing a radial distance threshold, one might reasonably question
the scaling in the spatial extent of the infectious focal points
themselves. As was discussed in Sec. 3, the distribution of fo-
cal points of infection is likely to be both helminth species-
dependent and variable according to local geographic consid-
erations. Due to this variability, the spatial scales associated
with clusters of focal points themselves will likely vary across a
map. However, this fact will not directly affect the conclusions
of the present work since we will consider distances between
each individual focal point and its neighbouring households in
turn. For specific case studies, the implementation of our algo-
rithm for binding focal points together should be well-defined
in most cases for STH and schistosomiasis transmission set-
tings, as long as the distribution of focal points can be inferred.
Recall also our previously stated point that, for STH, our house-
hold distribution itself could potentially be a tracer for many of
the focal points of infection [32, 33].

A more important effect, which is indirectly related to the
variability in spatial scale of focal point clusters, is how variable
cluster sizes might induce internal variability in the exposure of
each individual to new infections. In models of helminth trans-
mission [37], it is most common to account for this variabil-
ity in exposure through assigning each individual an additional
predisposition factor λ_i which is drawn from a gamma distri-
bution, $\text{Gamma}(\lambda_i; k, k)$, where k is the ‘aggregation parameter’
which modifies the variance-to-mean ratio of the distribution
of worms within hosts. Due to the epidemiological processes
that it aims to capture — and assuming that the basic reproduc-
tion number R_0 , which accounts for the transmission intensity,
remains unchanged — this aggregation parameter will likely
vary to capture a change in exposure according to the spatial
scale r_Λ chosen for the reservoir of infection. One simple model
to capture this variability in exposure might be to consider the
value of k itself to be drawn from a gamma distribution, i.e.,
 $p(k) = \text{Gamma}(k; a_k, b_k)$ from which samples are drawn across
the map for the individuals closest to each focal point of infec-
tion. When binding such focal points together to get the global

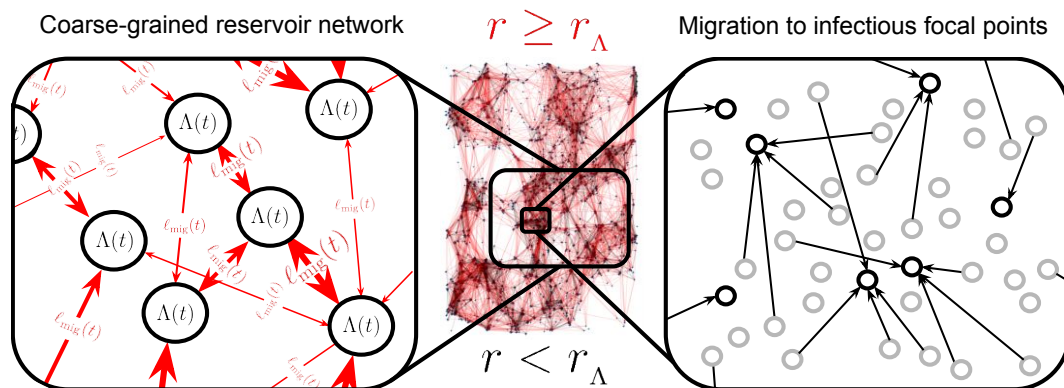


Figure 6: A diagram of real map data (based on buildings in central Malawi) and a zoomed illustration which indicate the coarse-graining procedure of Eq. (21) — in which the black arrows are ‘removed’. Arrows depict individual movements from households (grey hollow dots on the right hand side zoomed illustration) to focal points of infection (black hollow dots on the right hand side zoomed illustration) where black arrows on both the map and its zoomed counterpart correspond to jumps over a radial distance $r < r_\Lambda$ (where r_Λ is the spatial coarse-graining scale) and red arrows correspond to jumps over a radial distance $r \geq r_\Lambda$. On the left hand side zoomed illustration we see the emergence of the spatially coarse-grained reservoir network ‘nodes’ (nodes are neither drawn to scale nor have geometrically-accurate reservoir spatial shapes) connected by red migratory ‘links’ in time.

behaviour of the reservoir of infection, one would then have a mixture model of separate behaviours to consider in the distribution of worms within hosts, affecting the summation term of Eq. (16). Since variability of exposure with spatial scaling is not the main focus of the present work, we shall not investigate this any further here. However, due to its obvious importance to the dynamical behaviour of the reservoir by association, it will be an important component to be cognisant of in future work.

In Ref. [14] it was shown that, to good approximation for moderate-to-large population numbers, the worm burden distributions of individuals $p(w_i, t)$ evolve according to the following reservoir birth-death process master equation

$$\begin{aligned} \frac{d}{dt} p(w_i, t) = & - [\lambda_i \Lambda(t) + (d + d_w) w_i] p(w_i, t) \\ & + (d + d_w) (w_i + 1) p(w_i + 1, t) \\ & + \lambda_i \Lambda(t) p(w_i - 1, t). \end{aligned} \quad (17)$$

The solution to this equation is a Poisson distribution $p(w_i, t) = \text{Pois}[w_i; \mathcal{I}_i(t)]$ with time-dependent intensity

$$\mathcal{I}_i(t) = \mathcal{I}_i(t_0) e^{-(d+d_w)(t-t_0)} + \int_{t_0}^t \lambda_i \Lambda(t') e^{-(d+d_w)(t-t')} dt', \quad (18)$$

in which we have inserted the solution to Eq. (16). Summing over an ensemble of these individuals to get an overall distribution of worms within hosts, many epidemiological variables, e.g., the prevalence of infection or mean parasite burden of hosts associated to the reservoir can be calculated (up to specifying the additional gamma-distributed predisposition to infection λ_i for each individual) while maintaining the finite population variance neglected by deterministic disease models. Therefore, in order to understand how these variables which are associated with each reservoir of infection are affected by spatial movements of human hosts, it is essential to correctly assess how $\Lambda(t)$ is modified by inward migration.

Due to the movement patterns we have discussed, there may be individuals who contribute to one of the focal points of the

reservoir but come from a region further away such that they are not counted in S_Λ . There may be individuals who are counted in S_Λ but instead contribute to a focal point outside of their own associated cluster, e.g., on a particular day. Eq. (16) will hence be perturbed by these movements in the following way

$$\begin{aligned} d\Lambda(t) = & -d_{\text{res}} \Lambda(t) dt \\ & + \frac{d_{\text{res}}(d + d_w) R_0}{N_\Lambda} \left[\sum_{i \in S_\Lambda} d\ell_i(t) + d\ell_{\text{mig}}^+(t) - d\ell_{\text{mig}}^-(t) \right], \end{aligned} \quad (19)$$

where $\ell_{\text{mig}}^\pm(t)$ are also compound Poisson processes which sum over the amount of infectious material generated by these migrating individuals that is either net entering (+) or leaving (−) the reservoir focal points.

Note here that Eqs. (16) and (19) currently assume that N_Λ is static over the time period of interest, which may not be the case if one considers changes that take place over the course of, e.g., a year. This is because long-term effects can cause the population number to vary, including periodic seasonal migrant labour or family visits at particular times of year. Such long-term changes are likely to not have too great a dynamical significance on the terms of Eqs. (16) and (19), however, since these terms are associated to variations in the reservoir of infection which typically occur on much shorter timescales ($1/d_{\text{res}}$ is typically on the order of weeks or months depending on helminth species).

Eq. (19) implicitly describes a network of spatially coarse-grained reservoirs of infection, each of which varying in spatial extent according to a combination of the distribution of their infectious foci, as well as the spatial coarse-graining scale choice r_Λ . This network of coarse-grained reservoir ‘nodes’ is connected by migratory ‘pulses’, continually affecting each reservoir’s temporal stability (which we shall discuss later on) as well as potentially non-negligible statistical cross-correlations in local epidemiological indicators (e.g., prevalence or intensity of infection) with respect to other reservoirs. Note also that the choice of r_Λ can therefore control the strength of these

536 migratory interactions between reservoirs in the network. In
537 Fig. 6 we have illustrated this point explicitly on a digram with
538 real map data based on buildings in central Malawi. The black
539 (red) arrows on the map and zoomed illustrations show migratory
540 movements of a radial distance below (above or equal to)
541 r_Λ , which has been arbitrarily chosen for clear visual effect.

Let us now define T_Λ as the average rate of individuals travelling greater than a distance of r_Λ to a focal point of the reservoir of infection and contributing material to it. Following Ref. [14], this translates into the following decomposition of the positive migratory process in Eq. (19)

$$\ell_{\text{mig}}^+(t) = \sum_{j=0}^{\infty} y_j^+ \mathbb{1}_{[t_j^+, \infty)}(t),$$

$$p(t_j^+ - t_{j-1}^+) = \text{Exp}(t_j^+ - t_{j-1}^+; 1/T_\Lambda), \quad (20)$$

542 where y_j^+ is the contribution to the infectious material quantity
543 of a focal point by an individual: a random variable with a
544 distribution which varies between disease, but likely follows
545 a negative binomial character in some cases. This distribution
546 may be specifically derived by transforming the travelling individual's
547 worm burden w_i (e.g., drawn from a Poisson distribution and Eq. (18) or an individual-based simulation) into an
548 expected count of fertilised eggs or larvae using a helminth-specific
549 mating function [36]. Note that this compound Poisson process is exact since it tracks the contribution entering the
550 reservoir directly. If one wishes to temporally coarse-grain over
551 the timescale $1/d_{\text{res}}$ and see the effect on the mean worm burden
552 deterministic ODE model (as is done in Ref. [13]), an additional
553 non-Markovian component arises to account for the reservoir
554 pulse decay.
555
556

557 When an individual contributes to an infectious focal point
558 which is further than r_Λ away from their household, they do not
559 just make an additional contribution to the latter's reservoir of
560 infection, but they also remove their contribution from one of
561 the focal points which is within r_Λ of their household (their local
562 reservoir). In this sense, each random pulse within the $\ell_{\text{mig}}^+(t)$
563 process of one reservoir is exactly correlated to a pulse within
564 the $\ell_{\text{mig}}^-(t)$ process of another's. So, for model completeness, it
565 is sufficient to define only the $\ell_{\text{mig}}^+(t)$ process (attributing an intended
566 location of travel) for each reservoir of infection defined
567 over a map. This is tantamount to defining 'links' between the
568 coarse-grained nodes of the network illustrated by Fig. 6.

Let us now consider the influence that the coarse-graining procedure has on Eq. (20). By analogy to physical theories in nature, note that the fundamental dynamical theory should not change — e.g., individual contributions to each reservoir should not 'disappear' — simply because of an arbitrary choice for r_Λ , and so one must define a procedure whereby the effective dynamics in Eq. (20) are always describing the same fundamental phenomenon, for any choice of r_Λ . An obvious way to do this is to consider the variation in probability mass associated to individual movement rates as r_Λ is varied. Based on the arguments of Sec. 2, we hence point out that the following scaling should exist for T_Λ which accounts for this difference in probability

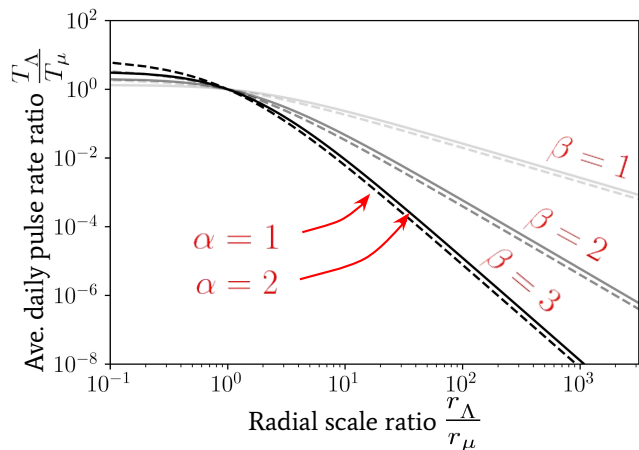


Figure 7: The spatially coarse-grained average daily pulse rate T_Λ into a focal point of infection as a fraction of its value T_μ from individuals arriving a distance of r_μ or greater away. This value is plotted as a function of the radial coarse-graining scale ratio r_Λ/r_μ used. The relationship is given by Eq. (22) for a range of α and β power-law parameters.

mass arising from a rescaling of distances

$$\frac{T_\Lambda}{T_\mu} = \frac{\int_{r_\Lambda}^{\infty} dr p(r; \alpha, \beta, r_\mu)}{\int_{r_\mu}^{\infty} dr p(r; \alpha, \beta, r_\mu)} = \frac{1 - \int_0^{r_\Lambda} dr \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{r_\mu^{\beta} r^{\alpha-1}}{(r_\mu+r)^{\alpha+\beta}}}{1 - \int_0^{r_\mu} dr \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{r_\mu^{\beta} r^{\alpha-1}}{(r_\mu+r)^{\alpha+\beta}}}, \quad (21)$$

569 where we have defined T_μ as the average rate of individuals
570 travelling greater than a distance of r_μ to a focal point of the
571 reservoir of infection and contributing material to it.

In order to derive Eq. (21) we have assumed that all individuals jump once per day, so that the marginalised jump PDF $p(r; \alpha, \beta, r_\mu)$ of Sec. 2 may be used directly. By integration of Eq. (21), one generally obtains

$$\frac{T_\Lambda}{T_\mu} = \frac{1 - \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+1)\Gamma(\beta)} \left(\frac{r_\Lambda}{r_\mu}\right)^\alpha {}_2F_1\left(\alpha, \alpha + \beta; \alpha + 1; -\frac{r_\Lambda}{r_\mu}\right)}{1 - \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+1)\Gamma(\beta)} {}_2F_1\left(\alpha, \alpha + \beta; \alpha + 1; -1\right)}, \quad (22)$$

572 where ${}_2F_1(\cdot, \cdot; \cdot; \cdot)$ is the (Gauss) hypergeometric function. In
573 Fig. 7 we use the solution given by Eq. (22) to compute the
574 quantity T_Λ/T_μ as a function of r_Λ/r_μ for a range of parameter
575 choices.

576 The variation of T_Λ/T_μ as a function of r_Λ provides a mapping
577 between the average pulse rate for Eq. (20) that is associated to each
578 focal point within the reservoir of infection (whose pulse amplitudes can be easily summed over to get a total contribution to and from the reservoir) and the distribution of households at different spatial coarse-graining scales. The value of r_Λ may hence be fixed to identify movement of individuals between structures at different scales, e.g., buildings, villages, towns, cities, etc. We may visualise this coarse-graining of movement distances by referring back to Fig. 6 — there is an emergence of black clusters of connected locations which are connected by red arrows at longer distances. By then clustering

the focal points on this map (note this is a mock illustration and these are not real focal points) according to whether or not they are within a radial distance r_Λ of one another, we can then build coarse-grained reservoirs of infection associated to the scale r_Λ .

It is important to note that infectious reservoir clusters may themselves have a significant spatial extent due to the particular geometry of the nearby focal points, and so T_Λ should be considered to apply to each and every focal point (or small clusters of focal points) within the reservoir individually. Hence, for reservoirs of infection which correspond to clusters of focal points which, when bound together, have a radial spatial extent that is much smaller than r_Λ then the results of this section may be used directly for the entire reservoir. In situations where this is not the case — e.g., long and thin geometric relationships binding focal points around the edge of a lake — then our results may still be applied to each focal point (or small cluster of focal points) in turn. The application of our more general formalism is then subject to change depending on the specific relationships between focal points and hence, indirectly, on the species of helminth.

4.2. Reservoir Wiener process approximation

Reconsidering Eq. (16), note that the following Gaussian sample mean approximation applies for the terms summing over all (non-migratory) infectious material input into the reservoir

$$\frac{d_{\text{res}}(d + d_w)R_0}{N_\Lambda} \sum_{i \in S_\Lambda} d\ell_i(t) \approx \mathcal{M}_{\text{res}}(t)dt + [\mathcal{V}_{\text{res}}(t)]^{1/2}dW_t, \quad (23)$$

where W_t is a Wiener process, $\mathcal{M}_{\text{res}}(t)$ is a time-dependent mean and $\mathcal{V}_{\text{res}}(t)$ is a time-dependent variance for the reservoir inputs. Such an approximation is motivated by the central limit theorem and will be most accurate in the limit of large population number N_Λ . By inserting Eq. (23) into Eq. (19), one finds a drift-diffusion which satisfies the following Gaussian distribution for $\Lambda(t)$

$$p[\Lambda(t) = z] = \mathcal{N}[z; \mathcal{M}_{\text{res}}(t), \mathcal{V}_{\text{res}}(t)]. \quad (24)$$

Calculating an exact analytic form for $\mathcal{V}_{\text{res}}(t)$ in the case of each helminth species is still an open research question since this is very likely to be quite complex to model. For some analytic insight, however, in Ref. [13] it was shown that, with a negative binomial distribution of worms within hosts, the distribution over typical egg outputs from hosts for hookworm is well-approximated by another negative binomial (with different mean and variance). Under these conditions, one may infer, from the sum of negative binomial variances, that

$$\mathcal{V}_{\text{res}}(t) = N_\Lambda \mathcal{M}_{\text{res}}(t) \left[1 + \frac{\mathcal{M}_{\text{res}}(t)}{k_{\text{res}}(t)} \right], \quad (25)$$

where $k_{\text{res}}(t)$ is the aggregation parameter for the reservoir input negative binomial, which can be time-dependent before the system relaxes to the endemic steady state configuration.

Eq. (24) provides a significant improvement in computational efficiency over a full stochastic simulation. In particular, for endemic steady-state regions where the constant values

of $\mathcal{V}_{\text{res}}(t)$ and $\mathcal{M}_{\text{res}}(t)$ can be computed quickly for each parameter configuration, it could be used to generate a simulation likelihood for statistical inference over an entire map affected by helminth transmission and migrating infected individuals.

Due to migratory inputs from (and outputs to) other reservoirs, applying Eq. (24) consistently to each node of our reservoir network will require the inclusion of spatial covariances (or higher-order statistics) between samples. The strength of these spatial covariances can, in principle, be computed by including the migratory terms of Eq. (19) into the approximation above. We can see this by recalling that the rate of these migratory compound Poisson processes (see Eq. (20)) is fundamentally connected to the spatial distance scales via Eq. (22). This ‘fundamentals-based’ method of spatial epidemiological inference for helminth transmission would be distinct from other methods since it would not only include important effects which are inherent to a simulation, such as finite population variance of the underlying stochastic process, but also a theoretical understanding of how the effective dynamical description itself changes with spatial coarse-graining into separate EUs. Note that the ‘scale-invariance’ of the theoretical model we have suggested here can therefore, in principle, be inferred with data collected at one EU scale and then interpolated/extrapolated for model predictions at another.

Although it is beyond the scope of this paper, in future work, it will be of interest to compare the results obtained from other NTD spatial epidemiological models — which typically employ deterministic models a particular spatial scale, e.g., Refs. [9, 10] — with the spatial correlations induced in the diagnostic output from our suggested approach. Such a comparison would help quantify the importance of finite population variance and EU choice on reservoir temporal stability. Other NTD spatial modelling approaches one might consider a comparison to are spatial GRF-based models [6] — these are to be contrasted with the temporally Gaussian, but spatially complex, power-law-like, reservoir network we have proposed in this section. In addition, it would be interesting to explore how the approximate temporal Gaussianity of the model we have proposed in Eq. (24) changes when considering: dynamics further from endemic steady states; configurations closer to transmission breakpoints; or systems with significantly smaller population sizes [14].

4.3. The critical spatial scale

In order to evaluate the success of control measures through the calculation of epidemiological observables, such as the prevalence or intensity of infection, it is common for an EU to correspond to a particular spatial scale for the reservoir of infection — see, e.g., [4, 38, 39]. In Sec. 1 we discussed the possibility of defining EUs such that the effect that human movement has on their implied reservoirs of infection was minimal. Note that this is not just of importance to the temporal stability of the reservoirs, but also advantageous because it nullifies any statistical inference bias that may arise through strong migration effects. Such biases may occur when there is a large migration pulse rate of infectious material into a given region where the local intensity and prevalence are increased, but the

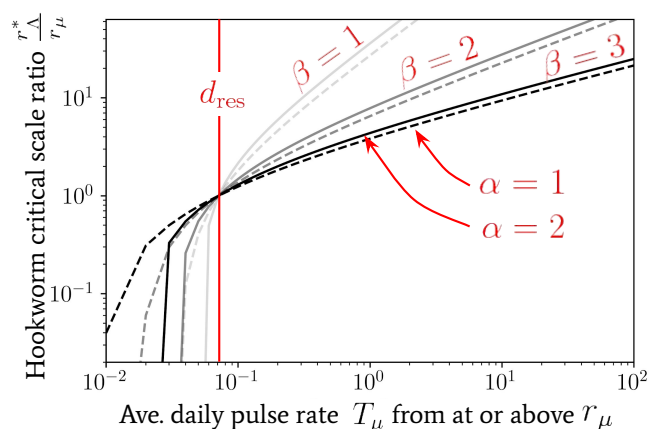


Figure 8: The critical spatial scale as a ratio r_{Λ}^* at which $T_{\Lambda}^* = d_{\text{res}}$, for hookworm — where $d_{\text{res}} \approx 0.071$ [40] — plotted as a ratio of r_{μ} . The value of this scale is shown against the value of the average daily pulse rate T_{μ} at or above r_{μ} . We have used Eq. (22) to generate this relationship with a (bisection) root-finding algorithm.

671 statistical models using these diagnostics for inference do not
672 include migration rates in their description.

673 It was shown in Refs. [13, 14] that if the average rate at which
674 reservoir pulses occur exceeds the average death rate of the infectious
675 material in the reservoir, d_{res} , then the effect of infected
676 human migration on the transmission dynamics becomes particularly
677 strong. In particular, the authors of Ref. [14] demonstrate
678 that, for a range of finite population sizes, migration rates at or
679 above this critical rate into a reservoir of infection can restart
680 transmission in regions where a mass treatment programme has
681 already achieved transmission elimination. In addition, it was
682 shown that this rate of migration can stabilise the stronger ‘fade-
683 out’ effects exhibited near transmission breakpoints. By combin-
684 ing the concept of a critical rate with our proposed spatial
685 model for migration between reservoirs, one may derive a new
686 and important critical spatial scale r_{Λ}^* around the focal points of
687 infectious reservoirs which depends on: the type of helminth,
688 the small-scale geometry of locations and the specific patterns
689 of human movement. This spatial scale can be used to define
690 the geographical size of EUs.

691 If the spatial scale of an EU is chosen which corresponds to a
692 region smaller than the critical scale r_{Λ}^* around each of its focal
693 points of infection, the average daily rate of pulses will always
694 be larger in magnitude than the value T_{Λ}^* , calculated using the
695 critical scale. This is because the curves shown in Fig. 7 are
696 always decreasing with increasing radial scale. This implies
697 that when regions around focal points are defined at this critical
698 scale or smaller for an EU, pulses of the form given by Eq. (19)
699 cannot be safely neglected and should become important to take
700 into account when modelling the transmission dynamics to cor-
701 rectly assess reservoir stability and possible observation biases.
702 This means that if EUs are at this critical scale of smaller NTD
703 programmes need to have measures in place to mitigate the im-
704 pact of human movement on programme targets. For example,

705 individuals who regularly move beyond the boundaries of EUs
706 should be specifically targeted for treatment. When defining an
707 EU in the opposite case, however, effects from migration be-
708 tween reservoirs of infection may be safely neglected.

709 Using Eq. (22) the critical spatial scale r_{Λ}^* at which T_{Λ}^* be-
710 comes equal to the death rate of the infectious material in the
711 reservoir per day, i.e., $T_{\Lambda}^* = d_{\text{res}}$, has been plotted for hookworm
712 (which has $d_{\text{res}} \approx 0.071$ [40]) in Fig. 8 as a function of the av-
713 erage daily pulse rate T_{μ} from individuals travelling distances
714 at or above r_{μ} . In this plot we note that nearly all parameter
715 combinations indicate a sharp decline in the critical scale ratio
716 if T_{μ} is found to be below d_{res} — where, in particular, for values
717 of $\alpha = 2$ the scale ratio appears to fall extremely sharply and
718 hence one can no longer find a finite critical scale below which
719 migration becomes important.

720 In contrast, for values $T_{\mu} > d_{\text{res}}$ of increasing orders of mag-
721 nitude, one infers from Fig. 8 that the value of α is nearly ir-
722 relevant (which follows from the construction of T_{μ}) and, par-
723 ticularly for $\beta = 2, 3$, the increase in the critical scale ratio is
724 very gradual. Such a relationship is also consistent with our
725 expectation as the decline in the average pulse rate from peo-
726 ple travelling from distances $r_{\Lambda} \gg r_{\mu}$ is particularly sharp for
727 $\beta = 2, 3$ — see, e.g., Fig. 7 — and so one requires a significant
728 increase in the amplitude of T_{μ} to achieve a significant change
729 in the critical scale.

730 We acknowledge that there are important caveats to the rela-
731 tionship above (and the one shown in Fig. 7) which arise from
732 the assumptions made in obtaining Eq. (21), i.e., the statistical
733 homogeneity and isotropy of people living in the surrounding
734 households who all jump only once per movement event/day di-
735 rectly to an infectious focal point. By relaxing the assumptions
736 made in obtaining Eq. (21), the rate of pulses will vary between
737 different reservoirs of infection. In addition to these modifying
738 effects, the inclusion of multiple unidirectional jumps with dif-
739 ferent initial preferred directions may no longer be isotropically
740 configured when viewed collectively over the course of the day
741 — which would be the case when multiple individuals all con-
742 verge to a globally preferred location such as a place of work or
743 a school. We propose to consider such modifications in future
744 work on specific case studies.

745 Perhaps an even more important model extension to consider
746 was noted in Sec. 4.1, where the potential for the aggregation
747 parameter k (for the distribution of worms within hosts) was
748 hypothesised to also vary with spatial scale due to differences
749 in infectious exposure. Although including this potential effect
750 will not directly influence our conclusions with regard to migra-
751 tion, it is possible that another critical spatial scale exists which
752 may be found by considering the value of k required to make
753 endemic steady state stability possible. We leave the investiga-
754 tion into the possibility of this additional critical scale to future
755 work.

756 5. Discussion and conclusions

757 In this work we have developed and studied spatial models
758 of human movement to infectious focal points — the impor-
759 tant drivers of new infections for many parasitic organisms in

humans. Our movement models have been motivated by the observed patterns of human mobility in, e.g., Refs. [24, 26], but remain robust to setting-specific parametric variations should these observations prove unsuitable for all settings.

At its most basic, our model of human mobility assigns a single daily journey (or ‘jump’) for each individual, however, we have discussed many extensions to this description, which include: multiple Poisson-distributed jumps, each of which may either be in random directions or unidirectional; and substantial local heterogeneity in the distribution of distances between available travel locations. By combining these extensions into a generalised framework, one can generate a wide variety of possible daily human movement behaviours to/between focal points of infection under a unifying description which offers a model-focused explanation for the observed patterns that is also computationally efficient. In future work, it will be interesting to extend our approach to include time dependence in the location of infectious contact events. Such a model will be necessary for diseases such as lymphatic filariasis or onchocerciasis, where bites from mosquito or black fly vectors infected by filarial larvae can vary spatially over time, and may not be Markovian (i.e. memoryless). An investigation into the longer-term temporal changes in population number, e.g., due to seasonal migrant labour patterns or population displacement due to conflict or climate change, may also be interesting.

We note here that the human movement models we have studied are comparable to those developed in Refs. [27, 28] for vector-borne diseases, which label the single daily journey models as so-called ‘Lagrangian’ mobility models and the multiple successive movement models as ‘Eulerian’. We consider our stochastic approach to generating frequency-distance distributions useful and informative in understanding the underlying probability distributions for these works, especially in relation to demonstrating the regimes in radial spatial scales which are relevant to location-specific geometric effects and human movement distance predispositions.

By clustering focal points into reservoirs of infection, our human movement models have also highlighted the most important information necessary to efficiently (by which we mean low numbers of parameters) describe the effect of migration patterns on these reservoirs. Our model defines a stochastic network of dynamical reservoirs linked by migratory ‘pulses’ (see Eq. (19)) which accounts for these migratory movements at any defined spatial evaluation scale around each focal point using the movement models that we have developed. The necessary data to infer such a ‘reservoir network’ can, in principle, be collected by well-designed field studies. Based on our work here, the information to obtain for a particular study may include:

1. Obtaining the appropriate (likely radially-dependent) power-law $\alpha(r)$ for the expected cumulative number of locations available to travel to by an individual — see, e.g., Fig. 2. These data can be obtained from geographic maps.
2. Evaluating the amplitude of the effect of small-scale heterogeneity (and perhaps anisotropy) on the power-law cumulative distribution of locations available to travel to by an individual. One can obtain this information from high

resolution maps of households.

3. Obtaining the average number of journeys performed within a single day per individual. In this case some metric for having ‘completed’ a journey associated to the time spent at each location will likely need to be determined. In principle, this information can be obtained from questionnaires on human movement that can be integrated into a regular field survey.
4. Evaluating the average number of people entering or leaving the relevant focal points for a given reservoir of infection at some particular spatial scale, e.g., this can be T_μ or evaluated at some other spatial scale since the reasoning of Eq. (21) can be easily adapted. Note that the true spatial scale of the reservoir does not need to be known for collecting these data. This information might be obtained through questionnaires.
5. Potentially obtaining an appropriate scaling value β for the distribution of predispositions to jump a certain distance (see Eq. (4)). This may either be consistent with the observed work-home commute patterns in Ref. [24], or deviate to a different power-law relationship entirely, depending on local access to vehicles and other forms of transportation. This information might also be obtained through questionnaires.

While developing our stochastic reservoir network model, we also discussed a computationally-efficient approximate description for its dynamical behaviour near or at endemic steady states. In particular, a Gaussian sample mean approximation was made for the sum of infectious material in each reservoir over a coarse-grained temporal scale, rendering its dynamical behaviour the same as a drift-diffusion process with migratory ‘jumps’ — see Eq. (24). The drastic improvement in computational efficiency that such an approximation offers makes it an attractive option to explore when statistically inferring fully-stochastic infectious reservoirs from diagnostic data. We leave the implementation of this approach, and the analysis of its robustness for reservoir network states far from endemicity, to future work.

The last, and perhaps most significant of our findings in this work concerns the spatial scales associated to EUs. An EU is commonly used to define epidemiological observables such as prevalence or intensity of infection for a given region, often in the context of assessing the performance of mass drug administration programmes [21, 20]. By combining our stochastic reservoir network model with a known critical rate for migration into/out of reservoirs of infection, we have been able to demonstrate that a critical radial spatial scale r_Λ^* exists around the focal points of infection inside a defined EU. When an EU is spatially defined such that the spatial region of consideration around each of its included focal points is larger than this critical scale, the effects of migration may be safely neglected and our reservoir network may effectively remove its ‘links’. In the opposing limit at scales smaller than r_Λ^* , however, we have described how: the temporal stability of the dynamics of infectious reservoirs defined at this scale may be affected through migratory jumps; statistical inference biases may occur through

the use of models without migratory effects; and mass drug administration applied only to the population within an EU without considering individuals moving between EUs may experience strong bounce-back effects.

The critical spatial scale we have found in this work is of biological and practical importance to many helminth disease transmission processes. We have also discussed how this model may need further modifications when the relative infectious exposure of individuals is varied according to spatial scale (through the worms-within-hosts aggregation parameter k) in addition to many other extensions we have proposed throughout. We therefore consider the model we have presented here to be a useful starting point for many interesting directions for future research.

Acknowledgements

RJH, JET and RMA gratefully thank the Bill and Melinda Gates Foundation for research grant support via the DeWorm3 (OPP1129535) award to the Natural History Museum in London (<http://www.gatesfoundation.org/>). CV gratefully acknowledges funding from the NTD Modelling Consortium (OPP1184344) by the Bill and Melinda Gates Foundation in partnership with the Task Force for Global Health (<http://www.taskforce.org/>). The views, opinions, assumptions or any other information set out in this article are solely those of the authors. All authors acknowledge joint Centre funding from the UK Medical Research Council and Department for International Development.

References

- [1] R. L. Pullan, P. W. Gething, J. L. Smith, C. S. Mwandawiro, H. J. Sturrock, C. W. Gitonga, S. I. Hay, S. Brooker, Spatial modelling of soil-transmitted helminth infections in kenya: a disease control planning tool, *PLoS neglected tropical diseases* 5 (2) (2011).
- [2] N. Schur, E. Hürilimann, A.-S. Stensgaard, K. Chimfwembe, G. Mushinge, C. Simoonga, N. B. Kabatereine, T. K. Kristensen, J. Utzinger, P. Vounatsou, Spatially explicit schistosoma infection risk in eastern africa using bayesian geostatistical modelling, *Acta tropica* 128 (2) (2013) 365–377 (2013).
- [3] R. L. Pullan, J. L. Smith, R. Jasrasaria, S. J. Brooker, Global numbers of infection and disease burden of soil transmitted helminth infections in 2010, *Parasites & vectors* 7 (1) (2014) 37 (2014).
- [4] C. L. Moyes, A. J. Henry, N. Golding, Z. Huang, B. Singh, J. K. Baird, P. N. Newton, M. Huffman, K. A. Duda, C. J. Drakeley, et al., Defining the geographical range of the plasmodium knowlesi reservoir, *PLoS neglected tropical diseases* 8 (3) (2014) e2780 (2014).
- [5] D.-A. Karagiannis-Voules, P. Biedermann, U. F. Ekpo, A. Garba, E. Langer, E. Mathieu, N. Midzi, P. Mwinzi, A. M. Polderman, G. Raso, et al., Spatial and temporal distribution of soil-transmitted helminth infection in sub-saharan africa: a systematic review and geostatistical meta-analysis, *The Lancet infectious diseases* 15 (1) (2015) 74–84 (2015).
- [6] P. Moraga, J. Cano, R. F. Baggaley, J. O. Gyapong, S. M. Njenga, B. Nikolay, E. Davies, M. P. Rebollo, R. L. Pullan, M. J. Bockarie, et al., Modelling the distribution and transmission intensity of lymphatic filariasis in sub-saharan africa prior to scaling up interventions: integrated use of geostatistical and mathematical modelling, *Parasites & vectors* 8 (1) (2015) 560 (2015).
- [7] N. A. Hamm, R. J. Soares Magalhães, A. C. Clements, Earth observation, spatial data quality, and neglected tropical diseases, *PLoS neglected tropical diseases* 9 (12) (2015) e0004164 (2015).
- [8] S. J. O'Hanlon, H. C. Slater, R. A. Cheke, B. A. Boatin, L. E. Cof-feng, S. D. S. Pion, M. Boussinesq, H. G. M. Zour, W. A. Stolk, M.-G. Basez, [Model-based geostatistical mapping of the prevalence of onchocerca volvulus in west africa](https://doi.org/10.1371/journal.pntd.0004328), *PLoS Neglected Tropical Diseases* 10 (1) (2016) 1–36 (01 2016). doi:10.1371/journal.pntd.0004328. URL <https://doi.org/10.1371/journal.pntd.0004328>
- [9] M. Ciddio, L. Mari, S. H. Sokolow, G. A. De Leo, R. Casagrandi, M. Gatto, The spatial spread of schistosomiasis: A multidimensional network model applied to saint-louis region, senegal, *Advances in water resources* 108 (2017) 406–415 (2017).
- [10] L. Mari, M. Gatto, M. Ciddio, E. D. Dia, S. H. Sokolow, G. A. De Leo, R. Casagrandi, Big-data-driven modeling unveils country-wide drivers of endemic schistosomiasis, *Scientific reports* 7 (1) (2017) 489 (2017).
- [11] R. M. Flueckiger, E. Giorgi, J. Cano, M. Abdala, O. N. Amiel, G. Baayenda, A. Bakhtiari, W. Batcho, K. H. Bennawi, M. Dejene, et al., Understanding the spatial distribution of trichiasis and its association with trachomatous inflammation/follicular, *BMC infectious diseases* 19 (1) (2019) 364 (2019).
- [12] C. Vegvari, J. E. Truscott, K. Kura, R. M. Anderson, Human population movement can impede the elimination of soil-transmitted helminth transmission in regions with heterogeneity in mass drug administration coverage and transmission potential between villages: a metapopulation analysis, *Parasites & vectors* 12 (1) (2019) 438 (2019).
- [13] R. J. Hardwick, C. Vegvari, J. E. Truscott, R. M. Anderson, The breakpoint of soil-transmitted helminths with infected human migration, *Journal of theoretical biology* (2019) 110076 (2019).
- [14] R. J. Hardwick, M. Werkman, J. E. Truscott, R. M. Anderson, Stochastic challenges to interrupting helminth transmission, *Epidemics* 34 (2021) 100435 (2021). doi:<https://doi.org/10.1016/j.epidem.2021.100435>.
- [15] S. Brooker, J. Bethony, P. J. Hotez, [Human hookworm infection in the 21st century](https://doi.org/10.1016/S0065-308X(04)58004-1), Vol. 58 of *Advances in Parasitology*, Academic Press, 2004, pp. 197 – 288 (2004). doi:[https://doi.org/10.1016/S0065-308X\(04\)58004-1](https://doi.org/10.1016/S0065-308X(04)58004-1). URL <http://www.sciencedirect.com/science/article/pii/S0065308X04580041>
- [16] D. W. T. Crompton, [Preventive chemotherapy in human helminthiasis : coordinated use of anthelmintic drugs in control interventions : a manual for health professionals and programme managers](https://doi.org/10.1181/whc.2006.06), World Health Organization, 2006 (2006). URL <http://www.who.int/iris/handle/10665/43545>
- [17] [Ending the neglect and reaching 2020 goals](https://unitingtocombatntds.org/london-declaration-neglected-tropical-diseases/), <https://unitingtocombatntds.org/london-declaration-neglected-tropical-diseases/>.
- [18] [Progress reports of the london declaration](https://unitingtocombatntds.org/reports/), <https://unitingtocombatntds.org/reports/>.
- [19] [Accelerating work to overcome the global impact of neglected tropical diseases : a roadmap for implementation : executive summary](https://www.who.int/iris/handle/10665/70809), World Health Organization, 2012 (2012). URL <http://www.who.int/iris/handle/10665/70809>
- [20] K. E. Halliday, W. E. Oswald, M. Mcharo, E. Beaumont, P. M. Gichuki, S. Kepha, S. S. Witek-McManus, S. H. Matendecheo, H. El-Busaidy, R. Muendo, et al., Community-level epidemiology of soil-transmitted helminths in the context of school-based deworming: Baseline results of a cluster randomised trial on the coast of kenya, *PLoS neglected tropical diseases* 13 (8) (2019) e0007427 (2019).
- [21] K. H. Ásbjörnsdóttir, S. S. R. Ajjampur, R. M. Anderson, R. Bailey, I. Gardiner, K. E. Halliday, M. Ibikounle, K. Kalua, G. Kang, D. T. J. Littlewood, A. J. F. Luty, A. R. Means, W. Oswald, R. L. Pullan, R. Sarkar, F. Schr, A. Szpiro, J. E. Truscott, M. Werkman, E. Yard, J. L. Walson, T. D. T. Team, [Assessing the feasibility of interrupting the transmission of soil-transmitted helminths through mass drug administration: The deworm3 cluster randomized trial protocol](https://doi.org/10.1371/journal.pntd.0006166), *PLoS Neglected Tropical Diseases* 12 (1) (2018) 1–16 (01 2018). doi:10.1371/journal.pntd.0006166. URL <https://doi.org/10.1371/journal.pntd.0006166>
- [22] D. A. Raichlen, B. M. Wood, A. D. Gordon, A. Z. Mabulla, F. W. Marlowe, H. Pontzer, Evidence of lévy walk foraging patterns in human hunter-gatherers, *Proceedings of the National Academy of Sciences* 111 (2) (2014) 728–733 (2014).
- [23] I. Rhee, M. Shin, S. Hong, K. Lee, S. J. Kim, S. Chong, On the levy-walk

- 1001 nature of human mobility, *IEEE/ACM transactions on networking (TON)*
1002 19 (3) (2011) 630–643 (2011).
- 1003 [24] K. S. Kung, K. Greco, S. Sobolevsky, C. Ratti, Exploring universal pat-
1004 terns in human home-work commuting from mobile phone data, *PLoS one*
1005 9 (6) (2014) e96180 (2014).
- 1006 [25] M. Lenormand, B. Gonçalves, A. Tugores, J. J. Ramasco, Human diffu-
1007 sion and city influence, *Journal of The Royal Society Interface* 12 (109)
1008 (2015) 20150473 (2015).
- 1009 [26] K. Zhao, M. Musolesi, P. Hui, W. Rao, S. Tarkoma, Explaining the power-
1010 law distribution of human mobility through transportation modality de-
1011 composition, *Scientific reports* 5 (1) (2015) 1–7 (2015).
- 1012 [27] C. Cosner, J. C. Beier, R. S. Cantrell, D. Impoinvil, L. Kapitanski, M. D.
1013 Potts, A. Troyo, S. Ruan, The effects of human movement on the per-
1014 sistence of vector-borne diseases, *Journal of theoretical biology* 258 (4)
1015 (2009) 550–560 (2009).
- 1016 [28] D. T. Citron, D. L. Smith, S. L. Wu, H. M. Sanchez Castellanos,
1017 A. J. Dolgert, J. M. Henry, C. A. Guerra, [Comparing metapopulation](#)
1018 [dynamics of infectious diseases under different models of human](#)
1019 [movement](#), medRxiv (2020). [arXiv:https://www.medrxiv.org/](https://www.medrxiv.org/content/early/2020/04/07/2020.04.05.20054304.full.pdf)
1020 [content/early/2020/04/07/2020.04.05.20054304](https://www.medrxiv.org/content/early/2020/04/07/2020.04.05.20054304),
1021 [doi:10.1101/2020.04.05.20054304](https://www.medrxiv.org/content/early/2020/04/07/2020.04.05.20054304).
1022 URL [https://www.medrxiv.org/content/early/2020/04/07/](https://www.medrxiv.org/content/early/2020/04/07/2020.04.05.20054304)
1023 [2020.04.05.20054304](https://www.medrxiv.org/content/early/2020/04/07/2020.04.05.20054304)
- 1024 [29] W. S.-t. Helminthiasis, Eliminating soil-transmitted helminthiasis as a
1025 public health problem in children: progress report 2001–2010 and strate-
1026 gic plan 2011–2020, France: World Health Organization (2012) 19–29
1027 (2012).
- 1028 [30] L. Taylor, No place to hide? the ethics and analytics of tracking mobil-
1029 ity using mobile phone data, *Environment and Planning D: Society and*
1030 *Space* 34 (2) (2016) 319–336 (2016).
- 1031 [31] [High Resolution Settlement Layer \(HRSL\)](#), Facebook Connectivity Lab
1032 and Center for International Earth Science Information Network - CIESIN
1033 - Columbia University, 2016, Source imagery for HRSL ©2016 Digital-
1034 Globe, Accessed (31/10/2019).
- 1035 [32] L. Steinbaum, S. M. Njenga, J. Kihara, A. B. Boehm, J. Davis, C. Null,
1036 A. J. Pickering, Soil-transmitted helminth eggs are present in soil at mul-
1037 tiple locations within households in rural kenya, *PLoS One* 11 (6) (2016)
1038 e0157780 (2016).
- 1039 [33] L. Steinbaum, J. Mboya, R. Mahoney, S. M. Njenga, C. Null, A. J.
1040 Pickering, Effect of a sanitation intervention on soil-transmitted helminth
1041 prevalence and concentration in household soil: A cluster-randomized
1042 controlled trial and risk factor analysis, *PLoS neglected tropical diseases*
1043 13 (2) (2019) e0007180 (2019).
- 1044 [34] A. Reynolds, E. Ceccon, C. Baldauf, T. Karina Medeiros, O. Miramontes,
1045 Lévy foraging patterns of rural humans, *PLoS one* 13 (6) (2018) e0199099
1046 (2018).
- 1047 [35] A. D. Mirlin, Y. V. Fyodorov, F.-M. Dittes, J. Quezada, T. H. Selig-
1048 man, Transition from localized to extended eigenstates in the ensemble
1049 of power-law random banded matrices, *Physical Review E* 54 (4) (1996)
1050 3221 (1996).
- 1051 [36] R. Anderson, R. May, [Infectious Diseases of Humans: Dynamics and](#)
1052 [Control](#), Dynamics and Control, OUP Oxford, 1992 (1992).
1053 URL <https://books.google.co.uk/books?id=HT0--xXBguQC>
- 1054 [37] R. M. Anderson, R. M. May, Population dynamics of human helminth in-
1055 fections: control by chemotherapy, *Nature* 297 (5867) (1982) 557 (1982).
- 1056 [38] C. D. Criscione, J. D. Anderson, D. Sudimack, J. Subedi, R. P. Upad-
1057 hayay, B. Jha, K. D. Williams, S. Williams-Blangero, T. J. Anderson,
1058 Landscape genetics reveals focal transmission of a human macroparasite,
1059 *PLoS Neglected Tropical Diseases* 4 (4) (2010) e665 (2010).
- 1060 [39] L. Steinbaum, L. H. Kwong, A. Ercumen, M. S. Negash, A. J. Lovely,
1061 S. M. Njenga, A. B. Boehm, A. J. Pickering, K. L. Nelson, Detecting
1062 and enumerating soil-transmitted helminth eggs in soil: New method de-
1063 velopment and results from field testing in kenya and bangladesh, *PLoS*
1064 *neglected tropical diseases* 11 (4) (2017) e0005522 (2017).
- 1065 [40] J. Truscott, H. Turner, S. Farrell, R. Anderson, [Chapter three - soil-](#)
1066 [transmitted helminths: Mathematical models of transmission, the](#)
1067 [impact of mass drug administration and transmission elimination](#)
1068 [criteria](#), in: M. G. Basez, R. M. Anderson (Eds.), *Mathematical*
1069 *Models for Neglected Tropical Diseases*, Vol. 94 of *Advances*
1070 *in Parasitology*, Academic Press, 2016, pp. 133 – 198 (2016).
1071 [doi:https://doi.org/10.1016/bs.apar.2016.08.002](https://doi.org/10.1016/bs.apar.2016.08.002).

URL <http://www.sciencedirect.com/science/article/pii/S0065308X1630077X>

1072

1073