

Spatial epidemiology and adaptive targeted sampling to manage the Chagas disease vector *Triatoma dimidiata*

B. K. M. Case^{1,2,*}, Jean-Gabriel Young^{1,3}, Daniel Penados⁴, Carlota Monroy⁴, Laurent Hébert-Dufresne^{1,2}, and Lori Stevens⁵

¹Vermont Complex Systems Center, University of Vermont, USA

²Department of Computer Science, University of Vermont, USA

³Department of Mathematics & Statistics, University of Vermont, USA

⁴Laboratorio de Entomología Aplicada y Parasitología, Universidad de San Carlos de Guatemala, Guatemala

⁵Department of Biology, University of Vermont, USA

*Corresponding author: bcase@uvm.edu

February 10, 2022

Abstract

Widespread application of insecticide remains the primary form of control for Chagas disease in Central America, despite only temporarily reducing domestic levels of the endemic vector *Triatoma dimidiata* and having little long-term impact. Recently, an approach emphasizing community feedback and housing improvements has been shown to yield lasting results. However, the additional resources and personnel required by such an intervention likely hinders its widespread adoption. One solution to this problem would be to target only a subset of houses in a community while still eliminating enough infestations to interrupt disease transfer. Here we develop a sequential sampling framework that adapts to information specific to a community as more houses are visited, thereby allowing us to efficiently find homes with domiciliary vectors while minimizing sampling bias. The method fits Bayesian geostatistical models to make spatially informed predictions, while gradually transitioning from prioritizing houses based on prediction uncertainty to targeting houses with a high risk of infestation. A key feature of the method is the use of a single exploration parameter, α , to control the rate of transition between these two design targets. In a simulation study using empirical data from five villages in southeastern Guatemala, we test our method using a range of values for α , and find it can consistently select fewer homes than random sampling, while still bringing the village infestation rate below a given threshold. We further find that when additional socioeconomic information is available, much larger savings are possible, but that meeting the target infestation rate is less consistent, particularly among the less exploratory strategies. Our results suggest new options for implementing long-term *T. dimidiata* control.

1 Introduction

Chagas disease is a vector-borne neglected tropical disease (NTD) endemic to all countries in Latin America [1]. It is the most serious parasitic disease in the region, with a 2005 estimate of disease burden 5 to 10 times greater than malaria [2], and is mainly a threat to people living in poverty [3, 4]. The disease, which can lead to potentially fatal cardiovascular or gastrointestinal issues, is caused by the parasite *Trypanosoma cruzi* and transmitted by insects in the Triatominae subfamily [5]. Control initiatives for Chagas primarily focus on disrupting the transmission pathway to humans by reducing domestic Triatomine levels, which is the primary mode of infection [6, 7]. A common control target is to reduce the proportion of infested households in a community to below 5% [8, 9], and it has been shown that reduction past 8% is sufficient to eliminate *T. cruzi* seroprevalence [10].

In Central America, the prevalence of Chagas disease has declined significantly since the 1980s, thanks in part to the near-elimination of the invasive vector *Rhodnius prolixus*. However, the species-complex *Triatoma dimidiata* still poses a significant health risk to millions of people in many areas [11]. Unlike the invasive *R. prolixus*, *T. dimidiata* is endemic to the continent and infests peridomestic and sylvatic environments [12]. Unfortunately, efforts to control domestic *T. dimidiata* populations are complicated by the resilience of *T. dimidiata* to traditional methods of vector control. Domestic populations of the insect can rebound within several months of insecticide spraying [13–15], and continuing to spray houses which report reinfestation appears to have little long-term effect [16]. Control measures for *T. dimidiata* are further complicated due to its significant variation in habitat, morphology, feeding patterns, and genetics [17–19], as well as variation in the domiciliary risk factors associated with its presence [9, 20]. Further, the sustainability of a given control strategy depends critically on cultural practices in the area [21–23]. Thus, meeting the goal of long-term *T. dimidiata* reduction requires adaptive, locale-specific strategies for surveillance and control [24, 25].

The limitations of insecticide for control of *T. dimidiata* in Guatemala have led to the gradual adoption of additional measures, mostly in the departments of Jutiapa and Chiquimula in southeastern Guatemala, which represent the majority of the country’s cases reported to the Ministry of Health [11, 26]. A promising multidisciplinary approach, often referred to as the EcoHealth approach, applies cost-effective, locally-tailored house and peridomestic improvements by collaborating with villagers and health personnel, in conjunction with initial insecticide application [21]. A pilot study of two villages found the method led to very low (< 5%) infestation rates 5 years after housing improvements [20, 27], and an expansion of the project to five villages in Chiquimula led to a sustained four-fold reduction in infestation [24]. This suggests house improvements following the EcoHealth approach can effectively prevent reinfestation in the long term.

Barriers to the widespread adoption of community engagement-based interventions include frequent shortages in insecticides [24] and the need for research experts and trained personnel to work with residents and identify infested houses. One possible solution to these issues is to more efficiently select homes in need of insecticides and improvements while still meeting necessary control targets. For example, by treating only a sample of homes such that overall the *T. dimidiata* infestation rate in the village goes below 5%, residual dispersal and non-domestic migration will be limited to homes that were recently improved, or were already unlikely to be suitable for infestation. Further, the EcoHealth approach’s emphasis on practical improvements and community participation could help ensure that the risk factors identified from this sample continue to be addressed throughout the entire village. This control strategy would in turn free up resources to be applied in other communities.

To be successful, such a strategy must balance the incentive to target houses that are believed to be infested, with the need to correctly identify the infestation status of unvisited homes. Selecting houses to treat based on perceived infestation risk will quickly bring the village’s infestation rate closer to the 5% goal, at least in the short term. However, a diverse range of samples throughout the area and across combinations of possible risk factors is required to reliably find the remaining infested houses, and to correctly predict whether the remaining number of infested houses is below that required to meet the 5% threshold [9, 28]. This was the conclusion in King et al. (2011), where a subset of households was inspected for Triatomines in villages across Guatemala using either random sampling or sampling based on pre-defined risk factors, and was used to predict whether the village infestation rate was below or above the 5% threshold. The authors found random sampling to consistently have higher prediction accuracy, noting that sampling based on a fixed set of factors failed to explore the ways factors associated with infection vary among different villages and regions [9]. In short, the dual objectives of prediction and targeted sampling lead to an exploration vs. exploitation tradeoff, where the space of houses to target must be searched to find configurations that are both of minimal sample size, and which contain enough information to correctly predict that the 5% target has been met.

The quantity and quality of covariate information available for exploration is a second key factor in the success of an intervention strategy. A number of studies have identified various socioeconomic factors associated with infested houses, such as the material and condition of house walls [4, 20, 29]. Therefore, if

additional dependent variables are available prior to selecting houses for treatment, fewer observations may be needed to make accurate estimates. Another option is to rely only on variables available remotely, such as elevation. While this sacrifices potentially useful information, it may ultimately be more cost effective, since collecting socioeconomic risk factors for inference and prediction requires additional labor and logistical planning [30].

Here we aim to address the problem of treating a subset of houses to reduce *T. dimidiata* infestation to a target threshold while minimizing the necessary resources. Using 5 villages of varying size and baseline infestation rates in Chiquimula, Guatemala as a case study, we employ *adaptive geostatistical design* strategies which sequentially select houses based on observations from previous iterations, and use inherent spatial autocorrelation in the observed data to improve prediction and inference.

While historically quite theoretically driven, the principles of geostatistical design have recently been applied to other problems of survey design and analysis in spatial epidemiology [31]. Chipeta et al. (2016) developed an adaptive sampling method which targets locations with high spatial uncertainty, and applied the approach to a cross-sectional malaria survey [32, 33]. For identifying hotspots of lymphatic filariasis, adaptive sampling based on prediction entropy and spatial exploration was shown to be effective [34]. Fronterre et al. (2020) used a non-adaptive, lattice-like sampling design combined with close pairs of points (proposed first in [35]) to predict whether an area’s disease prevalence exceeds a certain threshold, and found the method outperformed a current WHO assessment protocol on a simulated dataset [36].

In this work, we develop a class of adaptive strategies which transition from prioritizing houses based on prediction uncertainty to houses based on perceived risk of infestation. We compare these strategies to random sampling with empirical data, and assess their ability to efficiently locate infested houses while correctly predicting whether the current selection meets the reduction target. Additionally, we examine the effect of including socioeconomic covariates on the performance of each strategy. In the context of Chagas vector control with the EcoHealth approach, our methods address two key questions: 1) how can houses be more efficiently targeted for treatment to sufficiently reduce village-wide vector incidence? and 2) can further efficiency gains be made by collecting additional socioeconomic information? More generally, our methods provide a formal, statistical framework for targeted control strategies in a resource-limited setting, and hence are particularly relevant to the control of NTDs.

2 Methods

2.1 Data preparation

Our data come from the follow-up EcoHealth project discussed in the introduction, which was conducted in the countries Honduras, El Salvador, and Guatemala, between August and December 2011, and is described in detail in Bustamante et al. (2015) [20]. We focus only on the five villages in Chiquimula, Guatemala, since infestation rates were low in the other countries. After removing observations with missing information, 172 housing structures were recorded in El Amatillo, 147 in El Cerrón, 251 in El Guayabo, 108 in El Paternito, and 207 in La Prensa, for a total of 885 observations.

Each data entry was obtained as follows. After the informed consent of the residents, houses were searched for 35-45 minutes by one team member, while another performed interviews and assessed aspects of tidiness in the home. The home’s geocoordinates were also recorded. These surveys produce a binary response indicating Triatomine presence in the home, and 26 covariates. A positive response indicates that adult or juvenile insects, dead insects, or eggs were found. The covariates, listed in Table , include items related to socioeconomic factors, the number and type of domestic animals, and the house’s structure and cleanliness. We added two more covariates based on the house’s geocoordinates. The *distance to perimeter* is the shortest distance of the home to the village’s convex hull, plus 50m, while the *density* is the number of other houses within 100m. Covariates were checked for multicollinearity, and all continuous variables were

centered and scaled. Additionally, for convenience in setting priors, the coordinates of the houses were scaled such that the diameter of the village (maximum distance between any two points) was one.

2.2 Hierarchical modeling for geostatistics

Geostatistics is a field which studies spatial autocorrelation in point-referenced data, and leverages this information for inference and prediction [37]. Geostatistical models incorporate a spatial phenomena $Z = \{z(\mathbf{s}) \in \mathbb{R} \mid \mathbf{s} \in \mathcal{D}\}$ over a domain of possible locations \mathcal{D} , where $n = |\mathcal{D}|$ when \mathcal{D} is discrete. The closer two points \mathbf{s}_i and \mathbf{s}_j are to each other, the more similar the values $z(\mathbf{s}_i)$ and $z(\mathbf{s}_j)$ will tend to be. This spatial surface Z is itself a function of possibly unknown *spatial parameters*, which control how the covariance between points behaves. Rather than observing Z directly, for each location \mathbf{s} there is typically a measurable response $y(\mathbf{s})$, which is assumed to be a function of Z and some covariates $x_1(\mathbf{s}), \dots, x_p(\mathbf{s})$.

Following the general hierarchical framework first outlined in [38], we assume the response at each $\mathbf{s} \in \mathcal{D}$ follows a generalized linear model with spatially correlated random effects. In our setting, this amounts to $y(\mathbf{s})$ being a binary variable indicating the infestation status of a home at position \mathbf{s} , which has a probability $r(\mathbf{s})$ of being infested. The probability $r(\mathbf{s})$, or the *risk* of having an infestation at location \mathbf{s} , will then depend on the home's covariates and the risks of other homes in close proximity. More formally, the likelihood of $y(\mathbf{s})$ is as follows:

$$y(\mathbf{s}) \mid r(\mathbf{s}) \sim \text{Bernoulli}(r(\mathbf{s})) \quad (1)$$

$$\text{logit}(r(\mathbf{s})) = \eta(\mathbf{s}) = \mathbf{x}(\mathbf{s})\boldsymbol{\beta} + z(\mathbf{s}) + \varepsilon(\mathbf{s}), \quad (2)$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ is a vector of fixed effect coefficients. The spatial surface Z follows a zero-centered Gaussian distribution with Matérn covariance function (defined in [Mathematical definitions](#)) with smoothness parameter $\nu = 1$ [39]. This spatial process has two parameters σ_s and ρ , which respectively control the variance and *effective range*, here defined as the distance at which the correlation between two points reaches 0.1. Finally, ε is an independent random effect representing non-spatial latent variability at each location, which helps avoid finding spurious spatial correlation [36]. We assume $\varepsilon \sim N(0, \sigma_e^2)$.

The equations above describe the probability a house is infested, given some parameters and covariate information. In other words, they specify an assumption about how our response is generated as a function of these parameters. However, we are interested in reversing this process: given some finite set $\mathcal{S} \subset \mathcal{D}$ of locations, we will use the household data from these locations to make inferences about possible values of the parameters. Following convention, we write $\mathbf{y} = \{y(\mathbf{s}) \mid \mathbf{s} \in \mathcal{S}\}$ as the observed response, and $\boldsymbol{\theta} = (\boldsymbol{\beta}, \rho, \sigma_s, \sigma_e)$ as the parameters to be estimated. Under the Bayesian paradigm, one treats $\boldsymbol{\theta}$ as a random variable and assigns priors based on domain knowledge or on hypotheses formed prior to data collection. Bayes' theorem then gives the posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{y})$, which represents our updated beliefs about $\boldsymbol{\theta}$ given the data we have observed at \mathcal{S} . The posterior distribution can further be used to make predictions at unmeasured locations.

For our analysis, we use weakly-informative $N(0, 3.3)$ priors for $\boldsymbol{\beta}$, while for the spatial range and standard deviation we use the penalized-complexity prior of [40], set to induce tail probabilities of $\Pr(\rho < 0.1) = 0.05$ and $\Pr(\sigma_s > 3) = 0.1$. The variance for the non-spatial random effects σ_e^2 follows an inverse-gamma distribution with location 1 and scale 0.01.

Data preparation and analysis was performed in R version 4.1.0 [41]. For computational speed and convenience, all statistical models were fit using Integrated Nested Laplace Approximation (INLA) with the Stochastic Partial Differential Equation (SPDE) representation for the spatial effects, available from the R-INLA package [42, 43].

2.3 Model comparison and full-village analysis

To verify the suitability of the model outlined above, we compare its performance to two simpler alternatives. The first removes the correlated spatial random effects $Z(\mathbf{s})$ while the independent $\varepsilon(\mathbf{s})$ effects are removed from the other, but otherwise each model is the same. Models are evaluated based their *deviance information criterion* (DIC) and *marginal likelihood* (ML), formally defined in [Mathematical definitions](#). Both measure a model’s goodness-of-fit to the data while penalizing model complexity.

Each village is analyzed separately, using all available data within the village. Additionally, we consider two sets of covariates to be available. The *global* covariate set contains only variables obtainable from the geocoordinates, namely, the location’s density and distance to the perimeter, which represents a small amount covariate information that is convenient to collect. The other set contains these variables along with the socio-economic covariates listed in [Table 3](#).

2.4 Predicting out-of-sample infestation status

Let $\mathbf{y}_0 = \{y(\mathbf{s}_0) \mid \mathbf{s}_0 \in \mathcal{D} \setminus \mathcal{S}\}$ indicate the unknown infestation status at unvisited locations. Given a response \mathbf{y} observed at \mathcal{S} , the joint posterior predictive distribution for \mathbf{y}_0 is then

$$\Pr(\hat{\mathbf{y}}_0 \mid \mathbf{y}) = \int \Pr(\hat{\mathbf{y}}_0 \mid \mathbf{y}, \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathbf{y}) d\boldsymbol{\theta}. \quad (3)$$

We are interested in the total count of unvisited locations which are infested, defined as $I_0 = \mathbf{1} \cdot \mathbf{y}_0^\top$. To estimate the distribution $\Pr(\hat{\mathbf{y}}_0 \mid \mathbf{y})$, and hence $\Pr(\hat{I}_0 \mid \mathbf{y})$, from the posterior, we use INLA to generate 5,000 Monte Carlo samples from $p(\boldsymbol{\theta} \mid \mathbf{y})$, and then use each of these to draw a sample from $\Pr(\hat{\mathbf{y}}_0 \mid \mathbf{y}, \boldsymbol{\theta}^{(i)})$, resulting in 5,000 samples from (3).

2.5 An adaptive sampling strategy for infestation reduction

In the present study, our objective is not only to make predictions given data observed at $\mathcal{S} \subset \mathcal{D}$, but to choose the set \mathcal{S} itself to best control infestation. In this context, \mathcal{S} is referred to as the *sampling design*. To evaluate the quality of a sampling design, we specify a *design target*. In our context, this target is the number of houses selected for treatment (i.e. the size of \mathcal{S}), subject to the constraint that the true infestation rate among unvisited houses is below 5% (i.e. $I_0/n < 0.05$).

One possible strategy for choosing effective sampling designs is adaptive (or sequential) sampling. Rather than specifying our set of houses to sample completely before collecting data at these houses, we sample houses in batches: a collection of houses is selected, the data is gathered at these houses, a model is fit to the data so far, and a new batch of houses is chosen based on the model’s predictions. At the end of this process, the set of houses we have visited becomes our final sampling design.

Implementing our adaptive strategy requires the following [\[32\]](#): 1) an initial design \mathcal{S}_1 from which to fit the first model, 2) a *batch size* b as the number of new houses to add to the existing observations each iteration, 3) a *utility function* U to rank unobserved houses to target. For our application, we additionally require 4) a *termination condition* to predict whether the current design meets the infestation target.

For the utility function, a natural choice might be to rank a location $\mathbf{s}_0 \in \mathcal{D} \setminus \mathcal{S}$ according to its posterior predicted risk $\Pr(\hat{y}(\mathbf{s}_0) = 1 \mid \mathbf{y}) = \mathbf{E}[r(\mathbf{s}_0) \mid \mathbf{y}]$. However, this strategy has high sampling bias and fails to prioritize locations with potentially new information, hence ignoring possibly large amounts of the geospatial or covariate search space.

To better explore this space, we therefore balance this strategy with prioritizing locations of high uncertainty. Since predictions are generally less reliable with smaller sample sizes, the proposed utility function

transitions smoothly from prioritizing a location’s variance $\tilde{\nu}(\mathbf{s}_0)$, to prioritizing its expected risk $\tilde{r}(\mathbf{s}_0)$, where $\tilde{\nu}(\mathbf{s}_0)$ and $\tilde{r}(\mathbf{s}_0)$ have each been centered and scaled over the unvisited locations. If m_i is the current number of locations observed at iteration i of sampling, we rank locations according to the utility function

$$U(\mathbf{s}_0) = t(i; \alpha) \tilde{r}(\mathbf{s}_0) + (1 - t(i; \alpha)) \tilde{\nu}(\mathbf{s}_0) \quad (4)$$

where $t(i; \alpha) = ((m_i - m_1)/(n - m_1))^\alpha$ is a weighting function interpolating between 0 and 1. The *exploration parameter* $\alpha > 0$ controls the speed at which the predicted risk is prioritized, with $\alpha < 1$ transitioning to risk-based targeting more quickly and $\alpha > 1$ favoring exploration for longer.

We propose a termination condition based on our confidence that the current design satisfies the reduction target:

$$\Pr(\hat{I}_0 < \kappa n) \geq \gamma, \quad (5)$$

where κ is the desired infestation rate, and γ the desired confidence level. If \mathcal{S}_i is the current design, we can compute this probability using draws from $\Pr(\hat{I}_0 | \mathbf{y}_i)$ as described above. We fix $\kappa = 0.05$ and $\gamma = 0.95$ throughout the manuscript. Finally, we set the batch size $b = 3$, and sample the initial design \mathcal{S}_1 uniformly at random.

In summary, the adaptive sampling algorithm is as follows.

1. Sample initial design \mathcal{S}_1 uniformly and set $i = 1$
2. Fit the posterior distribution $p(\boldsymbol{\theta} | \mathbf{y}_i)$ as described in Section 2.2
3. If $\mathcal{D} \setminus \mathcal{S}_i$ satisfies (5), return \mathcal{S}_i . Otherwise, go to step 4
4. Assign each location $\mathbf{s}_0 \in \mathcal{D} \setminus \mathcal{S}_i$ a utility $U(\mathbf{s}_0)$ using (4)
5. Choose the b locations with highest utility and add them to \mathcal{S}_{i+1}
6. Set $i = i + 1$ and repeat steps 2-6.

2.6 Simulation study

We compare the performance of the adaptive sampling procedure from above using several values of the exploration parameter α , along with a random sampling procedure, on each of the five Guatemalan villages. To evaluate each resulting design, we record the size m of the final design, and the true infestation rate, defined as the number of truly infested houses not in the design divided by the total number of houses in the village. For all experiments, we used a batch size $b = 3$.

For each village, the following experiment is repeated 50 times. An initial set of 10 locations is chosen at random. Then, adaptive sampling using each of $\alpha = 0, 0.15, 0.3, 0.7, 1, 2$ is performed, as well as a random procedure which uniformly samples b new houses each iteration, resulting in 7 final designs to be evaluated. In this way, we apply each of the different procedures to the same set of randomized initial samples.

3 Results

3.1 Full-village analysis

A comparison of the model given in Section 2.2 to the two simpler alternatives is shown in Figure 1 for each village. The proposed model had a lower (better) DIC than the alternatives in all five villages, except for El Cerrón with the full covariate set, where the model with the spatial effects removed did slightly better. The spatial-only model ($\varepsilon(\mathbf{s})$ removed) had higher ML in all but 3 cases. However, the two models including a

spatial effect had similar performance according to both measures. In contrast, in El Guayabo, El Paternito, and La Prensa, these two models had substantially better performance than the model with spatial effects removed. Figure 1 also shows the computation time of the proposed model was considerably faster than the spatial-only model, while the model with no spatial effects was fastest.



Figure 1: **Model comparison based on two goodness-of-fit measures and computation time, by village and level of covariate information.** The best-performing model is indicated with a star in each case. “Global only” means only covariates available from the coordinates are available, and “All” indicates all 28 covariates were used. “ $Z(s)$ removed” refers to the model outlined in Sec. 2.2 but with no spatial effects, “ $\varepsilon(s)$ removed” refers to the independent effects removed, and “Both” refers to the full model. DIC and ML are defined in Mathematical definitions.

The proposed model, fit to all data in the village, also allowed us to draw inferences about the nature of *T. dimidiata* infestation at a local level. Table 1 summarizes the effective range after rescaling back to meters, while Figure 4 shows the full posterior distribution. The posterior mode was between 150m and 874m, or between 10% and 40% of the village diameter. Although there was notable variation in the effective range between the villages, it appears fairly similar between the two covariate sets for a given village, with a difference in mode between 6m and 223m. In the villages El Cerrón and El Paternito, the 95% highest posterior density interval (HPDI) for the effective range was quite large (spanning several kilometers), which suggests the spatial signal was weakest in these villages.

Examining the coefficients for the fixed effects, we found variation between the villages in the majority of the covariates in the full covariate set (Figure 2). The factors which consistently had a negative association with infestation, defined here as having a 50% HPDI fully below 0, were not having rats in the home (5 villages), having bedroom walls in good condition, and not keeping construction materials around the home (4 villages). The factors with a consistent positive association were having a kitchen outside the home (4 villages), evidence of bird nests inside, and keeping chickens outside the home (3 villages).

Village	Covariate set	Mode & 95% HPDI (meters)
El Amatillo	Global only	156 (37-931)
	All	150 (62-420)
El Cerrón	Global only	728 (4-8201)
	All	553 (6-7370)
El Guayabo	Global only	774 (334-1910)
	All	823 (358-2062)
El Paternito	Global only	651 (161-2860)
	All	874 (171-3751)
La Prensa	Global only	480 (212-1138)
	All	508 (205-1281)

Table 1: Posterior of the effective range from the full-village analysis. Each posterior is fit according to the model of Section 2.2, using two different sets of covariate information. The highest posterior density interval (HPDI) is the smallest interval such that 95% of the posterior mass is contained within it.

3.2 Simulation results

The results of the simulation experiment are shown in Figure 3. From the 50 final designs obtained from each group of strategy \times village \times covariate set, we calculated the mean and 90% confidence interval (CI) for the percentage of houses in the final design, and for the true infestation rate remaining in the village. In all villages and both covariate sets, all adaptive strategies had a smaller mean design size (final number of sampled houses) than random. Moreover, when the exploration parameter was less than $\alpha = 2$, the 90% CI for the design size was entirely below (i.e. not overlapping) the CI for random sampling. Comparing the top and bottom rows of Figure 3, we further find that, except for when $\alpha = 2$ and random sampling was used, having the full covariate set tended to lead to a smaller design size compared to having the global covariate set.

Although the adaptive strategies tended to produce smaller designs, they were also more likely to produce designs which failed to satisfy the design target of reducing the true infestation rate below 5%, especially when the exploration parameter α was too low. With the global covariate set, in four villages all adaptive strategies had a 90% confidence interval below the target threshold, i.e. at least 95% of the designs had a true infestation rate below 5% in the final sample, while in La Prensa only $\alpha \geq 0.7$ gave a CI below the target. With the full covariate set, the strategies with lower α missed the target threshold more frequently. The mean infestation rate was above 5% for $\alpha \leq 0.3$ in two villages, and in one village for $\alpha = 0.7$. Moreover, the CI contained 5% in four villages for $\alpha \leq 1$, and one village for $\alpha = 2$. The CI was always below 5% with random sampling.

Table 2 provides a higher-level perspective of the relative performance of the adaptive strategies, with data from all villages pooled together. Again, we find that the accuracy (percentage of observations which met the 5% infestation target) was lowest for lower values of α , especially with the full covariate effect, and that accuracy was inversely proportional to the difference in design size compared to random.

4 Discussion

The resilience of NTDs, combined with the resource-limited context in which they inherently reside, presents unique challenges for their control and elimination. Interventions must not only provide effective and long-term disruption of the disease, but also be logistically feasible to allow widespread application [44]. This leads to a difficult tradeoff between preserving resources while actually making a difference at the community level. Because of the complex environmental, ecological, and cultural aspects of NTDs, a further challenge

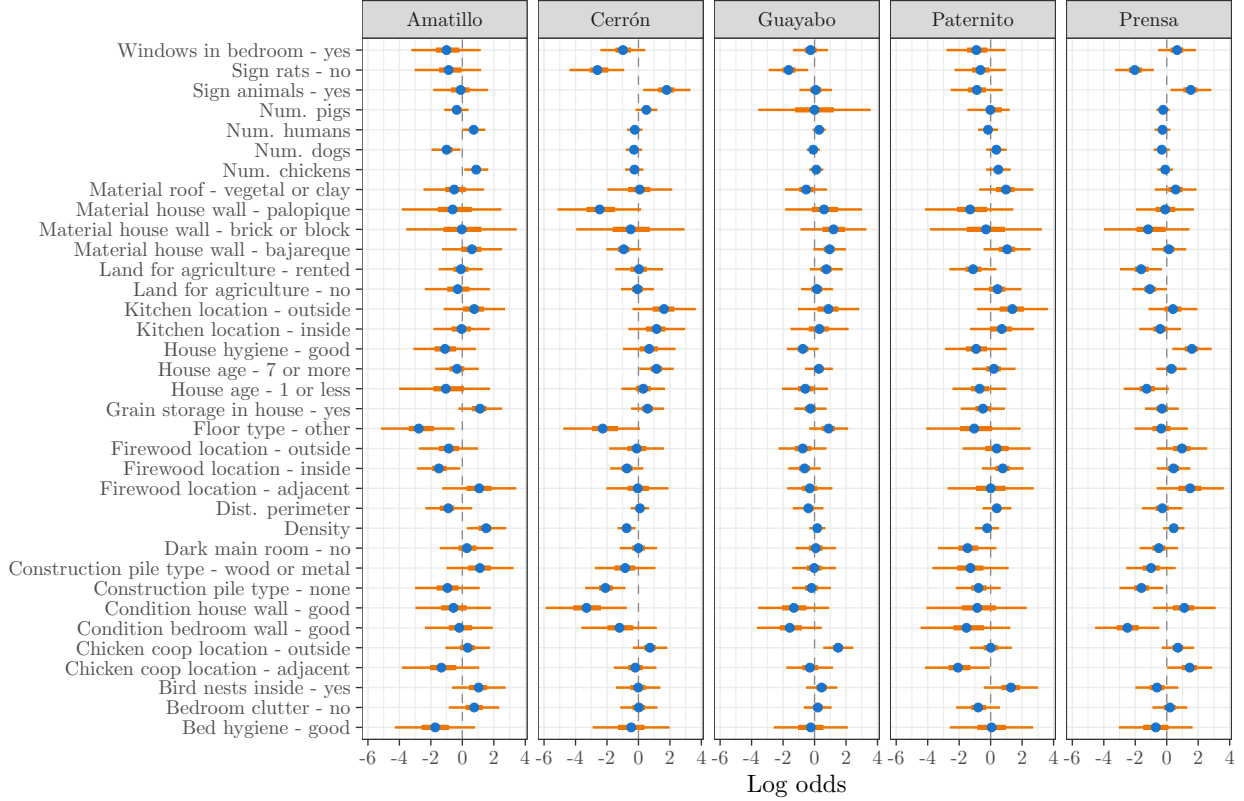


Figure 2: Fixed effect coefficients from fitting the model of Section 2.2 to all available data in each village. Coefficients are on log odds scale. The blue point is the posterior mean, the bold orange line is the 50% HPDI, and the thin orange line is the 95% HPDI.

is the need for spatially and temporally adaptive solutions for efficient control [25, 45].

We have proposed an algorithm which balances this tradeoff between efficiency and efficacy in the context of Chagas disease vector control in southeastern Guatemala. With the goal of reducing a village’s domiciliary Triatomine infestation rate below the government target of 5%, we consider an experimental design problem where only a subset of houses are treated for long-term vector elimination using locale-specific housing improvements. The algorithm uses a combination of adaptive sampling and Bayesian geostatistical modeling to minimize the number of treated houses while correctly predicting when the reduction target has been met.

The results from our simulation experiment show adaptive sampling strategies are universally more efficient than selecting houses at random, and are able to consistently predict whether the 5% infestation target has been achieved. In the case where only the spatial coordinates of houses are available, even the strategy which only prioritized perceived risk (i.e. the exploration parameter α was zero) was able to correctly predict the target was met with over 90% accuracy. This is surprising, as selecting houses based on current perceived risk will intuitively lead to biased predictions [31]. One possibility for this is that the spatial information alone leads to minimal shrinkage in the posterior during the initial stages of sampling, which would cause sampling to be nearly random even when $\alpha = 0$ (since unvisited locations will look very similar). A second explanation is that the termination condition we have used counteracts the effects of targeting perceived risk. In particular, sampling in this way can lead to higher variance at unvisited locations since we have neglected to gather relevant information about them, leading to a situation where \hat{I}_0 is low but the probability in (5) is high, and hence avoiding terminating prematurely.

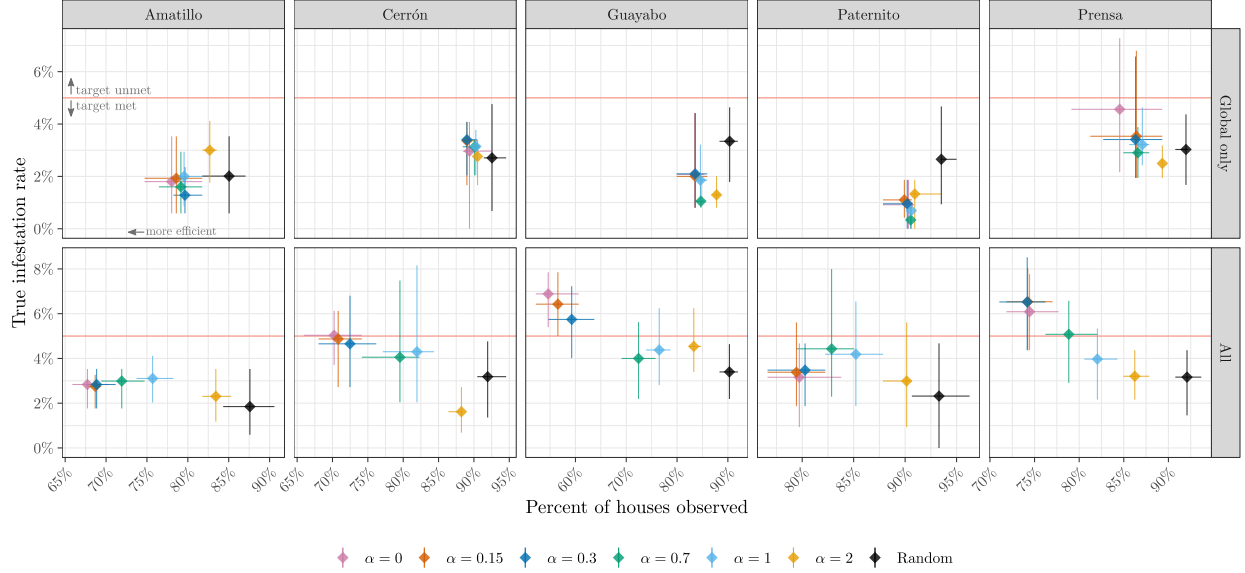


Figure 3: Results from the simulation experiment for each village and predictor set. The adaptive sampling procedure for varying α , along with random sampling, is performed for 50 initial sets of 10 houses. The x-axis is the percentage of houses in the final sampling design, and the y-axis is the true infestation rate remaining in the village. Diamonds show means, while cross-hairs show the 90% confidence intervals of the data corresponding to each axis. The red line corresponds to the 5% reduction target.

Our second major finding is that including additional socioeconomic covariates can significantly reduce the number of sampled houses, but only for the adaptive sampling strategies. However, this additional information also leads to a greater opportunity for bias in the adaptive strategies with lower exploration parameter, and even larger values of the parameter ($\alpha = 1, 2$) occasionally failed to meet the target, as did random sampling. While this is not ideal, it is reassuring that even with no exploration ($\alpha = 0$), the true infestation rate was below 8% in nearly all of the final designs (Table 2). Since this threshold has been shown to guarantee *T. cruzi* elimination throughout Central America [10], a viable strategy could be to set the target threshold slightly below the level actually desired, thus allowing maximum sampling efficiency while still meeting the actual target.

While the assumption that detailed socioeconomic information is available for the whole community beforehand is not necessarily realistic, the efficiency gained by including these covariates opens up new options for *T. dimidiata* control. For example, this information could be collected beforehand by community members or with a quick survey, thereby reducing the number of homes to search for infestation, which is more time-consuming to collect and requires trained personnel. Moreover, this could allow a more thorough search for Triatomine presence, which is a notoriously noisy measurement [20, 46].

On a practical level, our sampling strategy could compliment the EcoHealth approach in several ways. If local resources for housing improvements are readily available, then adaptive sampling could be used to conserve insecticide and the time of health personnel, while allowing house improvements for all residents who want them. If such resources are scarce, adaptive sampling could further indicate which homes are most in need of immediate improvement. Adaptive sampling could also be used to more quickly develop a localized improvement strategy, by identifying socioeconomic variables most associated with infestation risk from a subsample of relevant houses.

It has been previously suggested that adaptive geostatistical sampling based on prediction variance, followed by targeting areas of higher risk in later stages, could be a way to accurately identify areas for

Exploration parameter	Covariate set	Accuracy (%)		Difference (%) median (95% CI)
		5% target	8% target	
$\alpha = 0$	Global only	91.6	100	4.8 (1.6-11.7)
	All	54.4	97.6	19.4 (11.2-37.3)
$\alpha = 0.15$	Global only	97.6	100	4.8 (2.0-10.6)
	All	52.8	97.2	18.9 (11.2-36.1)
$\alpha = 0.3$	Global only	98.0	100	4.4 (2.0-9.6)
	All	57.2	97.2	18.4 (11.2-33.7)
$\alpha = 0.7$	Global only	99.6	100	4.4 (2.0-7.2)
	All	77.2	97.6	14.1 (8.4-19.4)
$\alpha = 1$	Global only	100	100	4.4 (2.0-7.3)
	All	81.2	97.6	10.6 (5.6-15.7)
$\alpha = 2$	Global only	100	100	2.8 (0-4.1)
	All	94.8	100	4.4 (1.8-8.4)
Random	Global only	99.2	100	-
	All	98.4	100	-

Table 2: Performance of adaptive sampling in the simulation study, compared to random sampling. Accuracy is the percentage of designs which met the control target, i.e. the out-of-sample infestation rate was below 5% (or 8%). Difference from random is the percentage of the village in the design, minus the corresponding percentage using random sampling, calculated for each initial design.

public health interventions while conserving resources [32]. Altogether, our results support this idea, while rigorously comparing different balances of prioritizing variance and risk. Our experiments show that with α sufficiently high, this strategy is robust to different communities and levels of covariate information. The most preferable setting for α , however, will depend on the particularities of a given intervention. For example, in the initial stages of an intervention it may be preferable to visit as many communities as possible, with the understanding that the control target may fail to be reached in a few communities, while in later stages a smaller number of remaining areas can be targeted with a higher level of precision.

4.1 Full-village analysis

The results from the full-village analysis, where several models were fit to all available data in each of the villages, provide further insight into factors associated with *T. dimidiata* infestation in southeastern Guatemala. First, the superior explanatory power of the spatial models emphasizes the spatial nature of infestation, something not accounted for in previous studies of infestation risk [4, 20]. There are several possible explanations for the spatial autocorrelation present in this data. One possibility is a limited migratory range from existing domestic populations, leading to local clustering following the insect’s dispersal season. Such a dynamic is supported by several population genetics studies, which have found insects in nearby houses and adjacent villages are more related [47, 48], as well as insects from highly-infested houses and their neighbors following insecticide application [15]. Another explanation is that spatial patterns are due to unmeasured covariates, which are themselves spatially autocorrelated. While our results show spatial effects remain even after accounting for a number of socioeconomic variables, environmental factors such as altitude, temperature, and precipitation have been shown to be mildly correlated with *T. dimidiata* infestation [49, 50].

A second important finding from the full-village analysis is the variation in fixed-effect coefficients among the five villages. While several covariates, such as the condition of bedroom walls, consistently had a meaningful association with infestation, many others varied in their explanatory power, and even whether there was a positive or negative association. Important fixed-effects also varied compared to previous *T.*

dimidiata studies in different areas. In particular, the distance from the village perimeter was found to lead to significantly higher infestation rates and vector abundance in Yucatan, Mexico [50]. However, not only was there no correlation ($\rho = 0.02$) between village perimeter and infestation overall in our data, but we found the importance of village perimeter as a covariate was negligible using either covariate set, which shows there is little association between these variables when accounting for spatial and various covariate effects as well. This could be due to the close proximity of several of the villages, or to the high level of deforestation in our study region, leading to a lack of sylvatic populations in the surrounding environment [51]. Ultimately, these results all add to the existing evidence that risk factors for *T. dimidiata* infestation must be considered in a local context [9, 22].

4.2 Limitations and future work

This study has a few limitations. First, our methods focus on bringing the infestation rate below 5% among untreated houses, with the logic that uninfested, unvisited homes already have factors not conducive to *T. dimidiata* infestation, and that the EcoHealth approach empowers communities to maintain these factors throughout the village [21]. However, it is certainly possible for the infestation rate to still rebound. For example, if a previous insecticide application was recent enough, some households may not be uninfested due to favorable conditions but rather that the vector population has had insufficient time to fully reestablish. Further, maintaining certain housing conditions alone may not be enough if changing environmental conditions alter the relationship between housing conditions and infestation risk. Additional research would therefore need to confirm the long-term effects of allowing a small fraction of houses to remain infested after an intervention.

We have also ignored any practical constraints when choosing locations during sampling, such as the travel time to selected points, or the logistics of centralizing incoming data after each iteration of sampling. It therefore may be more realistic to impose a penalty in (4) based on distance from current samples, or to increase the batch size b to allow a more natural sampling schedule. Finally, our decision to lump all signs of infestation into a single indicator ignores potentially useful information, including their different implications for disease risk. A recent study in Jutiapa found a large difference between adult and juvenile infestation rates [51], so it may be beneficial to separate these variables.

In the simulation study, several parameters of the adaptive sampling procedure were fixed throughout, such as the batch size b and size of the initial design, to limit computational overhead. While our results demonstrate the settings we chose can be effective across multiple villages, an analysis involving the interaction between these parameters would provide a more thorough summary of when our procedure works best, and its robustness to different parameter settings.

The methods developed in this work can be applied in contexts other than Chagas vector control. In particular, adaptive geostatistical sampling using the utility function (4) and termination condition (5) can be applied nearly directly to other NTD control initiatives seeking to efficiently meet a reduction target. For example, schistosomiasis prevalence in schools has been shown to have spatial autocorrelation, and covariate information for schools can be acquired through teacher-given surveys [52]. Adaptive geostatistical sampling could therefore be applied to efficiently target schools at greater risk. Our methods could also be applied to other Chagas endemic regions, and extended to account for more complex ecological data, such as jointly modeling several vector species and lifestages, or using spatiotemporal models [53] to improve the efficiency of follow-up surveys and adapt to seasonal effects. More generally, our framework could be used for monitoring multiple, possibly interacting, diseases [54]. This would allow sequential samples to take into account information both between and among pathogens.

In conclusion, we have proposed an adaptive strategy for public health interventions, which transitions from prioritizing areas of greatest uncertainty to those perceived to be most at risk. This would allow control initiatives to use resources more efficiently, by targeting areas of greatest need while still benefiting the entire community. We believe the methods used in this work are well-suited to address the complex

ecological, biological, and social factors inherent to disease spread, and hence are applicable to a wide range of epidemiological systems.

Acknowledgements

BC, LHD and JGY acknowledge support from the National Institutes of Health 1P20 GM125498-01 Centers of Biomedical Research Excellence Award. BC is also supported as a Fellow of the National Science Foundation under NRT award DGE-1735316, and LHD by the National Science Foundation award EPS-2019470. The data used in this work was supported by the tres países project grant IDRC 106531. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

The authors are grateful to the researchers, Ministry of Health personnel, and people of Chiquimula who participated in data collection.

References

- [1] P. J. Hotez, M. E. Bottazzi, C. Franco-Paredes, S. K. Ault, and M. R. Periago, “The neglected tropical diseases of Latin America and the Caribbean: a review of disease burden and distribution and a roadmap for control and elimination,” *PLOS Neglected Tropical Diseases*, vol. 2, no. 9, p. e300, 2008.
- [2] C. P. Bouillon, “The millennium development goals in Latin America and the Caribbean: progress, priorities and IDB support for their implementation (August 2005).” Available at SSRN: <https://ssrn.com/abstract=1543858>.
- [3] C. Franco-Paredes, A. Von, A. Hidron, A. J. Rodríguez-Morales, I. Tellez, M. Barragán, D. Jones, C. G. Náquira, and J. Mendez, “Chagas disease: an impediment in achieving the millennium development goals in Latin America,” *BMC International Health and Human Rights*, vol. 7, no. 1, pp. 1–6, 2007.
- [4] D. M. Bustamante, C. Monroy, S. Pineda, A. G. Rodas, X. Castro, V. Ayala, J. Quiñones, B. Moguel, and R. Trampe, “Risk factors for intradomiciliary infestation by the Chagas disease vector *Triatoma dimidiata* in Jutiapa, Guatemala,” *Cadernos de Saúde Pública*, vol. 25, pp. S83–S92, 2009.
- [5] World Health Organization, “Chagas disease in Latin America: an epidemiological update based on 2010 estimates,” *Weekly Epidemiological Record*, vol. 90, no. 06, pp. 33 – 44, 2015.
- [6] World Health Organization, “Chagas disease (American trypanosomiasis).” Available from: <https://www.who.int/health-topics/chagas-disease>, 2021 [cited 20 September 2021].
- [7] C. J. Schofield, *Triatominae: biology & control*. Bognor Regis: Eurocommunica Publications, 1994.
- [8] Y. Yamagata and J. Nakagawa, “Control of Chagas disease,” *Advances in Parasitology*, vol. 61, pp. 129–165, 2006.
- [9] R. King, C. Cordon-Rosales, J. Cox, C. Davies, and U. Kitron, “*Triatoma dimidiata* infestation in Chagas disease endemic regions of Guatemala: comparison of random and targeted cross-sectional surveys,” *PLoS Neglected Tropical Diseases*, vol. 5, no. 4, p. e1035, 2011.
- [10] H. Aiga, E. Sasagawa, K. Hashimoto, J. Nakamura, C. Zúniga, J. Chévez, H. Hernández, J. Nakagawa, and Y. Tabaru, “Chagas disease: assessing the existence of a threshold for bug infestation rate,” *American Journal of Tropical Medicine and Hygiene*, vol. 86, no. 6, pp. 972–979, 2012.
- [11] J. Peterson, K. Yoshioka, K. Hashimoto, A. Caranci, N. Gottdenker, C. Monroy, A. Saldaña, S. Rodriguez, P. Dorn, and C. Zúniga, “Chagas disease epidemiology in Central America: an update,” *Current Tropical Medicine Reports*, vol. 6, no. 2, pp. 92–105, 2019.
- [12] C. Monroy, D. M. Bustamante, A. G. Rodas, M. E. Enriquez, and R. G. Rosales, “Habitats, dispersion and invasion of sylvatic *Triatoma dimidiata* (Hemiptera: Reduviidae: Triatominae) in Peten, Guatemala,” *Journal of Medical Entomology*, vol. 40, no. 6, pp. 800–806, 2003.

- [13] E. Dumonteil, H. Ruiz-Piña, E. Rodríguez-Félix, M. Barrera-Pérez, M. J. Ramirez-Sierra, J. E. Rabinovich, and F. Menu, “Re-infestation of houses by *Triatoma dimidiata* after intra-domicile insecticide application in the Yucatan peninsula, Mexico,” *Memórias do Instituto Oswaldo Cruz*, vol. 99, no. 3, pp. 253–256, 2004.
- [14] J. Manne, J. Nakagawa, Y. Yamagata, A. Goehler, J. S. Brownstein, and M. C. Castro, “Triatomine infestation in Guatemala: spatial assessment after two rounds of vector control,” *American Journal of Tropical Medicine and Hygiene*, vol. 86, no. 3, pp. 446–454, 2012.
- [15] S. H. Cahan, L. Orantes, K. Wallin, J. Hanley, D. M. Rizzo, L. Stevens, P. Dorn, A. G. Rodas, and C. Monroy, “Residual survival and local dispersal drive reinfestation by *Triatoma dimidiata* following insecticide application in Guatemala,” *Infection, Genetics and Evolution*, vol. 74, p. 104000, 2019.
- [16] K. Yoshioka, E. Provedor, and J. Manne-Goehler, “The resilience of *Triatoma dimidiata*: an analysis of reinfestation in the Nicaraguan Chagas disease vector control program (2010–2016),” *PLoS ONE*, vol. 13, no. 8, pp. 1–18, 2018.
- [17] A. Gómez-Palacio, S. Arboleda, E. Dumonteil, and A. Townsend Peterson, “Ecological niche and geographic distribution of the Chagas disease vector, *Triatoma dimidiata* (Reduviidae: Triatominae): Evidence for niche differentiation among cryptic species,” *Infection, Genetics and Evolution*, vol. 36, pp. 15–22, 2015.
- [18] R. Lima-Cordón, L. Stevens, E. Solórzano Ortíz, G. A. Rodas, S. Castellanos, A. G. Rodas, V. Abrego, C. Zúñiga Valeriano, and C. Monroy, “Implementation science: epidemiology and feeding profiles of the Chagas vector *Triatoma dimidiata* prior to EcoHealth intervention for three locations in Central America,” *PLoS Neglected Tropical Diseases*, vol. 12, no. 11, p. e0006952, 2018.
- [19] S. Justi, S. Cahan, L. Stevens, C. Monroy, R. Lima-Cordón, and P. Dorn, “Vectors of diversity: genome wide diversity across the geographic range of the Chagas disease vector *Triatoma dimidiata* sensu lato (Hemiptera: Reduviidae),” *Molecular Phylogenetics and Evolution*, vol. 120, pp. 144–150, 2018.
- [20] D. M. Bustamante, M. M. Hernández, N. Torres, C. Zúñiga, W. Sosa, V. De Abrego, and C. Monroy, “Information to act: household characteristics are predictors of domestic infestation with the Chagas vector *Triatoma dimidiata* in Central America,” *American Journal of Tropical Medicine and Hygiene*, vol. 93, no. 1, pp. 97–107, 2015.
- [21] C. Monroy, D. M. Bustamante, S. Pineda, A. G. Rodas, X. Castro, V. Ayala, J. Quiñones, and B. Moguel, “House improvements and community participation in the control of *Triatoma dimidiata* re-infestation in Jutiapa, Guatemala,” *Cadernos de Saúde Pública*, vol. 25, no. Sup. 1, pp. S168–S178, 2009.
- [22] D. M. Bustamante, S. De Urioste-Stone, J. Juárez, and P. Pennington, “Ecological, social and biological risk factors for continued *Trypanosoma cruzi* transmission by *Triatoma dimidiata* in Guatemala,” *PLoS ONE*, vol. 9, no. 8, p. e104599, 2014.
- [23] L. Stevens, C. Monroy, A. G. Rodas, and P. Dorn, “Hunting, swimming, and worshipping: human cultural practices illuminate the blood meal sources of cave dwelling Chagas vectors (*Triatoma dimidiata*) in Guatemala and Belize,” *PLoS Neglected Tropical Diseases*, vol. 8, no. 9, p. e3047, 2014.
- [24] J. Peterson, K. Hashimoto, K. Yoshioka, P. Dorn, N. Gottdenker, A. Caranci, L. Stevens, C. Zuniga, A. Saldaña, S. Rodriguez, and C. Monroy, “Chagas disease in Central America: recent findings and current challenges in vector ecology and control,” *Current Tropical Medicine Reports*, vol. 6, no. 2, pp. 76–91, 2019.
- [25] M. Booth and A. Clements, “Neglected tropical disease control – the case for adaptive, location-specific solutions,” *Trends in Parasitology*, vol. 34, no. 4, pp. 272–282, 2018.
- [26] Guatemala Ministerio de Salud, “Sistema de información gerencial de salud (SIGSA) - enfermedades transmitidas por vectores, años 2012 al 2019.” Available from: <https://sigsa.mspas.gob.gt/datos-de-salud/morbilidad/enfermedades-transmitidas-por-vectores>, 2019 [cited 12 July 2021].
- [27] D. Lucero, L. Morrissey, D. Rizzo, A. G. Rodas, R. Garnica, L. Stevens, D. Bustamante, and C. Mon-

- roy, “Ecohealth interventions limit triatomine reinfestation following insecticide spraying in La Brea, Guatemala,” *American Journal of Tropical Medicine and Hygiene*, vol. 88, no. 4, pp. 630–637, 2013.
- [28] P. Diggle, R. Menezes, and T. li Su, “Geostatistical inference under preferential sampling,” *Journal of the Royal Statistical Society. Series C: Applied Statistics*, vol. 59, no. 2, pp. 191–232, 2010.
- [29] J. Hanley, D. Rizzo, L. Stevens, S. Helms Cahan, P. Dorn, L. Morrissey, A. G. Rodas, L. Orantes, and C. Monroy, “Novel evolutionary algorithm identifies interactions driving infestation of *Triatoma dimidiata*, a Chagas disease vector,” *The American Journal of Tropical Medicine and Hygiene*, vol. 103, no. 2, pp. 735–744, 2020.
- [30] E. N. I. Weeks, C. Córdón-Rosales, C. Davies, S. Gezan, M. Yeo, and M. M. Cameron, “Risk factors for domestic infestation by the Chagas disease vector, *Triatoma dimidiata* in Chiquimula, Guatemala,” *Bulletin of Entomological Research*, vol. 103, no. 6, pp. 634–643, 2013.
- [31] P. Diggle, B. Amoah, C. Fronterre, E. Giorgi, and O. Johnson, “Rethinking neglected tropical disease prevalence survey design and analysis: a geospatial paradigm,” *Transactions of the Royal Society of Tropical Medicine and Hygiene*, vol. 115, no. 3, pp. 208–210, 2021.
- [32] M. Chipeta, D. Terlouw, K. Phiri, and P. Diggle, “Adaptive geostatistical design and analysis for prevalence surveys,” *Spatial Statistics*, vol. 15, pp. 70–84, 2016.
- [33] A. N. Kabaghe, M. G. Chipeta, R. S. McCann, K. S. Phiri, M. Van Vugt, W. Takken, P. Diggle, and A. D. Terlouw, “Adaptive geostatistical sampling enables efficient identification of malaria hotspots in repeated cross-sectional surveys in rural Malawi,” *PLoS ONE*, vol. 12, no. 2, pp. 1–14, 2017.
- [34] R. Andrade-Pacheco, F. Rerolle, J. Lemoine, L. Hernandez, A. Meité, L. Juzziwelo, A. F. Bibaut, M. J. van der Laan, B. F. Arnold, and H. J. Sturrock, “Finding hotspots: development of an adaptive spatial sampling approach,” *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020.
- [35] M. Chipeta, D. Terlouw, K. Phiri, and P. Diggle, “Inhibitory geostatistical designs for spatial prediction taking account of uncertain covariance structure,” *Environmetrics*, vol. 28, no. 1, pp. 1–11, 2017.
- [36] C. Fronterre, B. Amoah, E. Giorgi, M. C. Stanton, and P. Diggle, “Design and analysis of elimination surveys for neglected tropical diseases,” *The Journal of Infectious Diseases*, vol. 221, no. Sup. 5, pp. S554–S560, 2020.
- [37] A. Gelfand, P. Diggle, M. Fuentes, and P. Guttorp, *Handbook of spatial statistics*. Boca Raton: CRC Press, 1 ed., 2010.
- [38] P. Diggle, J. Tawn, and R. Moyeed, “Model-based geostatistics,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 47, no. 3, pp. 299–350, 1998.
- [39] B. Matérn, “Spatial variation. phd thesis, Stockholm university,” 1960.
- [40] G.-A. Fuglstad, D. Simpson, F. Lindgren, and H. Rue, “Constructing priors that penalize the complexity of Gaussian random fields,” *Journal of the American Statistical Association*, vol. 114, no. 525, pp. 445–452, 2019.
- [41] R Core Team, *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021.
- [42] H. Rue, S. Martino, and N. Chopin, “Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion).,” *Journal of the Royal Statistical Society B*, vol. 71, pp. 319–392, 2009.
- [43] F. Lindgren and H. Rue, “Bayesian spatial modelling with R-INLA,” *Journal of Statistical Software*, vol. 63, no. 19, pp. 1–25, 2015.
- [44] World Health Organization, “Ending the neglect to attain the sustainable development goals: a sustainability framework for action against neglected tropical diseases 2021-2030,” *Geneva: World Health Organization*, vol. Licence: CC BY-NC-SA 3.0 IGO, 2021.
- [45] J. Cable, I. Barber, B. Boag, A. R. Ellison, E. R. Morgan, K. Murray, E. L. Pascoe, S. M. Sait, A. J. Wilson, and M. Booth, “Global change, parasite transmission and disease control: Lessons from

- ecology,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 372, no. 1719, p. 20160088, 2017.
- [46] C. Monroy, M. Mejia, A. G. Rodas, R. Rosales, M. Horio, and Y. Tabaru, “Comparison of indoor searches with whole house demolition collections of the vectors of Chagas disease and their indoor distribution,” *Medical Entomology and Zoology*, vol. 49, no. 3, pp. 195–200, 1998.
 - [47] P. Dorn, S. Melgar, V. Rouzier, A. Gutierrez, C. Combe, R. Rosales, A. G. Rodas, S. Kott, D. Salvia, and C. Monroy, “The Chagas vector, *Triatoma dimidiata* (Hemiptera: Reduviidae), is panmictic within and among adjacent villages in Guatemala,” *Journal of Medical Entomology*, vol. 40, no. 4, pp. 436–440, 2003.
 - [48] L. Stevens, C. Monroy, A. G. Rodas, R. Hicks, D. Lucero, L. Lyons, and P. Dorn, “Migration and gene flow among domestic populations of the chagas insect vector *Triatoma dimidiata* (Hemiptera: Reduviidae) detected by microsatellite loci,” *Journal of Medical Entomology*, vol. 52, no. 3, pp. 419–428, 2015.
 - [49] D. M. Bustamante, C. Monroy, A. G. Rodas, J. A. Juarez, and J. B. Malone, “Environmental determinants of the distribution of Chagas disease vectors in south-eastern Guatemala,” *Geospatial Health*, vol. 1, no. 2, pp. 199–211, 2007.
 - [50] M. J. Ramirez-Sierra, M. Herrera-Aguilar, S. Gourbière, and E. Dumonteil, “Patterns of house infestation dynamics by non-domiciliated *Triatoma dimidiata* reveal a spatial gradient of infestation in rural villages and potential insect manipulation by *Trypanosoma cruzi*,” *Tropical Medicine and International Health*, vol. 15, no. 1, pp. 77–86, 2010.
 - [51] D. Penados, J. Pineda, M. Catalan, M. Avila, L. Stevens, E. Agreda, and C. Monroy, “Infestation dynamics of *triatoma dimidiata* in highly deforested tropical dry forest regions of Guatemala,” *Memorias do Instituto Oswaldo Cruz*, vol. 115, no. 10, pp. 1–8, 2020.
 - [52] H. J. Sturrock, P. W. Gething, R. A. Ashton, J. H. Kolaczinski, N. B. Kabatereine, and S. Brooker, “Planning schistosomiasis control: Investigation of alternative sampling strategies for *Schistosoma mansoni* to target mass drug administration of praziquantel in East Africa,” *International Health*, vol. 3, no. 3, pp. 165–175, 2011.
 - [53] S. Banerjee, B. P. Carlin, and A. Gelfand, *Hierarchical modeling and analysis for spatial data*. Boca Raton: CRC Press, 2 ed., 2014.
 - [54] B. F. Arnold, H. M. Scobie, J. W. Priest, and P. J. Lammie, “Integrated serologic surveillance of population immunity and disease transmission,” *Emerging Infectious Diseases*, vol. 24, no. 7, p. 1188, 2018.
 - [55] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van Der Linde, “Bayesian measures of model complexity and fit,” *Journal of the Royal Statistical Society: Series B*, vol. 64, no. 4, pp. 583–639, 2002.

Supporting information

Mathematical definitions.

Matérn covariance function

For any pair of points at distance d from each other, the Matérn covariance between these points is

$$C(d) = \frac{\sigma_s}{2^{\nu-1}\Gamma(\nu)} (\kappa d)^\nu K_\nu(\kappa d),$$

where Γ is the Gamma function and K_ν the modified Bessel function. The parameters σ_s and ν are the spatial standard deviation and smoothness, respectively, while κ is implicitly defined via the effective range $\rho = \sqrt{8\nu}/\kappa$, which is the distance at which the correlation between points roughly becomes 0.1.

Deviance information criterion and marginal likelihood

Let \mathcal{M} denote a statistical model of interest. The deviance information criterion of \mathcal{M} is

$$D(\bar{\boldsymbol{\theta}}) + 2p_D,$$

where $D(\boldsymbol{\theta}) = -2\log(p(\mathbf{y} \mid \boldsymbol{\theta}))$ is the deviance, $\bar{\boldsymbol{\theta}}$ the posterior expectation of \mathcal{M} , and p_D the effective number of parameters [55].

The marginal likelihood is

$$p(\mathbf{y} \mid \mathcal{M}) = \int p(\mathbf{y} \mid \mathcal{M}, \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathcal{M}) d\boldsymbol{\theta}.$$

House factors	Values/range
Bed hygiene	Good, poor
Sign of bird nests inside	Yes, no
Location of chicken coop	Inside or adjacent to house, outside house, none
Clutter in bedroom	Yes, no
Poorly lit bedroom	Yes, no
Floor material	Dirt, other
Piles of construction material	Adobe or clay, wood or metal, none
Firewood location	Inside house, directly outside, outside, none
Grain storage in house	Yes, no
House age	Less than one year, 2-6 years, more than 7 years
House hygiene	Good, poor
Kitchen location	Inside house, outside, shared or none
Land for agriculture	Rented, owned, none
Sign of rats	Yes, no
Sign of other small animals	Yes, no
Bedroom wall condition	Good, deteriorated
Wall condition throughout home	Good, deteriorated
Material house wall	Adobe, bajareque, palopique, brick or other
Material roof	Aluminum or cement, clay or vegetal material, nylon panels
Windows in bedroom	Yes, no
Number residents	1-15
Number chickens	0-60
Number dogs	0-12
Number pigs	0-12

Table 3: Socioeconomic variables used for model fitting.

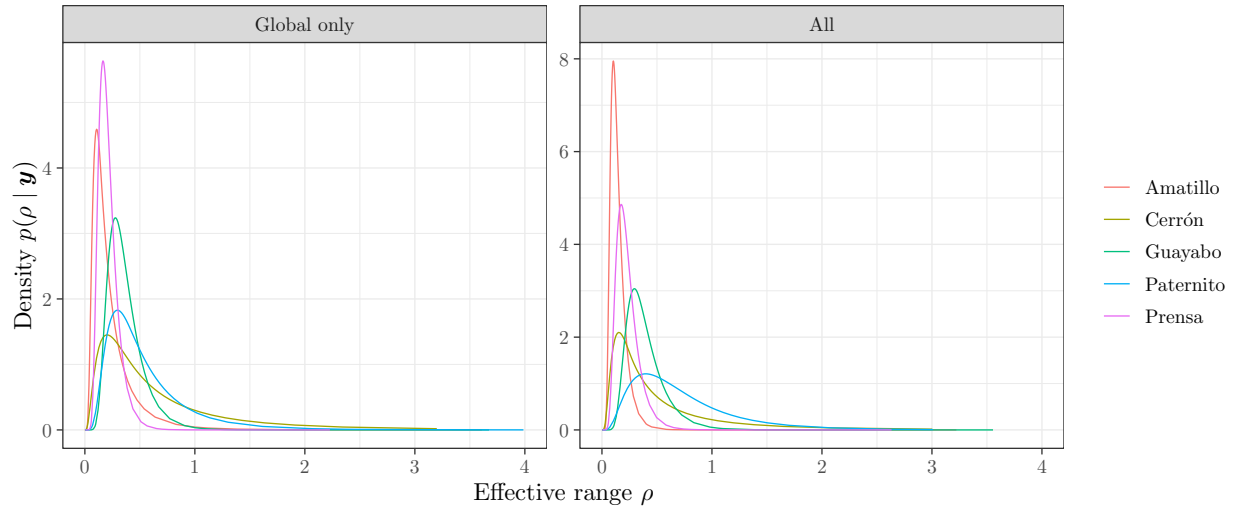


Figure 4: Effective range from full-village analysis. The posterior marginal $p(\rho | \mathbf{y})$ of the effective range, from fitting the model of Section 2.2 to all locations in each village. ‘Global only’ means only covariates available from the coordinates are available, and ‘All’ indicates all 28 covariates were used.