Predictive risk mapping of lymphatic filariasis residual hotspots in American Samoa using demographic and environmental factors

Short title: Spatial prediction of Lymphatic filariasis in American Samoa

Authors

Angela M. Cadavid Restrepo1*, Beatris Mario Martin1, Saipale Fuimaono2, Archie C.A. Clements3, Patricia M. Graves4, Colleen L. Lau1

Affiliations

1. School of Public Health, Faculty of Medicine, The University of Queensland, Brisbane, Queensland, Australia
2. American Samoa Department of Health, Pago Pago, American Samoa
3. Curtin School of Population Health, Faculty of Health Sciences, Curtin University, Perth, Western Australia, Australia
4. College of Public Health, Medical and Veterinary Sciences, James Cook University, Cairns, Queensland, Australia

*Corresponding author
E-mail: a.cadavidrestrepo@uq.edu.au (AMCR)

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

American Samoa successfully completed seven rounds of mass drug administration (MDA) for lymphatic filariasis (LF) from 2000-2006. The territory passed the school-based transmission assessment surveys in 2011 and 2015 but failed in 2016. One of the key challenges after the implementation of MDA is the identification of any residual hotspots of transmission.

Method

Based on data collected in a 2016 community survey in persons aged ≥8 years, Bayesian geostatistical models were developed for LF antigen (Ag), and Wb123, Bm14, Bm33 antibodies (Abs) to predict spatial variation in infection markers using demographic and environmental factors (including land cover, elevation, rainfall, distance to the coastline and distance to streams).

Results

In the Ag model, females had a 29.6% (95% CrI: 16.0–41.1%) lower risk of being Ag-positive than males. There was a 1.4% (95% CrI: 0.02–2.7%) increase in the odds of Ag positivity for every year of age. Also, the odds of Ag-positivity increased by 0.6% (95% CrI: 0.06–0.61%) for each 1% increase in tree cover. The models for Wb123, Bm14 and Bm33 showed similar significant associations as the Ag model for sex, age and tree coverage.

After accounting for the effect of covariates, the radii of the clusters were larger for Bm14 and Bm33 Abs compared to Ag and Wb123 Ab. The predictive maps showed that Ab-positivity was more widespread across the territory, while Ag-positivity was more confined to villages in the north-west of the main island.
Conclusion

The findings may facilitate more specific targeting of post-MDA surveillance activities by prioritising those areas at higher risk of ongoing transmission.

Key words

Wuchereria bancrofti; mass drug administration surveillance; environment; geographic information systems; Pacific Islands; geostatistics; predictive model
Author summary

The Global Programme to Eliminate Lymphatic filariasis (LF) aims to interrupt transmission by implementing mass drug administration (MDA) of antifilarial drugs in endemic areas; and to alleviate suffering of those affected through improved morbidity management and disability prevention. Significant progress has been made in the global efforts to eliminate LF. One of the main challenges faced by most LF-endemic countries that have implemented MDA is to effectively undertake post-validation surveillance to identify residual hotspots of ongoing transmission. American Samoa conducted seven rounds of MDA for LF between 2000 and 2006. Subsequently, the territory passed transmission assessment surveys in February 2011 (TAS-1) and April 2015 (TAS-2). However, the territory failed TAS-3 in September 2016, indicating resurgence. We implemented a Bayesian geostatistical analysis to predict LF prevalence estimates for American Samoa and examined the geographical distribution of the infection using sociodemographic and environmental factors. Our observations indicate that there are still areas with high prevalence of LF in the territory, particularly in the north-west of the main island of Tutuila. Bayesian geostatistical approaches have a promising role in guiding programmatic decision making by facilitating more specific targeting of post-MDA surveillance activities and prioritising those areas at higher risk of ongoing transmission.
Introduction

Lymphatic filariasis (LF) is a vector-borne parasitic disease caused by three species of filarial worms – *Wuchereria bancrofti*, *Brugia malayi*, and *B. timori* (1). The presence of adult worms in the lymphatic vessels leads to damage of the lymphatic system, causing clinical disease characterised by lymphoedema of the limbs or genitals, such as elephantiasis and scrotal hydrocoele (1). LF is one of the leading causes of chronic disability worldwide, being responsible for over 5 million disability-adjusted life years before the implementation of elimination strategies against the infection (2, 3).

In 1997, the World Health Organization (WHO) targeted LF for global elimination as a public health problem by 2020 (4). Subsequently, WHO launched the Global Programme to Eliminate Lymphatic Filariasis (GPELF) in 2000 that included two strategies: first, the implementation of mass drug administration (MDA) to interrupt the community-level transmission of LF, and second, management and prevention of morbidity and disability for people with chronic complications (5). By 2019, 72 countries were still considered endemic by the GPELF and 50 still required MDA (6). A number of countries have already achieved validation of LF elimination as a public health problem after intensive community-based MDA programs (including Cambodia, The Cook Islands, Egypt, Kiribati, Malawi, Maldives, Marshall Islands, Niue, Palau, Sri Lanka, Thailand, Togo, Tonga, Vanuatu, Viet Nam, Wallis and Futuna, and Yemen) (6). Some countries have stopped MDA and are under surveillance to determine if LF elimination criteria have been met. One of the main challenges faced by most LF-endemic countries that have implemented MDA is to effectively undertake post-MDA and post-validation surveillance (7).
American Samoa successfully completed seven rounds of MDA with a single dose of diethylcarbamazine (DEC) and albendazole from 2000-2006. Subsequently, the territory passed the WHO-recommended school-based transmission assessment surveys (TAS) conducted in 2011 (TAS-1) and 2015 (TAS-2) with crude prevalences of Ag-positive of 0.2% (95% confidence interval (CI) 0.0 to 0.8%) and 0.1% (95% CI 0.0 to 0.7%), respectively (8, 9). Despite this achievement, the territory failed TAS-3 in 2016 with an adjusted Ag prevalence of 0.7% (95% CI 0.3 to 1.8%), higher than the threshold and the recommended upper confidence limit of 1% (10). The findings in TAS-3 suggested potential resurgence of LF in the territory and were confirmed by a community-based survey conducted in the same year with an Ag prevalence of 6.2% (95%CI 4.5 to 8.6%) in individuals aged ≥8 years (10).

Evidence from other studies conducted in the territory in 2010 and 2014 also suggested ongoing LF transmission and the potential persistence of residual hotspots (11-14).

WHO recommends conducting follow-up surveys of nearby households of Ag-positive children identified through TAS to complement post-MDA surveillance (15). However, the recommendations are vague and lack a clear threshold for triggering a programmatic response (16). As LF prevalence decreases, the ability of the diagnostic methods, particularly in the TAS, to detect areas with residual transmission is also limited (7). This limitation is of particular importance in areas where the geographical distribution of LF has been demonstrated to be highly heterogeneous (17). In American Samoa, a recent study confirmed clustering of the infection in areas that were previously suspected as hotspots in 2010 and 2014 (Fagali’i village in the far north-west of Tutuila island, and also in the Ili’Ili-Vaitogi-Futiga area that is located on the south coast) and identified other potential areas where there is still potential residual infection (11, 18). Therefore, strategies for identifying foci of infection in low-prevalence settings are crucial in the context of the LF elimination efforts.
both from the perspective of targeting communities for MDA and also for understanding the
future of the post-MDA surveillance needs.

*W. bancrofti, B. malayi, and B. timori* require two hosts to complete its life cycle, the human
and the mosquito hosts. Therefore, sociodemographic, economic and environmental factors
that act at different spatial scales have the potential to influence the transmission pathways of
the parasites (19). The clustered distribution of LF has been associated with landscape
characteristics and climatic factors in several LF-endemic areas including the Pacific Islands
(20, 21). Bayesian model-based geostatistics combining socio-demographic and
environmental data with infection prevalence data proven to be able to predict disease
distribution in areas with scarce information (22-24). Hence, understanding how
environmental and sociodemographic factors interact to determine parasite transmission is
essential for the design and implementation of effective elimination strategies against LF.

The aim of this study was to identify areas where there is potential residual transmission of
*W. bancrofti* in American Samoa and produce LF predictive prevalence maps that can be
used to help guide and target future LF elimination strategies. A Bayesian model-based
diagnostics approach was used to: (i) assess and quantify the associations between LF
infection markers and sociodemographic and environmental factors at the household level
and (ii) develop spatial prediction of prevalence estimates of LF in American Samoa using
different infection markers – Ag and antibodies (Abs) against Wb123, Bm14, Bm33.

**Methods**

1. **Ethical considerations**
Ethical approval for the 2016 field survey was obtained from the American Samoa
Institutional Review Board and the Human Research Ethics Committee at the Australian
National University (protocol number 2016/482) and the University of Queensland
(2021/HE000896). After explaining the purpose and procedures of the survey, all adults and
parents/guardians of the minors (<18 years) who agreed to participate were asked to sign an
informed written consent form. Full details of local collaborations and official permissions to
visit villages have been previously described (10).

2. Study area

American Samoa is a United States territory in the South-central Pacific located
approximately between latitudes 11° North and 15° South and longitudes 168° East and 172°
West (Fig 1). The total land area of the territory is 200 km$^2$ and comprises five inhabited
volcanic islands Tutuila, Aunu’u, Ofu, Olosega and Ta’ū, and two remote coral atolls
(Swains Island and Rose Atoll). In 2010, the population of American Samoa was 55,519, the
majority of whom (95%) lived in Tutuila, the largest island (198.9 km$^2$), where the capital
Pago Pago is located.

American Samoa lies in the tropical savanna climate zone characterized by alternate wet
(October to May) and dry (June to September) seasons (25). Temperatures vary slightly
between the hottest period (December to April), when the average is approximately 31°C,
and the coolest period (June to August), when the average is 29 °C. The annual average
rainfall ranges from 3000 to 6000 mm, with 70% occurring during the hot and wet season.
The average elevation is 482 meters (m) with the highest point being Lata Mountain on the
island of Ta’ū (970 m).
Fig 1. Map of American Samoa and the distribution of the built-up areas in the main island of Tutuila

3. Data from community survey of lymphatic filariasis in 2016

Data on LF infection markers, Ag and Wb123, Bm14 and Bm33 Abs, were obtained from a two-stage equal probability cluster survey conducted in American Samoa in 2016. Full details about survey design and sampling methods have been previously reported (10). Briefly, 30 primary sampling units (PSUs) were randomly selected from a total of 70 villages/village
segments/village groups, that were defined based on a population size of less than 2000. Two
villages that were previously identified and confirmed as LF hotspots in 2010 and 2014,
respectively, were also added to the survey as PSUs (11). Within each PSU, a population
proportionate sampling method was implemented to randomly select households from a geo-
referenced list of buildings obtained from the American Samoa Department of Commerce
(26). In total, the survey included 32 PSUs (across 30 villages) and 754 households. A
household member was defined as an individual who considered the selected house as their
principal place of residence or who slept in that house the previous night. All consenting
household members aged \( \geq 8 \) years were surveyed and blood samples were tested for
circulating filarial Ag using the Alere TM Filariasis Test Strip (FTS) (Abbott, Scarborough,
ME) (27) and for Wb123, Bm14 and Bm33 Abs using multiplex bead assays (MBA) (28).

Standardised electronic questionnaires were administered by bilingual field research
assistants (in Samoan or English based on each participant’s preference). The demographics
data collected included sex, age and work location. Work location was categorised as indoor,
outdoor, tuna cannery (largest private employer in American Samoa), and other (including
mixed indoor/outdoor, unemployed, retired or unknown).

4. Geospatial data sources

We downloaded and assembled spatial and environmental data that have been found to be
associated with the geographical distribution of LF in other endemic regions (23, 29-32). The
boundary administrative maps and the covariate data consider for the analyses were derived
from the following datasets:
i) Village boundaries and buildings. A map of village boundaries and the distribution of all buildings in the territory were downloaded from the Fagatele Bay National Marine Sanctuary GIS data archive website (33).

ii) Coastline and streams. The American Samoa coastline and network of streams covering the entire territory were extracted in a shapefile format from the Fagatele Bay National Marine Sanctuary GIS data archive website (33).

iii) Population density. Data on population density for 2010/2011 were downloaded from the Pacific Data Hub website (34). A grid (i.e. raster surface) was available for American Samoa at the resolution of 100 m.

iv) Elevation. Data were obtained in a GeoTIFF format at the spatial resolution of 10 m from the United States Geological Survey (USGS) 10-m Digital Elevation Model (DEM): American Samoa: Tutuila (35).

v) Rainfall. Average monthly rainfall for 2016 were downloaded from the Pacific Environment Data Portal (36) in a raster format at the spatial resolution of 1 km. There was limited availability of spatial monthly rainfall datasets for the years prior to the survey. Therefore, the monthly rainfall layers from 2016 were used based on the assessment of the representativeness of the ten-year period prior the survey (Supplementary material).

vi) Land surface temperature. Satellite sensor data on land surface temperatures from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite were obtained from the USGS Earth Explorer website (37). These data were downloaded at 1 km resolution for every eight days from January 1 to December 31 2016.

vii) Land use/land cover map. Data were derived at 10m resolution from the Sentinel-2 Global Land Use/Land Cover (LULC) Timeseries produced by Impact Observatory, Microsoft, and the Environmental Systems Research Institute (Esri) (38).
5. Covariate data download and processing

The geo-referenced data sets that included the locations of the surveyed households, the covariates and the boundary map of American Samoa were imported into ArcGIS version 10.7.1 (39) to extract data (measured on a continuous scale) for the territory. The geographical distributions of the covariates are shown in Fig 2.

- Elevation estimates for the territory were extracted in meters (m) above sea level.
- A layer of the distance between each household location and the nearest coastline was developed (in m) using the Euclidean Distance Tool.
- The Euclidean Distance Tool was also used to estimate the distance (in m) between each household location and the nearest permanent surface stream.
- The monthly rainfall (mm) datasets were used to estimate the annual average rainfall and rainfall of the driest (August) and wettest (December) months in 2016.
- Annual average temperature and temperature of the hottest (December) and coolest (July) months in 2016 were estimated from the fortnightly temperature layers.
- The global LULC cover map with 11 LULC classes was used to generate four separate rasters for the LULC categories that cover the territory of American Samoa: crops, rangelands, trees, and built/urban area (Table 1).

Table 1. Land cover class definitions

<table>
<thead>
<tr>
<th>Land Use/Land cover Class</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crops</td>
<td>Human planted/plotted cereals, grasses, and crops not at tree height. Examples: corn, wheat, soy, fallow plots of structured land.</td>
</tr>
<tr>
<td>Category</td>
<td>Description</td>
</tr>
<tr>
<td>------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Rangeland</td>
<td>Open areas covered in homogenous grasses with little to no taller vegetation; wild cereals and grasses with no obvious human plotting (i.e., not a plotted field). Examples: natural meadows and fields with sparse to no tree cover, open savanna with few to no trees, parks/golf courses/lawns, pastures. Mix of small clusters of plants or single plants dispersed on a landscape that shows exposed soil or rock; scrub-filled clearings within dense forests that are clearly not taller than trees; examples: moderate to sparse cover of bushes, shrubs and tufts of grass, savannas with very sparse grasses, trees or other plants.</td>
</tr>
<tr>
<td>Trees</td>
<td>Any significant clustering of tall (~15 feet or higher) dense vegetation, typically with a closed or dense canopy. Examples: wooded vegetation, clusters of dense tall vegetation within savannas, plantations, swamp or mangroves (dense/tall vegetation with ephemeral water or canopy too thick to detect water underneath).</td>
</tr>
<tr>
<td>Built/Urban</td>
<td>Human made structures; major road and rail networks; large homogenous impervious surfaces including parking structures, office buildings and residential housing. Examples: houses, dense villages / towns / cities, paved roads, asphalt.</td>
</tr>
</tbody>
</table>
Fig 2. The geographical distributions of the covariates in American Samoa (a)
Population density (people/m²), (b) Annual average rainfall (mm), (c) Land cover, (d) streams and coastline, (e) temperature (K x 0.02) and (f) Elevation (m).

6. Buffer zones
The GPS locations of the surveyed households were used to delineate a buffer zone of 20 m around the households in ArcGIS (39). The buffer size was selected to represent an approximate distance within which the participants would spend extensive periods of time, and therefore have greatest exposure to the environmental conditions within the buffers (40). For each surveyed location, the data extracted within the buffer zone included the spatial mean values of population density, distance to coastline and streams, elevation, annual average rainfall, rainfall in the wettest (December) and driest (August) months in 2016, annual average temperature, and temperature of hottest (January) and coolest (July) months in the same year. Each of the four land cover classes covering the American Samoa territory were summarised as percentages of area within the 20 m buffer.

7. Descriptive analyses
For each infection marker and the covariates, summary statistics were calculated in R software R-4.0.3 (41). Crude prevalence of Ag, and Wb123, Bm14 and Bm33 Abs were estimated and mapped at the village level, and binomial exact methods were applied to estimate 95% confidence intervals (95% CI). Of note, in all subsequent analyses data were examined at the individual level and the respective household locations.

8. Variable selection

Collinearity between covariates was assessed using Spearman’s correlation. Non-spatial univariate logistic regression models were developed using R software R-4.0.3 (41) to examine the association of each LF infection marker (outcome variables) with the sociodemographic and environmental factors (covariates). For the strongly correlated covariates (Spearman’s correlation coefficient $\rho > 0.9$), the ones with the highest value of Akaike Information Criterion (AIC) in the univariate regression models was excluded. For each infection marker, multivariate logistic regression models were developed incorporating the remaining covariates. From these models, covariates were sequentially removed to assess AIC and p-values. The models with the lowest AIC were selected for further analyses and covariables with $p < 0.05$ were retained.

9. Multivariable non-spatial and spatial regression models

Bayesian geostatistical multivariate regression models were fitted using the OpenBUGS software version 3.2.3 rev 1012 (42). For each infection marker, separate logistic regression models were developed based on the binary outcome of the laboratory results. First, Bayesian geostatistical models were developed with the sociodemographic and environmental covariates as fixed effects but without considering the spatial dependence of the data. Then, Bayesian geostatistical models for each infection marker were fitted using Markov chain
Monte Carlo (MCMC) methods. The MCMC approach was selected for the geostatistical analyses because it allows the model to incorporate spatial dependence in both the infection and covariate data, and also enables full representation of uncertainty in model outputs (43).

The deviance information criterion (DIC) statistic was calculated to assess if the inclusion of spatial dependence in the data improved the fit of the models. Low DIC values indicate a better fit. Covariates in the models were considered statistically significant if the 95% credible intervals (95% CrI) of the estimated odds ratios (OR) excluded 1.

The mathematical notation of the spatial model is provided below, and contains all of the components of the non-spatial model. Assuming a Bernoulli-distributed dependent variable, \( Y_{ij} \), corresponding to the results of the infection markers (0=negative, 1=positive) of the \( i \)th participant \((I = 1 \ldots 2,671)\) the \( j \)th location \((j=1 \ldots 736)\), the model structures were as follows:

\[
Y_{ij} \sim \text{Bern}(p_{ij})
\]

\[
\text{logit}(p_{ij}) = \alpha + \gamma \times \text{age}_i + \delta \times \text{female}_i + \epsilon \times \text{outdoor}_i + \eta \times \text{mixed}_i + \rho \times \text{starkist}_i + \nu \times \text{student}_i + \theta \times \text{other}_i + \sigma \times \text{unemployed}_i + \sum_{z=1}^{z} \beta_z \times \lambda_{sj} + s_j
\]

where \( \alpha \) is the intercept, \( \gamma \) and \( \delta \) are coefficients for age and females, and \( \epsilon, \eta, \rho, \nu, \theta \) and \( \sigma \) are coefficient for the occupation categories. \( \beta \) is a matrix of \( z \) coefficients, \( \lambda \) is a matrix of \( z \) environmental variables and population density, and \( s_j \) a geostatistical random effect. The correlation structure of the geostatistical random effect was assumed to be an exponential function of the distance between points:

\[
f(d_{kj}; \phi) = \exp[-\phi d_{kj}]
\]
where \(d_{kl}\) are the distances between pairs of points \(k\) and \(l\), and \(\phi\) is the rate of decline of spatial correlation per unit of distance. A normal distribution was used for the priors for the intercept and the coefficients (mean = 0 and precision, the inverse of variance, = \(1 \times 10^{-3}\)), whereas a uniform distribution was specified for \(\phi\) (with upper and lower bounds \(s= 0.03\) and 100; the lower bound set to ensure spatial correlation at the maximum separating distance between survey locations was <0.5). A non-informative gamma distribution was used to specify the priors for the precision (shape and scale parameters = 0.001, 0.001).

A burn-in of 1,000 iterations were run first and discarded. Sets of 10,000 iterations were then run and examined for convergence. Convergence was assessed by visual inspection of history and density plots and by examining autocorrelation of the model parameters. In each model, convergence was achieved for all variables at approximately 30,000 iterations. The last 10,000 values from the posterior distributions of each model parameter were recorded. The rate of decay of spatial correlation between locations (\(\phi\)) with distance and the variance of the spatial structured random effect (\(\sigma^2\)) were also stored.

10. Predicted prevalence of lymphatic filariasis

To predict LF prevalence at unsampled locations, a regular 150 m × 150 m grid was overlaid on a map of American Samoa to extract the average environmental data for each grid cell. The predicted probabilities at the unsampled locations were estimated using the \textit{spatial.unipred} function in OpenBUGS. The function applies the model equation at each unsampled location using the covariates values extracted for them and the distance between those locations and the surveyed locations. Bayesian kriging was applied in ArcGIS to generate smooth risk maps of the posterior distributions of predicted prevalence of each LF infection marker.
Results

1. Sample description and sample site locations

The final dataset used for analyses included 754 households in 736 unique locations (some households shared the same building structure) from 32 PSUs in 30 villages. The total number of participants was 2,671 with a mean age of 33.5 years (range 8–93), and 54.7% (n = 1462) were female. Figs 3a and 3b show the locations of sampled villages and the geographical distribution of the survey locations, respectively. The highest overall crude prevalence was observed for Bm33 Ab (45.6%, 95% CI 43.7–47.5%), followed by Wb123 Ab (25.6%, 95% CI 24.0–27.3%), Bm14 Ab (13.1%, 95% CI 11.8–14.4%) and Ag (5.1, 95% CI 4.2–5.9%). At the village level, Fagali’i (n=81) and Fagamalo (n=13), located in the far north-west of Tutuila Island, consistently showed high overall crude prevalence of all infection markers. A detailed description of the Ag and Ab results has been presented elsewhere (7, 10, 18). Figs 3c, 3d, 3e and 3f display the observed geographical distributions of the prevalence of Ag and Wb123, Bm14 and Bm33 Abs, respectively, by village. The maps confirm that villages with high prevalence of Bm33 Ab were more widespread across the territory, while the distribution of villages with high prevalence of Ag, Wb123 and Bm14 Ab was more confined to the north-west of the territory.
Fig 3. a) Distribution of the sampled villages on the island of Tutuila, American Samoa 2016, b) surveyed household locations, and village-level seroprevalence of c) Ag, d) Wb123 Ab, e) Bm14 Ab and f) Bm33 Ab.

2. Variable selection and univariate regression analyses
The descriptive statistics and maps of the covariates considered for the analyses are presented in Table 2 and Supplementary Fig 1, respectively. Because temperature data were not available for large areas of American Samoa, this covariate was excluded from analyses. We identified four pairs of variables with Spearman’s rank >0.9 that were assessed with the univariate regression models. After comparing the AIC of the stepwise multivariate logistic regression models, the selected variables for the Bayesian non-spatial and spatial analyses included: sex, age, work location, population density, elevation, rainfall in the wettest month (December), distance to streams, cropland, tree coverage and urban areas.

Table 2. Descriptive statistics of environmental covariates within 20 m buffers of surveyed household locations in American Samoa in 2016

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>Standard deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population density (people/m²)</td>
<td>2.47</td>
<td>2.26</td>
<td>2.06</td>
<td>0.01</td>
<td>12.53</td>
</tr>
<tr>
<td>Elevation (m)</td>
<td>77.72</td>
<td>43.96</td>
<td>92.99</td>
<td>0</td>
<td>479.86</td>
</tr>
<tr>
<td>Distance to the coastline (m)</td>
<td>1081.97</td>
<td>644.25</td>
<td>1117.78</td>
<td>3.06</td>
<td>4371.57</td>
</tr>
<tr>
<td>Distance to streams (m)</td>
<td>358.25</td>
<td>151.12</td>
<td>434.39</td>
<td>0.32</td>
<td>1825.00</td>
</tr>
<tr>
<td>Average annual rainfall (mm)</td>
<td>3469.69</td>
<td>3374.57</td>
<td>698.28</td>
<td>2095.10</td>
<td>4630.68</td>
</tr>
<tr>
<td>Rainfall in the driest month - August (mm)</td>
<td>232.46</td>
<td>232.97</td>
<td>51.99</td>
<td>125.27</td>
<td>324.61</td>
</tr>
<tr>
<td>Rainfall in the wettest month - December (mm)</td>
<td>381.14</td>
<td>369.73</td>
<td>60.28</td>
<td>255.96</td>
<td>528.90</td>
</tr>
<tr>
<td>Land cover</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cropland (%)</td>
<td>0.01</td>
<td>0</td>
<td>0.03</td>
<td>0</td>
<td>0.91</td>
</tr>
<tr>
<td>Rangeland (%)</td>
<td>0.85</td>
<td>0</td>
<td>6.83</td>
<td>0</td>
<td>0.99</td>
</tr>
<tr>
<td>Tree coverage (%)</td>
<td>5.14</td>
<td>0</td>
<td>19.93</td>
<td>0</td>
<td>99.00</td>
</tr>
<tr>
<td>Urban (%)</td>
<td>76.36</td>
<td>100</td>
<td>34.38</td>
<td>0</td>
<td>100.00</td>
</tr>
</tbody>
</table>

3. Bayesian non-spatial and spatial models

For all infection markers, the best-fit model included the spatial random effect. Tables 3 and 4 show the odds ratios (ORs) and 95% CrI from the Bayesian non-spatial and spatial models for Ag and Wb123, Bm14 and Bm33 Abs, respectively.
3.1 Multivariate non-spatial and spatial models for Ag

The DICs of the models of Ag with and without accounting for spatial correlation were 830.3 and 1122.3, respectively. In the spatial model, females had a 29.6% (95% CrI: 16.0–41.1%) lower risk of being Ag-positive than males. There was a 1.4% (95% CrI: 0.02–2.7%) decrease in the odds of Ag positivity for every year of age. Tree coverage was also positively associated with Ag-positivity, with an estimated increase of 0.3% (95% CrI: 0.06–0.6%) in the odds of Ag-positivity for each 1% increase in the extent of tree coverage in the 20 m buffers.

After accounting for the effect of the statistically significant variables, the variance of the spatially structured random effect was 1.64 (0.55 to 4.92). The values of the decay parameter for spatial correlation ($\phi$), was 86.03. This means that, after accounting for the effect of covariates, the radius of the clusters was approximately 3.9 km. ($\phi$ is measured in decimal degrees, therefore, the cluster size is calculated dividing 3 by $\phi$; at the equator, one decimal degree is approximately 111 km).

3.2 Multivariate geostatistical model for Wb123, Bm14 and Bm33 Abs

The DICs of the models of positivity for Wb123 Ab with and without accounting for spatial correlation were 2707.9 and 2861.1, respectively. In the spatial model, females had a 52.6% (95% CrI: 41.4–61.4%) lower risk of Wb123 Ab positivity than males. There was also an estimated increase of 2.4% (95% CrI: 1.9%–2.9%) in Wb123 Ab-positivity for every year of age (Table 4). Also, there was an increase in prevalence of being positive for Wb123 Ab of 145.2% (95% CrI: 67.1–268.4%) and 46.9% (95% CrI: 16.3–83.3%) for tuna cannery workers and those who work in other locations (excluding outdoors and tuna cannery), respectively, compared to indoor workers. Additionally, there was a significant increase of
0.3% (95% CrI: 0.02–0.6%) in the prevalence of Wb123 Ab-positivity for each 1% increase in the coverage of trees in the 20 m buffers.

The spatial Bm14 Ab model had a DIC of 1896.6, while the model without the spatial component had a DIC of 1898.6. In the spatial model, there was a decrease in the prevalence of Bm14 Ab of 53.8% (95% CrI: 40.3–64.5%) for females compared to males. Age was also as significant covariate with an increase in the prevalence of Bm14 Ab of 2.8% (95% CrI: 2.2–3.5%) per every year of age. The prevalence of positive for Bm14 Ab was higher for those who worked outdoor and tuna cannery locations compare to those working indoors. The increase in the prevalence was 221.2% (95% CrI: 55.2–558.6%) and 96.3% (95% CrI: 24.3–225.4%), respectively. Tree coverage had a significant positive association with positivity for Bm14 Ab, with an estimated increase of 0.7% (95% CrI: 0.4–0.9%) in Bm14 positivity for each 1% increase in tree coverage in the 20 m buffer area. Population density had a significant negative association with Bm14 Ab prevalence, with a decrease of 12.0% (95% CrI: 2.7–21.0%) for every person/m².

The spatial model for Bm33 Ab also had a lower DIC, 3492.4, compared with the nonspatial model, 3505.6. Similar to all the other infection markers, the decrease in prevalence of Bm33 Ab was 28% (95% CrI: 14.3–39.1%) in females compared to males, and the increase per every year of age was 2.4% (95% CrI: 1.9–2.9%). Also, workers in outdoor and tuna cannery locations had an increase of 119.1% (95% CrI: 1.3–406.8%) and 74.8% (95% CrI: 18.1–165.5%) compared to workers in indoor areas. The prevalence of Bm33 positivity was found to increase by 0.3% (95% CrI: 0.07–0.6%) with a 1% increase in the extent of tree coverage in the 20 m buffers.
In the model of Wb123 Ab the variance of the spatially structured random effect was 1.1 (0.5 to 2.1) and in the models of Bm14 and Bm33 these parameters were 0.002 (0.001 to 0.004) and 0.9 (0.5 to 1.4), respectively, meaning that the residual spatial variation was higher for the model of Wb123 Ab. The value of the decay parameter for spatial correlation ($\phi$) was 83.4 for Wb123 Ab, 73.7 for Bm14 Ab, and 78.3 for Bm33 Ab. These estimates indicate that after accounting for the effect of covariates, the radii of the clusters were approximately 3.9, 4.5 and 4.2 km, respectively.
Table 3 Odds ratios (ORs) and 95% CrI from the Bayesian non-spatial models for Ag and Wb123, Bm14 and Bm33 antibodies in a community survey in American Samoa in 2016

<table>
<thead>
<tr>
<th>Model</th>
<th>Participants N (%)</th>
<th>Antigen</th>
<th>Wb123 antibody</th>
<th>Bm14 antibody</th>
<th>Bm33 antibody</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ORs, posterior mean (95% CrI)</td>
<td>ORs, posterior mean (95% CrI)</td>
<td>ORs, posterior mean (95% CrI)</td>
<td>ORs, posterior mean (95% CrI)</td>
</tr>
<tr>
<td>Total samples</td>
<td>2671</td>
<td>2671</td>
<td>2671</td>
<td>2671</td>
<td>2671</td>
</tr>
<tr>
<td>Total positives, N (%)</td>
<td>-</td>
<td>135 (5.05)</td>
<td>684 (25.60)</td>
<td>350 (13.10)</td>
<td>1219 (45.63)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1209 (45.26)</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td></td>
<td>1462 (54.74)</td>
<td>0.58 (0.49 to 0.68)</td>
<td>0.49 (0.41 to 0.59)</td>
<td>0.46 (0.36 to 0.59)</td>
<td>0.72 (0.61 to 0.85)</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td>1.02 (0.96 to 0.98)</td>
<td>1.03 (1.02 to 1.03)</td>
<td>1.02 (1.02 to 1.04)</td>
<td>1.02 (1.02 to 1.03)</td>
</tr>
<tr>
<td>Age (per year)</td>
<td>-</td>
<td>0.97 (0.96 to 0.98)</td>
<td>1.02 (1.02 to 1.03)</td>
<td>1.03 (1.02 to 1.04)</td>
<td>1.02 (1.02 to 1.03)</td>
</tr>
<tr>
<td>Work location</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indoor</td>
<td>727 (27.22)</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>Outdoor</td>
<td>40 (1.50)</td>
<td>1.01 (0.83 to 1.22)</td>
<td>2.78 (1.41 to 5.47)</td>
<td>3.24 (1.56 to 6.68)</td>
<td>2.87 (1.39 to 6.10)</td>
</tr>
<tr>
<td>Tuna cannery</td>
<td>131 (4.90)</td>
<td>0.99 (0.82 to 0.19)</td>
<td>2.21 (1.46 to 3.32)</td>
<td>1.94 (1.18 to 3.15)</td>
<td>1.69 (1.14 to 2.51)</td>
</tr>
<tr>
<td>Others</td>
<td>1773 (66.40)</td>
<td>0.51 (0.44 to 0.60)</td>
<td>1.50 (1.18 to 1.91)</td>
<td>1.27 (0.94 to 1.73)</td>
<td>1.10 (0.89 to 1.35)</td>
</tr>
<tr>
<td>Population density (people/m²)</td>
<td>-</td>
<td>0.89 (0.81 to 0.98)</td>
<td>0.95 (0.88 to 1.02)</td>
<td>0.88 (0.79 to 0.97)</td>
<td>0.95 (0.89 to 1.02)</td>
</tr>
<tr>
<td>Elevation (m)</td>
<td>-</td>
<td>1.00 (0.99 to 1.00)</td>
<td>1.00 (0.99 to 1.00)</td>
<td>1.00 (0.99 to 1.00)</td>
<td>1.00 (0.99 to 1.00)</td>
</tr>
<tr>
<td>Distance to streams (m)</td>
<td>-</td>
<td>1.00 (1.00 to 1.00)</td>
<td>0.99 (0.99 to 0.99)</td>
<td>0.99 (0.99 to 1.00)</td>
<td>0.99 (0.99 to 1.00)</td>
</tr>
<tr>
<td>Rainfall in the wettest month - December (mm)</td>
<td>-</td>
<td>1.00 (0.99 to 1.00)</td>
<td>1.00 (0.99 to 1.00)</td>
<td>1.00 (0.99 to 1.00)</td>
<td>1.00 (0.99 to 1.00)</td>
</tr>
<tr>
<td>Land Cover</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cropland (%)</td>
<td>-</td>
<td>1.00</td>
<td>1.04</td>
<td>0.96</td>
<td>1.03</td>
</tr>
<tr>
<td>Model</td>
<td>Participants N (%)</td>
<td>Antigen positive positive</td>
<td>Wb123 antibody positive</td>
<td>Bm14 antibody positive</td>
<td>Bm33 antibody positive</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------------</td>
<td>---------------------------</td>
<td>-------------------------</td>
<td>------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td></td>
<td>ORs, posterior mean (95% CrI)</td>
<td>ORs, posterior mean (95% CrI)</td>
<td>ORs, posterior mean (95% CrI)</td>
<td>ORs, posterior mean (95% CrI)</td>
<td></td>
</tr>
<tr>
<td>Total samples</td>
<td>2671</td>
<td>2671</td>
<td>2671</td>
<td>2671</td>
<td>2671</td>
</tr>
<tr>
<td>Total positives, N (%)</td>
<td>135 (5.05)</td>
<td>684 (25.60)</td>
<td>350 (13.10)</td>
<td>1219 (45.63)</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>Ref 0.70 (0.59 to 0.84)</td>
<td>Ref 0.47 (0.39 to 0.59)</td>
<td>Ref 0.46 (0.35 to 0.60)</td>
<td>Ref 0.72 (0.61 to 0.86)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1209 (45.26)</td>
<td>1462 (54.74)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (per year)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.01 (1.00 to 1.03)</td>
<td>1.02 (1.02 to 1.03)</td>
<td>1.03 (1.02 to 1.04)</td>
<td>1.02 (1.02 to 1.03)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1.05 (0.83 to 1.23)</td>
<td>1.53 (0.70 to 3.18)</td>
<td>3.21 (1.55 to 6.59)</td>
<td>2.19 (1.01 to 5.07)</td>
<td></td>
</tr>
<tr>
<td>Work location</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indoor</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td></td>
</tr>
<tr>
<td>Outdoor</td>
<td>40 (1.50)</td>
<td>1.01 (0.83 to 1.23)</td>
<td>1.53 (0.70 to 3.18)</td>
<td>3.21 (1.55 to 6.59)</td>
<td>2.19 (1.01 to 5.07)</td>
</tr>
<tr>
<td>Tuna cannery</td>
<td>131 (4.90)</td>
<td>1.05 (0.87 to 1.27)</td>
<td>2.45 (1.67 to 3.68)</td>
<td>1.96 (1.24 to 3.25)</td>
<td>1.75 (1.18 to 2.66)</td>
</tr>
<tr>
<td>Others</td>
<td>1773 (66.40)</td>
<td>0.86 (0.70 to 1.23)</td>
<td>1.47 (1.24 to 3.25)</td>
<td>1.28 (1.18 to 2.66)</td>
<td>1.08 (0.99 to 1.11)</td>
</tr>
</tbody>
</table>

Footnotes: ORs, Odds ratios; 95% CrI, 95% credible interval; DIC, deviance information criterion. Statistically significant ORs are highlighted in blue.
<table>
<thead>
<tr>
<th></th>
<th>(0.70 to 1.05)</th>
<th>(1.16 to 1.83)</th>
<th>(0.96 to 1.69)</th>
<th>(0.89 to 1.32)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population density (people/m²)</strong></td>
<td>0.96 (0.85 to 1.08)</td>
<td>0.99 (0.90 to 1.08)</td>
<td>0.88 (0.79 to 0.97)</td>
<td>0.95 (0.88 to 1.03)</td>
</tr>
<tr>
<td><strong>Elevation (m)</strong></td>
<td>1.00 (0.99 to 1.00)</td>
<td>1.00 (0.99 to 1.00)</td>
<td>1.00 (0.99 to 1.00)</td>
<td>1.00 (0.99 to 1.00)</td>
</tr>
<tr>
<td><strong>Distance to streams (m)</strong></td>
<td>1.00 (0.99 to 1.00)</td>
<td>1.00 (0.99 to 1.00)</td>
<td>0.99 (0.99 to 1.00)</td>
<td>0.99 (0.99 to 1.00)</td>
</tr>
<tr>
<td><strong>Rainfall in the wettest month - December (mm)</strong></td>
<td>1.00 (0.99 to 1.01)</td>
<td>1.00 (0.99 to 1.01)</td>
<td>1.00 (0.99 to 1.00)</td>
<td>1.00 (0.99 to 1.00)</td>
</tr>
<tr>
<td><strong>Land Cover</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cropland (%)</td>
<td>0.99 (0.93 to 1.04)</td>
<td>1.04 (0.97 to 1.10)</td>
<td>0.96 (0.80 to 1.05)</td>
<td>1.04 (0.98 to 1.11)</td>
</tr>
<tr>
<td>Trees (%)</td>
<td>1.00 (0.99 to 1.00)</td>
<td>1.00 (1.00 to 1.01)</td>
<td>1.01 (1.00 to 1.01)</td>
<td>1.01 (1.00 to 1.01)</td>
</tr>
<tr>
<td>Built/Urban (%)</td>
<td>1.00 (0.99 to 1.00)</td>
<td>0.99 (0.99 to 1.00)</td>
<td>0.99 (0.99 to 1.00)</td>
<td>0.99 (0.99 to 1.00)</td>
</tr>
<tr>
<td>Heterogeneity structured</td>
<td>1.64 (0.55 to 4.92)</td>
<td>1.12 (0.54 to 2.12)</td>
<td>0.002 (0.001 to 0.005)</td>
<td>0.87 (0.54 to 1.44)</td>
</tr>
<tr>
<td>φ (Decay of spatial correlation)</td>
<td>86.03 (54.95 to 97.37)</td>
<td>83.4 (51.32 to 99.36)</td>
<td>73.74 (42.51 to 98.37)</td>
<td>78.30 (47.77 to 98.59)</td>
</tr>
<tr>
<td>Deviance Information Criterion</td>
<td>830.28</td>
<td>2707.95</td>
<td>1896.61</td>
<td>3492.37</td>
</tr>
</tbody>
</table>

Footnotes: ORs, Odds ratios; 95% CrI, 95% credible interval; DIC, deviance information criterion. Statistically significant ORs are highlighted in blue.
3.3 Spatial predictions

Maps of the mean and standard deviation (SD) of the posterior distributions of predicted prevalence of each of the LF infection markers are shown in Fig 4. The highest predicted prevalence of all infection markers (≥0.61%) was mainly confined in the north-west part, an area that corresponds largely to the coastal villages of Fagali'i and Fagamalo. There were also predicted residual foci of high prevalence of all Abs (higher for Bm33 and Wb123 Abs than Bm14) in the southwest part of Tutuila, in areas that belong to Vaitogi and Futiga villages (0.21% and 0.49%), and high predicted prevalence estimates of Wb123 and Bm33 Ab in the western part of Tafuna village. High prevalence of Bm33 Ab covered larger areas compared to the other infection markers (≥0.21%), with higher prevalence estimates in confined areas in the north-west (≥ 0.61%), south-west (≥0.51%), the north-east (≥0.51%) and the central part around the Pago Pago area (≥0.41%). The maps of the posterior SDs demonstrate that the level of uncertainty was higher in inhabited areas in the north that were predominantly covered by trees (Figs 2 and 4).
Fig 4. Spatial distribution of predicted prevalence and standard deviations of Ag (a and b), Wb123 Ab (c and d), Bm14 Ab (e and f), and Bm33 Ab (g and h) in American Samoa 2016. Note that the scale for SD for Bm14 Ab is different to the others maps.
Discussion

In this study, we conducted a Bayesian geostatistical analysis of LF infection markers at the household level and produced predictive prevalence maps for American Samoa in 2016. In addition, this study examined potential sociodemographic and environmental factors that may influence the geographical distribution of LF in the territory. To our knowledge, this is the first time that the distributions of LF infection markers have been examined at such high spatial resolution to predict prevalence. Our results suggest that there are still areas with high prevalence of LF infection markers (including Ag) in American Samoa, particularly in the north-west of the main island of Tutuila. Also, we found that there are sociodemographic and environmental factors that may underly the geographical distribution of LF and potentially contribute to persistent transmission. These predicted prevalence estimates of LF infection markers may help maximise the effectiveness of post-intervention surveillance by contributing to the identification of areas with highest probability of residual transmission (7).

The results showed that the predicted prevalence of Ag, Wb123, Bm14 and Bm33 Abs differed geographically across the territory. Areas around Fagali'i and Fagamalo villages in the north-west had the highest predicted prevalence of all infection markers. Also, in the south, high predicted prevalence, particularly for Bm33 Ab, were observed in localised areas in Vaitogi, Futiga and Tafuna villages. These findings concurred with the results of a previous research conducted in the territory that found significant spatial dependency for all infection markers, and confirmed the presence of LF clusters and hotspots in the north-west, south and central part of Tutuila (18). The high-risk area in the far north-west was also previously identified as potential hotspot of residual infection in cross-sectional surveys conducted in American Samoa in 2010, 2014 and 2016 (10, 11, 13). In this study, the cluster
sizes for all infection markers were larger compared to the previous findings (11, 18). This discrepancy in cluster size may be explained by the implementation of different spatial methods, and also by the incorporation of sociodemographic and environmental covariates into the geostatistical models (noting that the cluster size is in the residual component). These covariates may be associated with heterogeneous exposure to mosquito bites. In areas where the parasite is transmitted predominantly by night-biting mosquitoes, clustering of infection around household locations can be expected and has been demonstrated (29, 32, 44). The results of this study support recent evidence that the home environment may be also an important area for exposure in LF-endemic regions where *W. Bancrofti* is transmitted by the day-biting mosquito, *Ae. Polynesiensis*. (45, 46). The cluster size also suggests that transmission may be occurring not only around households, but also in surrounding areas where the household members are likely to frequent (such as bus stops, schools and workplaces). In Samoa, a multilevel hierarchical modelling found that the intraclass correlation coefficients for Ag-positive individuals was higher at households (0.46) compared to primary sampling units (0.18) and regions (0.01) (46). The timely identification of these small pockets of residual infection can be used to prioritise further interventions to reduce the risk of LF recrudescence or resurgence in the territory.

The predictive models developed for the different LF infection markers can help characterise the spatial patterns of serological responses to LF in American Samoa. In *W. bancrofti* endemic areas, WHO recommends the use of Ag testing to assess the impact of the MDA and determine when the elimination targets have been reached (47). However, there is increasing evidence that suggests that the use of Ag alone for post-MDA surveillance may not be sufficiently sensitive to detect residual infection (7, 48, 49). Therefore, antifilarial Ab testing is currently been examined as an alternative or complementary method of diagnosing LF in
post-MDA surveillance surveys (48-51). However, the dynamics of the Ab responses post-infection and post-treatment are still not well understood (52). In this study, the geographical distribution of the predicted prevalence of Wb123 and Bm14 Abs were more clustered compared to the widespread distribution of positive Bm33 Ab responses. This finding suggests that Bm33 Ab may not be the best indicator to identify areas of ongoing *W. bancrofti* transmission but may be used to provide information about levels of historical exposure and infection. Studies that have monitored the development of antifilarial immunity in LF endemic areas have shown that Bm33 Ab can be detected more than one year before the other Ab responses, and can decrease after MDA (50). Additional longitudinal studies are required to help monitor how the stage of the infection and magnitude of the immunological responses determine the spatial patterns of antifilarial Abs. Such information will have implications for the selection of the most suited LF diagnostic tools in low prevalence and post-MDA settings.

There were consistent associations between the infection markers and the sociodemographic variables included in the models. The observed differences among females and males and the positive association with age is most likely to be exposure-related. These findings support what has been observed previously in the territory and in most LF-endemic areas (11, 45, 53). Males spend more time working outdoors compared to females. However, it has also been suggested that immunological and hormonal gender differences may account for the lower infection rates in females (54). There was a consistent positive association between all Abs and tuna cannery workers. Also, a positive association between Bm14 and Bm33 Abs and individuals working in outdoor locations. In 2013, An average of 17.6% of the total employed population in the territory worked in tuna cannery which is the largest non-government employer in American Samoa (55). The nature and time of day when this work
takes place may increase the risk of the exposure to mosquitoes. Higher prevalence of Wb123 Ab was also previously observed in tuna cannery workers in the territory but no associations between Ab responses has been identified with other occupational groups (11).

The spatial models for all infection makers indicated that there was a positive association between the prevalence of LF and the extent of tree coverage in the 20 m buffers. This finding supports the hypothesis that the tree coverage may impact mosquito population dynamics and behaviours (56). Tree canopy may sustain *W. Bancrofti* life cycle in high temperature areas by facilitating the survival of mosquitos that move in response to food supply (14). Most of American Samoa is steep, with approximately half of the area covered by rainforest (14, 57). Trees primarily cover most areas in the northern part of the territory (Fig 2) where the highest prevalence of LF was observed. Rainfall has been shown to be associated with high prevalence of LF in several endemic countries where the infection is transmitted by different vectors (20, 23, 31, 58, 59). No associations between the prevalence of infection markers and rainfall were found in this study. This finding was unexpected and deserves further investigation. These findings raise the need for high-quality spatial environmental datasets that can be used in further studies to determine the association of LF and other potential environmental drivers.

The strengths of this study include the availability of data at the household level that allowed us to assess the geographical distribution of LF in American Samoa at a small spatial scale. In this way, it was possible to explore the home environment as an exposure area of importance. The study also developed geostatistical models for different infection markers that may be used as baseline information to characterise the spatial patterns of the antifilarial Ab responses in the long-term. Besides the predicted prevalence maps, the spatial models developed here also
provided outputs to determine the associated uncertainty of the prevalence estimates (60). The maps of the SD (uncertainty) highlight the areas where predictions were imprecise and that need to be explored in future studies.

The limitations of the study include the lack of high-quality spatial environmental datasets for the territory. As a result, it was not possible to include covariates such as temperature, that has consistently been associated positively with LF (20, 30, 31). Also, the rainfall data used in the study was only available in a spatial format for the year 2016. Based on the assessment presented in the supplementary files, data were found to be representative of the average rainfall estimates for the ten-year period prior the survey (most likely time period of potential exposure). Despite this limitation, we believe that our results provide valuable information about the potential sociodemographic and environmental factors that may be influencing the distribution of the infection in American Samoa.

In this study, the Bayesian geostatistical models incorporating sociodemographic and environmental covariates showed that the predicted prevalence of LF was not homogeneous in American Samoa. Small-scale spatial variation in LF prevalence was observed which indicates that there is scope for further spatial analyses to help inform spatially-targeted interventions in American Samoa. Areas of priority for further study include the north and south-western part of the territory. Also, longitudinal monitoring of the prevalence of Ag, Wb123, Bm14 and Bm33 Abs would be useful to better understand the dynamics and potential use of different LF infection markers to inform and support the ongoing post-MDA surveillance efforts.

**Funding**
This study was supported by the Coalition for Operational Research on Neglected Tropical Diseases (COR-NTD), which is funded at The Task Force for Global Health primarily by the Bill & Melinda Gates Foundation [OPP1053230], the United Kingdom Department for International Development, and by the United States Agency for International Development through its Neglected Tropical Diseases Program. CLL was supported by Australian National Health and Medical Research Council Fellowships (APP1193826).

**Author’s contributions**

AMC and CLL developed the study conception and design. Analyses were performed by AMCR and CLL. AMCR and CLL drafted the manuscript. All authors helped in the interpretation of results and critically reviewed the manuscript.

**Declaration of interests**

We declare no competing interests.

**Data sharing**

The data used in the present study are available from the corresponding author on reasonable request.
References


42. Group. MoOPM. OpenBUGS version 3.2.3 rev 1012. 2014.


Supporting information captions

S1 table. Average monthly rainfall (mm) in the Pago Pago area, in American Samoa from 2000-2020 (Data extracted from the National Weather

S1 Figure. Average monthly rainfall (mm) for the period 2000-2020 and average monthly rainfall (mm) in 2016 in American Samoa. The driest (August) and wettest (December) months in 2016 were representative of the average rainfall in the respective months in previous 20 years.

S2 Figure. Average annual rainfall (mm) for the period 2000-2020. The horizontal line indicates the average in the previous 10 years. Total rainfall in 2016 was representative of average rainfall in the previous 20 years.