

Table A1: Number of samples of different types in which NTM species were detected. Species are sorted in order of highest prevalence. The total number of samples of each type are given in parenthesis. The * indicates respiratory-relevant species.

Species	Soil (n=492)	Stream/aquatic rock biofilm (n=165)	Stream/aquatic vegetation biofilm (n=62)	Well biofilm (n=6)	Lava-rock biofilm (n=7)	Stream/river filter (n=35)	Well filter (n=4)	All samples (n=771)
<i>M. abscessus</i> *	27	4	0	0	2	0	1	34
<i>M. chelonae</i>	12	7	2	0	0	0	0	21
<i>M. porcinum</i> *	11	4	0	0	0	1	0	16
<i>M. avium</i> *	6	4	0	0	0	0	0	10
<i>M. chimaera</i> *	6	3	0	0	0	0	0	9
<i>M. gallinarum</i>	7	1	1	0	0	0	0	9
<i>M. mucogenicum</i>	7	0	0	0	0	2	0	9
<i>M. florentinum</i>	5	1	0	0	0	2	0	8
<i>M. rhodesiae</i>	6	1	0	0	0	0	0	7
<i>M. stomatepiae</i>	0	0	0	0	0	7	0	7
<i>M. triplex</i>	5	0	0	0	0	1	1	7
<i>M. flavescens</i>	5	1	0	0	0	0	0	6
<i>M. peregrinum</i>	5	0	0	0	0	0	0	5
<i>M. septicum</i>	4	0	1	0	0	0	0	5
<i>M. wolinskyi</i>	4	0	1	0	0	0	0	5
<i>M. litorale</i>	3	1	0	0	0	0	0	4
<i>M. moriokaense</i>	3	0	1	0	0	0	0	4
<i>M. bacteremicum</i>	2	0	1	0	0	0	0	3
<i>M. boenickei</i>	3	0	0	0	0	0	0	3
<i>M. genavense</i>	2	0	0	0	0	1	0	3
<i>M. intracellulare</i> *	1	1	0	0	0	1	0	3
<i>M. iranicum</i>	3	0	0	0	0	0	0	3
<i>M. monacense</i>	3	0	0	0	0	0	0	3
<i>M. neoaurum</i>	2	0	1	0	0	0	0	3

<i>M. aurum</i>	2	0	0	0	0	0	0	2
<i>M. fluoranthenivorans</i>	2	0	0	0	0	0	0	2
<i>M. fortuitum*</i>	2	0	0	0	0	0	0	2
<i>M. franklinii</i>	1	1	0	0	0	0	0	2
<i>M. gordonae</i>	2	0	0	0	0	0	0	2
<i>M. grossiae</i>	1	0	0	0	0	1	0	2
<i>M. paraffinicum</i>	1	0	1	0	0	0	0	2
<i>M. parmense</i>	1	0	0	0	0	1	0	2
<i>M. poriferae</i>	2	0	0	0	0	0	0	2
<i>M. rutilum</i>	2	0	0	0	0	0	0	2
<i>M. timonense</i>	2	1	0	0	0	0	0	3
<i>M. vaccae</i>	2	0	0	0	0	0	0	2
<i>M. vanbaalenii</i>	2	0	0	0	0	0	0	2
<i>M. yongonense</i>	2	0	0	0	0	0	0	2
<i>M. alvei</i>	1	0	0	0	0	0	0	1
<i>M. arceuilense</i>	1	0	0	0	0	0	0	1
<i>M. barrassiae</i>	1	0	0	0	0	0	0	1
<i>M. farcinogenes</i>	1	0	0	0	0	0	0	1
<i>M. frederiksbergense</i>	1	0	0	0	0	0	0	1
<i>M. gadium</i>	1	0	0	0	0	0	0	1
<i>M. goodii</i>	1	0	0	0	0	0	0	1
<i>M. kansasii</i>	1	0	0	0	0	0	0	1
<i>M. lentiflavum</i>	0	0	0	0	0	1	0	1
<i>M. mageritense</i>	1	0	0	0	0	0	0	1
<i>M. marinum</i>	1	0	0	0	0	0	0	1
<i>M. marseillense</i>	1	0	0	0	0	0	0	1
<i>M. parascrofulaceum</i>	1	0	0	0	0	0	0	1
<i>M. phocaicum</i>	1	0	0	0	0	0	0	1
<i>M. psychrotolerans</i>	0	1	0	0	0	0	0	1
<i>M. simiae</i>	1	0	0	0	0	0	0	1

<i>M. smegmatis</i>	1	0	0	0	0	0	0	1
<i>M. sp</i>	23	10	1	0	0	2	0	36

Table A2: Beta coefficient estimates with associated 95% credible interval (LCL = lower limit, UCL = upper limit) for predictors in a spatial logistic regression model for the probability of NTM presence across the Hawaiian Islands. All variables were included in the model, along with spatial random effects. Estimates are based on 771 samples collected from 422 geographic sites.

Predictor	mean	LCL	UCL
Intercept	-0.99	-1.16	-0.83
Elevation (m)	0.11	-0.06	0.29
Population density (people/km ²)	0.07	-0.09	0.24
Distance to coast (km)	-0.05	-0.22	0.11
Distance thermal spring (km)	-0.05	-0.22	0.14
Distance inland water (km)	-0.02	-0.19	0.15
Distance to ag (km)	-0.14	-0.33	0.06
Avg. annual vegetation height (mm)	0.10	-0.08	0.27
Soil age midpoint (years)	0.02	-0.17	0.19
Water balance (mm)	0.02	-0.18	0.20
Soil expansive ability (COLE)	0.12	-0.04	0.28
Soil expansive ability x Water balance	0.26	0.03	0.52

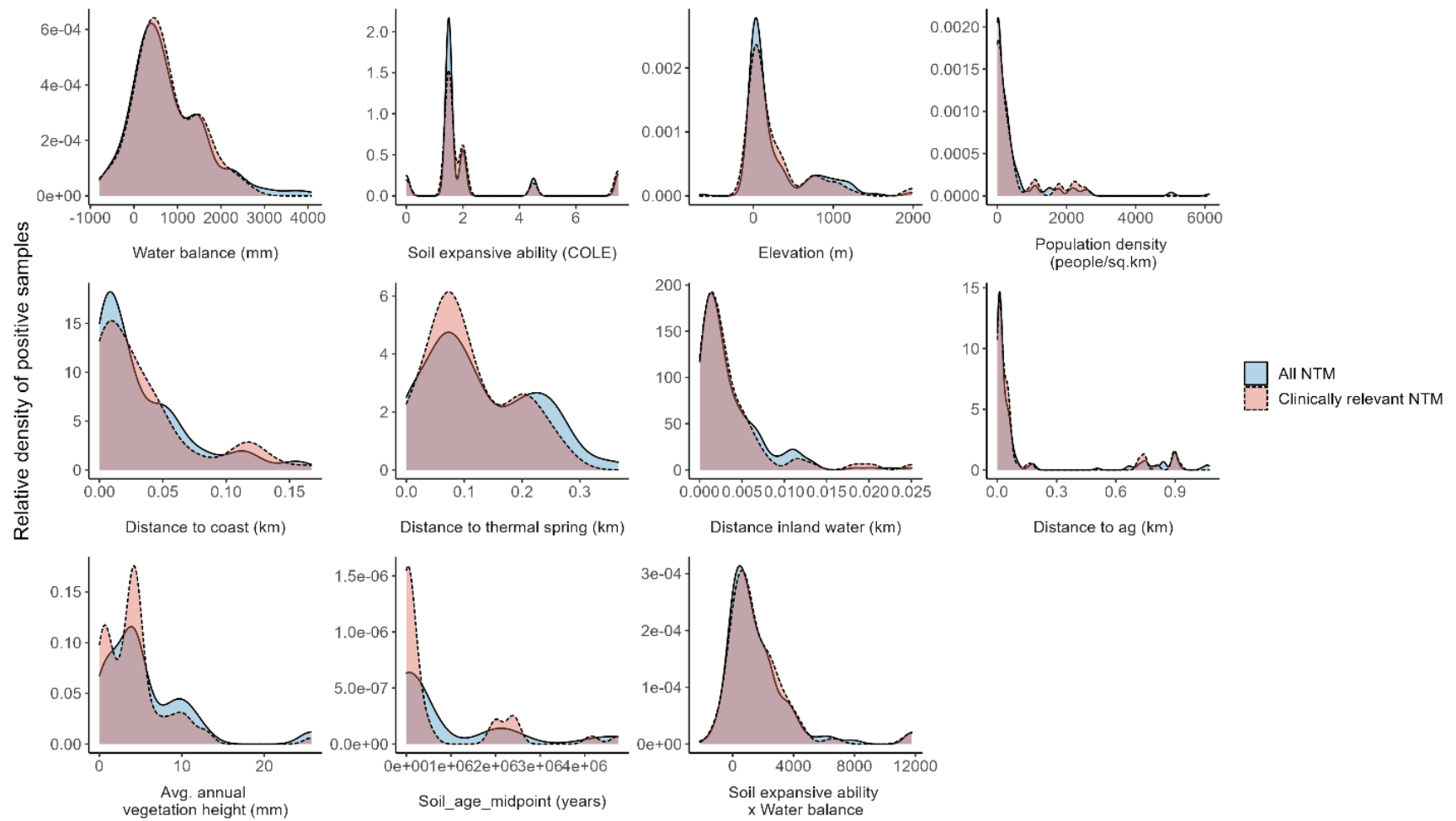


Figure A1: Distributions of samples positive for any species of NTM and samples positive only for clinically relevant species (i.e., *M. chimaera*, *M. avium*, *M. intracellulare*, *M. abscessus*, *M. porcinum*, *M. chelonae*) along gradients of 11 predictors used in spatial logistic regression of NTM in Hawai'i.

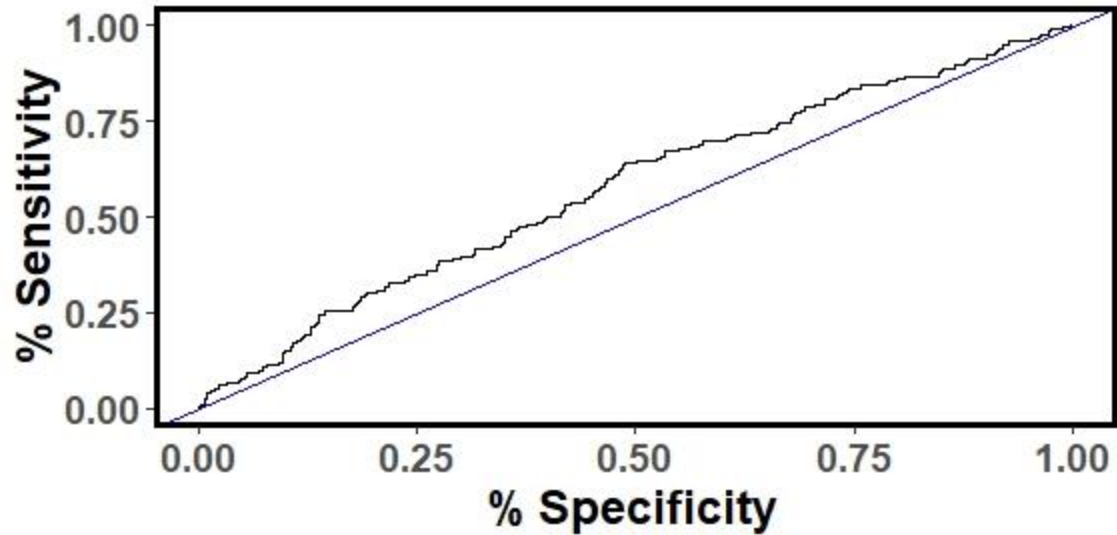


Figure A2: ROC curve for a spatial logistic regression model of the probability of NTM presence across Hawai'i with an AUC of 0.58. Sensitivity represents the number of model-predicted samples correctly identified as 1s divided by the total number observed 1s. Specificity represents the number of model-predicted samples correctly identified as 0s divided by the total number observed 0s. The area under the ROC curve, AUC, represents a measure of how much a model is capable of distinguishing between classes (0s and 1s). An AUC higher than 0.5 indicates the model is better at predicting 0 classes as 0 and 1 classes as 1 than random chance.

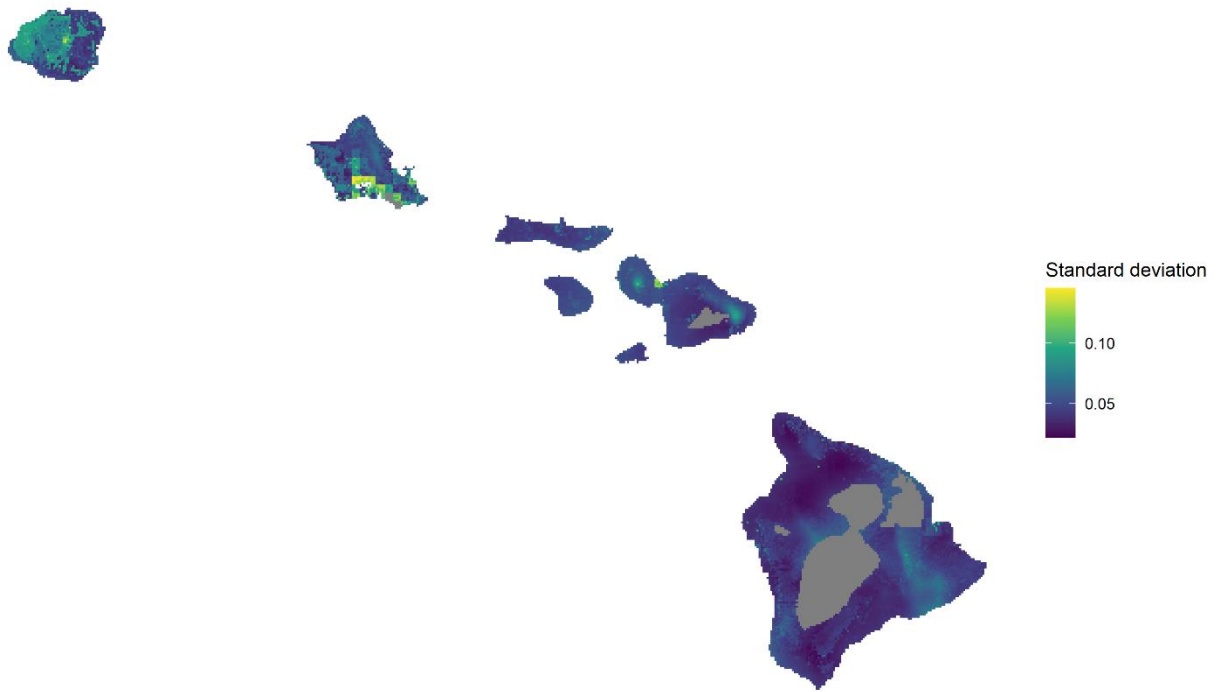


Figure A3: Uncertainty (standard deviation) in model predictions of the probability of NTM presence across the main Hawaiian Islands. Missing is the island of Ni'i'hau for which we were missing coverage of predictor variables. Predictions are generated from a spatial model using remotely sensed predictors and accounting for spatial bias in sampling effort. Gray areas are those with higher elevation, population density and/or soil expansive ability than measured by our sample data and thus have no associated predictions.