SI Appendix

Incorporating environmental heterogeneity and observation effort to predict host distribution and viral spillover from a bat reservoir.

https://doi.org/10.1098/rspb.2023.1739

Authors: Rita Ribeiro[1], Jason Matthiopoulos[1*], Finn Lindgren[2], Carlos Tello[3,4], Carlos M. Zariquiey[3], William Valderrama[3,5], Tonie E. Rocke[6], Daniel G. Streicker[1,7*]

∗These authors share senior authorship

Affiliations:

[1]School of Biodiversity, One Health and Veterinary Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK

[2]School of Mathematics, University of Edinburgh, Edinburgh, UK

[3]ILLARIY (Asociación para el Desarrollo y Conservación de los Recursos Naturales), Lima, Perú

[4]Yunkawasi, Lima, Perú

[5]Facultad de Medicina Veterinaria y Zootecnia, Universidad Peruana Cayetano Heredia, Lima, Perú

[6]U.S. Geological Survey, National Wildlife Health Center, Madison, Wisconsin, USA

[7]Medical Research Council - University of Glasgow Centre for Virus Research, Glasgow, UK

Contents:

S1 Spatial distribution and summary information of the variables initially assessed in this study.

S2 Constrained refined Delaunay triangulation.

S3 Pre-analysis of the pairwise distance between vampire bat roosts.

S4 Blocks for ten-fold spatial cross-validation.

S5 Number of rabies outbreaks in livestock in the inter-Andean valleys of Apurimac, Ayacucho, and Cusco between 2003 and 2021.

S6 Quadratic function of temperature.

S7 Posterior distribution of roost abundance.

S8 Maps of the predicted posterior mean of roost density in the linear scale.

S1 Spatial distribution and summary information of the variables initially assessed in this study.

Table S1: Environmental and anthropogenic variables initially assessed for the study, units, resolution, and source.

Variables (units)		**Original spatial and temporal resolutions**	**Final spatial resolution if rescaled**	**Meaning and source**
Climatic	Mean temperature (°C)	30*30 arcseconds (or ~ 1 km^2). Average for the years 1970-2000.	100 m^2	Annual mean temperature (BIO1) [1]
	Temperature seasonality (°C)			Temperature Seasonality (Standard Deviation) (BIO4) [1]
	Minimum temperature (°C)			Min Temperature of Coldest Month (BIO6) [1]
	Precipitation (mm)			Annual precipitation (BIO12) [1]
	Precipitation seasonality (fraction)			Precipitation seasonality (coefficient of variation) (BIO15) [1]
Topographic	Elevation (meters)	3*3 arcseconds (or ~ 90 m^2). Last update November 2018.		SRTM digital elevation data version 4 [2]
	Binary elevation			Derived from elevation. Binary variable defining preferred (coded as 1) and non-preferred (coded as 0) habitats for *D. rotundus* considering the threshold of 3600 m of altitude [3].
	Slope (degrees)			Derived from elevation using terrain () R package raster
	Terrain ruggedness (meters)			Terrain ruggedness index (TRI). Derived from elevation using SpatialEco R library
Landcover	Crop cover (%)	~100 m^2, 2019		Percent vegetation cover for cropland land cover class. Copernicus, PROBAV_LC100_globalv3.01-2019 [4]
	Tree cover (%)			Percent vegetation cover for forest land cover class. Copernicus, PROBAV_LC100_globalv3.01-2019 [4]
	Ground biomass (tons/ha i.e., Mg/ha)	~100 m^2, 2010		Above ground biomass (AGB) The mass, expressed as oven-dry weight of the woody parts (stem, bark, branches, and twigs) of all living trees excluding stump and roots [5].
	Proportion mines (proportion)	1 km^2, 2016		A polygon layer of the mines present in the study area was extracted from https://geocatmin.ingemmet.gob.pe/geocatmin/ (Geological and mining cadastral information system from Peru), and used to create a raster with the proportion of mines per 1 km^2.
Host related	Cattle density (number)	5*5 arcminutes (or ~ 10 km^2), 2010	5 km^2	Distribution of cattle in the study area. From: Gridded Livestock of the World (GLW3). Use of the dasymetric weighting version (DA) [6].

	Rural settlements (number)	2 km^2, 2017		Number of rural population centres. As a proxy of presence of livestock, we created a layer with the number of rural population centres (i.e., count of areas such as towns, villages, and communities). Data on the geographic locations of the rural population centres in the study area, and of the census population on those rural centres were obtained from two departments of Peru Government (INEI and CEPLAN) (https://www.inei.gob.pe) and https://www.ceplan.gob.pe/informacion-sobre-zonas-y-departamentos-del-peru/). As there was no previous information on its resolution, we hypothesised that this layer could act within the foraging range of vampire bats, hence creating two layers, one at 2 km^2 and one at 5 km^2. We found the resolution of 2 km^2 to have a better fit to the data when compared with 5 km^2.
	Human footprint (%)	0.0083*0.0083 degrees (or ~1-km^2), 1995-2004		Global human footprint expressed as a percentage of the relative human influence in each terrestrial biome. https://sedac.ciesin.columbia.edu/data/set/wildareas-v2-human-footprint-geographic [7].
	Distance water (meters)	1 km^2 (no temporal information)		Distance to the nearest river. A polygon layer with the river network in the study area from https://mapcruzin.com was used to create a raster of 1 km^2 size, with the mean Euclidean distance to the nearest river.
Accessibility to SENASA offices	Travel time (minutes)	0.0083*0.0083 degrees (~1 km^2), 2019		Created from a global friction surface developed by Weiss et al. [8] (please see here further information of the friction layer), which calculates land-based travel speed at 1 km^2 resolution. The datasets used to produce the friction surface included roads, railways, rivers, lakes, oceans, topographic conditions, landcover types and national borders. The friction surface gives for every pixel an overall speed of travel, with the fastest travel mode intersecting the pixel being used to determine the speed of travel in that pixel (i.e., is a relative measure of how difficult it is to cross that grid cell). With this layer, we then used two functions of the R package gdistance. This package was first used to convert the friction surface to a transition matrix considering 8 neighbours (function transition) and then, using the GPS locations of the National Service of Agrarian Health (SENASA) offices, a layer of travel times to the nearest SENASA office (function accCost, which calculates the accumulated cost surface from any points). The accessibility to the SENASA offices is represented by the travel time in minutes (logarithm scale) from each grid cell to the nearest SENASA office. A possible limitation of this approach is that the 1 km^2 resolution of this friction layer might mask variation in terrain within a single grid cell, such that a finer scale would be more accurate to measure travel times.

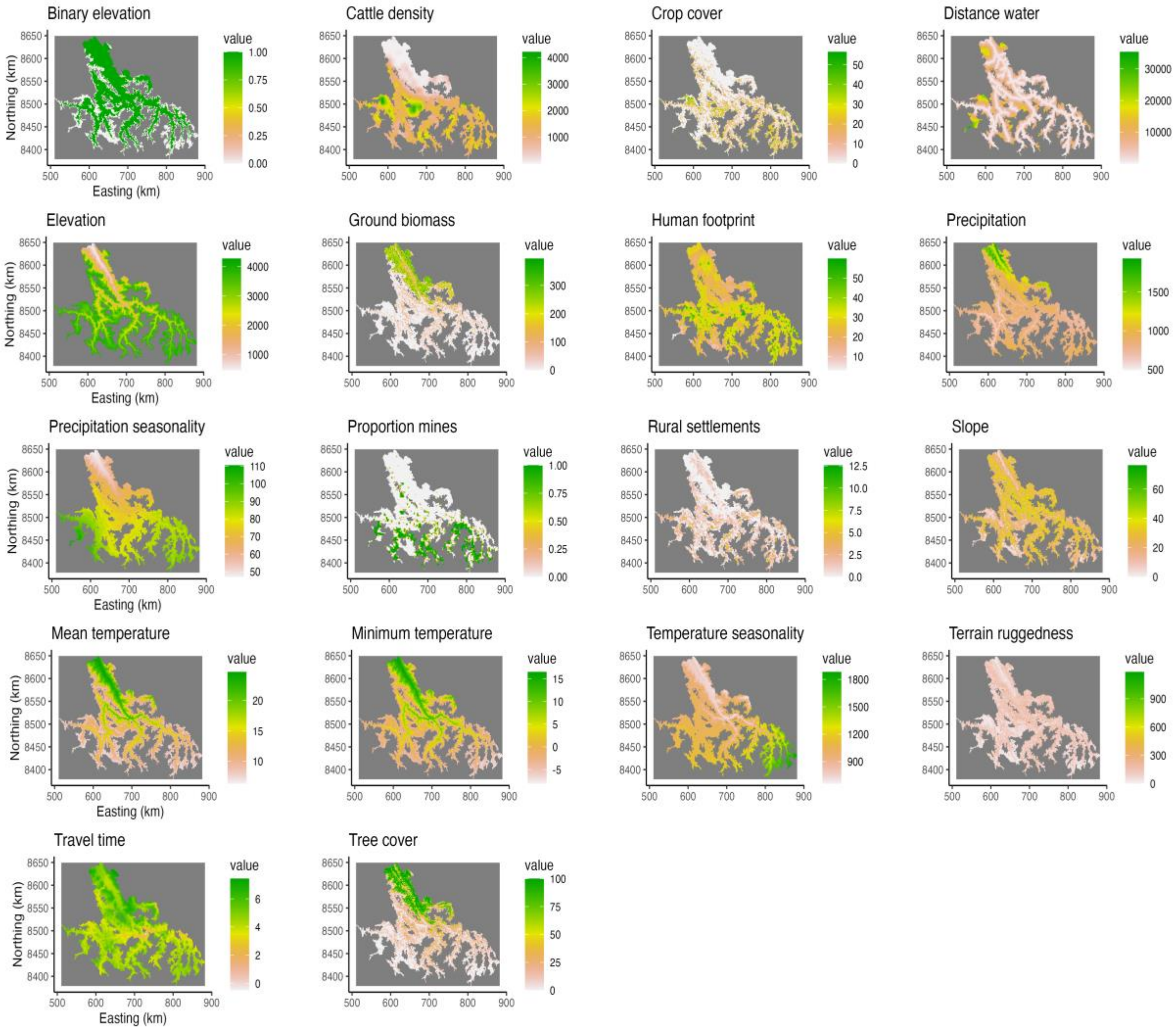

Figure S1: Spatial distribution of the environmental and anthropogenic variables initially assessed for the study, in the inter-Andean valleys of Apurimac, Ayacucho, and Cusco.

S2 Constrained refined Delaunay triangulation.

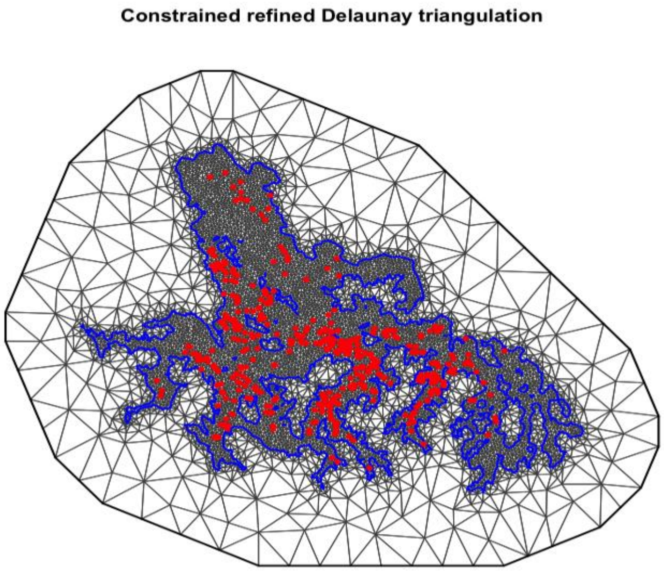

Figure S2: Constrained refined Delaunay triangulation. The ‘mesh’ consists of 7077 vertices, with small triangles with almost the same dimensions in the inner domain, where the predictions are important, and bigger triangles in the outer extension, to decrease the boundary effect. The blue line represents the smooth boundaries of the inter-Andean valleys of Apurimac, Ayacucho, and Cusco, and the red points are the 563 vampire bat roosts. The lengths of the mesh have the same units as the reference coordinate system.

S3 Pre-analysis of the pairwise distance between vampire bat roosts.

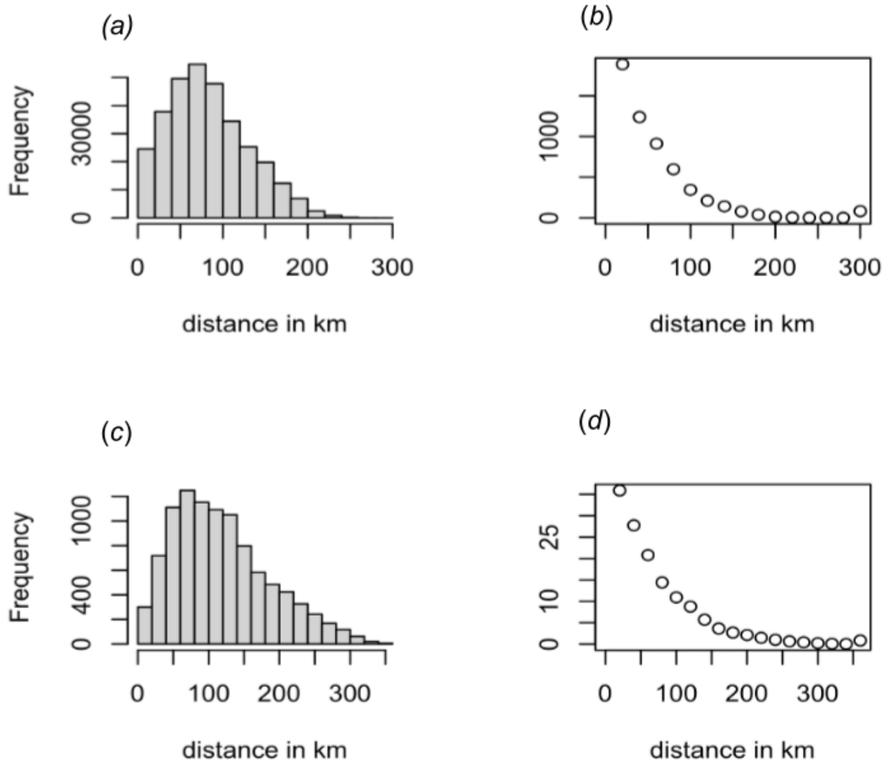

Figure S3: Analysis of pairwise distance between roosts excludes hypothesis of repulsion between roosts. (*a*) Histogram of the pairwise distance between the 563 roosts (units in km). (*b*) Y-axis is the count of pairwise distances divided by each distance break and X-axis is each distance break of the histogram. (*c*) and (*d*) represent the same as (*a*) and (*b*) but for 100 simulated roosts within the boundaries of the inter-Andean valleys of the study area, to verify the relationship kept the same.

S4 Blocks for ten-fold spatial cross-validation

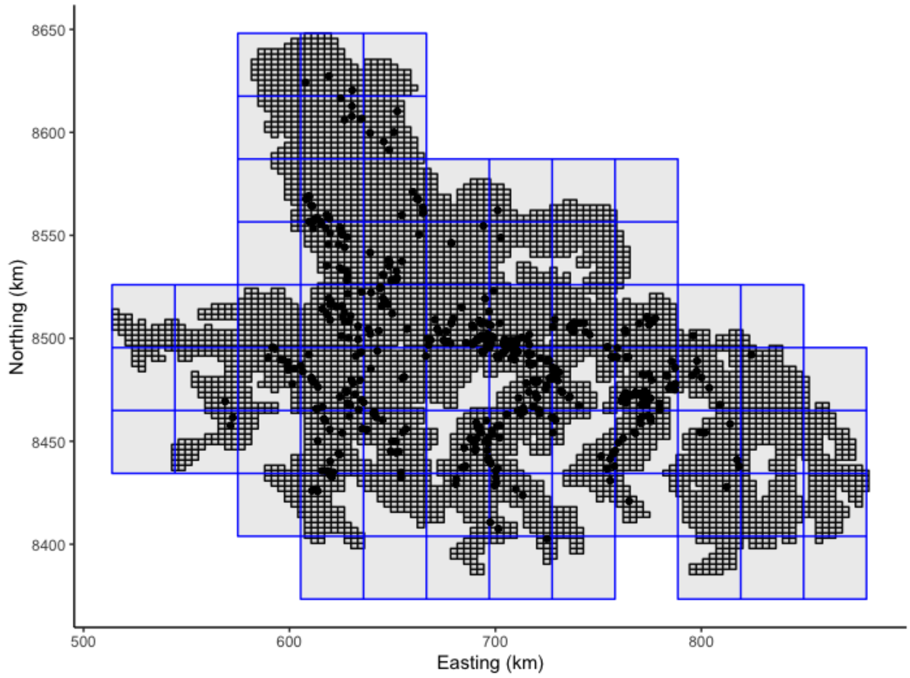

Figure S4: The blocks (blue) for aggregating the data and randomly dividing the study area in blocks for training and testing, the spatial grid of 8-km^2 (grey), and the 563 vampire bat roosts (black dots).

S5 Number of rabies outbreaks in livestock in the inter-Andean valleys of Apurimac, Ayacucho, and Cusco between 2003 and 2021.

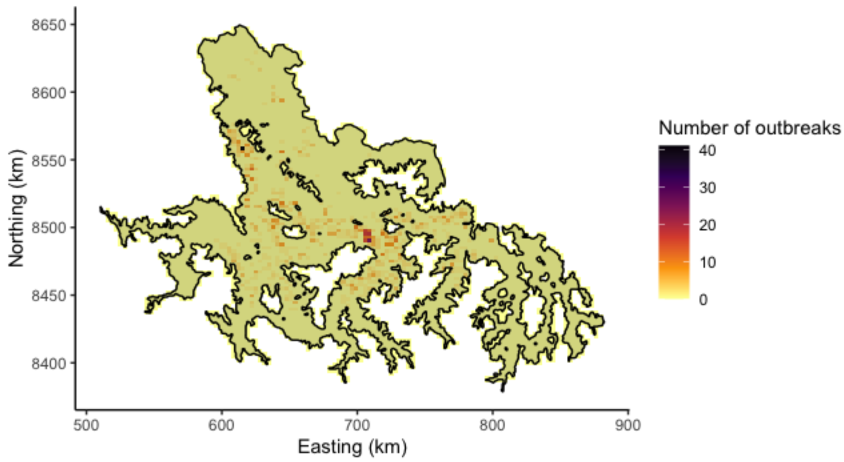

Figure S5: Number of rabies outbreaks per grid cell between 2003 and 2021 (1212 outbreaks). Darker colour means high number of outbreaks.

S6 Quadratic function of temperature.

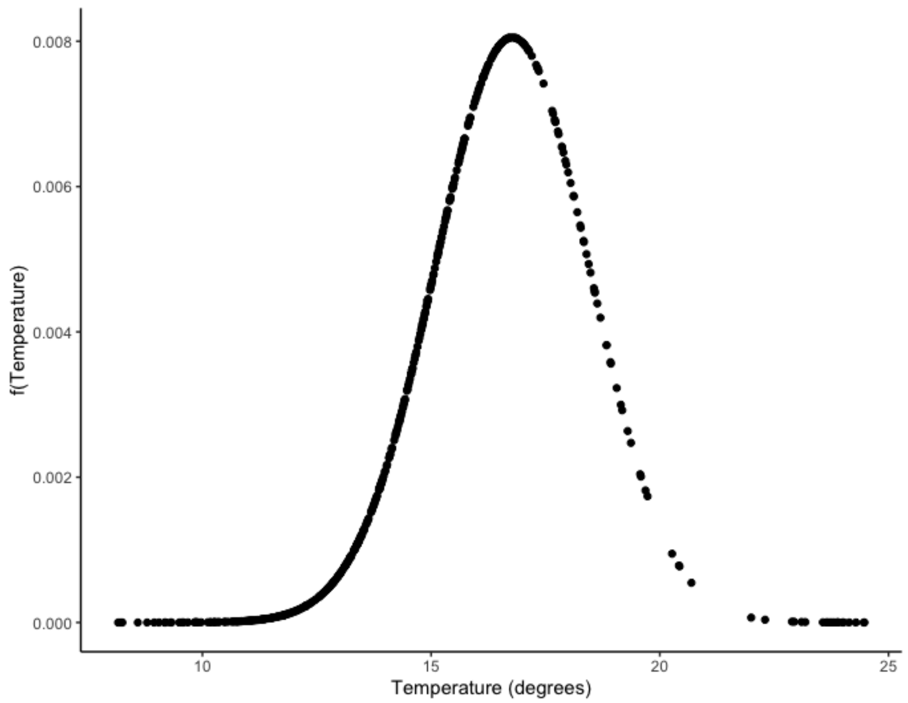

Figure S6: Quadratic function of temperature. The quadratic function of temperature is derived from the expression f(*temperature*) = exp(*a***temperature_std*^{2} + *b***temperature_std* + *c*). *Temperature_std* is the standardised temperature, *a* is the standardised coefficient for the quadratic term, *b* the standardised coefficient for the linear term, and *c* is the coefficient for the intercept. The optimal temperature (non-standardised temperature values) was computed based on: $X = ((-b)/2a) * \sigma + \mu$, where *X* is the parabola vertex, σ is the standard deviation of temperature in the non-standardized scale, and μ is the mean of temperature in the non-standardized scale.

S7 Posterior distribution of roost abundance.

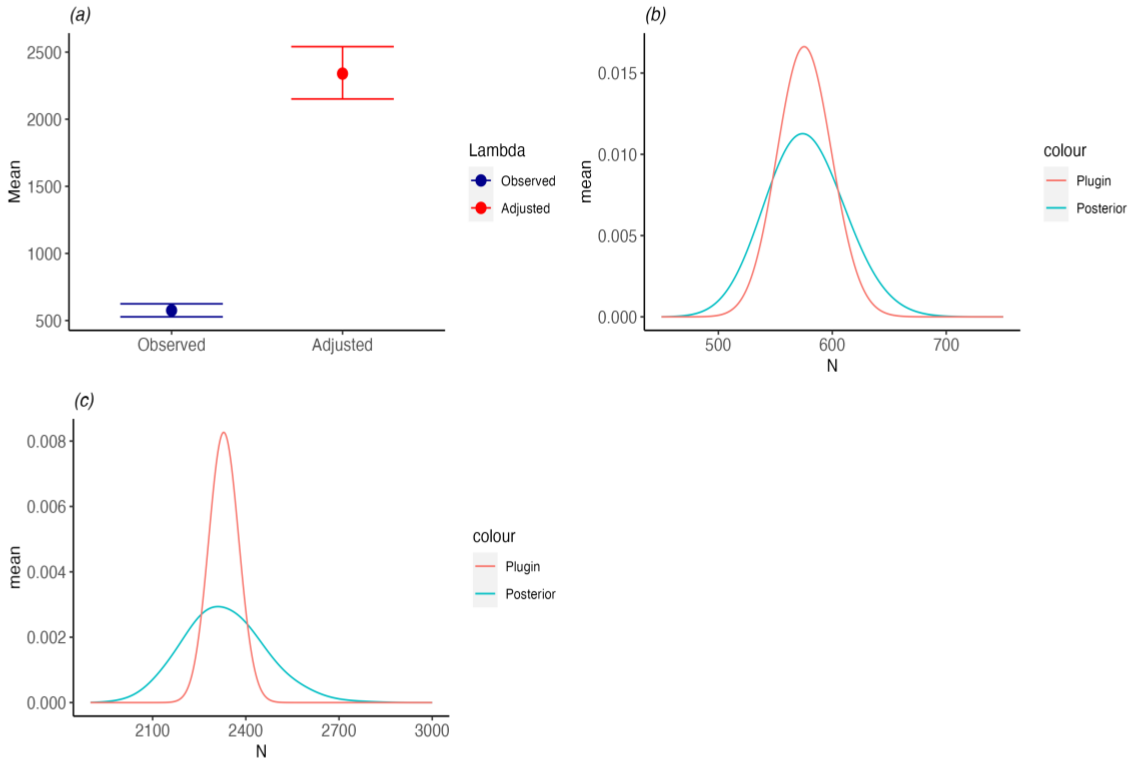

Figure S7: Posterior distribution of roost abundance, corrected or not for observation effort. (*a*) Predictions of the posterior mean of observed roosts ('Observed', blue dot), and predictions of total expected roosts ('Adjusted', red dot) and respective credible intervals (blue and red bars). (*b*) Posterior distribution for the total observed roosts (not correcting for the uneven effort). (*c*) Posterior distribution for the total expected roosts (correcting for the uneven effort. In both (*b*) and (*c*), the 'posterior' encompasses systematic stochasticity and uncertainty in parameter estimate, whereas the 'plugin' only considers model stochasticity. We can interpret the difference between the two curves by the parameter uncertainty.

S8 Maps of the predicted posterior mean of roost density in the linear scale.

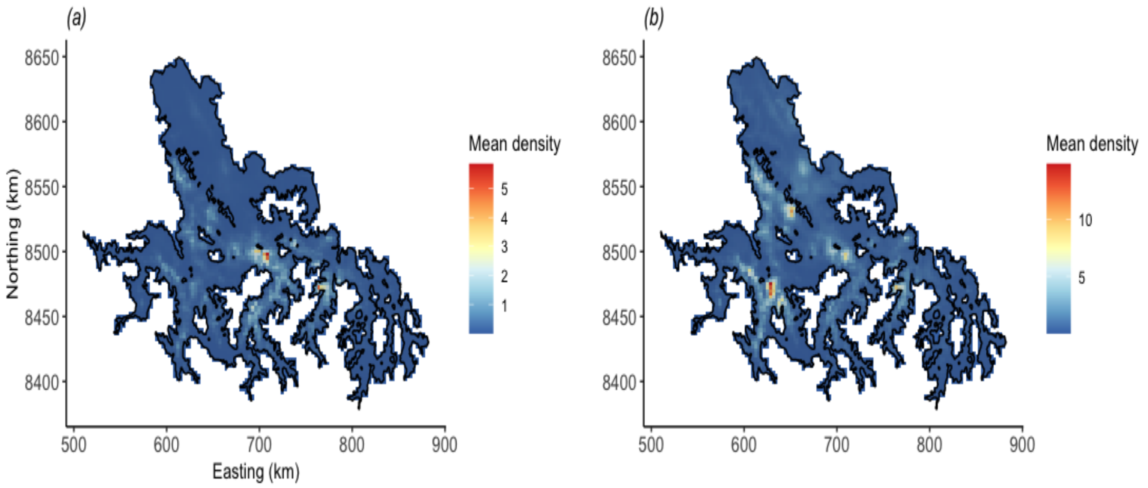

Figure S8: (*a*) Predictions of detected roosts (linear scale). (*b*) Predictions of the expected roosts when adjusting for the uneven effort (linear scale). Red means higher roost density.

References

1. Fick S, Hijmans R. 2017 WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315.
2. Jarvis A, Reuter HI, Nelson A, Guevara E. 2008 Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT).
3. Streicker DG *et al.* 2016 Host-pathogen evolutionary signatures reveal dynamics and future invasions of vampire bat rabies. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 10926–10931. (doi:10.1073/pnas.1606587113)
4. Buchhorn M, Lesiv M, Tsendbazar N-E, Herold M, Bertels L, Smets B. 2020 Copernicus Global Land Cover Layers—Collection 2. *Remote Sens.* **12**. (doi:10.3390/rs12061044)
5. Santoro M. 2018 GlobBiomass - global datasets of forest biomass. (doi:10.1594/PANGAEA.894711)
6. Gilbert M, Nicolas G, Cinardi G, Van Boeckel TP, Vanwambeke SO, Wint GRW, Robinson TP. 2018 Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. *Sci. Data* **5**, 1–11. (doi:10.1038/sdata.2018.227)
7. Wildlife Conservation Society - WCS, Center for International Earth Science Information Network - CIESIN, Columbia University. 2005 Last of the Wild Project, Version 2, 2005 (LWP-2): Global Human Footprint Dataset (Geographic). (doi:10.7927/H4M61H5F)
8. Weiss DJ *et al.* 2018 A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature* **553**, 333–336. (doi:10.1038/nature25181)