Spatial sampling for a rabies vaccination schedule in rural villages

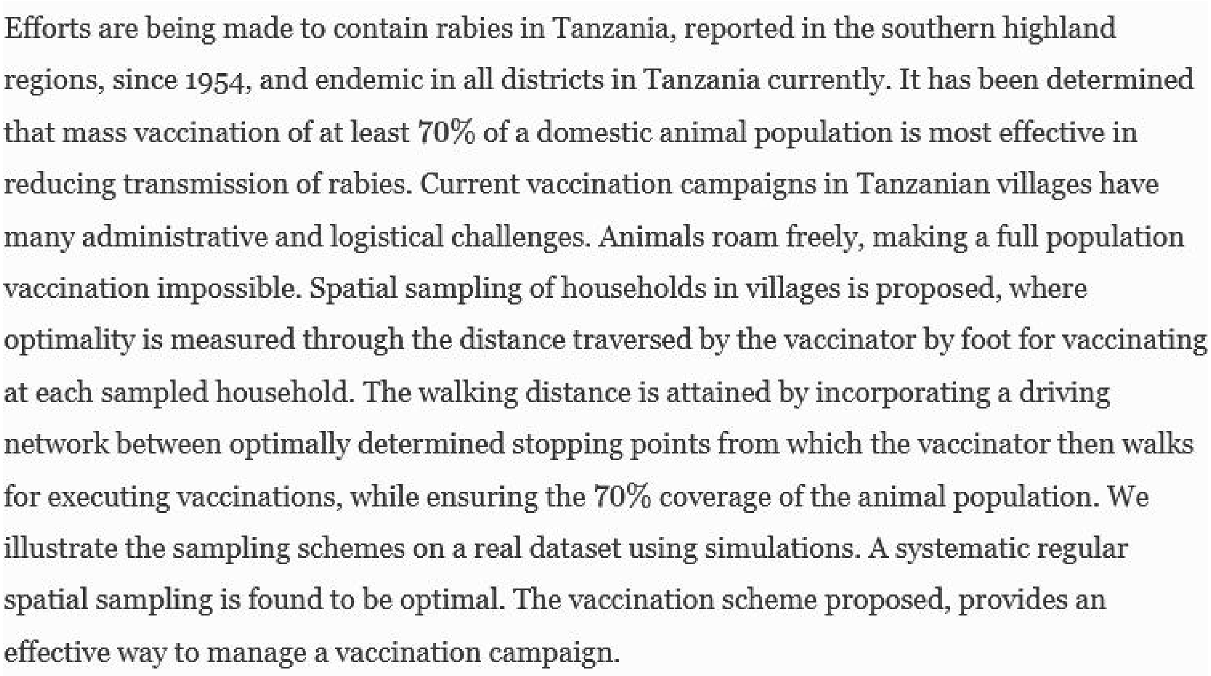
Inger Fabris-Rotelli[1], Hayley Reynolds[1], Alfred Stein [1,2],·Theodor Loots[1]

[1]University of Pretoria, Pretoria, South Africa

[2]University of Twente, Enschede, Netherlands

*Correspondence to inger.fabris-rotelli@up.ac.za

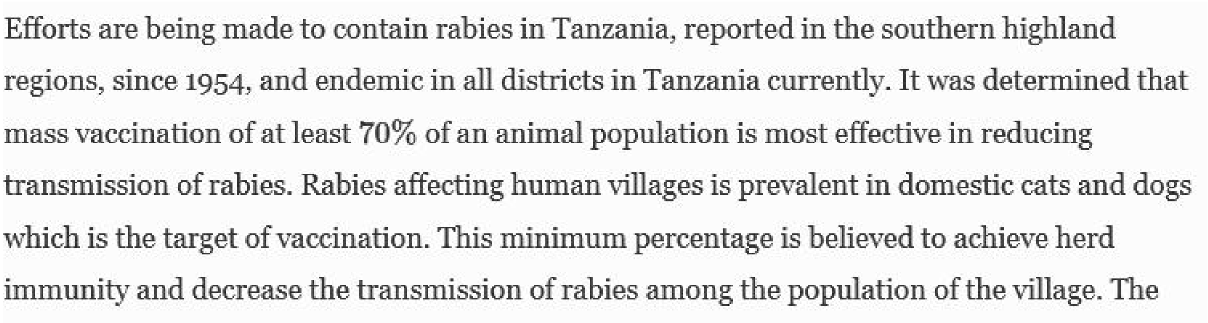
Abstract



Efforts are being made to contain rabies in Tanzania, reported in the southern highland regions, since 1954, and endemic in all districts in Tanzania currently. It has been determined that mass vaccination of at least 70% of a domestic animal population is most effective in reducing transmission of rabies. Current vaccination campaigns in Tanzanian villages have many administrative and logistical challenges. Animals roam freely, making a full population vaccination impossible. Spatial sampling of households in villages is proposed, where optimality is measured through the distance traversed by the vaccinator by foot for vaccinating at each sampled household. The walking distance is attained by incorporating a driving network between optimally determined stopping points from which the vaccinator then walks for executing vaccinations, while ensuring the 70% coverage of the animal population. We illustrate the sampling schemes on a real dataset using simulations. A systematic regular spatial sampling is found to be optimal. The vaccination scheme proposed, provides an effective way to manage a vaccination campaign.

Keywords: Environmental sampling, Rabies vaccination,·Spatial sampling,·Spatial data

Introduction

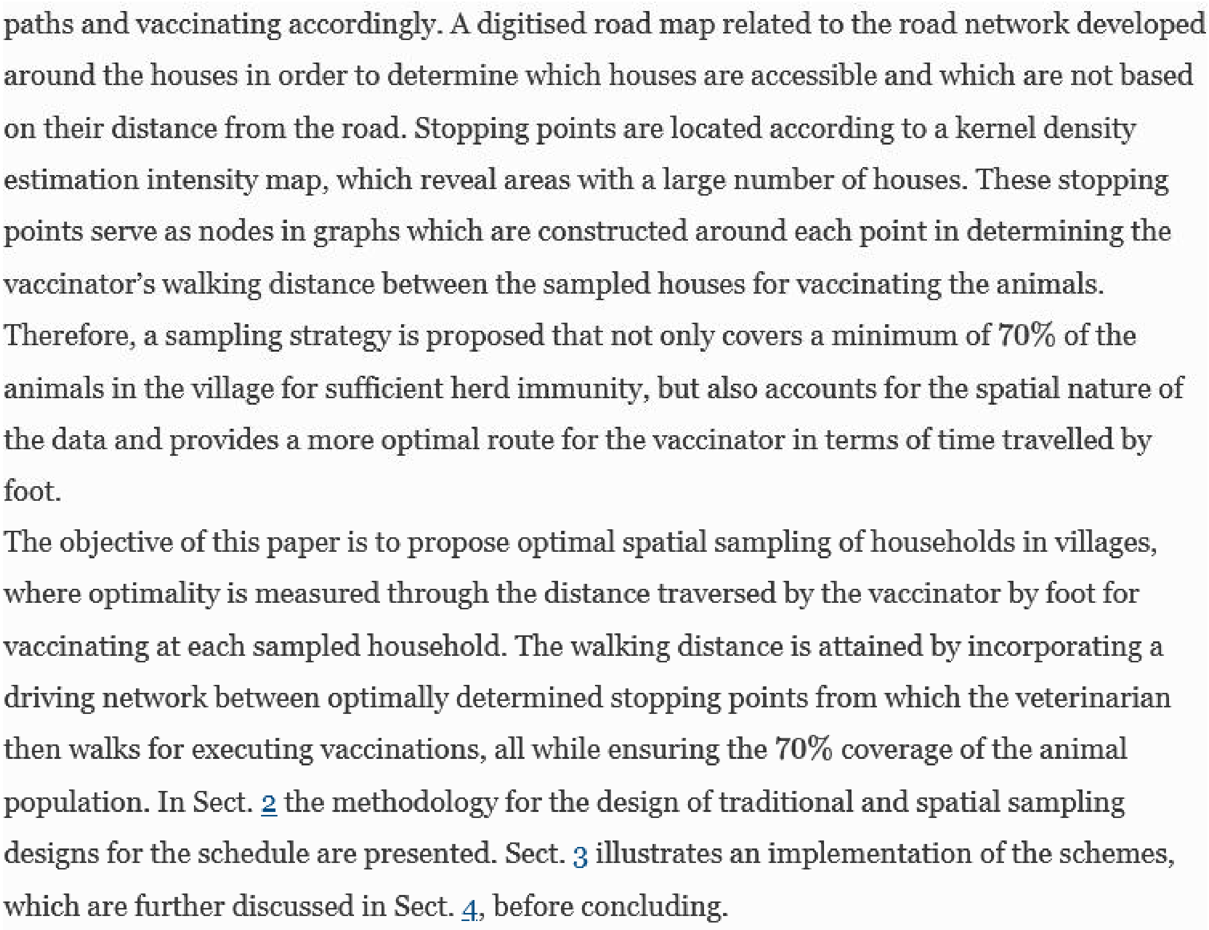


Efforts are being made to contain rabies in Tanzania, reported in the southern highland regions, since 1954, and endemic in all districts in Tanzania currently. It was determined that mass vaccination of at least 70% of an animal population is most effective in reducing transmission of rabies. Rabies affecting human villages is prevalent in domestic cats and dogs which is the target of vaccination. This minimum percentage is believed to achieve herd immunity and decrease the transmission of rabies among the population of the village. The minimum animal coverage of 70%, is suggested by the World Health Organisation (Kayali et al. 2003; Zinsstag et al. 2009; Coleman and Dye 1996; Cleaveland et al. 2003) as being most cost-effective. Achieving the 70% coverage has, however, been experienced (see Mpolya et al. 2017) as hindered by logistical and administrative challenges[1]. Currently the process for vaccinating animals in rural Tanzanian villages is that a vaccination station is set up in the middle of the village with the expectation that the villagers will bring their animals for the necessary treatment. Obviously, this approach is ineffective. An alternative used for vaccination in Tanzanian villages takes some features from the EPI cluster survey method (Bostoen and Chalabi 2006). The method is rather basic and unreliable, as the vaccinator starts in the middle of the village and then chooses a random direction. A random house is chosen in that direction and the vaccinator continues as such. This does not yield good results for rabies vaccination, shown in (Kraamwinkel et al. 2014).

For these reasons, the optimal road network and walking route amongst houses, over which the vaccinator will travel to the animals, is herein proposed. This road network is optimal for driving through the village while ensuring 70% coverage of the animals within a village. The concept is to drive along the optimal route from location to location, which will be referred to as stopping points. At each stopping point, the vaccinator will exit the vehicle and optimally walk among the houses so as to minimise the walking distance, while vaccinating the animals at the sampled houses. The vaccinator will then drive to the next stopping point, walking among the houses, vaccinating the animals and so on.

The focus for optimisation is to minimise the distance walked by the vaccinator rather than the money spent, thereby focusing on the time constraint of the approach only. Driving time is considered to be negligible, but the walking time at each stopping point, costly. The time to administer the vaccination per animal is considered to be deterministic and not taken into account in calculating the cost. The animals also roam freely and are thus difficult to include in the model. However, the administration of the vaccination is still not considered to be difficult to perform in practice under these circumstances.

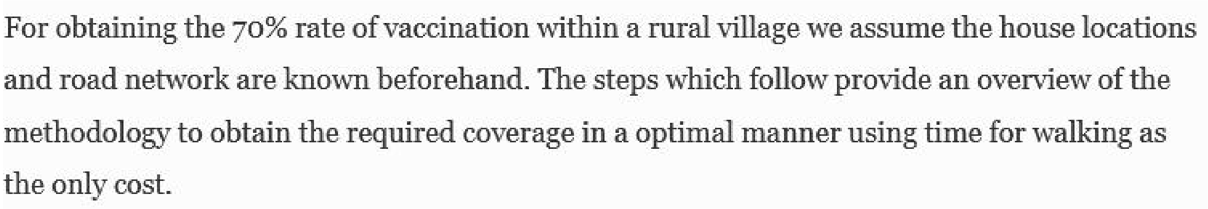
Using spatial statistics when spatial data are available has been growing in importance (Stein et al. 1998). Spatial sampling on this network is proposed here, optimally achieving the required coverage, including a comparison in which traditional sampling is used. The strategy draws spatial samples of the houses, incorporating graph theory in obtaining optimal walking



paths and vaccinating accordingly. A digitised road map related to the road network developed around the houses in order to determine which houses are accessible and which are not based on their distance from the road. Stopping points are located according to a kernel density estimation intensity map, which reveal areas with a large number of houses. These stopping points serve as nodes in graphs which are constructed around each point in determining the vaccinator's walking distance between the sampled houses for vaccinating the animals. Therefore, a sampling strategy is proposed that not only covers a minimum of 70% of the animals in the village for sufficient herd immunity, but also accounts for the spatial nature of the data and provides a more optimal route for the vaccinator in terms of time travelled by foot.

The objective of this paper is to propose optimal spatial sampling of households in villages, where optimality is measured through the distance traversed by the vaccinator by foot for vaccinating at each sampled household. The walking distance is attained by incorporating a driving network between optimally determined stopping points from which the veterinarian then walks for executing vaccinations, all while ensuring the 70% coverage of the animal population. In Sect. 2 the methodology for the design of traditional and spatial sampling designs for the schedule are presented. Sect. 3 illustrates an implementation of the schemes, which are further discussed in Sect. 4, before concluding.

Methodology



For obtaining the 70% rate of vaccination within a rural village we assume the house locations and road network are known beforehand. The steps which follow provide an overview of the methodology to obtain the required coverage in a optimal manner using time for walking as the only cost.

.

1. An appropriate spatial window is chosen at the start. This is important as it declares the area in which observations can be observed (Baddeley et al. 2015). Most often rectangular or convex hull domains are used (Ripley and Rasson 1977). The reason is mathematical - in such domains the Euclidean distance is defined and measurable. The convex hull is a good choice as a rectangular window won't match the locations of real village households. Villages have irregular shapes and develop as a result of terrain and community, see Fig. 1.
2. Areas of the village with highest density of houses are selected as targets for vaccination. These areas are not exhaustive of the village and result in an initial sampling. We refer to these as stopping points for the vaccinator.
3. Houses within a certain radius, which is determined as a reasonable walking distance, of the stopping points are deemed accessible and are targets for the sampling. At each stopping point, walking is restricted to houses within this

certain radius only and the number of houses is determined as those within walking distance of the respective stopping point. We assume that houses not within the determined radius of the stopping point are excluded. The selection of the number of stopping points is done so that more than the 70% of the accessible houses are still available for sampling. Note that the houses will be sampled and there will not be complete coverage.

4. The stopping points and the road network are modelled as a graph. An advised driving route is provided to the vaccinator between these stopping points. The time for driving is considered negligible being a few minutes, as the walking required between houses at each stopping point takes far more time depending on the radius chosen. The environment under consideration is rural, resulting in a limited, relatively small area with little likelihood for good driving conditions off the main road network. Walking, of course, provides the worst case scenario; if the route is drivable between houses at a stopping point this can be done and time saved.
5. The houses at each stopping point selected within a minimum radius from the stopping point, are then sampled. For this purpose we propose and simulate 8 possible sampling schemes, described in Sect. 2.6. This sampling is done so that the sample size *n* is 70% of the total number of houses at all the stopping points *N*.
6. The optimal walking route for vaccinator amongst the sampled houses is determined,
7. The cost of the scheme is determined as the total walking distance amongst the houses sampled at every stopping point for the vaccinator using the optimal walk route determined.
8. The sampling is repeated 1000 times for each scheme to determine a bootstrap distribution for the cost of each scheme.

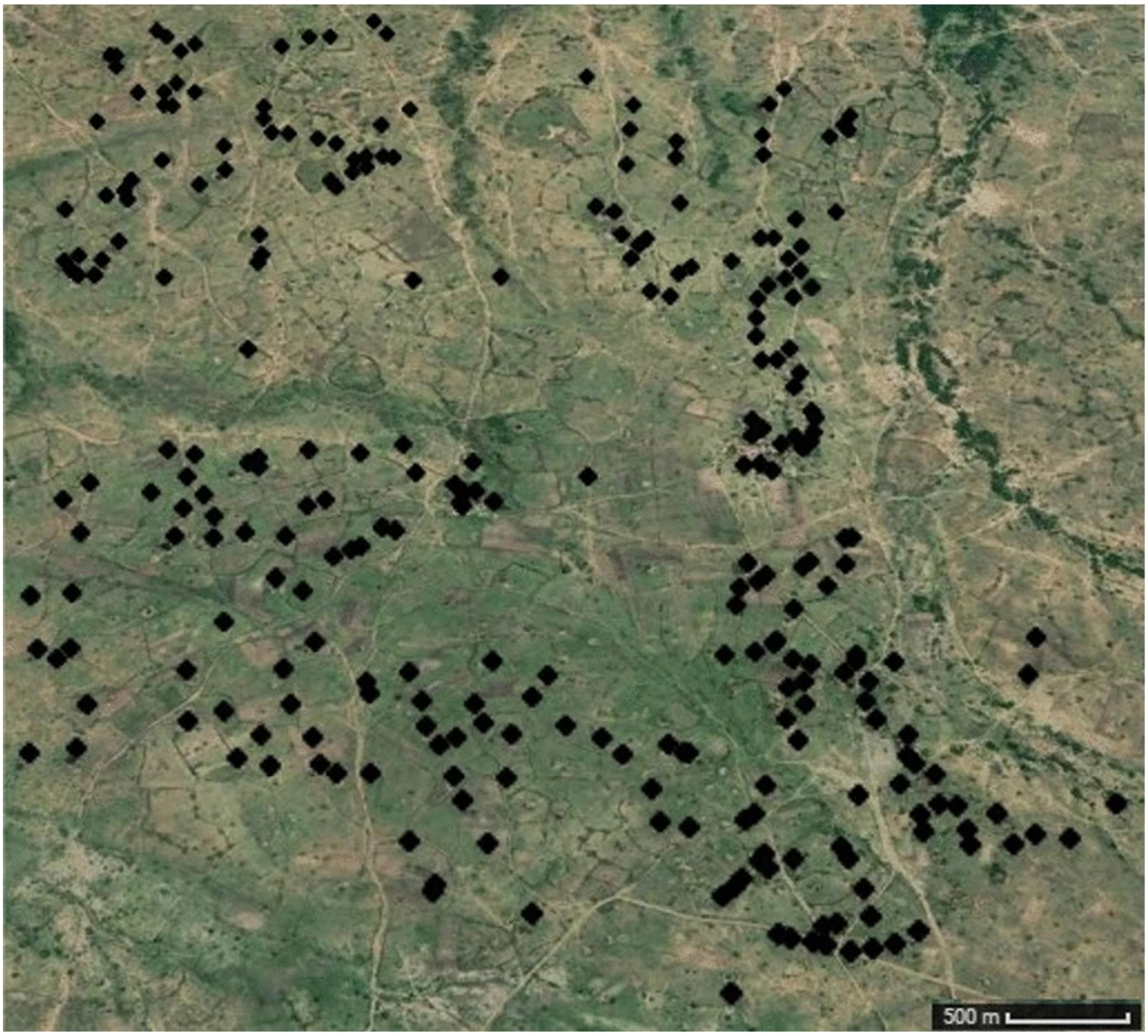
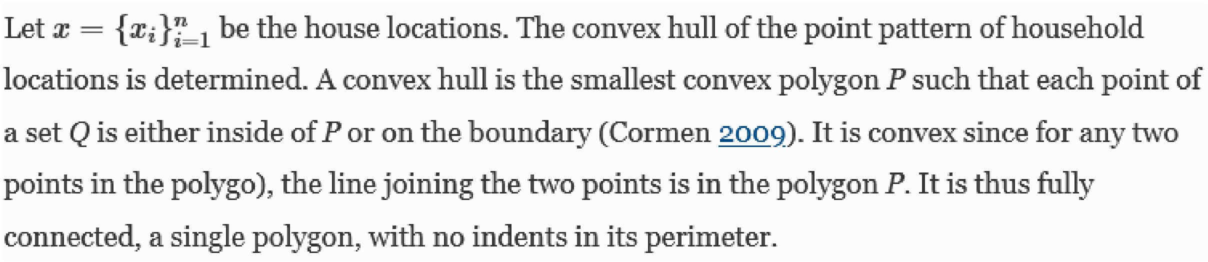


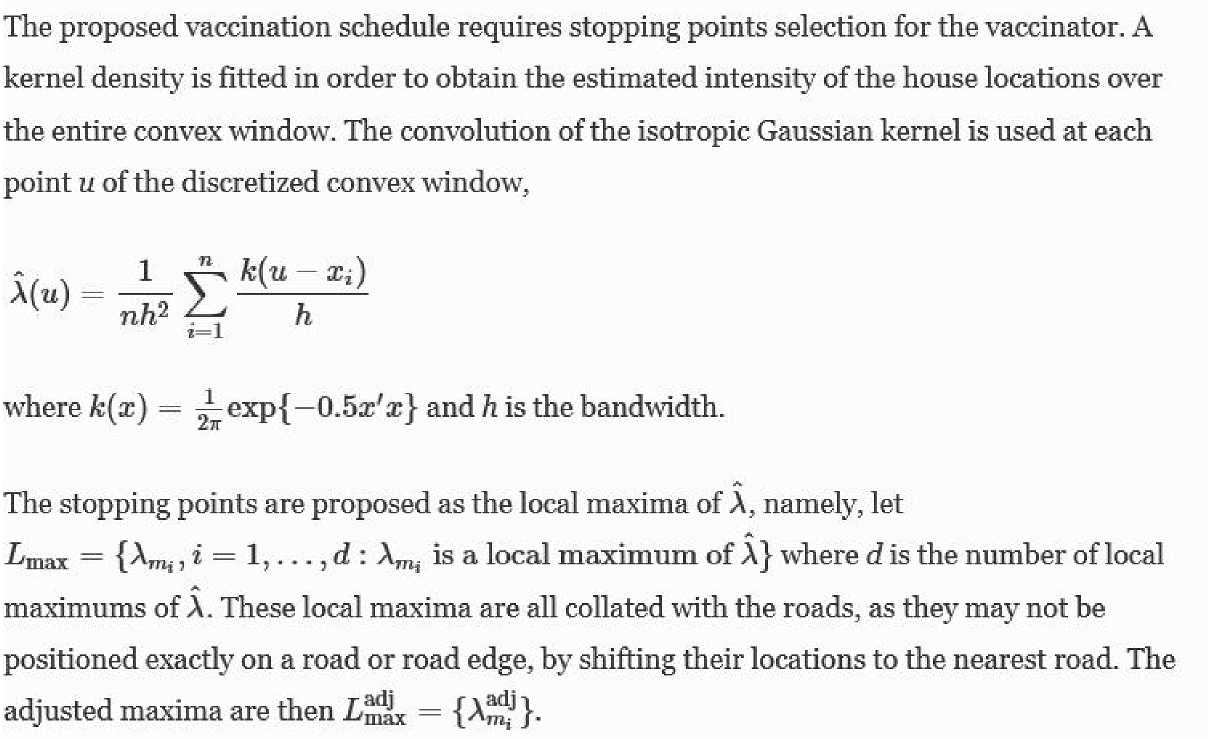
Fig. 1. A Google Earth image with the locations of the 280 houses in the village of Buchanchari, Tanzania

Window selection



Let $x = \{x_i\}_{i=1}^{n}$ be the house locations. The convex hull of the point pattern of household locations is determined. A convex hull is the smallest convex polygon P such that each point of a set Q is either inside of P or on the boundary (Cormen 2009). It is convex since for any two points in the polygo), the line joining the two points is in the polygon P. It is thus fully connected, a single polygon, with no indents in its perimeter.

Stopping point selection



The proposed vaccination schedule requires stopping points selection for the vaccinator. A kernel density is fitted in order to obtain the estimated intensity of the house locations over the entire convex window. The convolution of the isotropic Gaussian kernel is used at each point u of the discretized convex window,

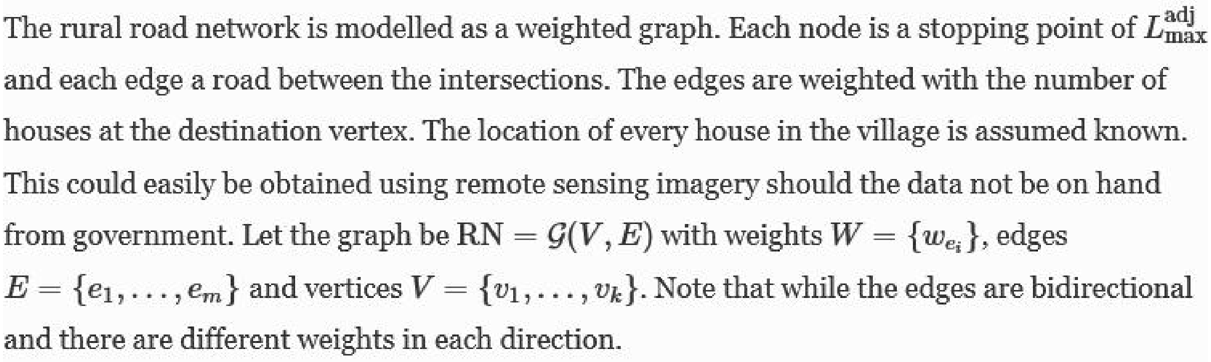
$$\hat{\lambda}(u) = \frac{1}{nh^2} \sum_{i=1}^{n} \frac{k(u - x_i)}{h}$$

where $k(x) = \frac{1}{2\pi}\exp\{-0.5x'x\}$ and h is the bandwidth.

The stopping points are proposed as the local maxima of $\hat{\lambda}$, namely, let $L_{\max} = \{\lambda_{m_i}, i = 1, \ldots, d : \lambda_{m_i} \text{ is a local maximum of } \hat{\lambda}\}$ where d is the number of local maximums of $\hat{\lambda}$. These local maxima are all collated with the roads, as they may not be positioned exactly on a road or road edge, by shifting their locations to the nearest road. The adjusted maxima are then $L_{\max}^{\text{adj}} = \{\lambda_{m_i}^{\text{adj}}\}$.

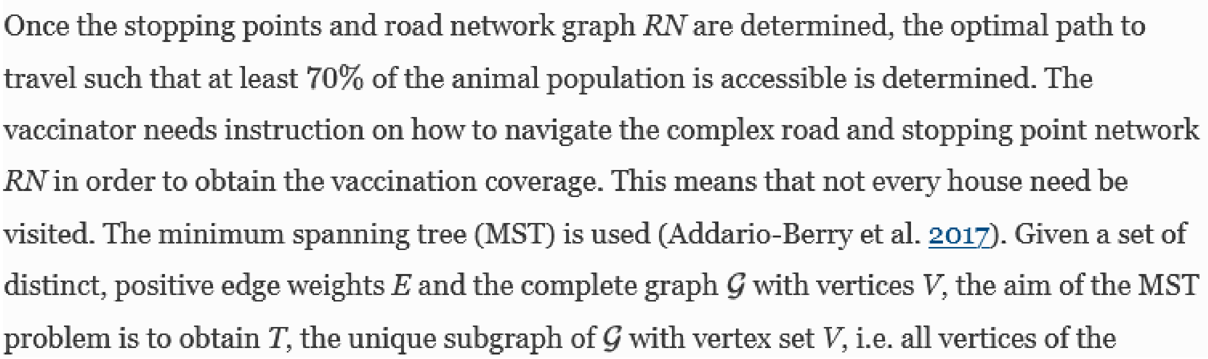
.

The rural road network

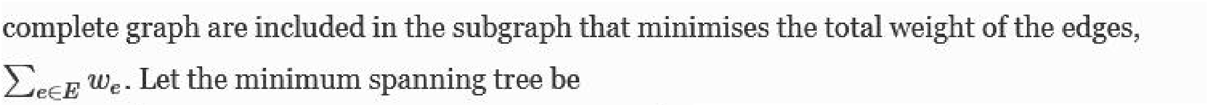


The rural road network is modelled as a weighted graph. Each node is a stopping point of $L_{\max}^{\text{adj}}$ and each edge a road between the intersections. The edges are weighted with the number of houses at the destination vertex. The location of every house in the village is assumed known. This could easily be obtained using remote sensing imagery should the data not be on hand from government. Let the graph be $\text{RN} = \mathcal{G}(V, E)$ with weights $W = \{w_{e_i}\}$, edges $E = \{e_1, \ldots, e_m\}$ and vertices $V = \{v_1, \ldots, v_k\}$. Note that while the edges are bidirectional and there are different weights in each direction.

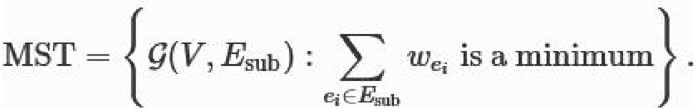
Optimal driving route construction



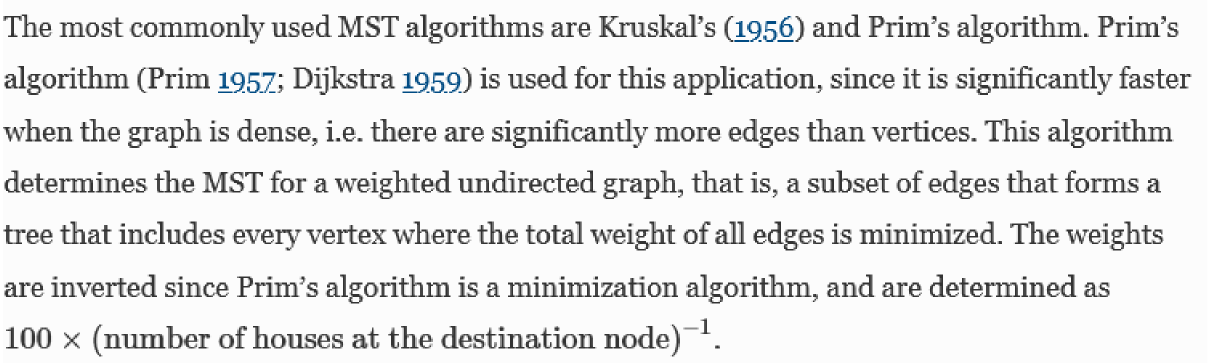
Once the stopping points and road network graph *RN* are determined, the optimal path to travel such that at least 70% of the animal population is accessible is determined. The vaccinator needs instruction on how to navigate the complex road and stopping point network *RN* in order to obtain the vaccination coverage. This means that not every house need be visited. The minimum spanning tree (MST) is used (Addario-Berry et al. 2017). Given a set of distinct, positive edge weights E and the complete graph $\mathcal{G}$ with vertices V, the aim of the MST problem is to obtain T, the unique subgraph of $\mathcal{G}$ with vertex set V, i.e. all vertices of the



complete graph are included in the subgraph that minimises the total weight of the edges, $\sum_{e \in E} w_e$. Let the minimum spanning tree be

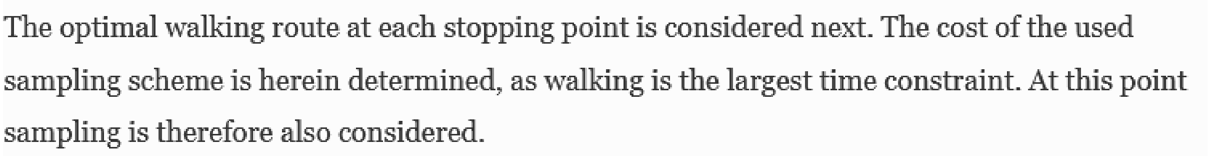


$$\text{MST} = \left\{ \mathcal{G}(V, E_{\text{sub}}) : \sum_{e_i \in E_{\text{sub}}} w_{e_i} \text{ is a minimum} \right\}.$$

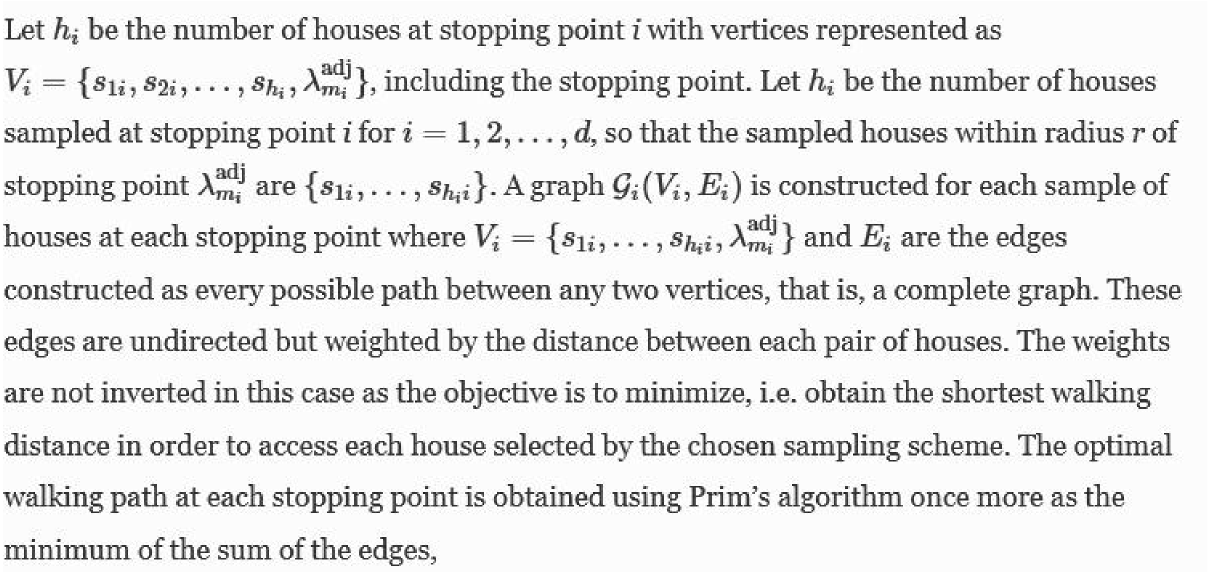


The most commonly used MST algorithms are Kruskal's (1956) and Prim's algorithm. Prim's algorithm (Prim 1957; Dijkstra 1959) is used for this application, since it is significantly faster when the graph is dense, i.e. there are significantly more edges than vertices. This algorithm determines the MST for a weighted undirected graph, that is, a subset of edges that forms a tree that includes every vertex where the total weight of all edges is minimized. The weights are inverted since Prim's algorithm is a minimization algorithm, and are determined as $100 \times (\text{number of houses at the destination node})^{-1}$.

Optimal walking route

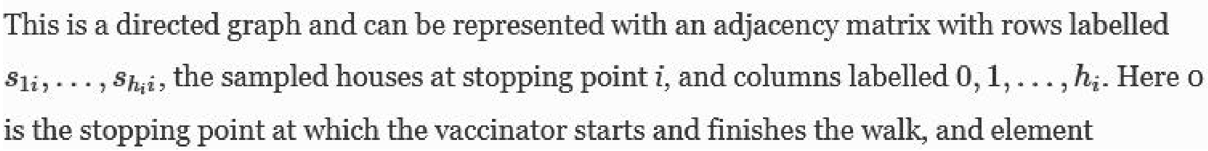


The optimal walking route at each stopping point is considered next. The cost of the used sampling scheme is herein determined, as walking is the largest time constraint. At this point sampling is therefore also considered.



Let h_i be the number of houses at stopping point i with vertices represented as $V_i = \{s_{1i}, s_{2i}, \ldots, s_{h_i}, \lambda_{m_i}^{\text{adj}}\}$, including the stopping point. Let h_i be the number of houses sampled at stopping point i for $i = 1, 2, \ldots, d$, so that the sampled houses within radius r of stopping point $\lambda_{m_i}^{\text{adj}}$ are $\{s_{1i}, \ldots, s_{h_i i}\}$. A graph $\mathcal{G}_i(V_i, E_i)$ is constructed for each sample of houses at each stopping point where $V_i = \{s_{1i}, \ldots, s_{h_i i}, \lambda_{m_i}^{\text{adj}}\}$ and E_i are the edges constructed as every possible path between any two vertices, that is, a complete graph. These edges are undirected but weighted by the distance between each pair of houses. The weights are not inverted in this case as the objective is to minimize, i.e. obtain the shortest walking distance in order to access each house selected by the chosen sampling scheme. The optimal walking path at each stopping point is obtained using Prim's algorithm once more as the minimum of the sum of the edges,

$$\mathcal{G}_{i,\text{opt}} = \mathcal{G}(\mathbf{V}_{i,\text{opt}}, E_{i,\text{opt}}).$$

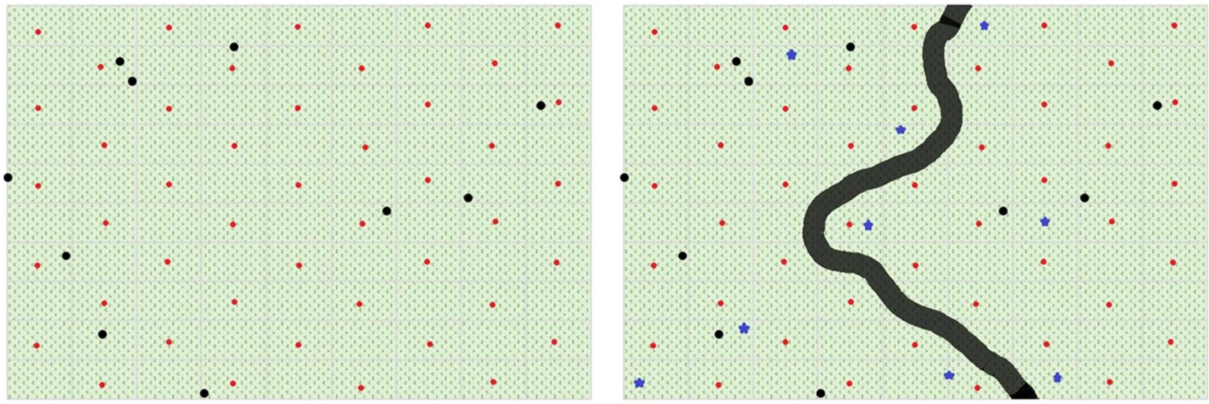


This is a directed graph and can be represented with an adjacency matrix with rows labelled $s_{1i}, \ldots, s_{h_i i}$, the sampled houses at stopping point i, and columns labelled $0, 1, \ldots, h_i$. Here 0 is the stopping point at which the vaccinator starts and finishes the walk, and element

$(s_{ji}, k) = 1$ iff vertex s_{ji} is visited at step k. The optimal route can also be exactly reversed, at the discretion of the vaccinator, with obtained knowledge of the terrain on arriving at the stopping point.

Sampling schemes

Spatial sampling aims to collect samples from higher dimensions by incorporating location into data (Wang et al. 2012). The focus here is on design-based sampling, allowing for an uncomplicated comparison between traditional (non-spatial) and spatial sampling techniques with the aim of achieving the 70% rabies vaccination rate. Traditional (non-spatial) and spatial sampling techniques are used, and their strengths and weaknesses discussed.



(a) Plot of independently selected uniform random pairs (black) and a spatial systematic regular sample (in red)

(b) A river and house locations (blue) added

Fig. 2. Traditional versus spatial sampling

The importance of using spatial sampling over traditional sampling can be illustrated with an example. Figure 2b shows location of village houses in blue and a river through the area. If one samples in a traditional random way (Fig. 2a black points), each point has an equal and independent probability of being sampled (Ripley 2005). Practically, two random numbers K_i and $K^!_i$ are generated from a $\text{Uniform}(0, 1)$ distribution. The point $u_i = (x_i, y_i)$ is then sampled such that $x_i = K_i L_x$ and $y_i = K^!_i L_y$, where L_x and L_y denote the lengths of the study area in both directions. This procedure is repeated until m samples are obtained (Delmelle 2011). The red points in Fig. 2a provide a spatial systematic regular sample. Without the knowledge of the river the black sampled points could be argued to be sufficient. With the knowledge of the river and the house locations the black points will not pick up accurate animal information from the house locations. The animals are free-roaming and likely to not cross a large river. As with most spatial data, nearby units interact with each other and this information is pertinent to the efficiency of the sampling plan for the region. The red points achieve more dispersion. Traditional sampling could result in local clustering and low dispersion.

Obtaining the sample of houses at each stopping point

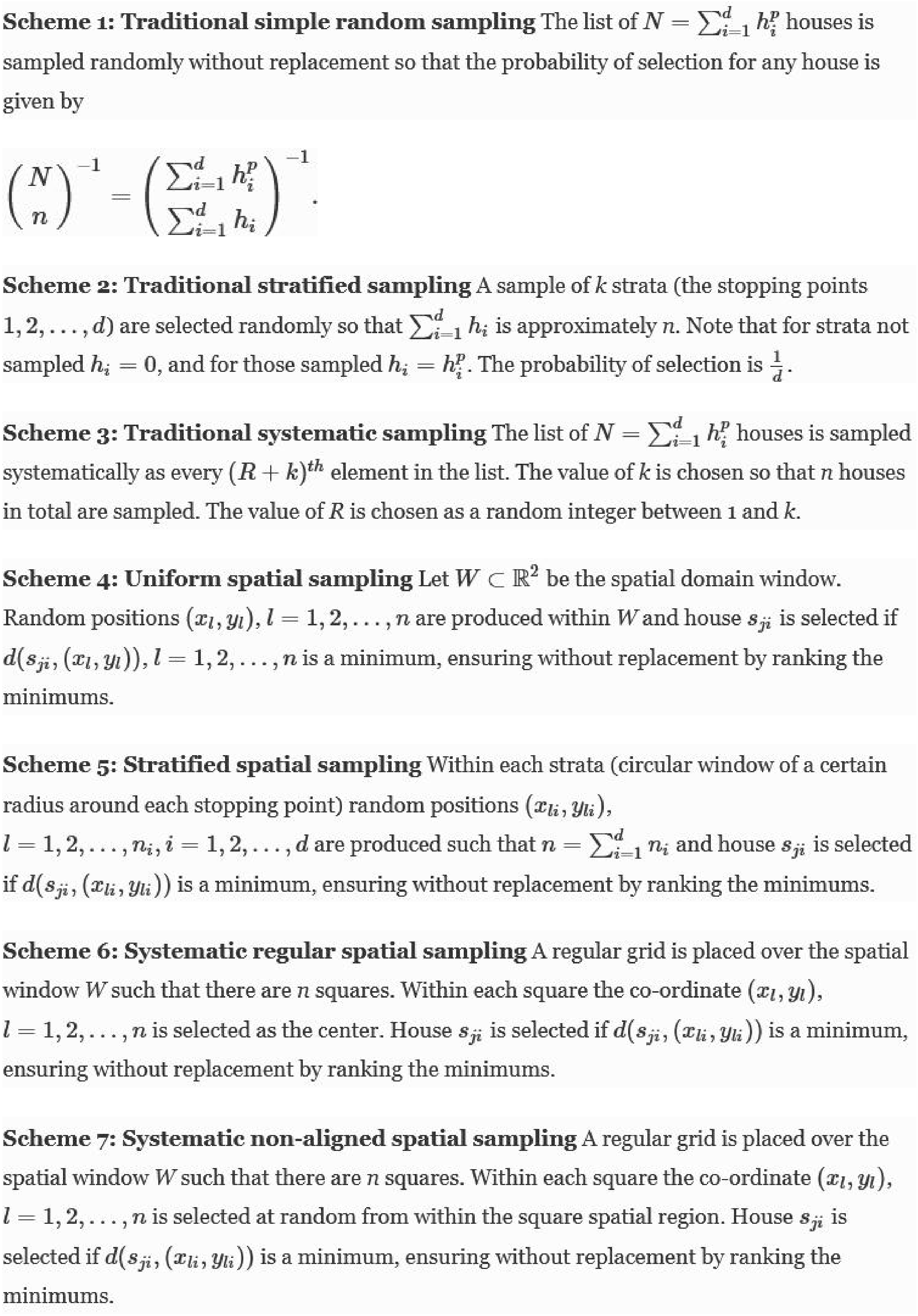
$$V_i = \left\{ s_{1i}, \ldots, s_{h_i i}, \lambda_{m_i}^{\text{adj}} \right\}$$

was done in 8 different ways; 3 traditional (non-spatial) techniques and 5 spatial techniques. The traditional sampling methods implemented were simple random, stratified and systematic sampling, with the stopping points and the houses within the specific radius the strata, and the houses as the sampling units. Houses are chosen as sampling units as in reality the locations of animals will be unknown. All animals at the sampled houses are vaccinated. Here, only a single animal is assumed to be present at each house for simplicity. The method can easily be adapted if more than one animal or less than one animal is assumed, or if prior information is available. Traditional sampling does not account for the spatial relationship between observations and the sampling is done directly on the list of houses, not the geographical window. These methods are thus predicted to result in a less efficient solution, in terms of the cost of the total walking distance.

The spatial sampling methods implemented are uniform, stratified and systematic sampling. Uniform random sampling draws spatial positions randomly from the specified window. Stratified spatial sampling uses the stopping points as strata and random coordinates are sampled from each strata. Systematic spatial sampling is applied in three ways. First, a regular grid is used and a single observation chosen at the center of each cell, also known as centric systematic sampling. Second, the regular grid is used but to have the single observation in each cell randomly positioned, non-aligned. Third, a hexagonal grid beehive-like structure with regular positioning within each cell, as for the first way of sampling.

In order to ensure that a house is selected to be sampled, the house closest to each sampled point on the continuous window is selected, and it is ensured that a house is not included more than once.

The sampling schemes under consideration sample *n* houses which achieve the required coverage are as follows. In all cases *n* is 70% of *N* to achieve the coverage. The list of houses that are within reach are referred to as the accessible population, and is the population *N* from which samples are drawn.



Scheme 1: Traditional simple random sampling The list of $N = \sum_{i=1}^{d} h_i^p$ houses is sampled randomly without replacement so that the probability of selection for any house is given by

$$\binom{N}{n}^{-1} = \binom{\sum_{i=1}^{d} h_i^p}{\sum_{i=1}^{d} h_i}^{-1}.$$

Scheme 2: Traditional stratified sampling A sample of *k* strata (the stopping points $1, 2, \ldots, d$) are selected randomly so that $\sum_{i=1}^{d} h_i$ is approximately *n*. Note that for strata not sampled $h_i = 0$, and for those sampled $h_i = h_i^p$. The probability of selection is $\frac{1}{d}$.

Scheme 3: Traditional systematic sampling The list of $N = \sum_{i=1}^{d} h_i^p$ houses is sampled systematically as every $(R + k)^{th}$ element in the list. The value of *k* is chosen so that *n* houses in total are sampled. The value of *R* is chosen as a random integer between 1 and *k*.

Scheme 4: Uniform spatial sampling Let $W \subset \mathbb{R}^2$ be the spatial domain window. Random positions $(x_l, y_l), l = 1, 2, \ldots, n$ are produced within *W* and house s_{ji} is selected if $d(s_{ji}, (x_l, y_l)), l = 1, 2, \ldots, n$ is a minimum, ensuring without replacement by ranking the minimums.

Scheme 5: Stratified spatial sampling Within each strata (circular window of a certain radius around each stopping point) random positions (x_{li}, y_{li}), $l = 1, 2, \ldots, n_i, i = 1, 2, \ldots, d$ are produced such that $n = \sum_{i=1}^{d} n_i$ and house s_{ji} is selected if $d(s_{ji}, (x_{li}, y_{li}))$ is a minimum, ensuring without replacement by ranking the minimums.

Scheme 6: Systematic regular spatial sampling A regular grid is placed over the spatial window *W* such that there are *n* squares. Within each square the co-ordinate (x_l, y_l), $l = 1, 2, \ldots, n$ is selected as the center. House s_{ji} is selected if $d(s_{ji}, (x_{li}, y_{li}))$ is a minimum, ensuring without replacement by ranking the minimums.

Scheme 7: Systematic non-aligned spatial sampling A regular grid is placed over the spatial window *W* such that there are *n* squares. Within each square the co-ordinate (x_l, y_l), $l = 1, 2, \ldots, n$ is selected at random from within the square spatial region. House s_{ji} is selected if $d(s_{ji}, (x_{li}, y_{li}))$ is a minimum, ensuring without replacement by ranking the minimums.

Scheme 8: Systematic hexagonal spatial sampling A hexagonal (beehive) grid is placed over the spatial window W such that there are n hexagons. Within each hexagon the co-ordinate $(x_l, y_l), l = 1, 2, \ldots, n$ is selected as the center. House s_{ji} is selected if $d(s_{ji}, (x_{li}, y_{li}))$ is a minimum, ensuring without replacement by ranking the minimums.

Schemes 5–8 force a spreading of the sample over the spatial window. They avoid the possibility of sampling occurring in a few local areas of the spatial window, unlike schemes 1–4.

Application

The data set considered here is an extensive census with information regarding the villages in Tanzania, the location of the houses within the villages and the number and type of animals at each of the houses (cat or dog, younger or older than 3 months, vaccinated or not vaccinated etc.). From this comprehensive data set, the validity of each of the applied sampling techniques could be verified, in that the true data on houses with animals is known. In order to achieve herd immunity, at least 70% of the animals in an area need to be vaccinated against rabies and it was determined that consideration of the spatial component of the data is useful in achieving this minimum coverage.

A number of traditional and spatial sampling techniques are proposed with emphasis on design-based sampling so as to draw a clear relationship between the sampling strategies under traditional and spatial sampling theories. These approaches to sampling are applied to a census data set regarding three villages of Tanzania; Buchanchari, Park Nyigoti and Rigicha[2]. (We limit graphical output to the Buchanchari village herein.) Within this extensive data set is the location of the houses within the villages, as well as the number of animals at each of the houses. It is desirable to access at least 70% of the animal population in a village in order to vaccinate them against rabies.

Consider Fig. 1 again, a plot of the houses of Buchanchari, Tanzania. It is not obvious how to identify the spatial distribution and where the stopping points should be. There are not obvious clusters to target for vaccination, nor is the pattern obviously regular. The terrain has a clear influence on the house locations. In order to design a vaccination scheme for the vaccinator, a kernel density map is used at the outset for identifying stopping points along the driving route. A kernel density map is useful for visualising the spatial distribution of point data (Ježek et al. 2017).

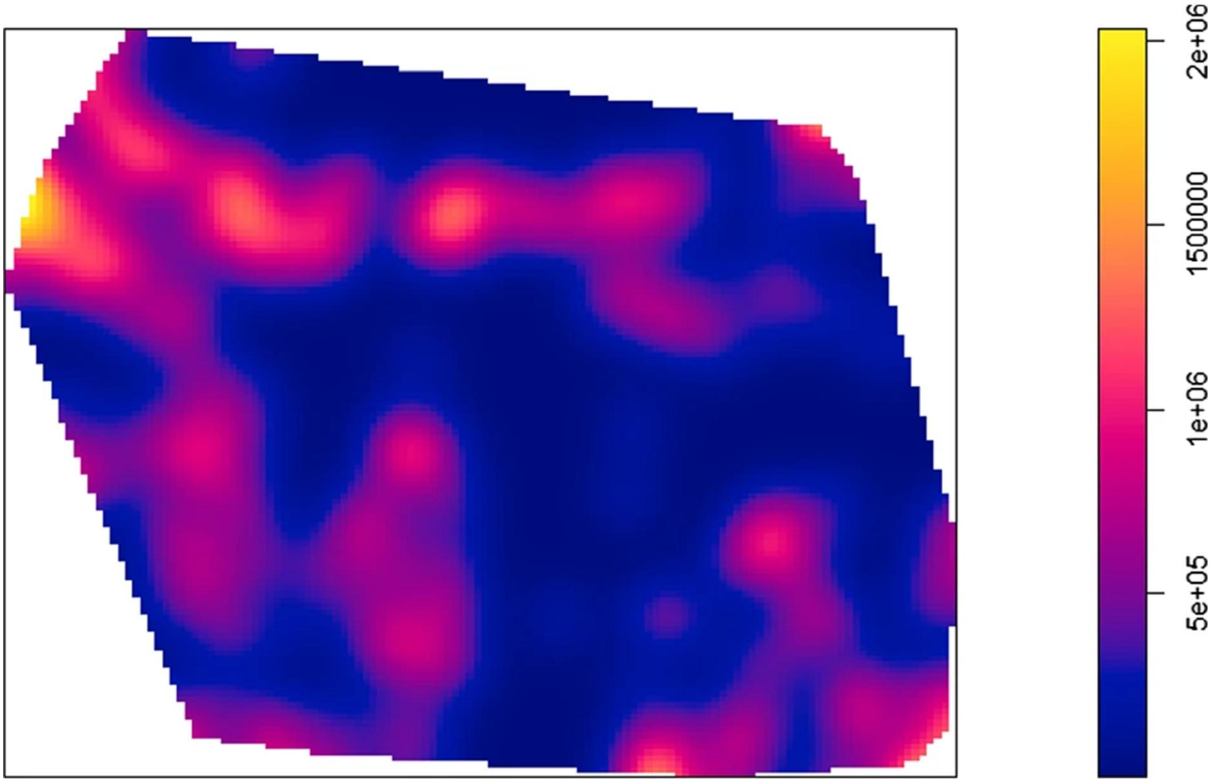
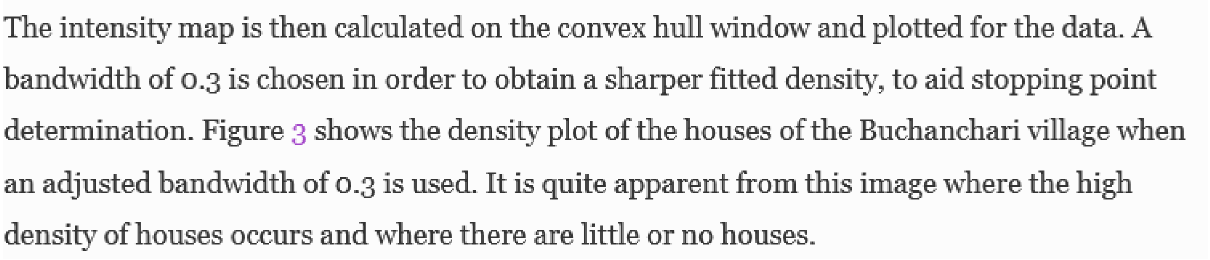


Fig. 3. Density map of houses in Buchanchari with an adjusted bandwith of 0.3, intensity values are counts per 100 square kilometers


The intensity map is then calculated on the convex hull window and plotted for the data. A bandwidth of 0.3 is chosen in order to obtain a sharper fitted density, to aid stopping point determination. Figure 3 shows the density plot of the houses of the Buchanchari village when an adjusted bandwidth of 0.3 is used. It is quite apparent from this image where the high density of houses occurs and where there are little or no houses.

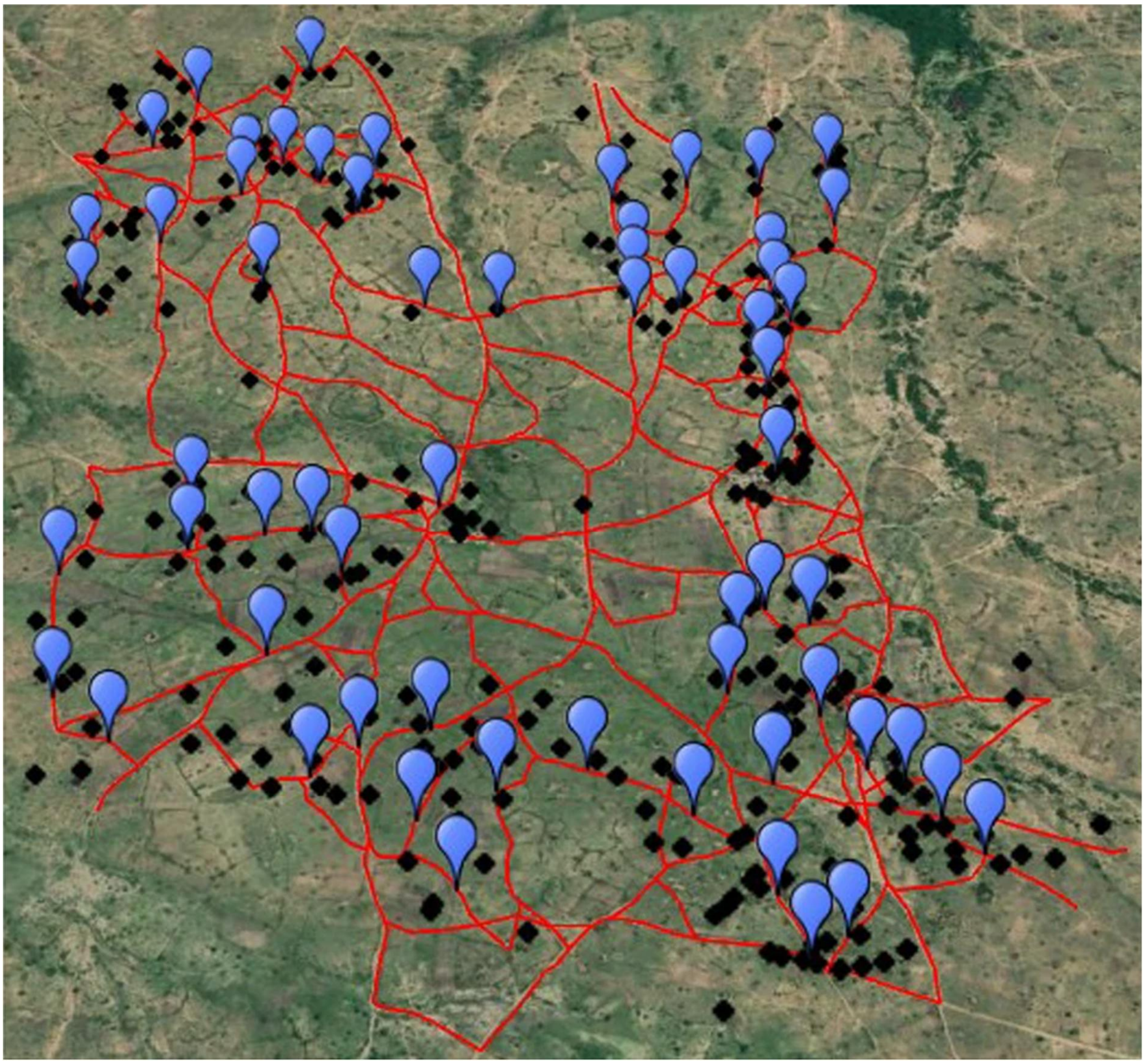


Fig. 4. Google Earth image of the Buchanchari houses, stopping points and the digitised road network

After obtaining these stopping points, it is necessary to determine the road network, as the roads within this village are all informal dirt roads. The digitising of the road network is achieved manually and the stopping points are then mapped along the road network to ensure that the coordinates align. The road network, houses and stopping points are all illustrated in Fig. 4.

Consider Fig. 5. On the left is a portion of the village, with the stopping point indicated in blue, the roads in red and the houses in black. On the right is the road network for the three stopping points. Every possible route between stopping points forms an edge.

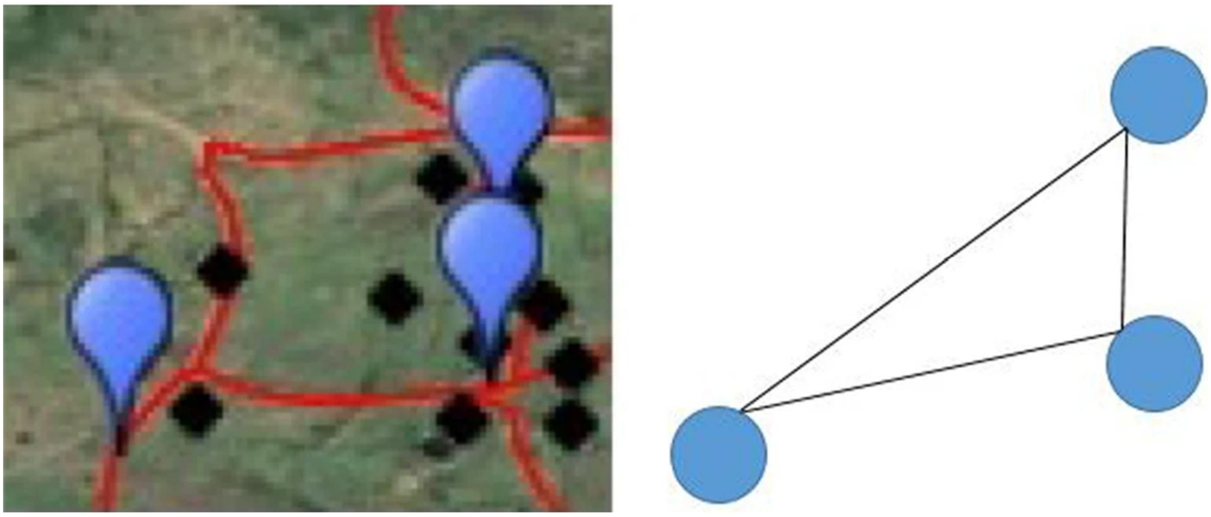
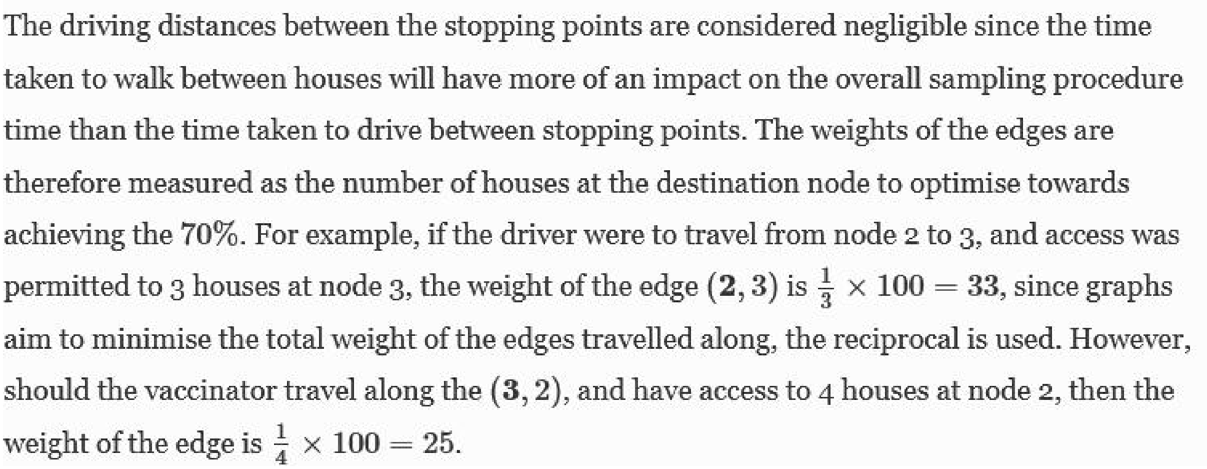
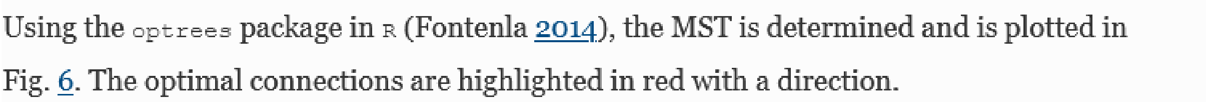


Fig. 5. The weighted graph model (right) for a portion of the village (left) (stopping points indicated in blue, the roads in red and the houses in black)



The driving distances between the stopping points are considered negligible since the time taken to walk between houses will have more of an impact on the overall sampling procedure time than the time taken to drive between stopping points. The weights of the edges are therefore measured as the number of houses at the destination node to optimise towards achieving the 70%. For example, if the driver were to travel from node 2 to 3, and access was permitted to 3 houses at node 3, the weight of the edge $(\mathbf{2},\mathbf{3})$ is $\frac{1}{3} \times 100 = 33$, since graphs aim to minimise the total weight of the edges travelled along, the reciprocal is used. However, should the vaccinator travel along the $(\mathbf{3},2)$, and have access to 4 houses at node 2, then the weight of the edge is $\frac{1}{4} \times 100 = 25$.



Using the `optrees` package in `R` (Fontenla 2014), the MST is determined and is plotted in Fig. 6. The optimal connections are highlighted in red with a direction.

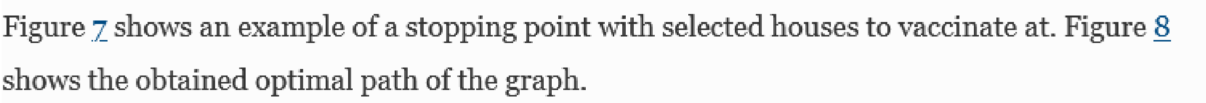
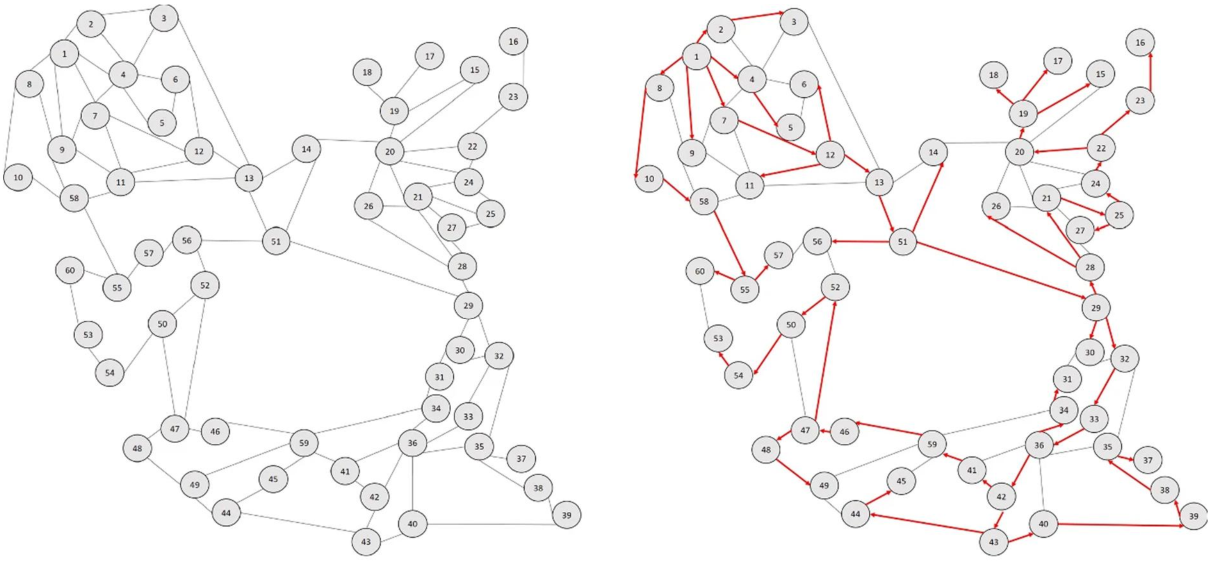


Figure 7 shows an example of a stopping point with selected houses to vaccinate at. Figure 8 shows the obtained optimal path of the graph.



(a) A graph of the stopping points and connections

(b) An illustration of the minimum spanning tree (MST) for the graph

Fig. 6. Graphical representation of the road network as well as the optimal route

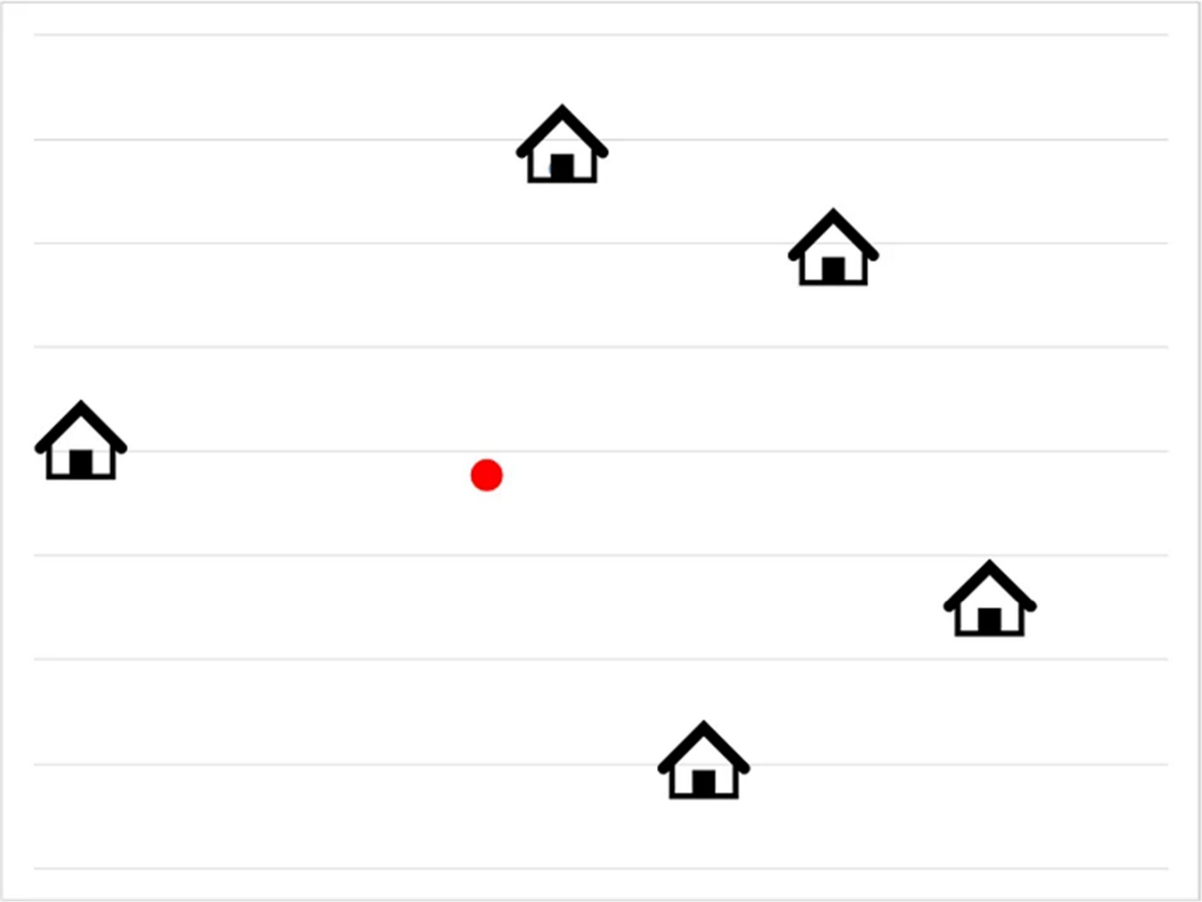


Fig. 7. Plot of the houses surrounding a stopping point

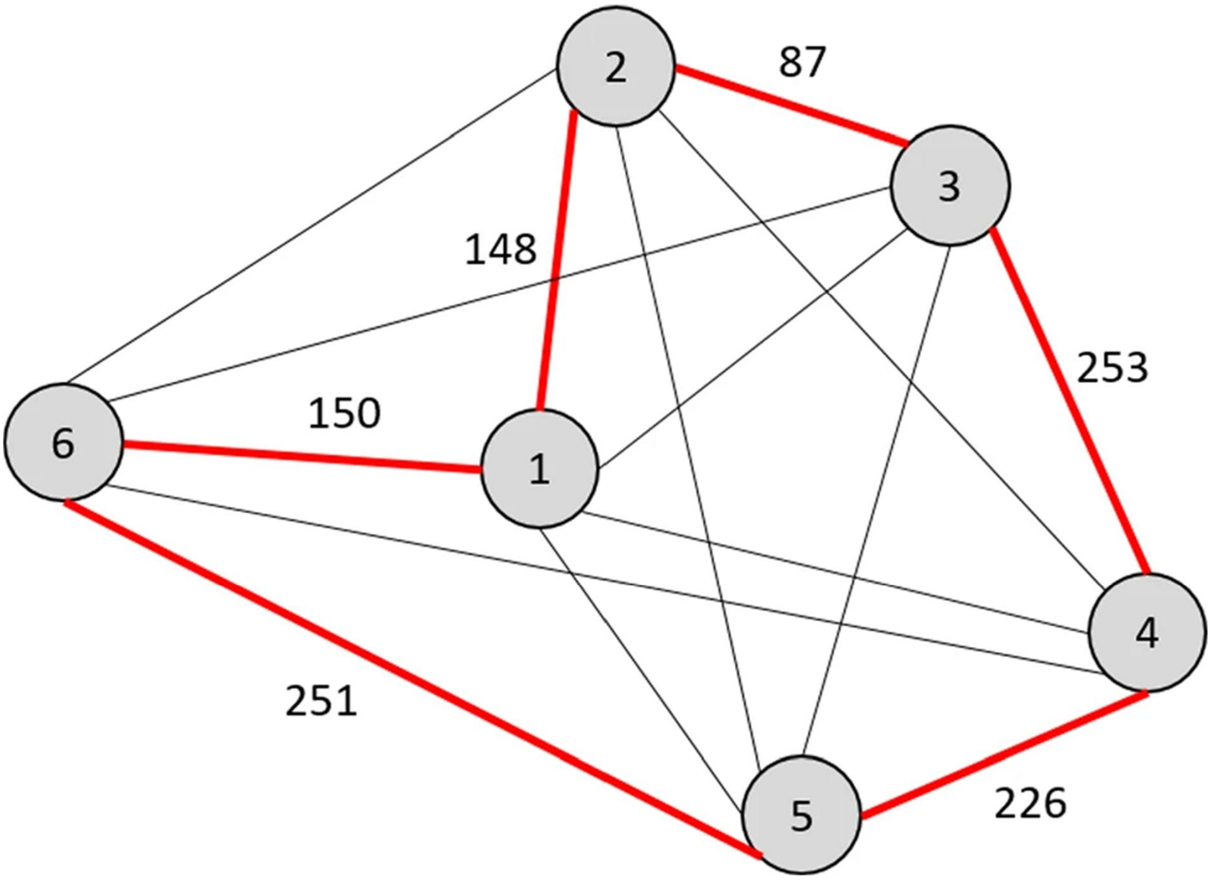
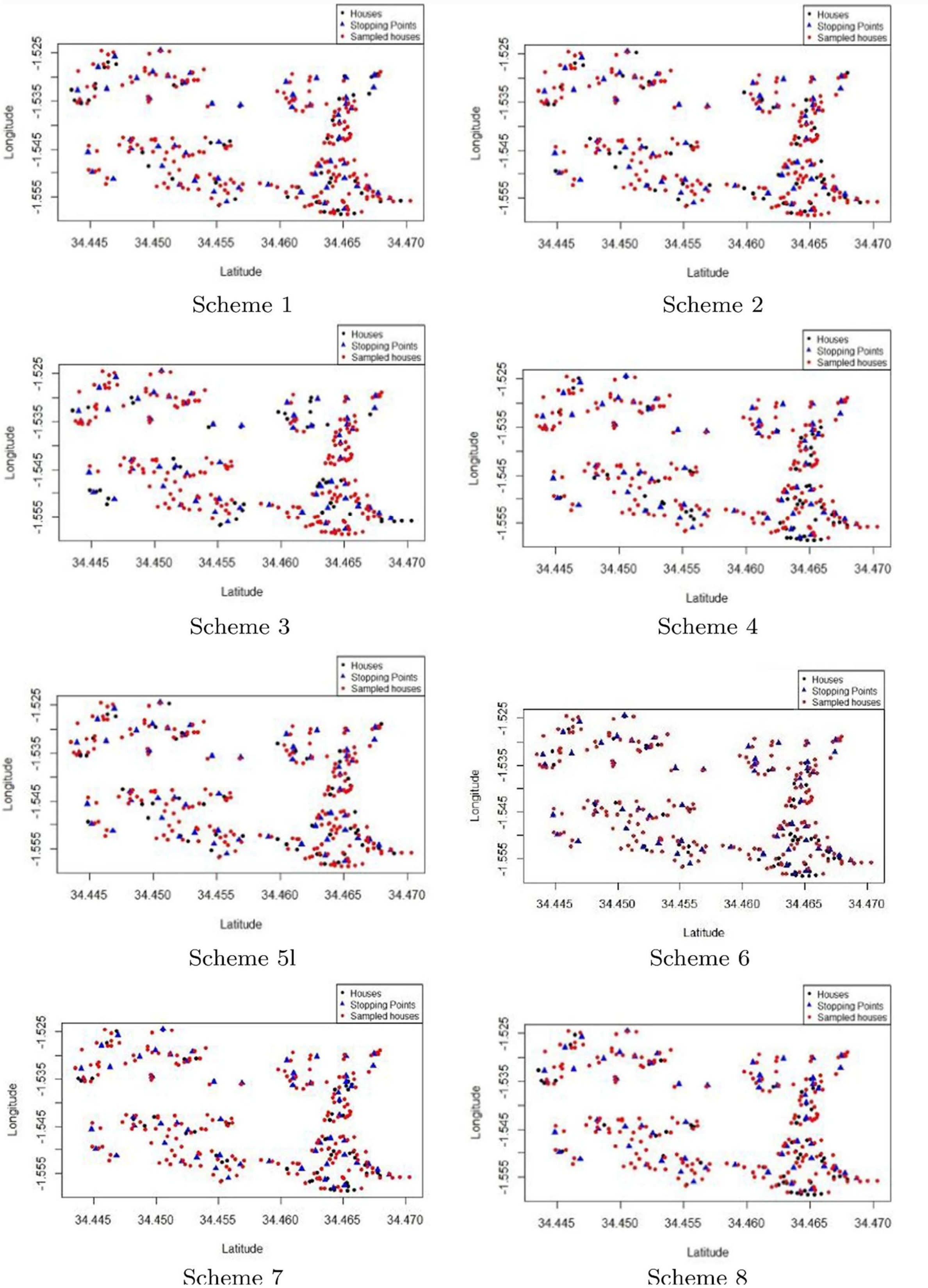


Fig. 8. Graph representation of the optimal walk (in metres) between the houses with the stopping points as vertices and the edges weighted according to the distance between the houses

The sampling techniques presented, not only ensure sufficient coverage of the area of interest, but also ensure that the walking distance of the vaccinator is minimised. Therefore, the stopping points that were developed are sufficient in allowing access to at least 70% of the animals. The attainable houses of the Buchanchari village have 614 animals, which is 88% of the population value of 701. Similar higher percentages were achieved for the other two villages. Only houses within 200*m* of a stopping point are included. Since the percentage of accessible animals in the village is sufficiently larger than the herd immunity value of 70%, samples are drawn from this accessible population using both traditional and spatial sampling techniques.

Each sampling scheme was repeated 1000 times in order to obtain a scheme specific cost distribution and make comparisons between traditional and spatial schemes. Figure 9 provides a single realisation of sampling by each scheme for the village of Buchanchari. The stopping points are not changed with each simulation, only the sampled points (in red). Such sampled points are repeated 1000 times to provide information on the variance of each scheme.



Scheme 1

Scheme 2

Scheme 3

Scheme 4

Scheme 5l

Scheme 6

Scheme 7

Scheme 8

Fig. 9. An example of a single realisation of each sampling scheme: houses in the village of Buchanchari (black), stopping points along the road network (blue) and the sampled houses (red) according to each of the 8 sampling schemes

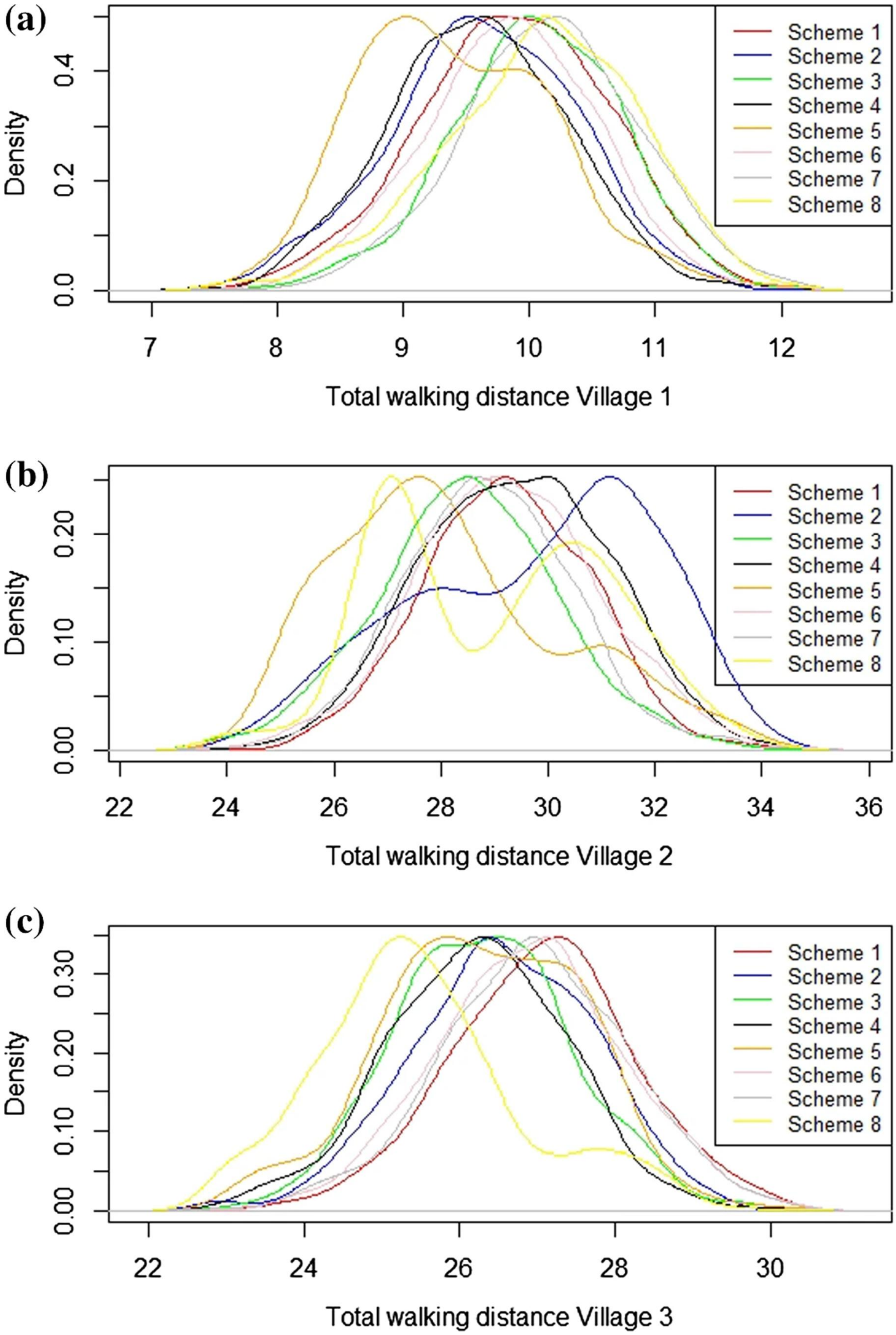
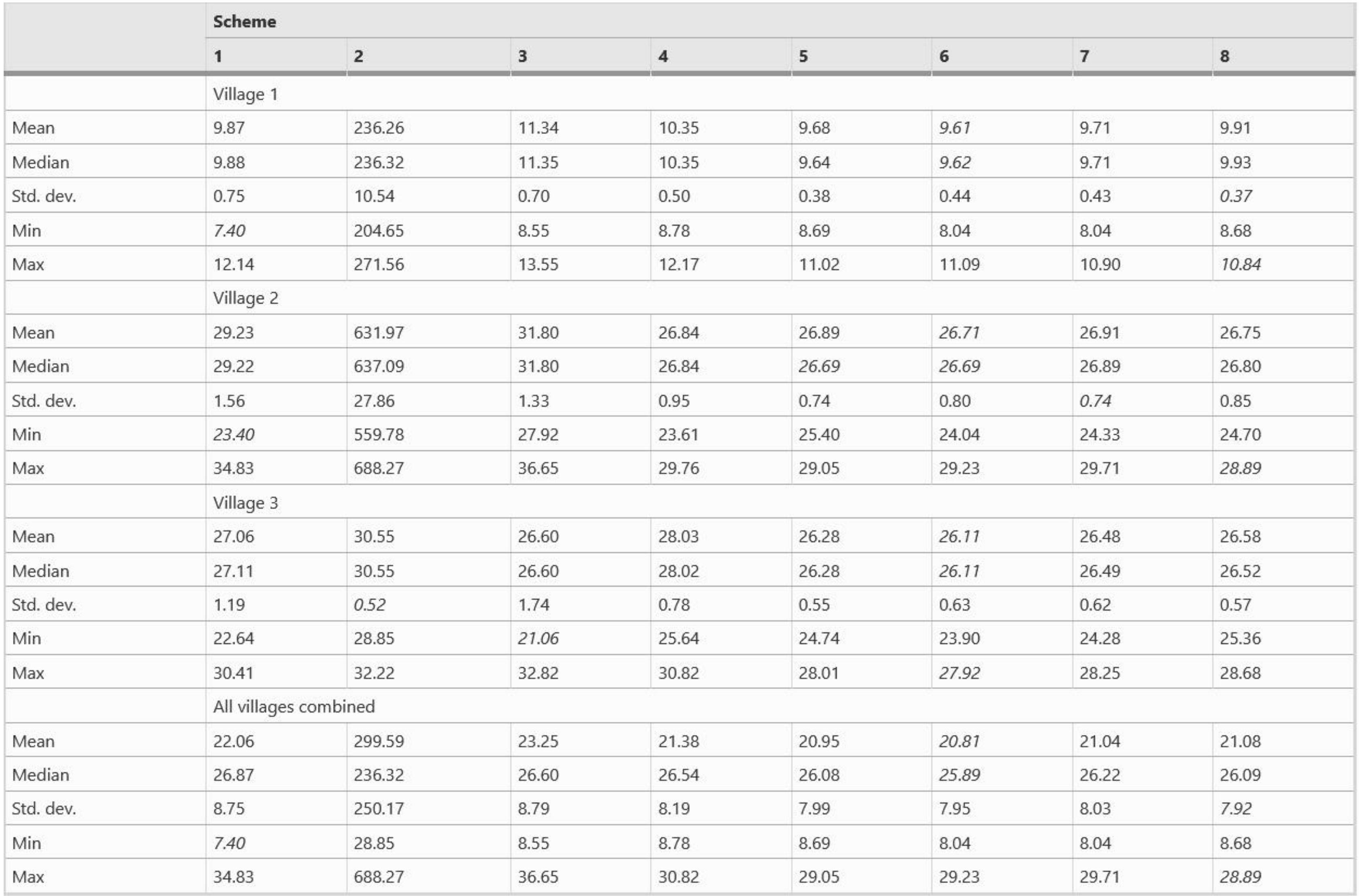


Fig. 10. Distributions of the costs (in kms) of traditional and spatial sampling strategies in the a Park Nyigoti village (village 1), b Rigicha village (village 2) and c Buchanchari village (village 3)

Table 1 Summary statistics of the cost distributions (in kms) of traditional and spatial sampling techniques for the three villages. Cells in italic are the smallest in their case



	Scheme							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
	Village 1							
Mean	9.87	236.26	11.34	10.35	9.68	*9.61*	9.71	9.91
Median	9.88	236.32	11.35	10.35	9.64	*9.62*	9.71	9.93
Std. dev.	0.75	10.54	0.70	0.50	0.38	0.44	0.43	*0.37*
Min	*7.40*	204.65	8.55	8.78	8.69	8.04	8.04	8.68
Max	12.14	271.56	13.55	12.17	11.02	11.09	10.90	*10.84*
	Village 2							
Mean	29.23	631.97	31.80	26.84	26.89	*26.71*	26.91	26.75
Median	29.22	637.09	31.80	26.84	*26.69*	*26.69*	26.89	26.80
Std. dev.	1.56	27.86	1.33	0.95	0.74	0.80	*0.74*	0.85
Min	*23.40*	559.78	27.92	23.61	25.40	24.04	24.33	24.70
Max	34.83	688.27	36.65	29.76	29.05	29.23	29.71	*28.89*
	Village 3							
Mean	27.06	30.55	26.60	28.03	26.28	*26.11*	26.48	26.58
Median	27.11	30.55	26.60	28.02	26.28	*26.11*	26.49	26.52
Std. dev.	1.19	*0.52*	1.74	0.78	0.55	0.63	0.62	0.57
Min	22.64	28.85	*21.06*	25.64	24.74	23.90	24.28	25.36
Max	30.41	32.22	32.82	30.82	28.01	*27.92*	28.25	28.68
	All villages combined							
Mean	22.06	299.59	23.25	21.38	20.95	*20.81*	21.04	21.08
Median	26.87	236.32	26.60	26.54	26.08	*25.89*	26.22	26.09
Std. dev.	8.75	250.17	8.79	8.19	7.99	7.95	8.03	*7.92*
Min	*7.40*	28.85	8.55	8.78	8.69	8.04	8.04	8.68
Max	34.83	688.27	36.65	30.82	29.05	29.23	29.71	*28.89*

Figure 10 contains the distribution of the costs of each of the sampling schemes that were applied to the three villages. Table 1 provides summary statistics for the distances travelled for each of the villages as well as combined summary statistics. Kruskall-Wallis rank sum tests for each Village yield p-values less than 0.005 indicating at least one scheme differs significantly. Conducting pairwise Wilcoxon rank sum tests for each village yields significant differences for every pair except for Scheme 1 and 9 for Village 1, Scheme 4 and 5, Scheme 4 and 7 and Scheme 6 and 8 in Village 2, and Scheme 3 and 7 and Scheme 3 and 8 in Village 3.

Discussion

In Fig. 10 it is seen that, in general, the average cost for spatial sampling (schemes 4–8) is less than the cost of traditional sampling (schemes 1–3) across all three villages. From Table 1[3], in the Buchanchari village, the sampling strategy which yielded the lowest cost after 1000 samples was the systematic regular spatial scheme which had an average walking distance of 26.11km. The stratified spatial scheme had the next lowest average total cost of 26.28km. Only the uniform spatial scheme has a larger average than the non-spatial schemes. This is to be expected as no mechanism is incorporated in uniform spatial sampling to disperse the sampling over the area. For the villages of Park Nyigoti and Rigicha all the average costs for the non-spatial schemes are larger. Notably, scheme 2 for both of these villages has a very large impractical distance due to the sparse layout of the houses in these villages. These results support the prediction that the spatial approach to this application is beneficial and appropriate. The standard deviations of all of the costs are around 1km. This shows that the costs of the samples do not vary too drastically through each iteration, which is beneficial as then only a single sample need be chosen and optimisation is not required. Tests of significant differences in the curves in Fig. 10 were not performed not added knowledge would result.

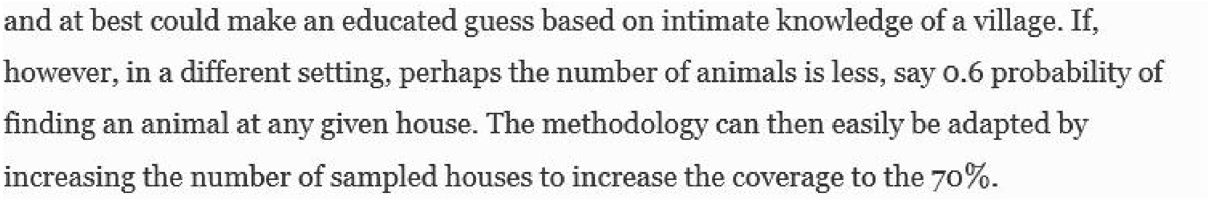
Combining the simulations for all three villages, identifies the systematic regular spatial scheme to be the optimal one with stratified spatial sampling being next in line. Again all spatial schemes are to be preferred above the non-spatial ones. The results reveal that spatial sampling results in the smallest walking distance between houses, with the minimum 70% coverage attained. The optimal route through the stopping points is an illustration of the possibility of optimisation, however the vaccinator will be given a list of locations (the stopping points) to reach and may naturally decide which route will be best.

Determining the stopping points $L_{\max}^{\text{adj}}$ could be determined in other ways, for example, clustering. A spatial cluster model could be fitted to determine cluster centres, or a standard clustering algorithm such as k-means used. These methods would most likely not result in the same determined stopping points but are also viable options for a practitioner to use. The clustering technique used should however provide compact clusters suitable for the purpose of the stopping points, i.e. to provide a useful position to access houses by minimal walking. Clustering does not yield any obvious advantages over the using kernel density estimation however. Estimating a two-dimensional intensity for a point pattern of house locations, such as available herein, results in local maxima being areas where houses tend to cluster in terms of their count. Kernel density estimation also captures the inhomogeneity of the point pattern of house locations.

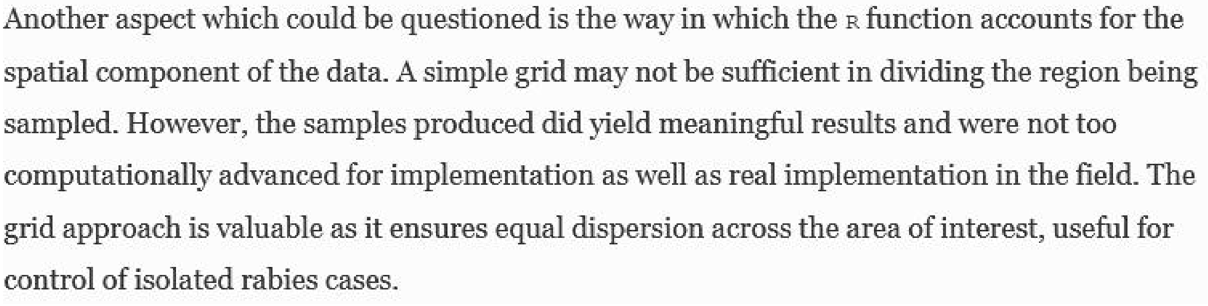
This approach was also applied to larger villages (more than 700 houses), but deemed computationally expensive. It is therefore suggested that such villages be broken up into smaller regions and then sampled. A driver could not walk such a case in a single day, nor would attaining 70% in each smaller region negatively impact the 70% in the whole village. This division will also ensure more even coverage of the village area as it was also apparent that those villages which have a very high density of houses in a small area with the rest being quite scattered did not reach the minimum 70% animal coverage.

Within the application of this work, driving distance between houses was deemed negligible since the distance the vaccinator drives is not as taxing as the distance walked to reach the animals. It is understood that the approach used currently does not require the veterinarian to drive between houses or even walk, but the addition of these factors allows for more efficient samples. Kitala et al. (2002) suggested that at any given time 59% of animals should be vaccinated in order to control rabies and only an annual coverage of 70%. Therefore, the strategy suggested here could be divided into trips over the whole year. By doing this, the veterinarian need not walk the full 20*km* in a day to vaccinate all the animals, but could rather do multiple trips, walking shorter distances and still achieving the desired population coverage.

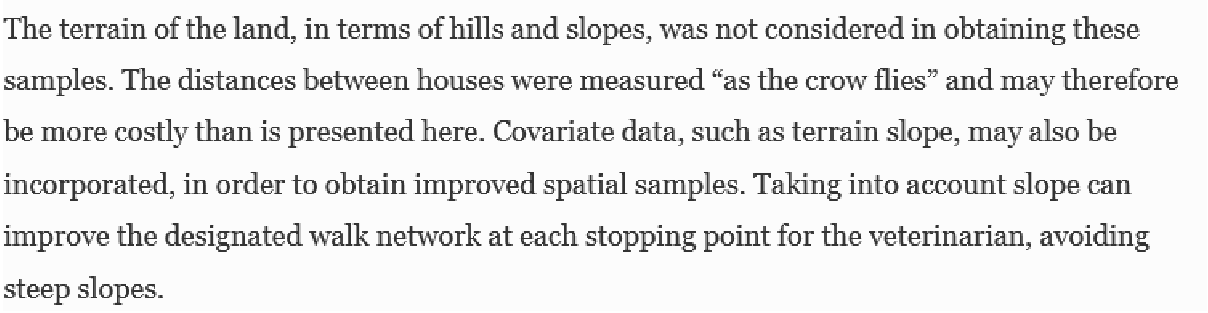
The methodology assumes one animal at each house. The data considered in this paper is a census collection from numerous villages in Tanzania and provides excellent information on the animal population. In the data there is an average of 1.39 dogs over three years old, 0.36 dogs less than three years old, 0.6 cats over three years old and 0.11 cats less than three years old each a house, giving an overall average of 2.36 animals per house, with a standard deviation of 2.27. Thus since the animals roam freely in such rural villages, it is likely to find an animal (at least one) near the sampled house even if it doesn’t belong to that house. This still contributes to the vaccination goal. In practice, one would never know the true occupancy



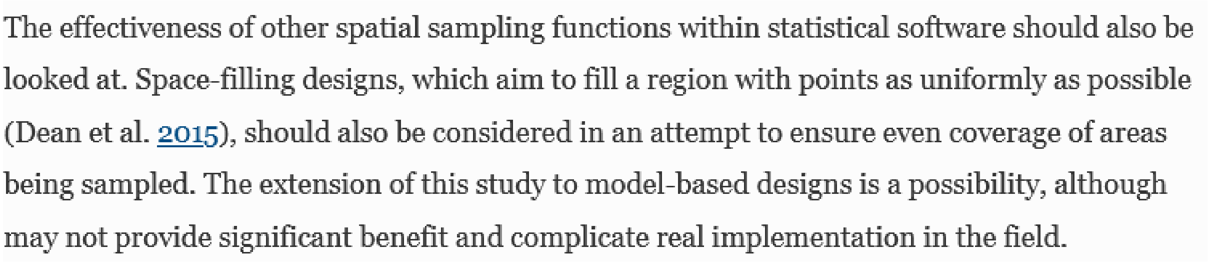
and at best could make an educated guess based on intimate knowledge of a village. If, however, in a different setting, perhaps the number of animals is less, say 0.6 probability of finding an animal at any given house. The methodology can then easily be adapted by increasing the number of sampled houses to increase the coverage to the 70%.



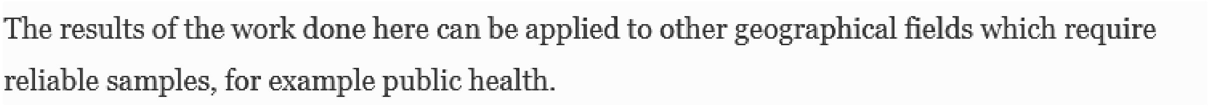
Another aspect which could be questioned is the way in which the R function accounts for the spatial component of the data. A simple grid may not be sufficient in dividing the region being sampled. However, the samples produced did yield meaningful results and were not too computationally advanced for implementation as well as real implementation in the field. The grid approach is valuable as it ensures equal dispersion across the area of interest, useful for control of isolated rabies cases.



The terrain of the land, in terms of hills and slopes, was not considered in obtaining these samples. The distances between houses were measured “as the crow flies” and may therefore be more costly than is presented here. Covariate data, such as terrain slope, may also be incorporated, in order to obtain improved spatial samples. Taking into account slope can improve the designated walk network at each stopping point for the veterinarian, avoiding steep slopes.

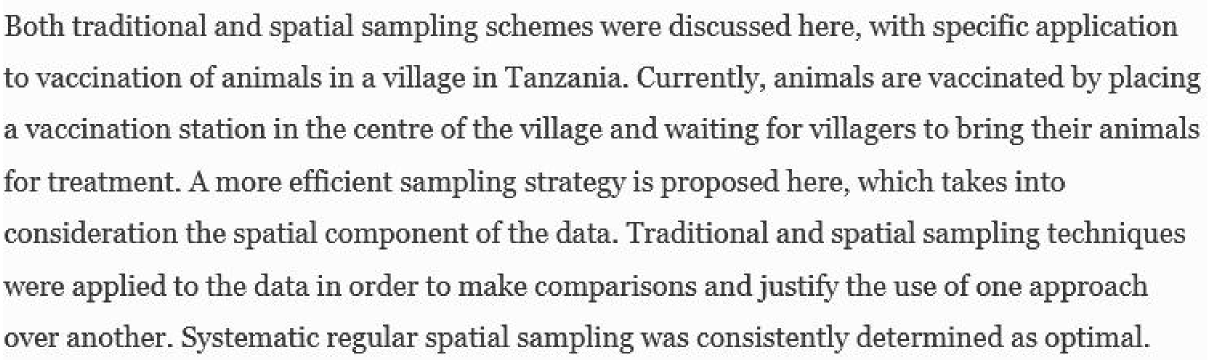


The effectiveness of other spatial sampling functions within statistical software should also be looked at. Space-filling designs, which aim to fill a region with points as uniformly as possible (Dean et al. 2015), should also be considered in an attempt to ensure even coverage of areas being sampled. The extension of this study to model-based designs is a possibility, although may not provide significant benefit and complicate real implementation in the field.

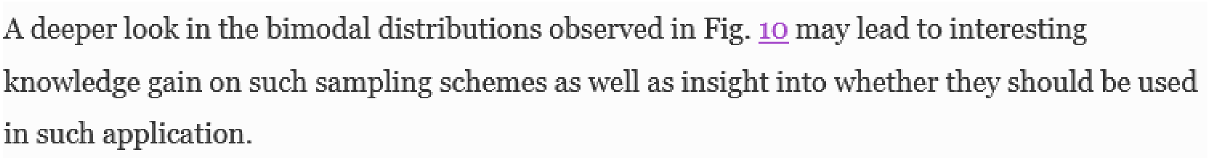


The results of the work done here can be applied to other geographical fields which require reliable samples, for example public health.

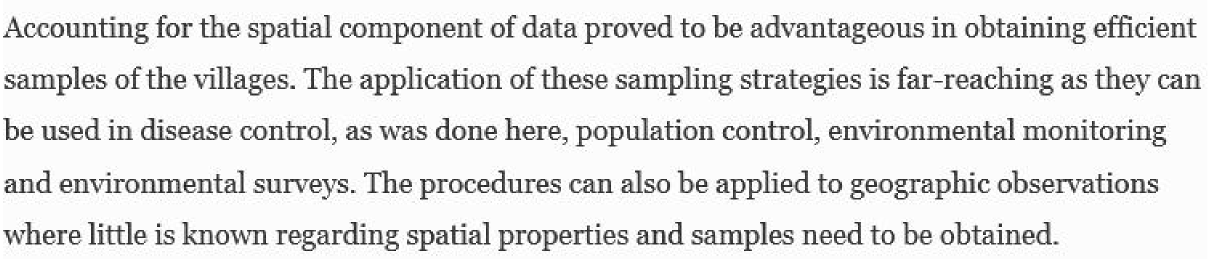
Conclusion



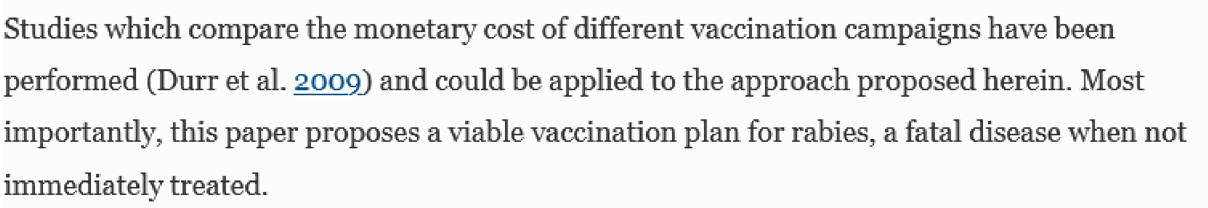
Both traditional and spatial sampling schemes were discussed here, with specific application to vaccination of animals in a village in Tanzania. Currently, animals are vaccinated by placing a vaccination station in the centre of the village and waiting for villagers to bring their animals for treatment. A more efficient sampling strategy is proposed here, which takes into consideration the spatial component of the data. Traditional and spatial sampling techniques were applied to the data in order to make comparisons and justify the use of one approach over another. Systematic regular spatial sampling was consistently determined as optimal.



A deeper look in the bimodal distributions observed in Fig. 10 may lead to interesting knowledge gain on such sampling schemes as well as insight into whether they should be used in such application.



Accounting for the spatial component of data proved to be advantageous in obtaining efficient samples of the villages. The application of these sampling strategies is far-reaching as they can be used in disease control, as was done here, population control, environmental monitoring and environmental surveys. The procedures can also be applied to geographic observations where little is known regarding spatial properties and samples need to be obtained.



Studies which compare the monetary cost of different vaccination campaigns have been performed (Durr et al. 2009) and could be applied to the approach proposed herein. Most importantly, this paper proposes a viable vaccination plan for rabies, a fatal disease when not immediately treated.

Notes

1. http://www.who.int/rabies/control/Tanzania_Project_Summary_310317.pdf?ua=1
2. http://www.gla.ac.uk/researchinstitutes/bahcm/staff/katiehampson, http://www.katiehampson.com.
3. Scheme 1: Traditional SRS, Scheme 2: Traditional Stratified, Scheme 3: Traditional Systematic, Scheme 4: Uniform Spatial, Scheme 5: Stratified Spatial, Scheme 6: Systematic Regular Spatial, Scheme 7: Systematic Non-aligned Spatial, Scheme 8: Systematic Hexagonal Spatial

References

Addario-Berry L, Broutin N, Goldschmidt C, Miermont G (2017) The scaling limit of the minimum spanning tree of the complete graph. Ann Probability 45(5):3075–3144

Baddeley A, Rubak E, Turner R (2015) Spatial point patterns: methodology and applications with R. Chapman and Hall/CRC, London

Bostoen K, Chalabi Z (2006) Optimization of household survey sampling without sample frames. Int J Epidemiol 35(3):751–755

Cleaveland S, Kaare M, Tiringa P, Mlengeya T, Barrat J (2003) A dog rabies vaccination campaign in rural Africa: impact on the incidence of dog rabies and human dog-bite injuries. Vaccine 21(17):1965–1973

Coleman PG, Dye C (1996) Immunization coverage required to prevent outbreaks of dog rabies. Vaccine 14(3):185–186

Cormen TH (2009) Introduction to algorithms. MIT Press, Cambridge

Dean A, Morris M, Stufken J, Bingham D (2015) Handbook of design and analysis of experiments, vol 7. CRC Press, London

Delmelle E (2011) Spatial sampling. In: Fotheringham A, Rogerson P (eds) The SAGE handbook of spatial analysis. SAGE Publications, Ltd, Thousand Oaks, pp 183–206 chap 10

Dijkstra EW (1959) A note on two problems in connexion with graphs. Numerische Mathematik 1(1):269–271

Durr S, Mindekem R, Kaninga Y, Moto DD, Meltzer M, Vounatsou P, Zinsstag J (2009) Effectiveness of dog rabies vaccination programmes: comparison of owner-charged and free vaccination campaigns. Epidemiol Infect 137(11):1558–1567

Fontenla M (2014) optrees: Optimal Trees in Weighted Graphs. https://CRAN.R-project.org/package=optrees, r package version 1.0

Ježek J, Jedlička K, Mildorf T, Kellar J, Beran D (2017) Design and evaluation of Web gl-based heat map visualization for big point data. In: The Rise of Big Spatial Data, Springer, pp 13–26

Kayali U, Mindekem R, Yemadji N, Vounatsou P, Kaninga Y, Ndoutamia A, Zinsstag J (2003) Coverage of pilot parenteral vaccination campaign against canine rabies in N'djamena, Chad. Bull World Health Organ 81(10):739–744

Kitala P, McDermott J, Coleman P, Dye C (2002) Comparison of vaccination strategies for the control of dog rabies in Machakos District, Kenya. Epidemiol Infect 129(1):215–222

Kraamwinkel C, Fabris-Rotelli I, Fosgate G, Knobel D, Hampson K (2014) A study on the apparent randomness of an animal sample. In: Proceedings of the 2014 SASA Conference, Grahamstown

Kruskal JB (1956) On the shortest spanning subtree of a graph and the traveling salesman problem. Proc Am Math Soc 7(1):48–50

Mpolya E, Lembo T, Lushasi K, Mancy R, Mbunda E, Mkungu S (2017) Towards elimination of dog-mediated human rabies: experiences from implementing a large-scale demonstration project in southern Tanzania. Front Vet Sci 4:21

Prim RC (1957) Shortest connection networks and some generalizations. Bell Syst Techn J 36(6):1389–1401

Ripley B, Rasson JP (1977) Finding the edge of a Poisson forest. J Appl Probability 14(3):483–491

Ripley BD (2005) Spatial statistics, vol 575. Wiley, New York

Stein A, Van Groenigen J, Jeger M, Hoosbeek M (1998) Space-time statistics for environmental and agricultural related phenomena. Environ Ecol Stat 5:155–172

Wang J, Stein A, Gao B, Ge Y (2012) A review of spatial sampling. Spat Stat 2:1–14

Zinsstag J, Dürr S, Penny M, Mindekem R, Roth F, Gonzalez SM, Naissengar S, Hattendorf J (2009) Transmission dynamics and economics of rabies control in dogs and humans in an African city. Proc Nat Acad Sci 106(35):14996–15001