

## Chapter 1

- ✓ Fig 1.3. Explain in caption why color gradient looks continuous, but maps are binary/categorical, or fix the color gradient in the legend to also be binned
  - The figure has been updated with a categorical legend for an index score of 0-25, 25-50, 50-75 and 75-100 to reflect that the index score was not truly continuous.
- ✓ Add a sentence or two about the apparent disconnect between data and WHO classification of countries as Lassa endemic or outbreaks
  - The following sentence has been added to paragraph 1 in section 1.6.2.  
"Countries designated as endemic by the WHO do not all regularly reported cases of Lassa fever (e.g., Guinea) and between endemic countries there are substantial differences in the number of annually reported cases (Simons, 2022b). "

## Chapter 2

- ✓ 2.1 (optional) add some discussion on why there are gaps in existing databases (CLOVER and GBIF) and whether the work done here helps fill those gaps
  - The following sentence was added to the results section 2.4.2.  
"Importantly GBIF data typically originates from different sources than primary trapping studies. The majority of occurrences (65.6%) included in the GBIF dataset were produced from museum catalogues of preserved specimens (GBIF, 2023). Despite a small proportion of small-mammal sampling data being subsequently incorporated into museum collections through the depositing of voucher specimens these two sources of data are complementary when considering the sampling effort throughout a small-mammal species' range."
  - The composition and data of CLOVER (and the constituent datasets) depended on literature reviews which can result in incorrect, incomplete or out-of-date host-pathogen associations being reported or not. This is a limitation of static datasets. To highlight this the following sentence was added in the discussion.  
"A limitation of host-pathogen datasets is produced through their construction (i.e., systematic reviews of scientific literature) which can introduce spurious associations, incidental associations or out-of-date associations being introduced into the dataset. Due to the shared method of construction, this limitation exists for the current dataset."

## Chapter 3

- ✓ Add details of biosafety procedures for working with non-inactivated samples that may contain a CL4 virus
- ✓ Add details of PPE used during rodent capture/sampling
  - The following paragraphs were added to chapter 3 section 3.4.

"Handling potentially hazardous material.

Samples obtained from small-mammals were potentially infected with LASV or other rodent-associated pathogens that can cause human disease. To minimise the risk of harm to the members of the study team and the communities in which sampling was taking place field biosecurity practices were implemented (Mills et al. 1995, Herbreteau et al. 2011). These precautions included training field workers on the safe handling of live-trapped small mammals, the provision and use of adequate personal protective equipment (i.e., disposable gowns, N95 respirators, eye-goggles and kevlar coated gloves) and ensuring that live small-mammals were only directly handled following deep anaesthesia and euthanasia. Surface disinfectants were used to clean sampling equipment, sampling surfaces and traps between each animal to reduce risk of infection (Fichet-Calvet, 2014). Following the

processing of samples at the preparation site the processing area was carefully decontaminated including the external surfaces of any sample collection vials prior to packaging for transportation to the laboratory (Mills, 1995).

Laboratory investigations and sample handling were conducted adhering to Biosafety Level 3 (BSL-3) criteria (CDC, 1994). Tissue samples including ear punches and eyes were preserved in 4% Formaldehyde, liver and spleen samples were preserved in 100% methanol and blood or dried-blood spot elute was inactivated with 1% Triton-X 100, this is expected to completely inactivate LASV and eliminate the risk of transmission (Olschewski et al, 2021)."

- ✓ Check bait composition is consistent across chapters
  - Discrepancies in the discription of bait composition have been corrected throughout. The correct composition was oats, peanut, palm oil and dried fish.
- ✓ (optional for later chapters) consider a measure of luminosity that accounts for night to night variation in moonrise/moonset
  - Thank you for this suggestion, this has not been incorporated in this thesis but will be considered for Chapter 4 when it is updated for submission.
- ✓ Expand on justification of the unbalanced habitat selection, explain data and methods underlying how the 1:4:2 allocation of trapping effort was derived
  - The following was added to section 3.3.5 Trapping grids within different land use types:

"To approximate representativeness of trapping effort weighted to the area of these different land uses and human activity within them, we opted for the following setup at each of the four selected villages: a single grid in forest, four grids in agriculture (two in settings proximal to the village and two in more distal locations) and two within the villages (one in outdoor and one in indoor settings). A data driven approach for this allocation of trapping effort was used to obtain a sample representative to expected hazard of human infection.

First, a raster of IUCN level-2 habitat type at 100m resolution was obtained and subsequently subset to the Eastern Province of Sierra Leone (Jung et al. 2020). I used the IUCN level-2 habitat categories to calculate the proportion of landuse type within the wider study region, with 65% of land categorised as forest, 34% as agriculture (incorporating shrubland, arable, pasture and plantations) and <1% as areas of human habitation. Second, a raster of human population density estimates from 2020 at 1km resolution was obtained, this was subset to the Eastern Province of Sierra Leone and merged to the same resolution as the habitat category raster using nearest neighbour values (SEDAC, 2021). Population density was then converted to population count by dividing by the raster cell area. This produced a measure of proportion landuse classification by population for forest (0.19), agriculture (0.45) and areas of human habitation (0.34) which was used to approximate the 1-to-4-to-2 balance of habitat types selected for trapping. "

## Chapter 4

- ✓ Revise language on the putative interaction between Mus and Mastomys: 'led to', 'the finding of interactions', 'regulates occupancy' might be too strong.
  - Changes have been made throughout to highlight the correlational analysis cannot be used for causal inference. This complements the wording currently in the thesis.

"A statistically significant correlation was interpreted as one species being more (if a positive correlation) or less (if a negative correlation) likely to occur in a grid cell of the specific land use type if the other species were present. The causal mechanism and direction behind any observed correlations cannot be inferred from the current analysis."

- ✓ Also, would be nice to recommend some future research that could test whether these 2 species are competing/displacing each other, discuss any hypothetical ecological mechanisms for interaction.
  - This was previously included in Section 4.5.2 with the following sentence:

"It is not possible to ascertain from the current study whether *M. musculus* is expanding into rural settings and what effect this may have on rodent communities and LASV transmission. To identify the causal processes of changes in rodent species community structures in response to invasive species' range expansion longer term monitoring of rodent communities would be beneficial, similar to that conducted in Senegal (Dalecky et al., 2015)."

and has been expanded by adding the following paragraph.

"To identify a causal interaction between *M. natalensis* and *M. musculus* occupancy and any biotic interactions between these species that may moderate their occupancy, investigations at the invasion front of *M. musculus* in Sierra Leonean settings would be informative. Possible ecological mechanisms for these processes, beyond anthropogenic factors such as landuse change, urbanisation and environmental modifications, could include resource competition, competition for breeding habitats, interference competition (i.e., disruption of reproductive behaviours) or territoriality (Harrington et al. 2009, Polito et al. 2022, Granjon et al. 2023)."

- ✓ Be explicit about the alternative to biotic interaction: within a given land use type, there is a gradient of some other factor which influences both species, so their occurrence is governed by that, not each other.
  - The following paragraph has been added to the discussion in 4.5.2

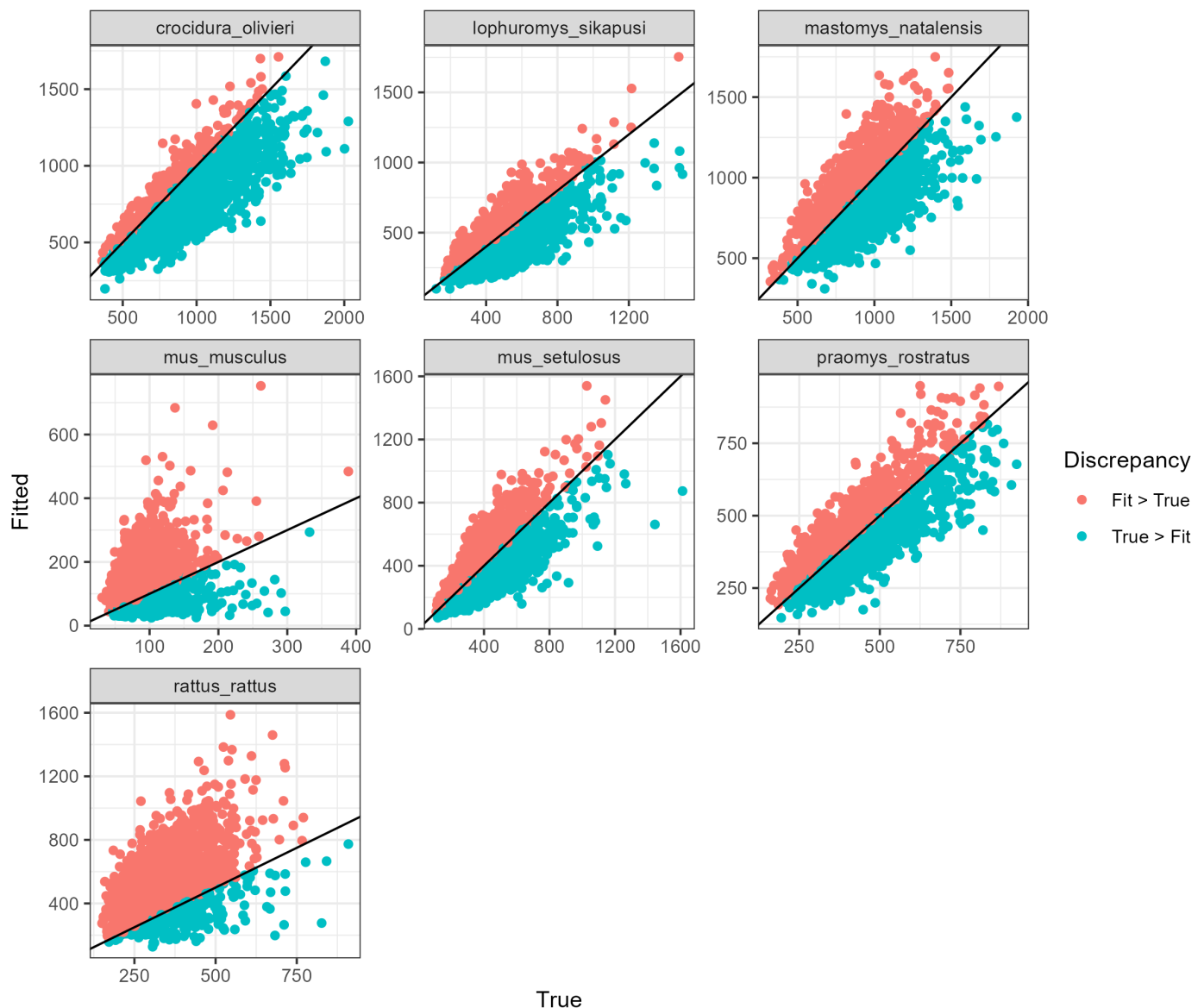
"It is possible that the observed correlations in the probability of species co-occurrence are not driven solely by biotic interactions and that gradients of not included abiotic factors may be associated with changes in small-mammal co-occurrence. For example, if elevation is important to the occurrence of *M. natalensis* but not *M. musculus* and this was systematically different between landuse types this could potentially produce the observed effect of negative associations of co-occurrence between these species."

- ✓ On p78, clarify if *Mastomys* was included in the cited study.
  - I am not too sure which citation is being referenced here. Apart from Ostfeld and Holt (rodents generally) all cited studies include some assessment of *Mastomys natalensis* either individually or as part of a meta-analysis of host vs. non-host species.
- ✓ Discuss how sample size may have affected multi-species occupancy model and challenges of model validation
  - The following section has been added to 4.5.3 below the more general limitations of the study

"The Bayesian approach used to model species occupancy allowed for the propagation of uncertainty throughout the analysis, however, limitations due to small sample sizes remain. First, small sample sizes in some landuse types and for some species likely increase the uncertainty in modelled parameter estimates, impact the precision of the model. Model checks, including Bayesian p-values at the community (0.5) and individual species (0.16-0.86) level suggested adequate model fit and sample sizes but it would be informative to validate these results on studies conducted elsewhere in Sierra Leone or the same sites over a longer time period. Posterior predictive checks were used for further model validation. Results from this suggest a greater discrepancy in the fit statistic for *C. olivieri*, *M. musculus* and *R. rattus* than the four other species (Appendix C.4). It is likely that the differences in sample sizes, from a more limited number of landuse types and sampling sites led to a greater impact on model accuracy for these species than others. Finally, the threshold of 25 detections used for including a species in the model may influence inference drawn from a model which does not include present, but relatively rare species. However, an observation of improved Bayesian p-values for species with more than 40 detections supports the approach of not including rarely detected species."

## Appendix C.4

## Comparison of fit statistic for the observed data (true) and fitted model

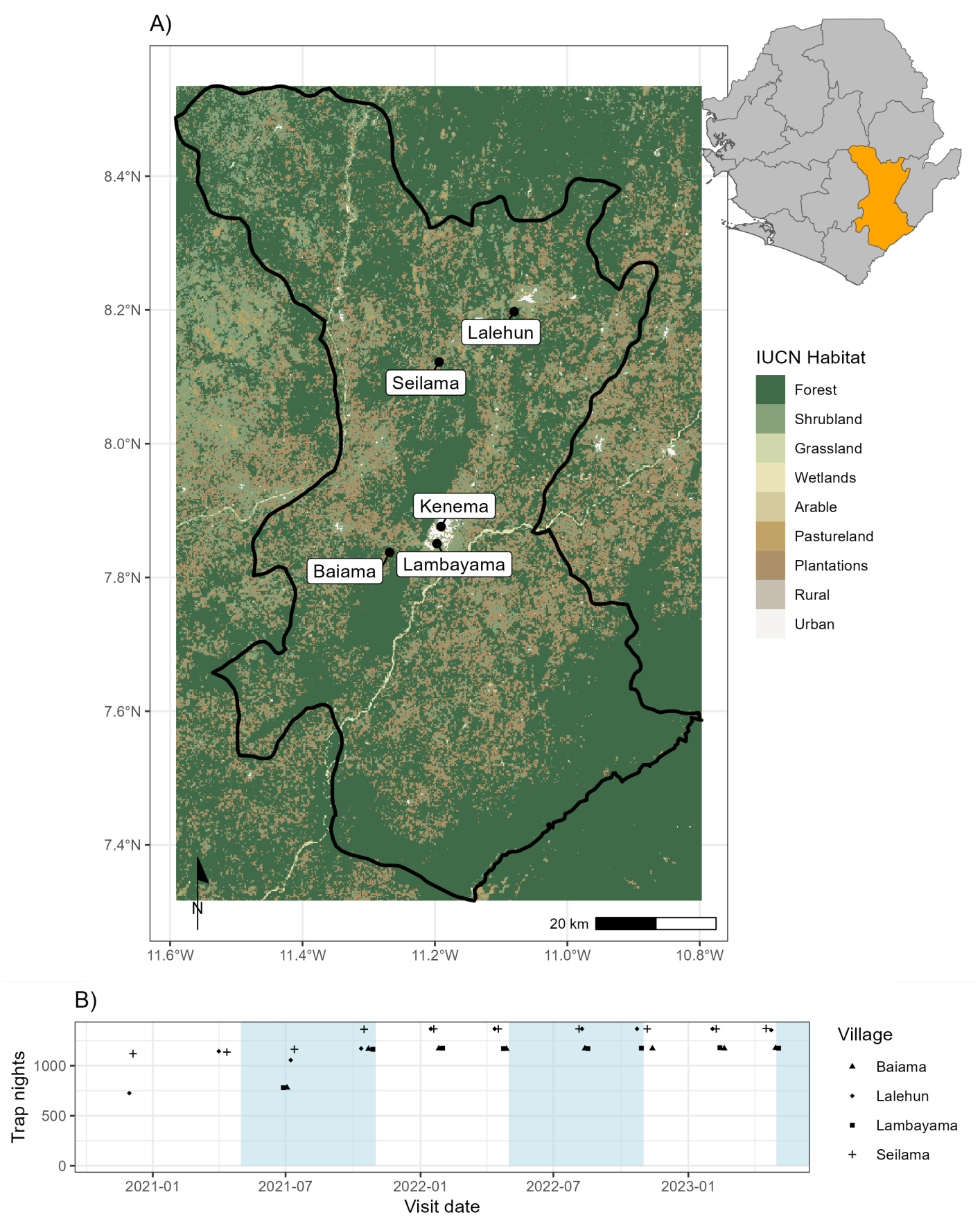


Posterior predictive checks of the final Bayesian multi-species occupancy model. The fit statistic for the observed data and fitted model is shown. Pink points represent a positive discrepancy in the fit statistic with blue points representing a negative discrepancy in the fit statistic. Fit statistics clustered around the black line (representing no discrepancy) indicate better model validation to data.

- ✓ should the title be 'small mammal' rather than 'rodent'?
  - Changed to small mammal as appropriate throughout
- ✓ Fig 4.3 Labelling is confusing. Should the 'village only' data have different colors and labels?
  - Unclear what change is requested. Fig 4.3 shows the probability of occupancy of different small-mammal species occurrence in three landuse types, forest, agriculture and village. The output of the model is stratified by whether these landuse types are located in rural  $<500$  people per  $\text{km}^2$  or peri-urban  $>500$  people per  $\text{km}^2$  settings. The labels and colours appear appropriate, to me, to convey this.
- ✓ Mention temporal dynamics in the discussion
  - I think this has already been covered in the following paragraph in the discussion:

"The finding of similar or increased prevalence of *M. natalensis* between seasons, not accounting for imperfect detection, were consistent across village study sites. It is possible that in this region different agricultural processes or food storage practices by village communities results in different rodent behaviour to elsewhere in its range (Kelly et al., 2013, Leach et al., 2017). Alternatively, increased trap-shyness during periods of increased abundance may mask replication of previous findings. Additional rodent community studies, incorporating local human

- ☒ Replace/add a raster map of habitat type to figure 4.1
- This has been updated to a raster background layer.





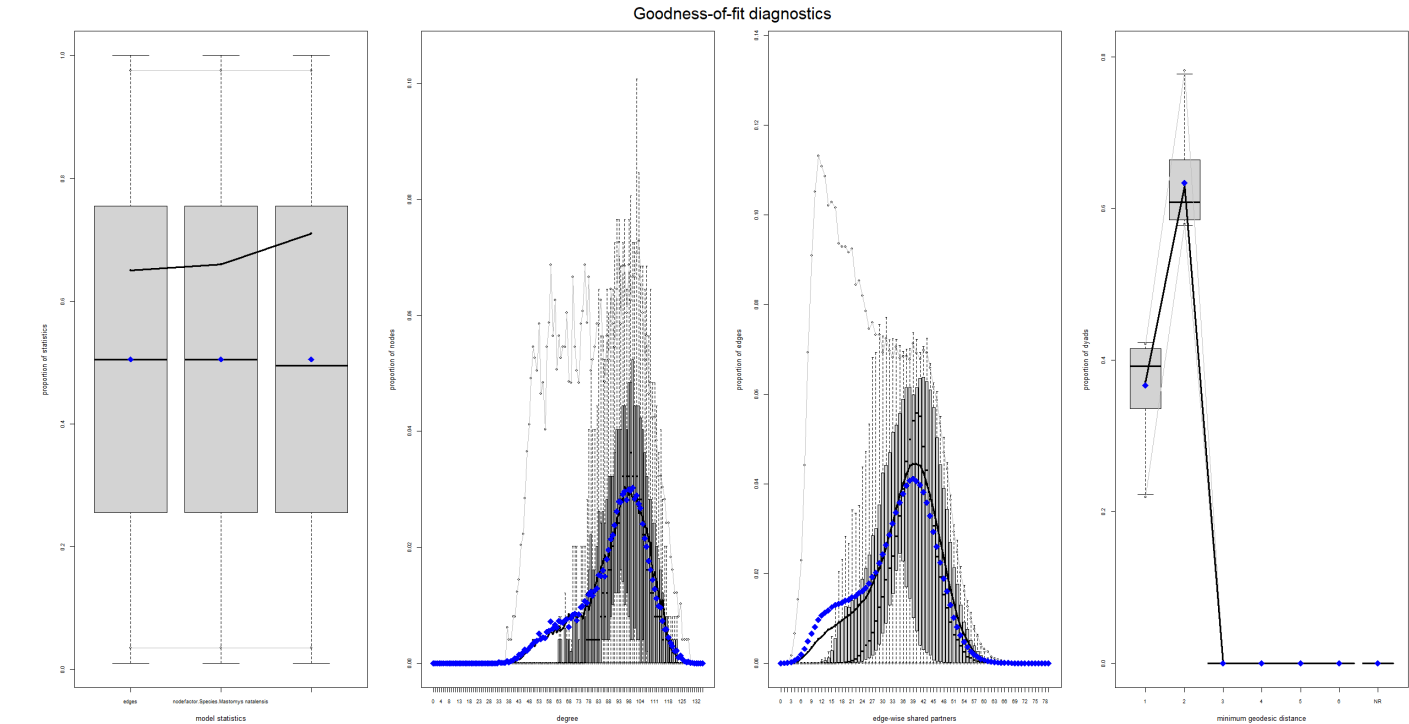
similar capture rates or discussing what % of posteriors diverged from the prior distribution according to a defined cutoff in distribution overlap or Kolmogorov Smirnov tests

Consider a grid-level (rather than grid cell level) analysis of the correlation in species occurrence. This seems consistent with the aims of the analysis (occupancy is more relevant at the grid level than the grid cell level), would reduce p inflation from having massive sample sizes, and might enable incorporation of uncertainty if pooling posteriors for a given species across a grid and sampling from that distribution iteratively.

I appreciate these comments and think that including these suggested changes will improve the final manuscript. They have not been implemented for this version of the thesis.

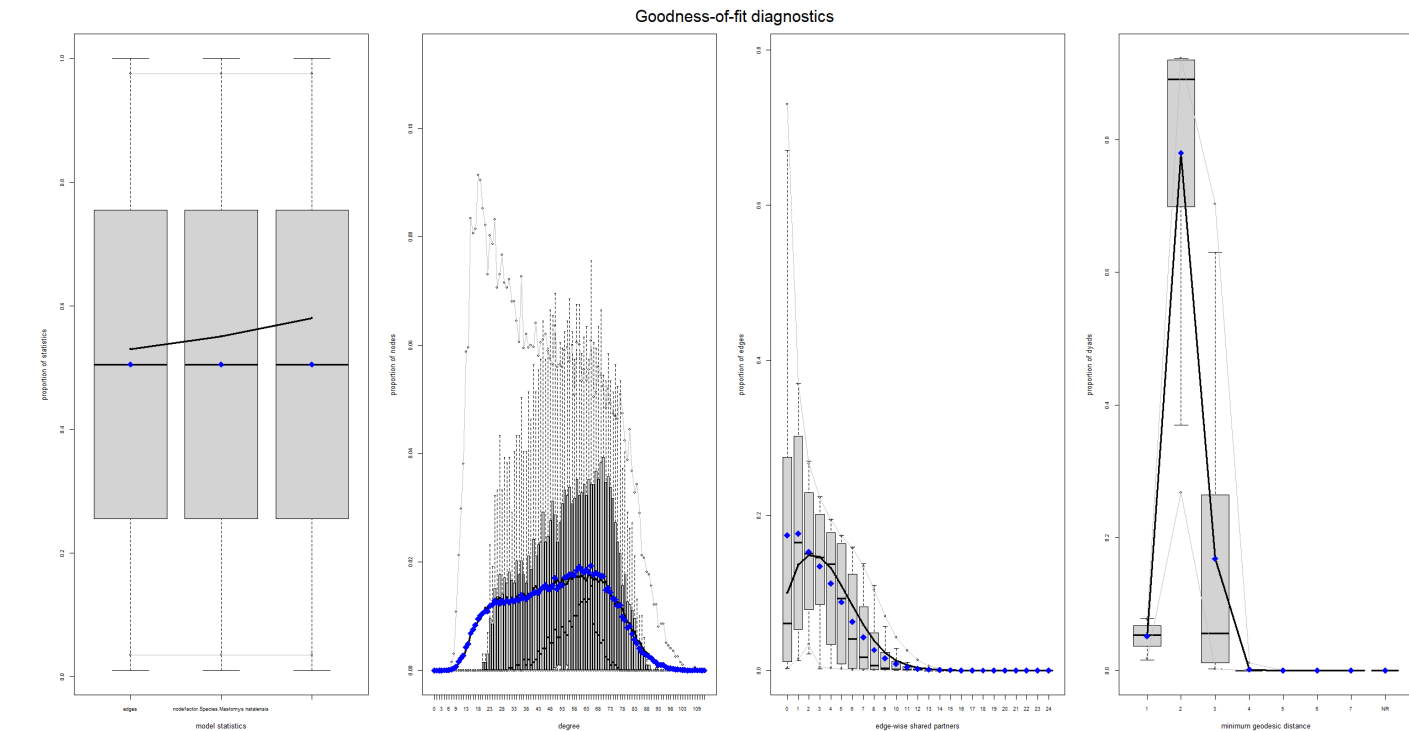
## Chapter 5

- ☒ Add discussion of cross-reactivity, especially LCMV and how this influences conclusions
  - An additional complication to the interpretation of the results of the sero-epidemiological analysis arises from potential cross-reactivity to other arenaviruses within the small-mammal communities. In the current interpretation I have assumed that LASV is the only circulating pathogen within these communities that can lead to detection of antibodies using the BLACKBOX ELISA assay. This may not be reasonable, despite a reported sensitivity and specificity of 95% and 100% respectively compared to an Immunofluorescence Assay, other circulating arenaviruses (e.g. Lymphocytic choriomeningitis virus and Mobala virus) may lead to cross-reactivity and false positive results for our assay (Soubrier, et al. 2022, Zapata, et al. 2011, Simo Tchegnana, et al. 2021). Further work is required to investigate the impact of other circulating Arenaviridae on calculations of LASV seroprevalence.
- ☒ Discuss value/risk of using inferred abundance in network models– recommended for future studies?
  - Typically, all nodes and edges within a network are not explicitly observed, therefore, statistical models developed from observed data require a measure of the proportion of all possible existing nodes and edges to aid inference (Keeling and Eames, 2005). In this chapter I achieved this by estimating the total population size of all detected small-mammals, based on spatially replicated counts, within the spatial confines of each produced network (Royle, 2004). These estimations were derived from point-prevalence sampling and therefore, do not take into account the population dynamics of small-mammal communities expected in these settings (Fichet-Calvet et al., 2010, Leirs et al., 1996). Further, for species with few detections there is a wide ranging distribution for the estimated abundance. The use of these estimates of total population size are expected to improve inference from these network models although an accurate estimate of the total population size for each detected species at the time point relevant for each network would be ideal. Improving confidence in the estimates of abundance through capture-mark-recapture studies would provide an estimate of the proportion of individuals being modelled from within the entire population.
- ☒ Add explicit statements on how confidence intervals in meta-analysis help evaluate fit of network models
  - The fit of each network model was inspected prior to inclusion in meta-analysis. The confidence intervals produced for the meta-analysis are based on the standard error of the effect size from the ERGM and so do not directly assess the fit of the network model, but do take into account the variability of the effect size from these models. The evaluation of the fit of the network model is through the goodness-of-fit assessment. It is challenging to produce a quantitative assessment of these and graphically would take up a lot of space. I have included two examples below to demonstrate the output of the GOF assessment for a model that had relatively poor fit (village 4) and relatively good fit (agriculture 3).



Goodness-of-fit tests for the village 4 model

For GOF assessments 100 samples of the network are produced for the model and compared. Blue diamonds represent expected values for the different metrics based on the underlying data informing the model. Black lines are the modelled output. Output falling outside of the errorbars represent a statistically significant deviation in the fit-statistic.



Goodness-of-fit tests for the agriculture 3 model

- Point estimates and standard errors for the odds of a contact, odds of an inter-specific contact and odds of an intra-specific contact derived from the network models were included in meta-analysis. The wide confidence intervals shown in the forest plots of the random effects meta-analysis suggest that the small sample sizes limit the inference that can be drawn. To assess whether any individual model was having an outsize effect on the direction of the association of the model outcomes I used several sensitivity analyses.
- ☑ Include sensitivity analyses as an appendix to the thesis
  - These have been added and presented in greater detail in the results section (5.4.4).

- The following has been added to the discussion.

"Sensitivity analysis 1 repeated the network construction and modelling based on a contact range of 15m and 50m to supplement the 30m used in the primary analysis. Results of the random-effects meta-analysis for these two other scenarios are shown in Appendix C4. I found that for the 15 meter scenario headline results remained the same. Differences relate to magnitude of the effect sizes of reduced odds of a contact being observed for *M. natalensis*, reduced odds of an inter-specific contact and increased odds for an intra-specific contact. Similar patterns but with reduced effect sizes were seen for the networks and models for the 50m scenario (Appendix C4). However, for this sensitivity analysis the odds of an interspecific contact being observed for *M. natalensis* were more similar to those seen for other rodent species and this was similar in both agricultural and village settings. Despite this, the probability of an intra-specific contact within *M. natalensis* remained high (5.6 OR, 95% C.I. = 3.2-9.6,  $p = <0.001$ ) in agricultural settings but not significantly different in village settings (1.8 OR, 95% C.I. = 0.97-3.3,  $p = 0.06$ ). These sensitivity analysis suggest that the inference of increased intra-specific contacts among *M. natalensis* in agricultural settings is not sensitive to the selection of defining contacts at 15m, 30m and 50m. Although the use of these cut-offs for defining a potential contact has an effect on the frequency of contacts and proportion of contacts that are inter-specific. This is expected and highlights that species-specific and ecologically plausible contact definitions are important for inference (i.e., using a species expected home range).

In the second sensitivity analysis I investigated the role of the constituent networks and models on the final effect sizes calculated through random-effects meta-analysis. This was only performed on the networks and models of the primary analysis (i.e., 30m definition of a contact). The results of these sensitivity analyses suggests that no individual network or model is having an outsize effect on the results. This was expected due to the incorporation of modelled variance in effect sizes being incorporated into the analysis. Agriculture 2, relating to the sampling in Agricultural settings during visit 2 had the largest impact on the final effect sizes, removing this network had no overall effect on the direction of effect size but modified the magnitude by -2.2 on the Odds Ratio scale. All other networks had a lesser impact on the final results."

- ☑ P101. Hypothesize that intra>inter specific contract rates in human dominated settings, but don't say what the hypothesis is for other settings. Also presumably intra>inter. This hypothesis should be better framed so the differences between habitat types are identified.
  - I would have expected inter-species interactions to be as common as intra in more diverse settings or at least more common than in human dominated landscapes where anthropogenic barriers to interaction may be present. I have added the following to the last section of the introduction.

"I hypothesised that spatial clustering of conspecifics and the increased abundance of commensal species in anthropogenically dominated settings will result in greater intra-specific contact rates compared to inter-specific contact rates within these communities. Conversely, I hypothesised that in agricultural or forest settings where small-mammal species richness and abundance is potentially increased inter-specific contacts will predominate."

- ☑ P101. Mention the number of grids per land type in the main text as this might be perceived to be equal
  - This has been added to section 5.3.1.
- ☑ P106. only 13 networks total were retained. Provide proportions of possible networks that were kept for ag vs village.
  - This has been added to section 5.3.5
- ☑ That's a lot of data loss. Might be worth a few sentences in the discussion over whether models were too complex for the sparse data?
  - I have added a section to the discussion.

"Thirty-two networks could potentially be constructed from the trapping design. However, due to low detection rates of small-mammals in these settings 13 networks were not included for modelling as they contained no individuals (3) or no individuals of the species of interest (10). Of the remaining 19 networks, six were not included in the random effects meta-analysis as no estimates for the homophily term could be produced as few individuals of the species of interest were detected during the trapping session. Given the limited sample size and prior expectation that it would not be possible to explore time-varying changes in rodent contact rates, a random effects



meta-analytic approach was chosen. This prioritises a simpler model, limiting overfitting as may occur in a mixed-effects model. It is not expected that model complexity alone was the limiting factor in the number of networks included in the random effects meta-analysis. The species interactions modelled are for the most commonly detected species. Networks not contributing to the meta analysis were uninformative and tended to have low sample sizes with a null model not able to converge for the six excluded models."

(For peer reviewed publication)

Sampling methods: be sure you include things like dosages of anesthesia and how exactly blood samples were collected; any buffer used to preserve samples? What filter paper was used for blood spots, what was the negative control for the ELISA? Did you use a positive control? etc.

P117 - paragraph on virus extinction seems a bit arbitrary. Can you make the exact opposite argument that high contacts are necessary to maintain LASV in agricultural settings and it is bound to go extinct from villages due to low contact rates? Serology might help work this out - expect more variability in the setting that is more prone to virus extinction.

L118 - might mention automated RFID systems or contact tracing tags (see work by Sarah Knowles, Lucy Kirkpatrick) and discuss whether you can get permission to release them. Is radiotelemetry going to be more specific than what you did?

Thank you for the above comments. These will be addressed alongside submission of the manuscript.

Chapter 6+ appendices

- ✓ 6.1 Discuss the advantages and disadvantages of having an unbalanced study design
  - The following has been added.

"I was interested in using the findings on small-mammal ecology in Sierra Leone to better understand the hazard to local human populations from LASV spillover. Due to this I designed the sampling effort to be relative to human exposure risk (i.e., a combined measure of landuse type and human population counts) rather than sampling relative to landuse type or equivalent sampling effort across landuse categories. I anticipate that this improves our understanding of potential contact between humans and the hosts of LASV based on species occupancy and points of contact within the small-mammal community. This study design may limit the usefulness of these data for other purposes where this design may represent a source of bias. "

- ✓ 6.2. Ensure that all figures have captions. Make sure that figures are referenced in the main text.

- I was unable to find the non-pdf version of the images in Appendix B.1 and so have removed this appendix from the thesis. I have removed reference to it in the main text.

- ✓ Several typos across the thesis: 2.4.1 (of); p66 (effect, affect), p78 (missing a zero, extra '-'), p83 (incomplete sentence ",where"), p86 (extra period), p103 (missing period), p104 (networks)
  - Corrected throughout.