Modelling the impact of MAUP on environmental drivers for *Schistosoma japonicum* prevalence

CURRENT STATUS: ACCEPTED

Parasites & Vectors BMC

Andrea Araujo Navas
Universiteit Twente Faculteit Geo-Informatie Wetenschappen en Aardobservatie
a.l.araujonavas@utwente.nl*Corresponding Author*

Frank Osei
Universiteit Twente Faculteit Geo-Informatie Wetenschappen en Aardobservatie

Ricardo J. Soares Magalhães
University of Queensland

Lydia R. Leonardo
University of the Philippines Diliman

Alfred Stein
Universiteit Twente Faculteit Geo-Informatie Wetenschappen en Aardobservatie

DOI:

10.21203/rs.2.20917/v1

SUBJECT AREAS

Parasitology

KEYWORDS

Schistosomiasis modelling, modifiable areal unit problem, uncertainty, Bayesian statistics, convolution model

Abstract

Background: The modifiable areal unit problem (MAUP) arises when the support size of a spatial variable affects the relationship between prevalence and environmental risk factors. Its effect on Schistosomiasis modelling studies could lead to unreliable parameter estimates. The present research aims to quantify MAUP effects on environmental drivers of *Schistosoma japonicum* infection by (i) bringing all covariates to the same spatial support, (ii) estimating individual-level regression parameters at 30 m, 90 m, 250 m, 500 m, and 1 km spatial supports, and (iii) quantifying the differences between parameter estimates using five models.

Methods: We modelled the prevalence of *Schistosoma japonicum* using sub-provinces health outcome data and pixel-level environmental data. We estimated and compared regression coefficients from convolution models using Bayesian statistics.

Results: Increasing the spatial support to 500 m gradually increased the parameter estimates and their associated uncertainties. Abrupt changes in the parameter estimates occur at 1 km spatial support, resulting in loss of significance of almost all the covariates. No significant differences were found between the predicted values and their uncertainties from the five models. We provide suggestions to define an appropriate spatial data structure for modelling that gives more reliable parameter estimates and a clear relationship between risk factors and the disease.

Conclusions: Inclusion of quantified MAUP effects was important in this study on schistosomiasis. This will support helminth control programs by providing reliable parameter estimates at the same spatial support, and suggesting the use of an adequate spatial data structure, to generate reliable maps that could guide efficient mass drug administration campaigns.

Full Text

Due to technical limitations, full-text HTML conversion of this manuscript could not be completed. However, the manuscript can be downloaded and accessed as a PDF.

Figures

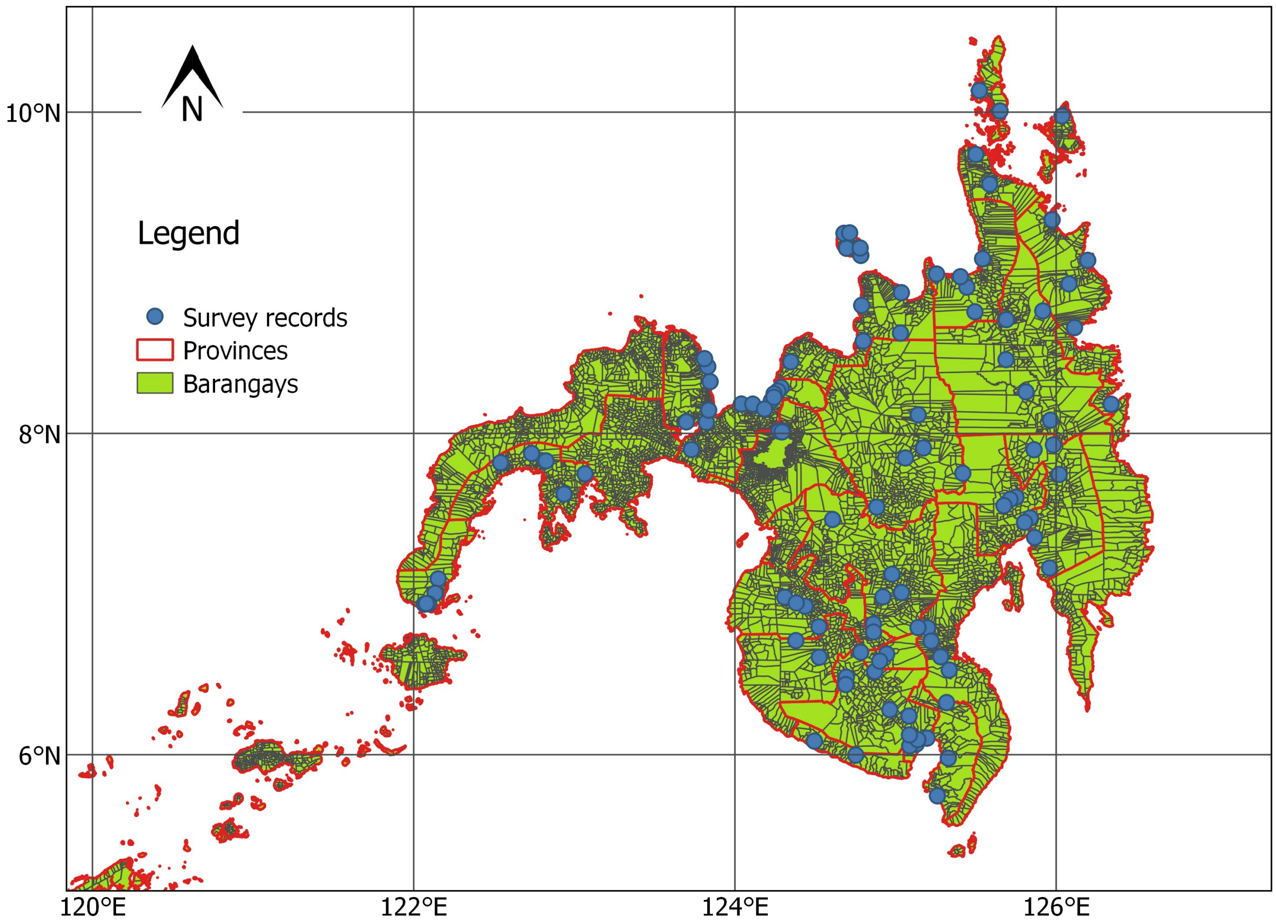


Figure 1

Study Area: The Mindanao region in The Philippines. Blue dots are the aggregated survey data at barangay-level.

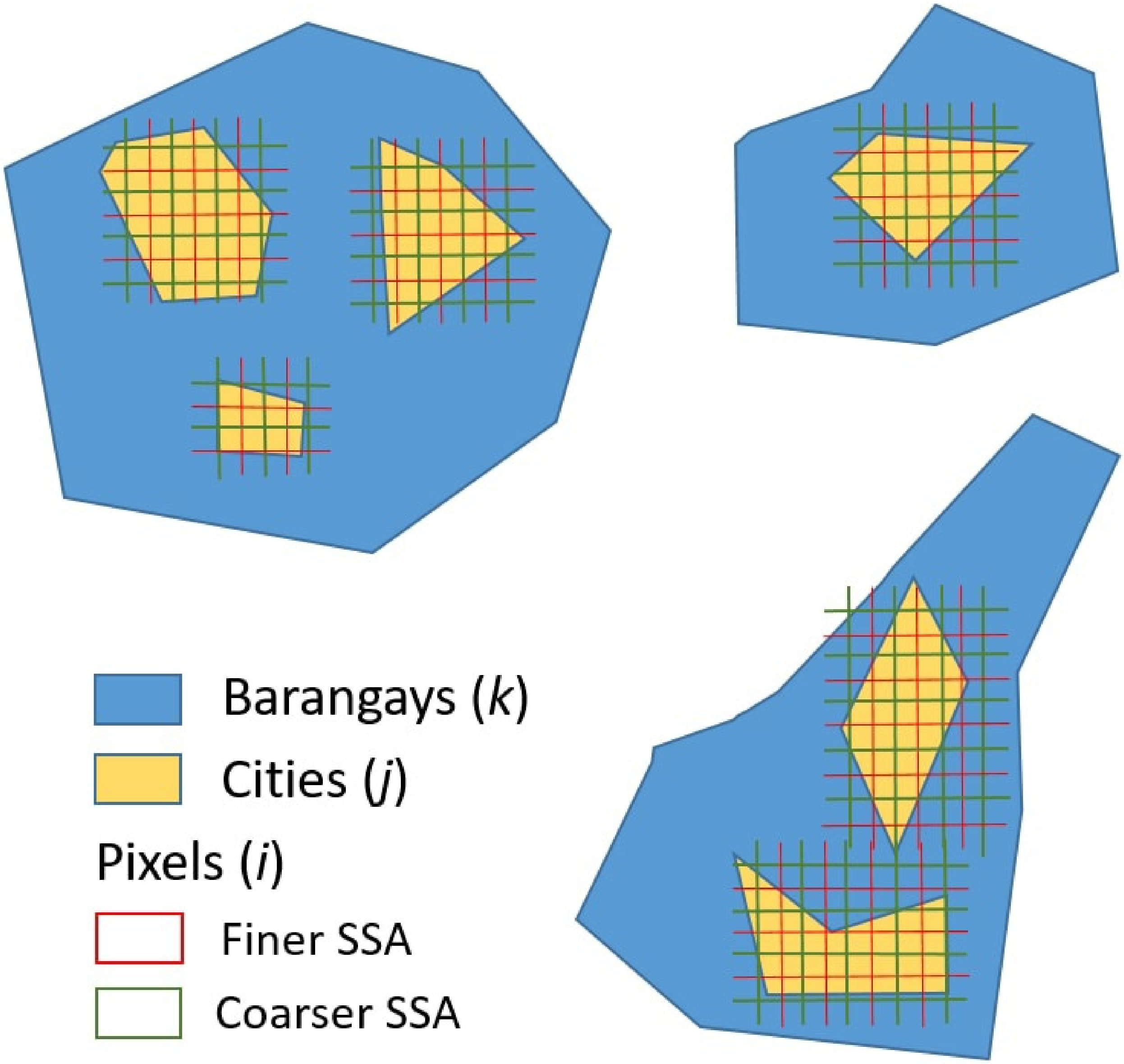


Figure 2

Environmental risk factors extraction at pixel-level from cities within barangays

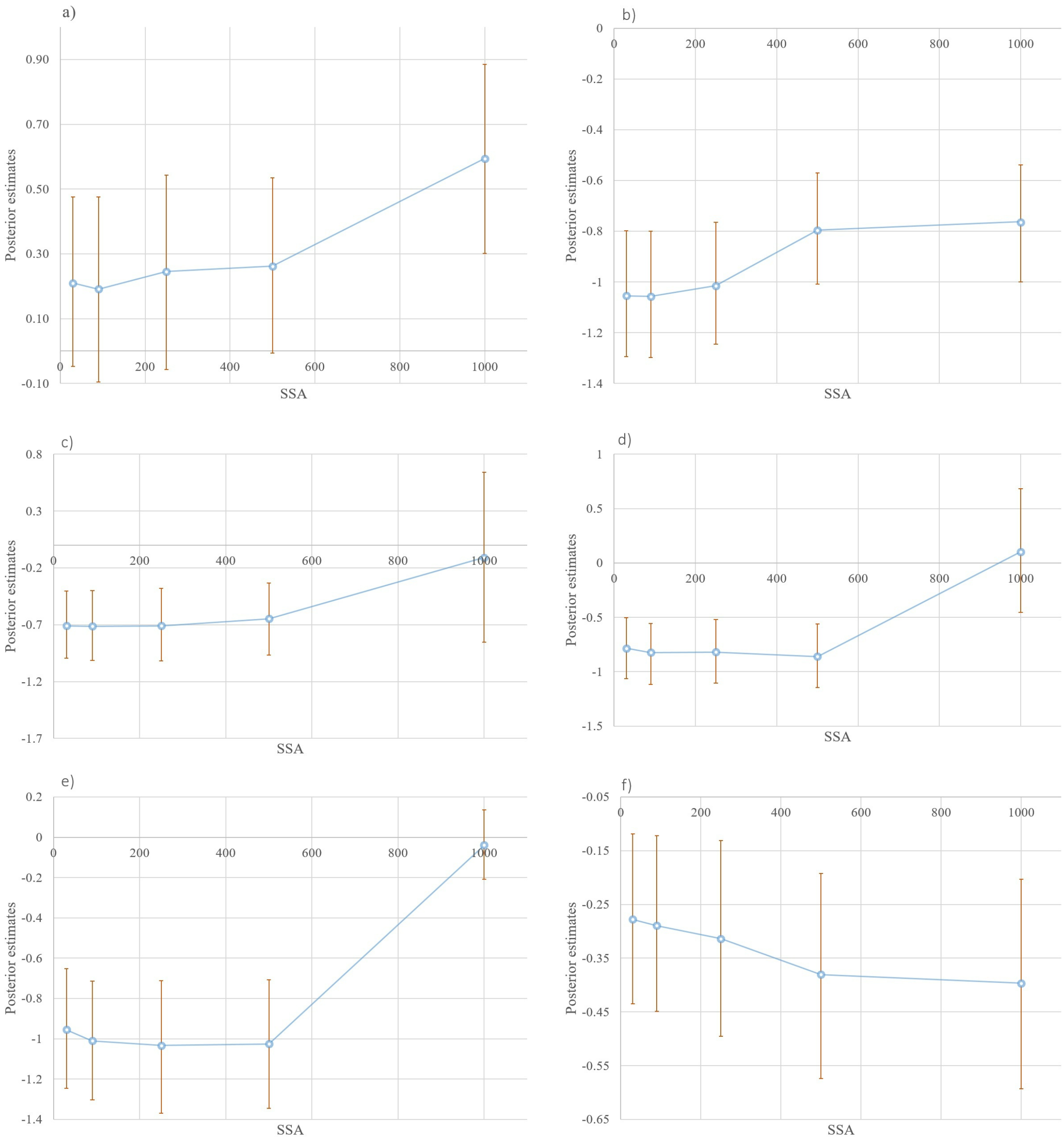
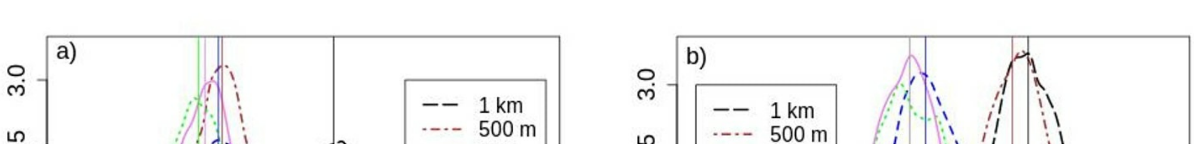


Figure 3

Posterior estimates and their credible intervals: a) Normalized difference vegetation index; b) Normalized difference water index; c) Land surface temperature day d) Land surface temperature night; e) Elevation; f) Nearest distance to water bodies. Abbreviations: SSA, Spatial support of analysis



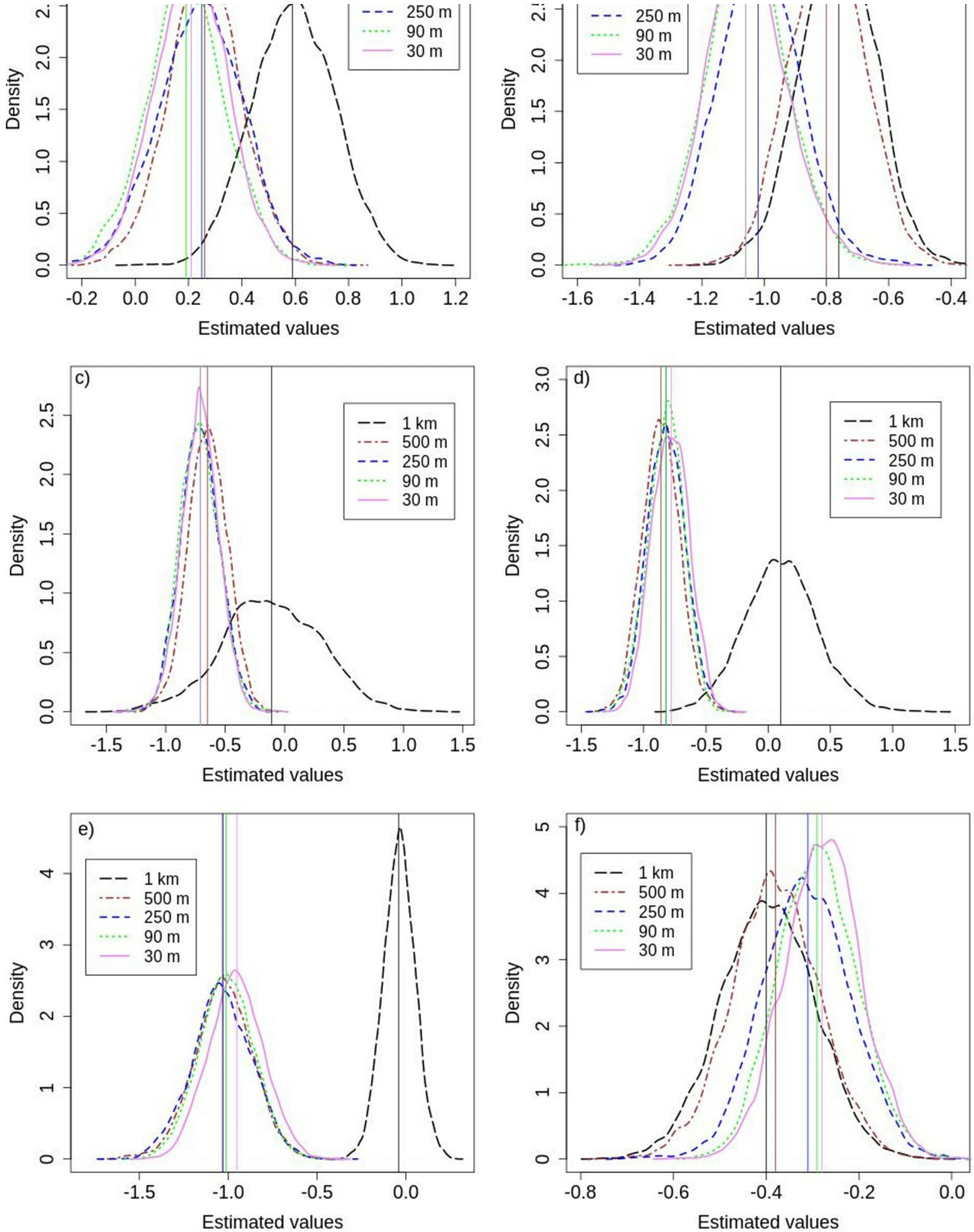


Figure 4

Density plots for the risk factors regression coefficients: a) Normalized difference vegetation index; b) Normalized difference water index; c) Land surface temperature day d) Land surface temperature night; e) Elevation; f) Nearest distance to water bodies

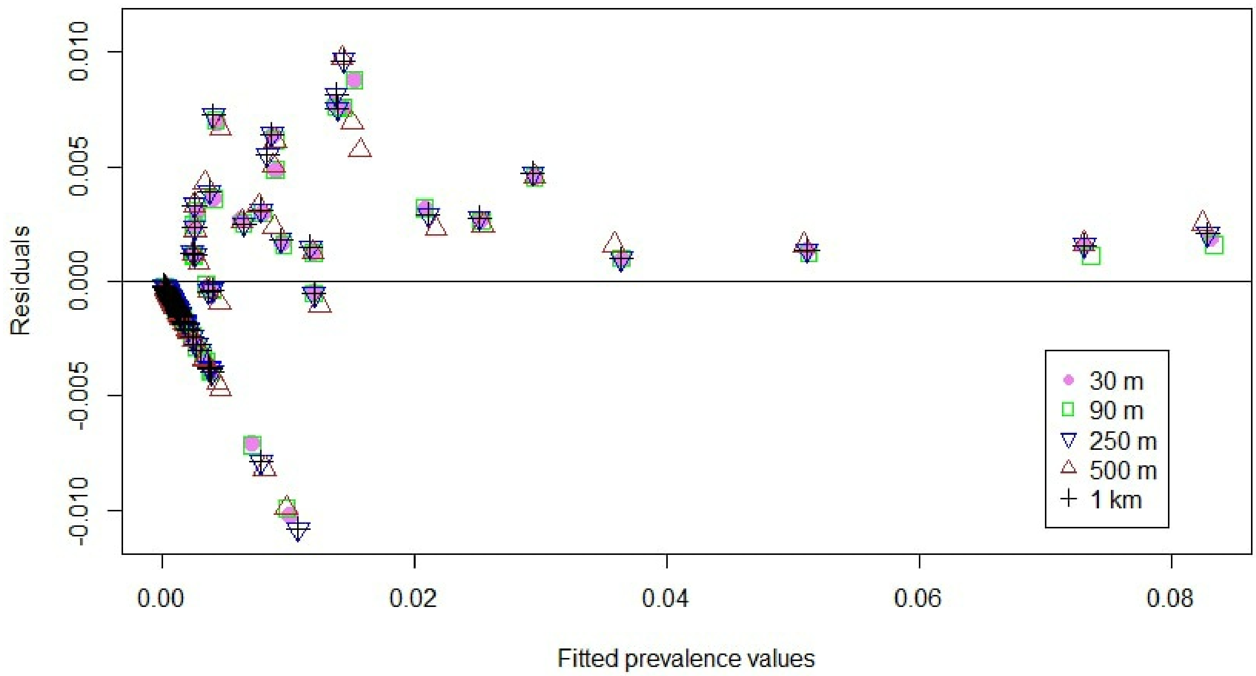


Figure 5

Residual plot for the five increasing spatial supports of analysis. The x axis represents the fitted prevalence values for the five spatial supports of analysis. The y axis represents the residuals calculated by the difference between the observed and predicted prevalence values.

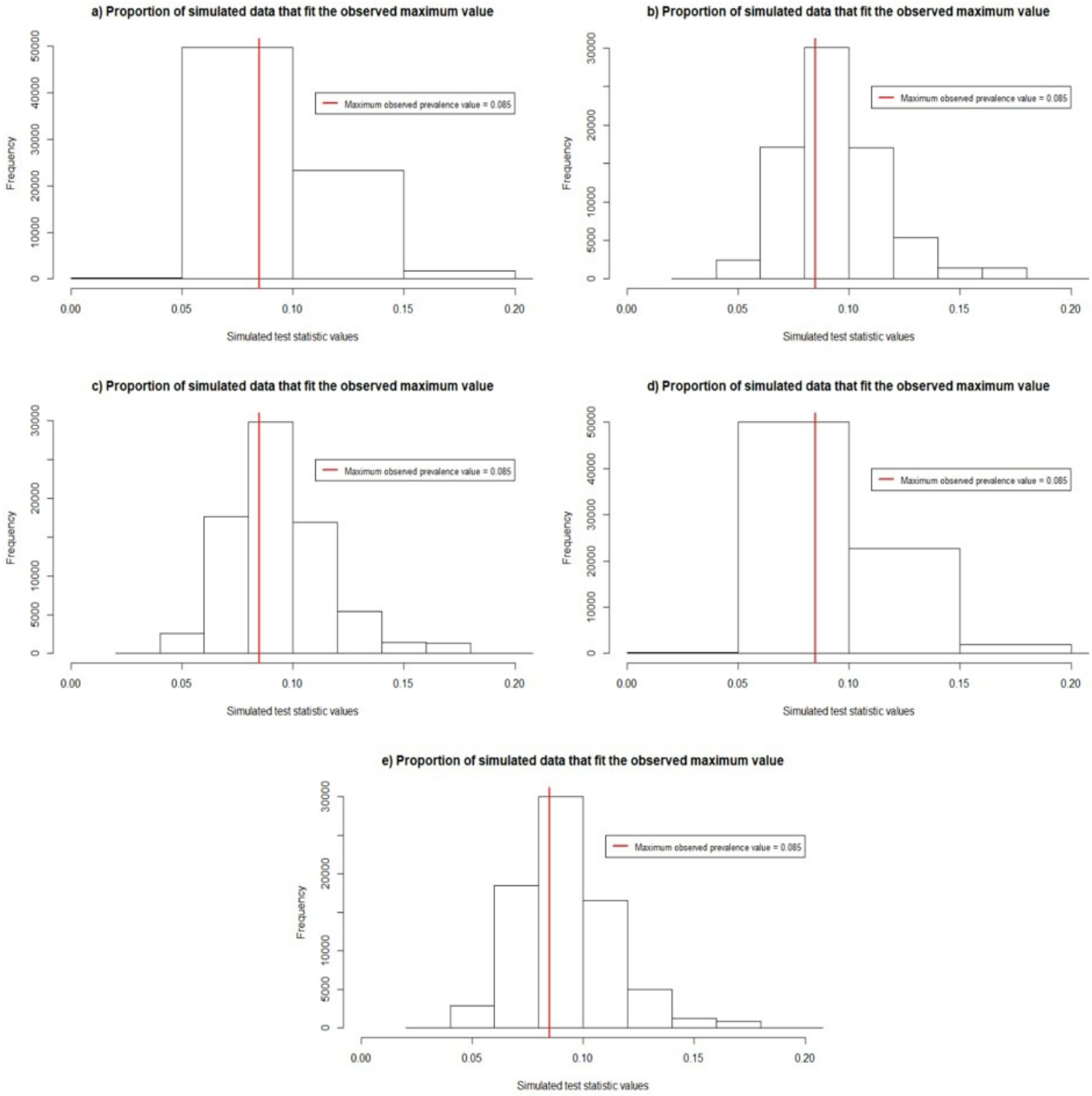


Figure 6

Proportion of simulated prevalence data that fit the observed maximum prevalence value.

a) SSA=30m, b) SSA=90m, c) SSA=250m, d) SSA=500m, e) SSA=1 km. Abbreviations: SSA, Spatial support of analysis.

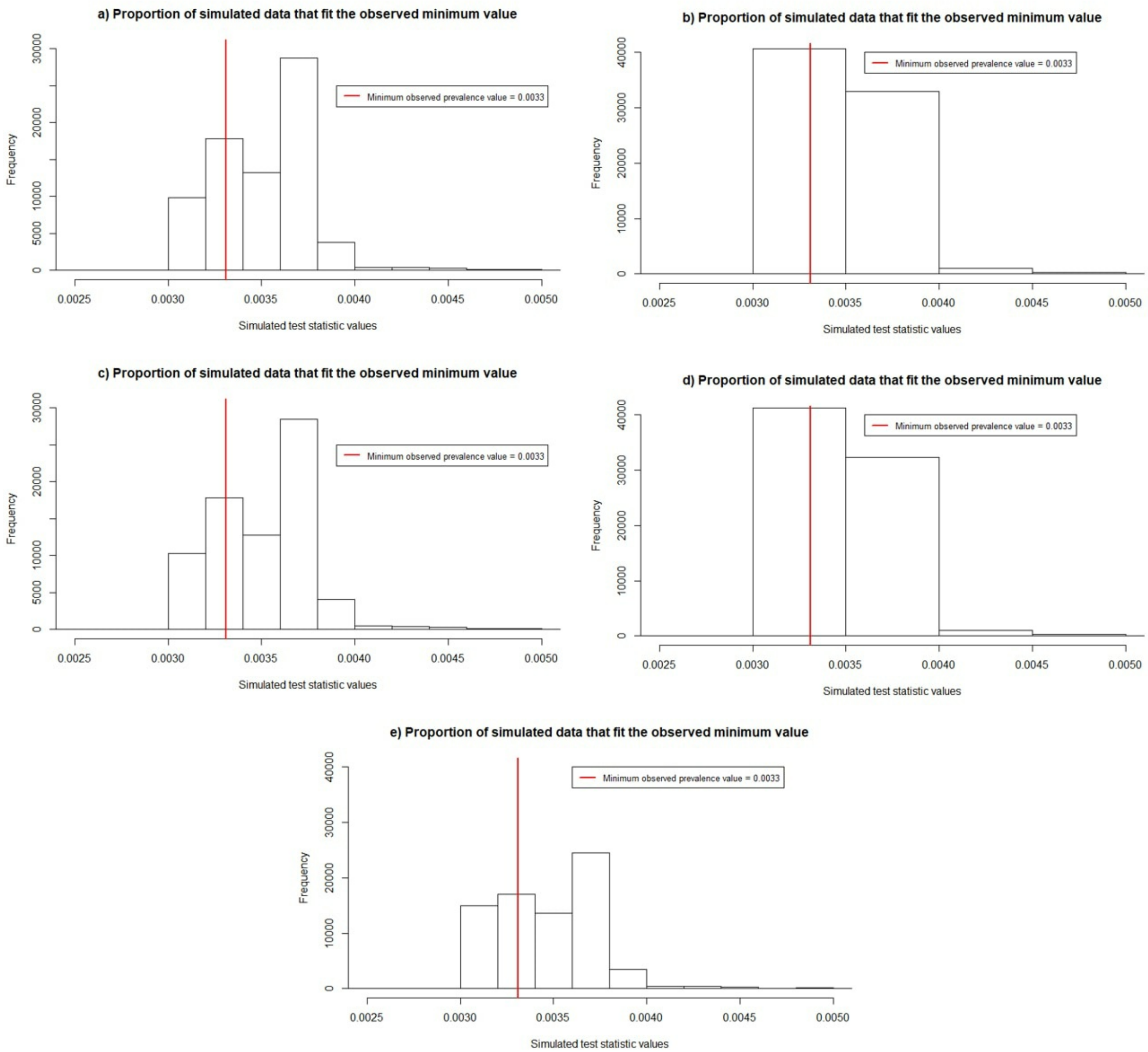


Figure 7

Proportion of simulated prevalence data that fit the observed minimum prevalence value. a) SSA=30m, b) SSA=90m, c) SSA=250m, d) SSA=500m, e) SSA=1 km. Abbreviations: SSA, Spatial support of analysis.

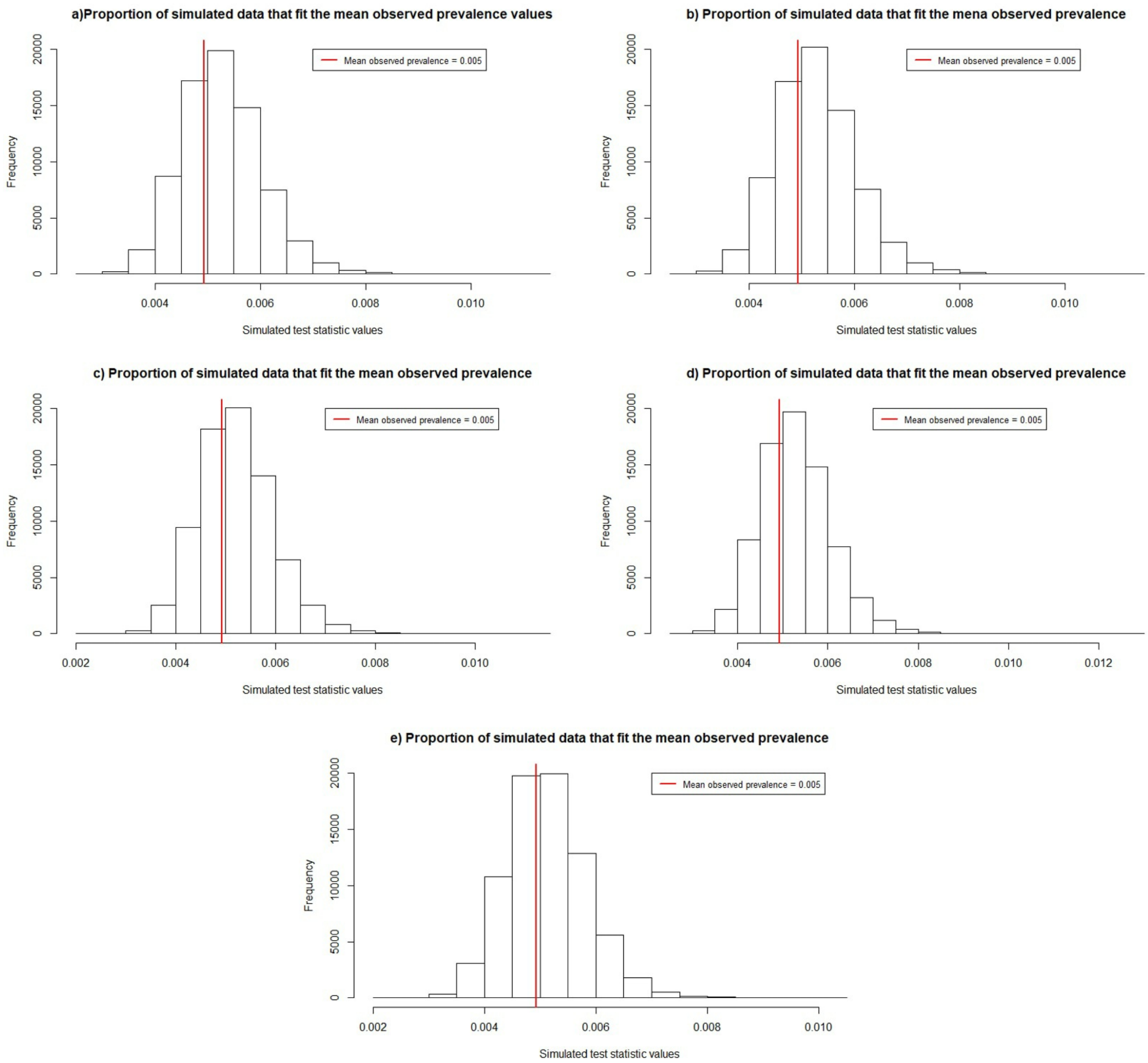


Figure 8

Proportion of simulated prevalence data that fit the observed mean prevalence value. a) SSA=30m, b) SSA=90m, c) SSA=250m, d) SSA=500m, e) SSA=1 km. Abbreviations: SSA, Spatial support of analysis.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

Additional file 3.tiff
Additional file 1.txt
Additional file 2.txt