

A protocol for biodiversity-informed wildlife disease surveillance

Michael D. Catchen¹, Francis Banville¹, Amélie C. Boutin², Cole B. Brookson³, Colin J. Carlson³, Gabriel Dansereau¹, Rory Gibb⁴, Marianne Houle¹, Benjamin Kaza^{1,5}, Hailey Robertson³, David Simons⁶, Stephanie N. Seifert⁷, Timothée Poisot¹

¹Université de Montréal, Département de sciences biologiques, Montréal, QC, Canada

²Carleton University, Ottawa, ON, Canada

³Yale University School of Public Health, Department of Epidemiology of Microbial Diseases, New Haven, CT, USA

⁴University College London, Department of Genetics, Evolution and Environment, London, UK

⁵Cornell University, Ithaca, NY, USA

⁶Pennsylvania State University, Department of Anthropology & Center for Infectious Disease Dynamics, State College, PA, USA

⁷Washington State University, College of Veterinary Medicine, Pullman, WA, USA

Corresponding Author: michael.catchen@umontreal.ca

Abstract

Land use and climate change are increasing the risk of spillover of zoonotic disease into human populations. However, we lack actionable information about the prevalence of pathogens in wildlife populations for most of the globe, challenging our ability to implement strategies to prevent zoonoses. Even when this data exists, it has historically been sampled opportunistically and without guidance based on known geographic distributions of hosts of zoonotic pathogens. Biosurveillance is essential to mitigating zoonotic spillover risk, but given the expensive nature of monitoring pathogens in wildlife, we need to be strategic about deciding where and what to sample to obtain as much useful information as possible. The field of biodiversity monitoring has established many practices that can directly inform optimal biosurveillance efforts. One such concept is the Biodiversity Observation Network (or BON), which aims to select monitoring locations that most effectively and efficiently capture the status and trends of biodiversity. We present a protocol for integrating data on host biodiversity into sampling priority for wildlife disease surveillance based on host species distribution models, with optional potential to integrate pathogen prevalence data (if available). This protocol has the flexibility to target different forms of sampling (collecting host occurrence vs pathogen prevalence data) to adapt to different levels of data availability, but still makes adaptive sampling recommendations based on a principled understanding of host distribution and pathogen biology. We illustrate this flexibility with two case studies, prioritizing sampling for *Hanta-* and *Arenaviridae* in rodents in India and South Korea, respectively representing data poor and data rich contexts. We view this framework as a basis for integrating long-term biosurveillance and biodiversity monitoring programs, and maximizing the useful information available for public health decision making.

Keywords: disease ecology, One Health, biosurveillance, biodiversity

Introduction

Despite the well-established links between biodiversity and the potential for zoonotic pathogen spillover, wildlife disease surveillance programs rarely integrate biodiversity data to guide sampling efforts. The relationship between biodiversity and disease risk is complex, as both biodiversity itself and biodiversity loss can influence the presence of pathogens in context-dependent ways (Halliday et al. 2020, Keesing and Ostfeld 2021, Carlson et al. 2025). Host and pathogen diversity are entwined through ecological and evolutionary processes (Johnson et al. 2015, Stephens et al. 2016, Plowright et al. 2017, Carlson et al. 2025), which together shape zoonotic spillover risk by affecting the prevalence of pathogens among wildlife that humans may encounter and from which infections may arise. Climate and land use change exacerbate this risk by modifying the geographic ranges of hosts of

zoonotic pathogens, leading to increased potential for cross-species transmission (Carlson et al. 2022a), including spillover into human populations.

As a result, monitoring infectious diseases in wildlife is an important goal of several international agreements, which all fall under the One Health approach. The interlinkages between biodiversity and health are recognized in the Global Action Plan on Biodiversity and Health from the Convention on Biological Diversity (CBD) and the Pandemic Agreement of the World Health Organization (WHO). Specifically, the CBD encourages Parties to “[reinforce] planning and surveillance of biodiversity, including for wildlife habitats and zoonotic pathogen spillover risk, to better assess and address health and disease risks in order to manage wild species sustainably” (Convention on Biological Diversity 2024) and the each party to the WHO agreement must “[coordinate] multisectoral surveillance to detect and conduct risk assessment of emerging or re-emerging pathogens with pandemic potential” (WHO 2025). To meet this goal, countries must design monitoring systems that efficiently allocate resources to maximize their knowledge of zoonotic disease prevalence and spillover risk. Improving spillover risk estimation therefore requires a multifaceted approach that incorporates biodiversity surveillance with data on landscape change and host ecology – including shifts in reservoir species’ habitats, community composition, and geographic distributions (Bell et al. 2025).

The field of biodiversity monitoring has established many practices and tools that can directly inform biosurveillance efforts (Poisot et al. 2025a). One such core concept is the Biodiversity Observation Network (or BON): a set of monitoring locations designed to best capture the status and trends of biodiversity (Scholes et al. 2012). Ideally, BONs establish a set of spatial locations at which biodiversity data are collected in standardized formats, which can then be easily aggregated for the detection and attribution of biodiversity change (Gonzalez et al. 2023a) and be used to inform decision-making at local, regional, and ultimately global scales (Gonzalez et al. 2023b). Here, we explore how the BON perspective would be useful to prioritize locations for which wildlife disease surveillance could be maximally informative, particularly when disease data are scarce.

Given the cost of effective biosurveillance, strategic allocation of resources toward where, when, what, and how much to sample is imperative to ensure sampling effort yields new and useful information. Efforts have been made to think about biosurveillance sampling prioritization from both a statistical (Farver et al. 1985, Nusser et al. 2008) and spatially explicit perspective (Andrade-Pacheco et al. 2020, Dumelle et al. 2022). Here, we develop a context-dependent protocol for adaptive sampling to maximize the reduction of our uncertainty of the status of a given pathogen or host. The type of data (i.e. disease prevalence among a population of susceptible hosts, or occurrence records of host species) useful for a particular context depends on existing data availability and what drives our uncertainty of the hazard level of zoonotic pathogen prevalence. To optimally allocate sampling effort, both for choosing sampling locations and the type of data sampled, it is necessary to prioritize resource allocation towards both improving estimates of host presence and pathogen prevalence within host species. However, the locations for efficient sampling for these two goals are intrinsically different: to improve host presence prediction we should go where we are most unsure about the host’s presence, whereas to improve prevalence estimation we should rather sample where we are already sure hosts are.

Here, we establish a set of guidelines to use biodiversity data to prioritize sampling locations for wildlife disease surveillance. Notably, we select these sampling locations by accounting both for the distribution of biodiversity and current uncertainty about this distribution, as well as utilizing existing prevalence data if available. As applying this framework requires numerous decisions, we provide guidance about where it can be adapted to reflect local priorities, sampling constraints, and the availability of existing data on species presence and the prevalence of pathogens among host

species. We present case studies illustrative of this framework in both data-poor (no prevalence data) and data-rich (spatially explicit estimates of prevalence) contexts on *Hanta-* and *Arenaviridae* in rodents.

The Role of Biodiversity Data in Biosurveillance

Conceptualising zoonotic spillover requires distinguishing *hazard* – the active circulation of a pathogen in a host reservoir in a state conducive to transmission to humans – from *risk*, which is the probability of that hazard being realised via spillover (Hosseini et al. 2017; Gibb et al. 2020a, 2025). The intensity of this hazard fluctuates with shifts in host distribution and abundance, infection prevalence, and pathogen biology, all of which occur independently of human presence, and thus the proximity of reservoir hosts to human populations does not alone constitute a significant risk.

Spillover risk is actualized when a susceptible human is exposed to a zoonotic disease hazard, through processes governed by sociological, behavioural, and economic drivers at the human-wildlife interface (Gibb et al. 2024, Friant et al. 2025). For instance, domestic or agricultural activities in environments contaminated by animal excreta elevate risk, whereas mitigations like improved sanitation or hygiene can reduce it towards zero. Monitoring the underlying hazard is critical because the conditions for exposure are not static. Anthropogenic changes, such as urban encroachment or ecological shifts, like climate-driven species migration, can rapidly intensify human-wildlife contact, transforming a latent hazard into an immediate risk (Daszak et al. 2000, Gottdenker et al. 2014, Gibb et al. 2020b, Carlson et al. 2022a, Eby et al. 2023).

Biodiversity data is also critical – but underutilised – for predicting probable hosts and geographic ranges of as-yet-undescribed pathogens which may pose future threats to health. A small proportion of mammalian viruses have been identified (Carlson et al. 2019) and most mammalian species have been poorly sampled with respect to viruses, with current patterns of viral diversity largely reflect historical priorities in discovery effort (Gibb et al. 2022). This means that the zoonotic hazards posed by unknown viruses that circulate in wild animal populations cannot be directly estimated. However, as we describe below, using models to predict the likely hosts of zoonotic pathogens prior to their emergence (Han et al. 2015, Becker et al. 2022, Cummings et al. 2025) can help to alleviate the uncertainty created by this viral “dark matter”.

When is biodiversity data relevant to guiding disease surveillance?

The diversity of host communities is a central determinant of pathogen dynamics and prevalence (Dobson 2004, Keesing et al. 2006). One of the leading ecological determinants of spillover risk is the density of infected hosts in a given location across all possible host species (Plowright et al. 2017). Unfortunately, this itself is often not actionable because it relies on spatially explicit data about abundance of – and pathogen prevalence in – all relevant hosts, which is rarely (if ever) available. Therefore, estimates of spillover risk are limited both by lack of knowledge on disease dynamics and the distribution and abundance of host species.

To circumvent this common data limitation, we rely on a proxy that uses weighted estimates of local host species composition to prioritize locations for sampling and identify the type of sampling to conduct. While estimates based only on the predicted presence of host species do not fully capture the risk of disease transmission, they hint at the potential for pathogen presence, especially if different species can be weighted by their relative importance as reservoir hosts, i.e. accounting for intraspecific variation in host competence, population density, and probable role in pathogen maintenance.

By contrast with a situation where all species are assumed to be equally important, prior knowledge about the relevance of different hosts for pathogen transmission can be incorporated in these models as a relative weight of host sampling priority. This data can come from phylogenetic similarity to known hosts, existing records of infection, or expert knowledge on the system in question (Tseng et al. 2025). In the extreme case where there is no prior information to establish weights for host species, this approach lends itself to an adaptation of the well-established practice of estimating local host diversity by stacking predicted species distributions (Thuiller et al. 2015), which is particularly appropriate when using quantitative predictions of species presence (Grenié et al. 2020). Although stacking species distribution models (SDMs) is often biased towards higher estimates of local species richness (Calabrese et al. 2014), in our context this bias is unobjectionable as it leads towards more cautious recommendations that in the worst case overestimate the ranges of reservoirs, and the stacking of species-specific SDMs has been shown to produce better estimates of richness compared to joint (multi-species) SDMs (Zurell et al. 2020). For a full set of guidelines for interpreting the outputs of SDMs, see Box 1.

Box 1: Guidelines for the interpretation of Species Distribution Models

Species distribution models (SDMs) are widely used tools in ecology that relate species occurrences with environmental variables to identify how species are geographically constrained by different ecological factors (Elith et al. 2011). SDMs can then be used to estimate the realized ecological niche of a species – the environmental space (rather than geographic space) that is suitable for the species to occupy. Because occurrence data is often presence-only, meaning there are few/no records of the verified *absence* of a species, the maps generated from SDMs should be interpreted as the relative environmental suitability for a species in a given location compared to other sites, rather than a probability of occurrence at a particular site – though there are methods that allow for this translation (Smith and Levine 2025).

In disease ecology, SDMs are often used to understand the current (and future) distribution of wildlife hosts and vectors (e.g. Lippi et al. (2023), Kopsco et al. (2022), Bussières-Fournel and Poisot (2025)) and determine which areas have the potential to carry spillover risk. While SDMs can help in planning and management decisions (e.g., sampling prioritization), their interpretation can prove challenging for end users (Guisan et al. 2013). For example, predictions for each species (e.g. in a multi-host system) are frequently stacked (i.e. the suitability scores for each species are summed at each location) to understand their overlapping distribution. However, SDMs do not typically incorporate biotic interactions (for discussion of joint species distribution models, which incorporate co-occurrence data and are generally used to infer relationships rather than create spatially-explicit predictions, see Pollock et al. (2014) & Wilkinson et al. (2021)), which means they cannot predict species abundance or the definitive interactions between species within a pixel where species are predicted to co-occur (Blanchet et al. 2020). Without the influence of other species and other latent variables, SDMs likely overestimate the “true” suitable area (false positives) and overall number of co-occurring species. Additionally, SDMs typically use predictors that are treated as functionally static in time (e.g. WorldClim bioclimatic variables), which bakes in the assumption that species are in equilibrium with their environment (Milanesi et al. 2020). In reality, species distribution patterns likely shift through time, varying with seasonal patterns, ecosystem disturbance, and resource availability (Milanesi et al. 2020). Thus, when incorporating SDMs into a decision-making process, users must consider the plausibility of predictions and inherent uncertainty in the context of their particular expertise.

Further complicating the interpretation of SDMs are the many decisions made during data selection, pseudo-absence generation, model specification, and evaluation that proliferate through the outputs and generate errors (Barry and Elith 2006, Merow et al. 2013). Even a poorly designed model can successfully be trained and generate predictions that seem plausible, but are ultimately as unreliable as the underlying model (Fourcade et al. 2018). Thus, there’s no replacement for collaboration between modelers with technical expertise and local decision-makers, with a clearly defined purpose and standardized reporting and documentation (Guisan et al. 2013, Araújo et al. 2019).

What are host distribution models telling us about pathogens?

The circulation of pathogens (such as viruses or obligate intracellular bacterial infections) through an animal population can vary significantly over time and is influenced by host species behavior, immunology, and environmental factors, among other drivers. Over long time periods, pathogens are thought to circulate stably within their primary host populations throughout a landscape according to the contacts of individual members of a population. These pathogens can also circulate through other susceptible animal populations if the two competent animal populations overlap in the landscape and the pathogen is shared.

In places where no susceptible or potentially susceptible hosts are present, it is not possible to find infected individuals. Thus, the extent of a pathogen can be thought of as the aggregation of ranges for each host species, and a rough estimate of pathogen suitability can be obtained based on weighting host susceptibility to infection ([Cao et al. 2023](#), [Redding et al. 2024](#)). The ensemble distribution then represents the landscape that the pathogen faces, which is not directly related to the physical environment, but is instead determined by the distribution of hosts in a landscape.

A Protocol for Biodiversity-informed Biosurveillance

Here we present a protocol for adaptive sampling of wildlife disease that adjusts sampling priority, both across space but also in terms of the type of data to collect (host occurrence vs disease prevalence), based on the availability of existing data. This provides a standardized framework for practitioners to decide where to sample, reevaluate priorities after data is collected, and design further samples to most efficiently inform the predicted status of zoonotic disease in wildlife.

Biodiversity dose, host weighting, and uncertainty

Throughout this manuscript, we refer to the estimate of local host composition that reflects both host richness and *a priori* estimates of host importance as the “biodiversity dose” — the ecological capacity for a location to maintain the pathogen (not a measure of the viral load in the environment). This is related to the existing concept of community competence, but distinct as the biodiversity dose is calculated as a *weighted* average of SDM habitat suitability scores. Dose therefore captures both the likely presence of host reservoirs and their relative importance in disease transmission. Biodiversity dose is defined for a group of hosts, ideally encompassing the full set of hosts for a given pathogen, but also accounting for the subset of hosts of interest for a given monitoring program.

This still leaves the task of assigning species weights. If no information is available about species importance, a naive assumption of equal weights can be applied. If the capacity of a species to serve as a reservoir for a given pathogen is predicted with a model ([Becker et al. 2022](#)), the relative confidence in the prediction can be used as a weight. Hosts for a particular pathogen identified through methods with higher false-negative rates (e.g. serology vs. PCR) can be given lower weights. When there is uncertainty about the capacity of a host to be infected by a pathogen, methods that assign weights based on phylogenetic proximity to known hosts ([Elmasri et al. 2020](#)) can also provide usable information. Ideally, all aspects of the competence and relevance of hosts for disease transmission can be aggregated by experts to establish species weighting — we discuss the use of sensitivity analyses to better understand the consequences of these choices on the recommended configuration of the monitoring network in the final subsection of the protocol.

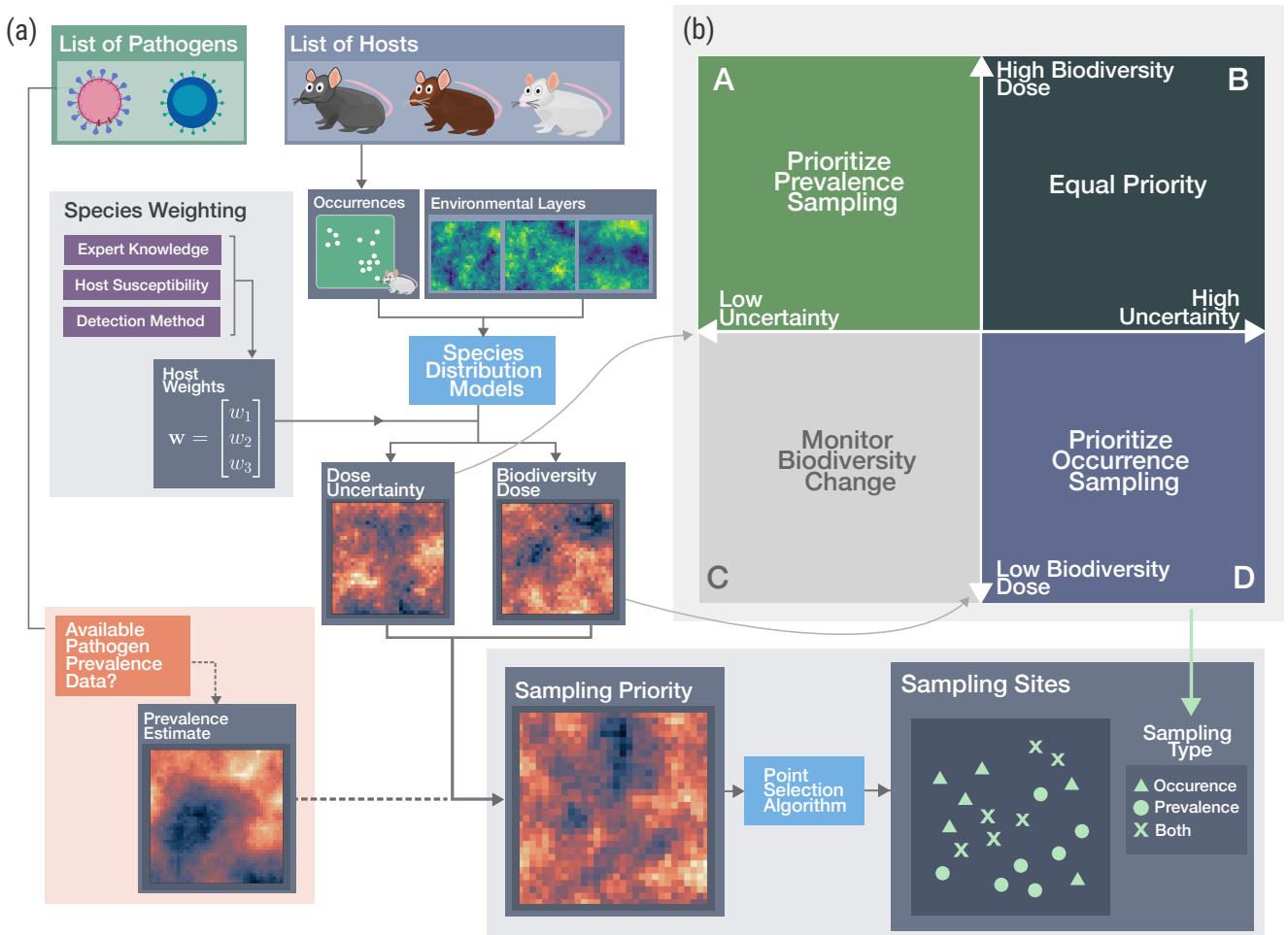


Figure 1: (a): The conceptual framework for optimizing sampling priority using host species distribution models, and pathogen prevalence data if available (left). **(b):** The type of sampling that is most informative (pathogen prevalence vs. host occurrence), depending on a location’s level of biodiversity dose (weighted host habitat suitability) and uncertainty about that dose.

Another crucial element for guiding sampling is the uncertainty about the distribution of the biodiversity dose, i.e. the weighted average of the uncertainty associated with the SDM for each host. There are several different ways to derive uncertainty from an SDM, and they tend to reflect different aspects of model fit – see Box 3 for more information on the various forms of SDM uncertainty that can be used. High uncertainty points to an increased priority for sampling host occurrence to better refine the prediction of host presence and overall biodiversity dose.

The information about biodiversity dose, dose uncertainty, and prevalence data (if available) can be used to make a map of sampling priority to decide where and what to sample – see Figure 1(b). It is critical to emphasize that while this diagram in Figure 1 is useful for planning the type of data to sample when prioritizing data collection, it cannot be used to infer risk or guide outbreak response. Instead, this protocol helps identify what type of sampling is most appropriate at a given stage of surveillance, based on host biodiversity distributions and assumptions about host importance. When the sampling objective is population surveillance or estimating pathogen prevalence within hosts, areas of high biodiversity dose (Figure 1(b), Quadrants A & B) should be prioritized. However, when the goal is to improve host occurrence data, sampling should focus on areas of high dose uncertainty (Figure 1(b), Quadrants B & D).

Based on the goal of sampling, we can combine dose and uncertainty predictions into an overall priority score P_x at each location x as

$$P_x = \sum_i w_i (w_d s_{ix} + w_u u_{ix})$$

where s_{ix} is the SDM score of species i at location x , u_{ix} is the uncertainty of this score, w_d is the weight for the biodiversity dose, w_u is the weight for the uncertainty such that $w_u + w_d = 1$, and w_i is the weight for species i , such that $\sum_i w_i = 1$.

Integration of prevalence data into the monitoring network design

When available, data on pathogen prevalence can refine sampling priorities, but the best way to use this information for sampling is contingent on the spatial and temporal scale of the prevalence data and the goals of sampling. For example, continued surveillance of populations known to have high prevalence may be useful if the goal is to determine if this population serves as a persistent reservoir for a given pathogen, or if it was undergoing a transient outbreak when the existing data was collected.

Alternatively, if existing prevalence data has high enough spatial coverage that it enables spatially explicit estimates of prevalence, sampling can most effectively improve spatial predictions of prevalence by targeting regions where the degree of prevalence is uncertain, but for which we are confident there is a high biodiversity dose (as we do in the second case study). This is useful for both precisely locating existing reservoirs across space to better map zoonotic hazard and spillover risk. For a full overview of potential methods to use depending on the scale of existing prevalence data see Supplemental Table 1, and for guidelines on how to target different forms of sampling regions based on their predicted dose, dose uncertainty, and predicted prevalence in Supplemental Table 2.

Another important consideration is the variability in what existing prevalence data is saying. For example, serological diagnostic tests indicate exposure to a pathogen *at any point in the past*, and do not directly indicate the individual's ability to transmit that pathogen, whereas PCR can detect both active infection and viral load as a measure of transmissibility. Together, these can provide complementary forms of information about both historical infection dynamics and the active prevalence. Fully representative sampling of wildlife populations is rarely feasible, so estimates of prevalence are obtained from a small subsample of the entire population of unknown size. Both of these factors induce error in the resulting prevalence estimates; however, there are numerous methods from population ecology for estimating abundance with imperfect detection (e.g. with N-Mixture Models [Royle \(2004\)](#)) and to infer population trends over time ([Kéry et al. 2009](#)). These can naturally be extended to the context of imperfect detection of both host abundance and pathogens within hosts in infectious disease modelling ([DiRenzo et al. 2019](#)).

Sampling Point Selection

Once we have a sampling priority map, we want to use it to guide sampling site selection. There are many options for point selection algorithms rooted in sampling theory, which itself has a well-developed theory of spatial sampling. A full review of spatial sampling theory is outside of the scope of this paper, but see [Wang et al. \(2012\)](#) and [Dumelle et al. \(2022\)](#), and for a comprehensive review of the general theory of sampling, see [Thompson \(2012\)](#).

A crucial point here is that because we are interested in targeting areas of high priority, we are constrained to a set of point-selection methods for unequal probability sampling, meaning each location in space can have a distinct probability of being included in the sample. A naive approach would be to draw samples with probability directly proportional to priority value, but due to the autocorrelation associated with biodiversity dose and uncertainty, this would result in many sampling points clumped close together, which would likely provide redundant information, and be an inefficient use of sampling effort. This is a well-established issue in spatial sampling, and as a

result there are many algorithms for selecting spatially balanced samples — for example: Generalized Random Tessellation Sampling (GRTS; [Stevens and Olsen \(2004\)](#)), the Pivotal Method ([Grafström et al. 2012](#)), and Balanced-Acceptance Sampling (BAS; [Robertson et al. \(2013\)](#)). These methods are all included in the `BiodiversityObservationNetworks.jl` package in Julia (which is maintained by the first author), which we utilize in the case studies.

Recent results suggest that most site selection algorithms achieve equivalent performance for the same problem ([Norman and Poisot 2025](#)), which allows users to pick an algorithm based on specific features, such as the support for auxiliary environmental data. For our purposes, we find BAS the most flexible method because it is both the most computationally efficient and achieves better spatial balance than the alternatives ([Robertson et al. 2013](#)).

In practice, algorithms for generating spatially balanced samples targeted toward regions of high priority may not be skewed “enough” toward regions of high priority for a given use case, so we also tend to tilt the distribution of priority scores to make the values of high priority even higher. This is done using a function $T(P_x, \alpha)$, defined as

$$T(P_x, \alpha) = \frac{\exp(\alpha P_x)}{1 + \exp(P_x)}$$

where P_x is the original priority at a given location x , and the parameter $\alpha \in (0, \infty)$ controls how much to tilt the adjusted map toward areas of high priority.

Another common method for designing samples is stratification. In the context of spatial sampling, stratification consists of splitting the spatial domain into discrete regions (corresponding with specific strata), which each contains a user-determined number of samples. In our case, stratification can be used to divide space into the different data collection regimes (sampling for host presences, prevalence or both; Figure 1(b)) and distribute sampling effort across these strata based on the goals of the monitoring program, with spatially balanced point selection methods being used within each stratum. For further guidelines for interpretation and practical considerations when using selected sampling points, see Box 3.

Sensitivity/Robustness Analysis

This protocol relies on several choices of weights based on the goals of sampling and *a priori* estimates of host relevance for the pathogen(s) of interest. Assessing the sensitivity of the priority map to weight choices is important to ensure the priority map is robust to small tweaks to the weights that do not reflect meaningful differences in host relevance.

For a given set of weights, \mathbf{w} , a natural way to assess the sensitivity of the overall priority is to examine the mean absolute change in the priority map, P , when a small random perturbation is applied to the weights. This gives us an interpretable metric of how sensitive the resulting priority map is at a given chosen weight value. We create a “nudged” version of the weights, $\mathbf{w}' = \mathbf{w} + \varepsilon$, by adding noise ε , which is a vector of i.i.d. $\text{Normal}(0, \sigma_{\text{noise}})$ of the same length as \mathbf{w} , and then renormalized (so \mathbf{w}' also sums to 1). We then construct the priority map, P , for the original weights, and the priority map P' from the “nudged” weights. The overall sensitivity, S , of the priority map at a given weight value \mathbf{w} can be computed as the sum of the absolute difference across all locations x :

$$S(\mathbf{w}) = \sum_x |P_x - P'_x|$$

If the selected weights for a given use case are close to high sensitivity regions in the space of possible weights, end users should consider the additional caveat that the selected weights are very sensitive to small changes, and ensure the selected weight values are reflective of meaningful

biological differences in host relevance. We provide an example of this analysis in the Supplemental Material, assessing the sensitivity of weights for the first case study.

Box 2: Guidelines for the refinement of various methodological steps

The workflow described in this manuscript involves many steps, and most of them call for decisions made by end users. In this box, we go through the different steps, and highlight key methodological considerations.

List of hosts: the list of potentially competent hosts can be assembled from biodiversity data (IUCN range maps, in-country checklist, GBIF data), or from past wildlife disease data (e.g. from VIRION ([Carlson et al. 2022b](#)), literature surveys, or previous surveillance programs). The list of hosts may be arbitrarily filtered or expanded to reflect local priorities.

Host weights: weights of hosts used in the biodiversity dose calculation should essentially serve as a quantification of their expected contribution to disease transmission, and will almost always be a compromise between strength of evidence (e.g. serology, PCR and pathogen isolation provide different levels of confidence in host species competence), and transmission risk (potentially estimated using phylogenetic similarity to well-known hosts). Weights can also ultimately attempt to capture the risk of transmission to human populations (e.g. by weighting synanthropic species higher).

Host SDMs: the usual recommendations about the training and validation for SDMs apply throughout this pipeline (see [Zurell et al. \(2020\)](#)), and in particular the choice of predictor data, spatial extent, the quality control of occurrences, and the selection of predictive variables are important.

Uncertainty: There are many ways to quantify SDM uncertainty. Some models (like Generalized Linear Models and the form of a Boosted Regression Tree we use in the case studies) have estimates of variance built into the model structure. Absent this, a common alternative is measuring the variance of the predicted suitability score at each site across many cross-validation folds, or methods using conformal prediction ([Poisot 2024](#)). An alternative is using Bayesian Additive Regression Trees (BART; [Carlson \(2020\)](#)), which directly obtains samples from parameters, and which can thereby be aggregated into uncertainty in predicted suitability.

Prevalence: Relevant factors for choosing what prevalence data to incorporate are the minimum number of individuals sampled, the time since data was collected, and the form of test used. These can each impact the reliability of the prevalence estimate, and therefore utility of prevalence data.

BON design: Many algorithms exist for spatially balanced sampling with unequal inclusion probabilities, and the ability of these algorithms to handle various forms of auxiliary data should be considered (see [Norman and Poisot \(2025\)](#)). It is necessary to make the typical considerations when planning sampling: the accessibility of a given site, the cost-effectiveness of sampling given how much time is required to reach a location, the ability to access private land, the sovereignty of indigenous land — these can all be used to further adjust the inclusion probability.

Sensitivity analysis: Assessing the sensitivity of the overall priority map to both selected species weights (as discussed in the section on sensitivity), as well as the weights toward uncertainty and dose are all key factors to ensure the priority map is robust to small changes in weighting. Further checks on the response of the overall priority map to the inclusion/exclusion of host species (particularly those for which there is no direct evidence they can host a particular pathogen) should also be considered.

Case Studies

To illustrate this approach, we consider two case studies on *Arena-* and *Hantaviridae* and their rodent hosts, using the global database compiled by (Simons et al. 2025). The first case study is on rodent hosts of *Arenaviridae* in India, and reflects the scenario where there is no prevalence data available. The second is for South Korea, where we have sufficient prevalence data on *Hantaviridae* to make spatially explicit predictions of prevalence and use this to guide sampling.

Creating Dose and Uncertainty Maps

For both case studies, we use the same SDM methodology outlined here. We emphasize that the specific approach to building SDMs is not the focus — sampling prioritization is only going to be as good as the SDM, and we present guidelines for the interpretation of SDM outputs in Box 1.

We obtain occurrence records for each rodent host species from the Global Biodiversity Information Facility (GBIF). As predictor variables, we use the 19 bioclimatic variables from CHELSA (Karger et al. 2017) at 1km x 1km resolution for South Korea, and from WorldClim (Fick and Hijmans 2017) at 2.5 arcminute resolution (approximately 5km x 5km at the equator) for India (to keep the raster size reasonable). We generate pseudoabsences (Barbet-Massin et al. 2012) using background thickening (Vollering et al. 2019), with a buffer around each occurrence record where no pseudoabsences can go (25km for India, 8km for South Korea).

Models were trained using SpeciesDistributionToolkit.jl (Poisot et al. 2025b) and EvoTrees.jl (Desgagne-Bouchard et al. 2025) in Julia. Specifically, we use a Boosted Regression Tree (BRT; Elith et al. (2008)) with a Gaussian loss metric, meaning the value of each node in the tree is fit to a Gaussian using maximum-likelihood estimation. Therefore, the BRT provides both a suitability prediction score, and an uncertainty value associated with it, for each pixel.

In order to standardize the dose values for each species, we compute the empirical cumulative distribution function (ECDF) on raw output predictions, ensuring all prediction layers are on the same scale.

Case Study 1: Designing a network without prevalence data

Our first case study considers the situation where we are primarily limited by a lack of prevalence data. This case study was inspired by Rodrigues et al. (1978), published 9 years after Lassa virus (*Mammarenavirus lassaense*) was first identified in Nigeria. In the immediate aftermath of the discovery of a virus of severe public health importance that circulates in hosts with widespread distributions, Rodrigues et al. (1978) were the (self-identified) first to sample for Lassa virus outside of Africa. At the time, Lassa itself was difficult to distinguish in clinical settings from other infectious diseases present in India (e.g. typhoid fever, malaria), and therefore it was not unreasonable to suspect it may be present in rodent populations in India.

Today we know that Lassa is confined to Western Africa, but this case study serves as an example for how we use biodiversity data to prioritize sampling in the context of a recently identified virus without existing reliable information about its geographic extent, prevalence, or known reservoir species. We construct a priority map for sampling using the host rodents tested by Rodrigues et al. (1978) : *Funambulus tristriatus*, *Hystrix indica*, *Rattus rattus*, and *Suncus murinus*. To generate a priority map for sampling, we follow the SDM procedure outlined above, and for the sake of example assign species weights at random.

Figure 2(a) shows a bivariate map of both the dose and uncertainty aggregated across host species. This highlights three large-scale regions of interest: the south, where the model is confident that many host species are present; the northwest, where the model is uncertain about many or all species; and the northeast, where the model is also relatively confident that host species are present.

We say a region in the top 50% of uncertainty for a given species is for “discovery sampling”, and a region in the top 50% of predicted suitability is for “prevalence sampling”. In Figure 2(b) we see the total number of species that fall into each of these categories across space. Figure 2(c) shows our priority map, with both w_u and w_d equal to 0.5 (reflecting equal priority toward discovery and surveillance). The white points reflect sampling locations generated using Balanced Acceptance Sampling applied to the logically tilted priority map with $\alpha = 5$ (see Sampling Point Selection subsection above). The teal markers reflect the four historical sampling locations from [Rodrigues et al. \(1978\)](#). Finally, in Figure 2(d), we see the host species that contributed the most to the priority score across space (computed as the maximum value of $w_i(w_u U_x + w_d P_x)$ at location x across species i).

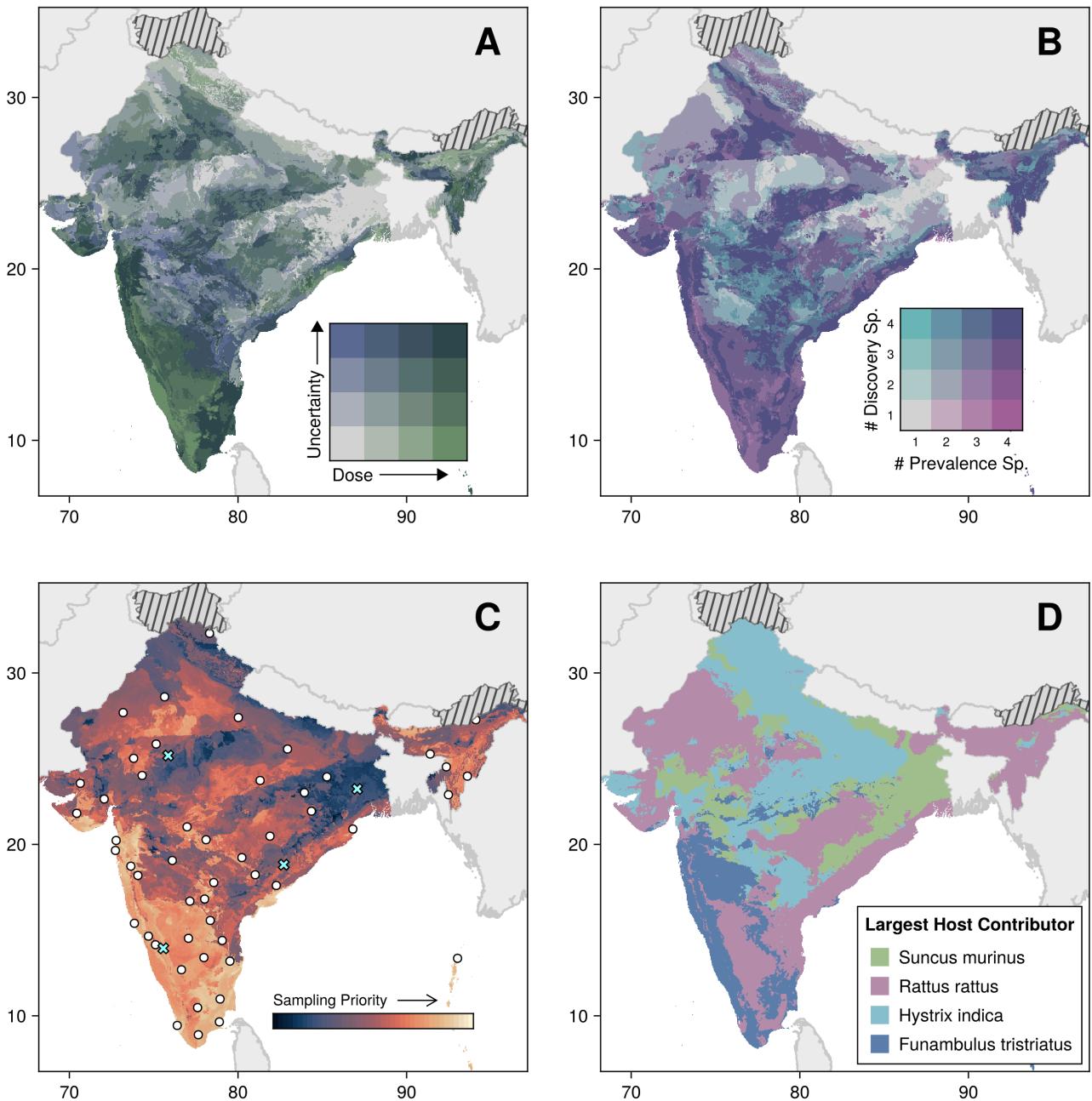


Figure 2: The case study for India. (a) Biodiversity dose + uncertainty bivariate plot. (b) Number of hosts in prevalence-regime (top 50% of habitat suitability for that species) vs. discovery (top 50% of uncertainty for that species), (c) Sampling priority + BON. X's are historical sampling locations from [Rodrigues et al. \(1978\)](#) (d) The largest host contributor to priority.

Case Study 2: Accounting for prevalence data in network design

In contrast to the prior example where very little is known about the prevalence of a given pathogen, we now demonstrate the protocol for the surveillance of two hantaviruses in South Korea (*Orthohantavirus hantanense* & *Orthohantavirus seoulense*), which are two of the main causative agents of hemorrhagic fever with renal syndrome (HFRS; [Park et al. \(2025\)](#)). Typically, the hosts of these viruses are rodents and shrews, and in our study we focus on three small-mammal hosts: *Rattus norvegicus*, *Apodemus agrarius*, and *Crocidura lasiura*. These host-virus pairs were selected as the database contained prevalence records across the country at a broad enough spatial scale to enable spatially explicit prediction of prevalence. The data comes from the following studies: [Gu et al. \(2011\)](#), [Kim et al. \(2017\)](#), [Kim et al. \(2007\)](#), [Kim et al. \(2020\)](#), [Klein et al. \(2015\)](#), [Lee et al. \(1978\)](#), [Lee et al. \(2017\)](#), [No et al. \(2016\)](#), [Park et al. \(2022\)](#), [Ryou et al. \(2011\)](#), [Sames et al. \(2009\)](#), [Seo et al. \(2022\)](#), and [Song et al. \(2009\)](#).

Prevalence across the extent is estimated using Gaussian Process regression (also known as kriging) for each host-virus pair with at least 5 populations sampled across space. This is done using a Squared-Exponential covariance kernel with a fixed length scale parameter using [GaussianProcesses.jl](#) ([Fairbrother et al. 2022](#)), and the hyperparameters for the model are optimized using the L-BFGS algorithm in [Optim.jl](#) ([Mogensen and Riseth 2018](#)). This provides both spatial estimates of prevalence, but crucially, spatially explicit uncertainty associated with this prevalence estimate.

Figure 3 shows both the bivariate plot of biodiversity dose and dose uncertainty (Figure 3A) and dose vs. prevalence uncertainty (Figure 3B). To isolate regions where we are confident hosts are present for prevalence sampling, and regions of high host presence uncertainty for discovery sampling, we stratify the spatial extent into discrete regions that prioritize prevalence sampling and occurrence sampling as follows: the region that is both in the top 50% of biodiversity dose and top 50% of prevalence uncertainty is the stratum for prevalence sampling, and the region in the top 50% of dose uncertainty is the stratum for discovery sampling (Figure 3C). Spatial regions that fall into both categories are then targets for both forms of sampling. The choice of 50% quantiles is arbitrary for this example. The sampling priority within each of these strata can be seen in Figure 3D. The overall priority map can be seen in Figure 3E, along with the points sampled using Balanced Acceptance Sampling with a tilting value of $\alpha = 2.0$, and the type of sampling that should take place at each location.

This case study demonstrates how we can integrate host SDMs with existing prevalence information to design sampling strategies and split geographic space into discrete regions where each type of sampling (occurrence and prevalence) should take place.

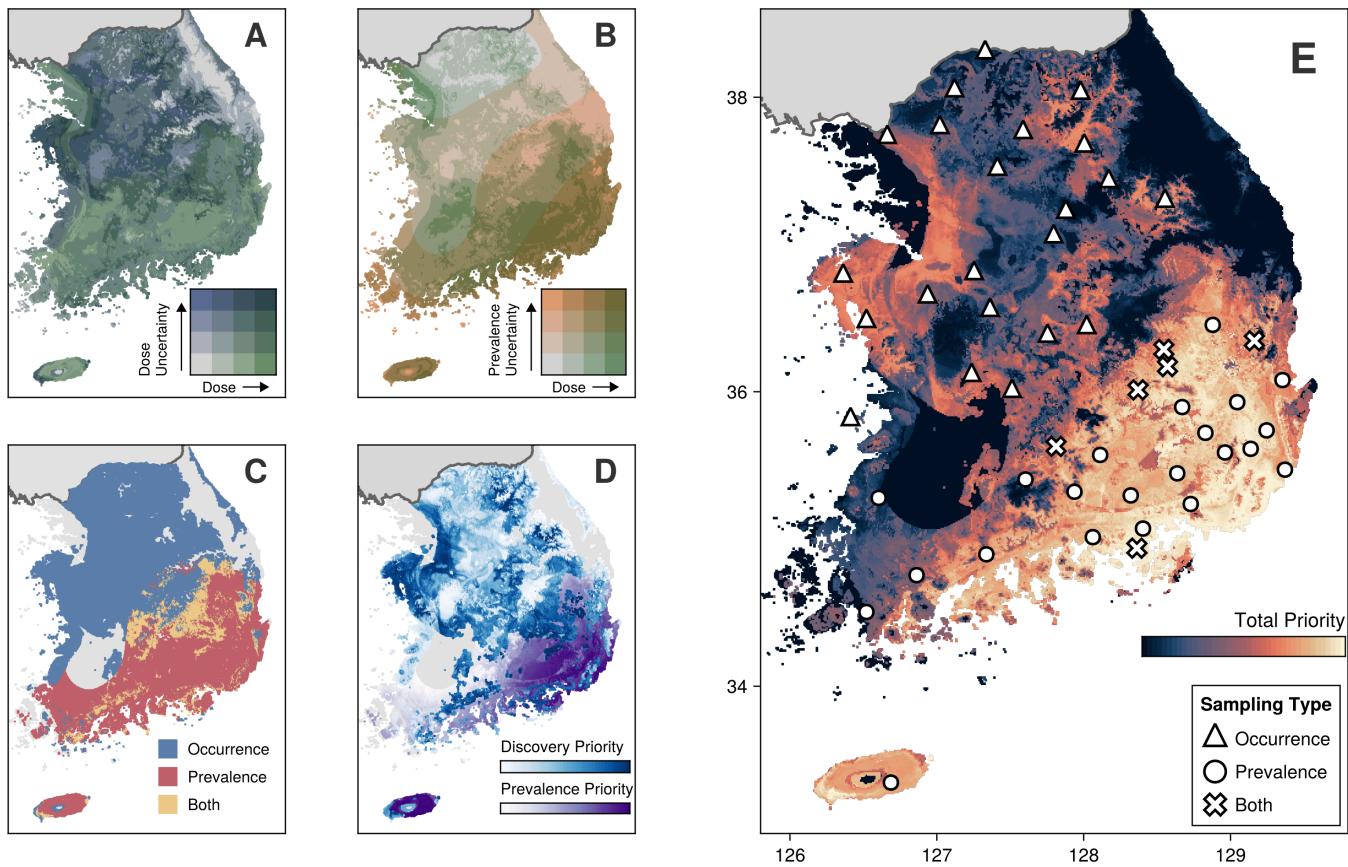


Figure 3: The case study for South Korea. (A): Bivariate plot of dose (increasing green) and dose uncertainty (increasing blue). (B): Bivariate plot of dose (increasing green) and prevalence uncertainty (increasing orange; measured as the sum of estimated variance from kriging across each host-virus pair). (C): The spatial strata for which occurrence, prevalence, and both forms of sampling should occur (see main text for how these are computed). (D) The priority scores for regions where discovery sampling is a priority (increasingly blue) and prevalence sampling is the priority (increasingly purple) in their respective strata. (E) The total priority map with sites selected using Balanced Acceptance Sampling. The marker shape of each site corresponds to the type of sampling that should occur there, based on the stratum in (C) that each point falls within.

Box 3: Guidelines for the interpretation of recommended sampling locations

We want to emphasize that sampling points, though given as exact (longitude, latitude) coordinates, are not to be interpreted without critical thought. Since the grain at which the underlying SDMs are operating is often significantly larger than the scale at which one may have to decide where to place a sampling trap, selected points should be thought of as most useful as the centroids of “local” regions (where “local” corresponds to the pixel size of the SDMs used for prioritization) wherein sampling effect could best improve knowledge about the ecological settings of viral transmission. To make this point explicit, if the model suggests sampling for *Mus musculus* in an urban area and the coordinate itself is in the middle of the center of a highway, the model’s output should be interpreted as indicating the general area where sampling should be done (see below figure).

For example, Figure 4 shows three selected sampling sites for the second case study on South Korea, with the satellite imagery corresponding to the cell in the priority map that was selected. As the original climate layers have pixels that are approximately 1 km^2 , there is considerable heterogeneity within these pixels. Experts on particular taxa should assess the scope of the local region near each sampling point, and use it to guide the practical logistics of where effort (e.g. traps) are most likely to result in useful information.

Beyond just the spatial scale at which recommendations are made, temporal variation in both prevalence and host occurrence is a well-known driver of disease dynamics, often driven by environmental variation tied to seasonality (Altizer et al. 2004, Cosgrove et al. 2008). Factors like migration and birth rates can affect these within-year cycles (van Dijk et al. 2014). Between-year fluctuations, particularly in rodents, are both dramatic and can also counter-intuitively inverse effect on populations' prevalence (Davis et al. 2005). Here, we do not consider temporal variation in host abundances in our sampling recommendation protocol, since to do so would require orders of magnitude more data than we consider available, and would likely require the use of a different class of models/approaches.

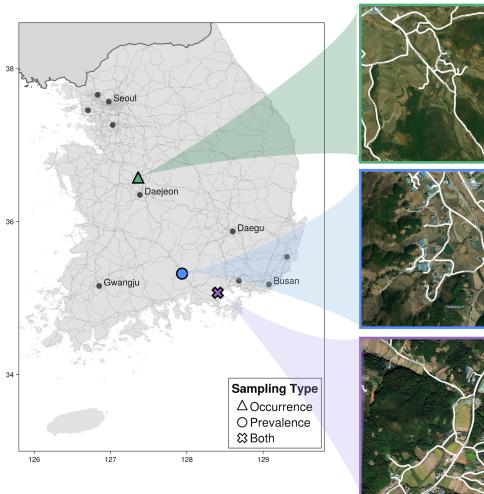


Figure 4: Three selected sites for the second case study (colored markers, left) overlaid on a map of South Korea with the 10 most populous cities shown in dark grey dots, and the major highway network shown in light grey (to get a sense of the accessibility of these sites). Right: Satellite imagery, derived from OpenStreetMap, for each site.

Discussion

Here, we have presented a protocol for adaptive sampling of wildlife disease based on weighted predictions of host richness and uncertainty, and guidelines for integrating prevalence data into sampling prioritization if available. The strength of this approach is prioritizing sampling locations for wildlife disease surveillance by accounting for varying availability of prevalence and host competence data, allowing it to be tailored to any particular stage of developing biosurveillance programs. When enough prevalence data are known to allow the spatially explicit modelling of prevalence (Gras et al. 2018), relying on biodiversity data becomes a lower priority. Still, because spatially explicit modelling of prevalence is mostly feasible at small spatial scales, and because it is important to keep track of larger-scale changes in species distribution (Carlson et al. 2022a, Lawlor et al. 2024), we anticipate that the design of a biosurveillance network that is aligned with biodiversity data collection should become a standard practice. A direction for future work is to explicitly assess how good current biodiversity observation networks are for jointly supporting pathogen surveillance and biodiversity monitoring, both to assess their existing coverage and its relevance for wildlife disease monitoring, and to make recommendations for cost effective extensions to improve their utility for both biosurveillance and monitoring biodiversity change.

While this protocol relies on the spatial distribution of suitable habitat for pathogen reservoirs, our approach is not designed to capture temporal variance in virus prevalence or transmission dynamics, nor is it particularly suited for guiding time-series based optimization of surveillance efforts. Instead, this framework identifies areas with suitable environmental conditions for pathogen hosts, but does not predict when virus prevalence, or spillover risk, is increased within those areas. To fully maximize the effectiveness of sampling, users must integrate domain expertise regarding viral

ecology, seasonal transmission patterns, and host-pathogen dynamics to interpret model outputs appropriately. Additionally, the static nature of SDMs may not adequately reflect rapidly changing environmental conditions, anthropogenic landscape modifications, or evolving host community structure that could influence reservoir distributions or pathogen prevalence in time and space.

There are numerous future directions for incorporating more sophisticated forms of SDMs to help alleviate these challenges: temporally-explicit SDMs are possible but require much more data at a high temporal resolution, which are typically not available for most species (Anselmetto et al. 2025). Similarly, mechanistic SDMs, which directly integrate biological processes into habitat suitability predictions, require far more data than is typically available. Another limitation of conventional correlative SDMs is that habitat suitability does not directly translate to species occupancy. Species may not be present in suitable habitat due to competition or dispersal limitation, and this can impact viral transmission if a host species is replaced by a less competent host (e.g. Lassa fever is reported less in cities where *Mus musculus* has displaced primary reservoir species). Joint SDMs (JSDMs), which model species composition across an entire species pool, thereby accounting for species interactions, are another possible frontier, but JSDMs require explicit co-occurrence data, typically not attainable from platforms like GBIF. Generally, these more informative SDMs are not possible without extensive, long-term monitoring, which is why the goal of this protocol is to get these long-term monitoring programs off the ground with efficient sampling in the low data context.

However, an increasing amount of pathogen prevalence data is only actionable when individual test outcomes, both positive and negative, are shared in a way that is both standardized and interoperable with the commonly used representations of biodiversity data (Schwantes et al. 2025). Further integration of open biosurveillance and biodiversity monitoring systems is essential to meet our common goals (Poisot et al. 2025a). Better biosurveillance itself is not spillover prevention, but it's a necessary component of it to make prevention possible — it's the best we can do in the context of limited information on prevalence — but a full approach to prevention of zoonotic spillover is broader than this alone (Ramakrishnan 2023). This protocol represents a necessary first step toward widespread biosurveillance integrated with global biodiversity monitoring systems (Gonzalez et al. 2023b) to ensure a just planetary future.

Software and Data Availability: The software and data used for this analysis is available in [this](#) GitHub repository.

Acknowledgements: MDC was funded through an IVADO postdoctoral fellowship. TP was funded by the Wellcome Trust (223764/Z/21/Z); CBB, CJC, RG, HR, DS, SS, and TP were supported by the US National Science Foundation (grant no. DBI 2213854). AB was funded by the CANOPY CIHR training program. CBB was supported by an NSERC PGSD (no. 56749). HR was funded by an NSF GRFP (grant no. DGE-2139841). RG was supported by the Wolfson Foundation, via a UCL Excellence Fellowship. BK was funded by CIHR DRA (BCB 201224). DS was funded by a joint NSF-NIH-NIFA EEID Award #2208034

References

Altizer S, Hochachka WM, Dhondt AA (2004) Seasonal Dynamics of Mycoplasmal Conjunctivitis in Eastern North American House Finches. *Journal of Animal Ecology* 73:309–322. <https://doi.org/10.1111/j.0021-8790.2004.00807.x>

Andrade-Pacheco R, Rerolle F, Lemoine J, et al (2020) Finding Hotspots: Development of an Adaptive Spatial Sampling Approach. *Scientific Reports* 10:10939. <https://doi.org/10.1038/s41598-020-67666-3>

Anselmetto N, Garbarino M, Weldy MJ, et al (2025) Leveraging Long-term Data to Improve Biodiversity Monitoring with Species Distribution Models. *Journal of Applied Ecology* 1365–2664.70177. <https://doi.org/10.1111/1365-2664.70177>

Araújo MB, Anderson RP, Márcia Barbosa A, et al (2019) Standards for Distribution Models in Biodiversity Assessments. *Science Advances* 5:eaat4858. <https://doi.org/10.1126/sciadv.aat4858>

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting Pseudo-Absences for Species Distribution Models: How, Where and How Many?. *Methods in Ecology and Evolution* 3:327–338. <https://doi.org/10.1111/j.2041-210X.2011.00172.x>

Barry S, Elith J (2006) Error and Uncertainty in Habitat Models. *Journal of Applied Ecology* 43:413–423. <https://doi.org/10.1111/j.1365-2664.2006.01136.x>

Becker DJ, Albery GF, Sjodin AR, et al (2022) Optimising Predictive Models to Prioritise Viral Discovery in Zoonotic Reservoirs. *The Lancet Microbe* 3:e625–e637. [https://doi.org/10.1016/S2666-5247\(21\)00245-7](https://doi.org/10.1016/S2666-5247(21)00245-7)

Bell D, von Agris J, Tacheva B, Brown GW (2025) Natural Spillover Risk and Disease Outbreaks: Is Over-Simplification Putting Public Health at Risk?. *Journal of Epidemiology and Global Health* 15:65. <https://doi.org/10.1007/s44197-025-00412-y>

Blanchet FG, Cazelles K, Gravel D (2020) Co-Occurrence Is Not Evidence of Ecological Interactions. *Ecology Letters* 23:1050–1063. <https://doi.org/10.1111/ele.13525>

Bussières-Fournel A, Poisot T (2025) Climate Change Increases the Distribution of Reservoirs of the Raccoon Rabies Virus in Quebec

Calabrese JM, Certain G, Kraan C, Dormann CF (2014) Stacking Species Distribution Models and Adjusting Bias by Linking Them to Macroecological Models. *Global Ecology and Biogeography* 23:99–112. <https://doi.org/10.1111/geb.12102>

Cao B, Bai C, Wu K, et al (2023) Tracing the Future of Epidemics: Coincident Niche Distribution of Host Animals and Disease Incidence Revealed Climate-Correlated Risk Shifts of Main Zoonotic Diseases in China. *Global Change Biology* 29:3723–3746. <https://doi.org/10.1111/gcb.16708>

Carlson CJ (2020) Embarcadero: Species Distribution Modelling with Bayesian Additive Regression Trees in r. *Methods in Ecology and Evolution* 11:850–858. <https://doi.org/10.1111/2041-210X.13389>

Carlson CJ, Albery GF, Merow C, et al (2022a) Climate Change Increases Cross-Species Viral Transmission Risk. *Nature* 607:555–562. <https://doi.org/10.1038/s41586-022-04788-w>

Carlson CJ, Brookson CB, Becker DJ, et al (2025) Pathogens and Planetary Change. *Nature Reviews Biodiversity* 1:32–49. <https://doi.org/10.1038/s44358-024-00005-w>

Carlson CJ, Gibb RJ, Albery GF, et al (2022b) The Global Virome in One Network (VIRION): An Atlas of Vertebrate-Virus Associations. *mBio* 13:e2985–21. <https://doi.org/10.1128/mbio.02985-21>

Carlson CJ, Zipfel CM, Garnier R, Bansal S (2019) Global Estimates of Mammalian Viral Diversity Accounting for Host Sharing. *Nature Ecology & Evolution* 3:1070–1075. <https://doi.org/10.1038/s41559-019-0910-6>

Convention on Biological Diversity (2024) Global Action Plan on Biodiversity and Health

Cosgrove CL, Wood MJ, Day KP, Sheldon BC (2008) Seasonal Variation in Plasmodium Prevalence in a Population of Blue Tits Cyanistes Caeruleus. *Journal of Animal Ecology* 77:540–548. <https://doi.org/10.1111/j.1365-2656.2008.01370.x>

Cummings CA, Vicente-Santos A, Carlson CJ, Becker DJ (2025) Viral Epidemic Potential Is Not Uniformly Distributed across the Bat Phylogeny. *Communications Biology* 8:1510. <https://doi.org/10.1038/s42003-025-08929-5>

Daszak P, Cunningham AA, Hyatt AD (2000) Emerging Infectious Diseases of Wildlife—Threats to Biodiversity and Human Health. *Science* 287:443–449. <https://doi.org/10.1126/science.287.5452.443>

Davis S, Calvet E, HL (2005) Fluctuating Rodent Populations and Risk to Humans from Rodent-Borne Zoonoses. *Vector-Borne and Zoonotic Diseases* 5:305–314. <https://doi.org/10.1089/vbz.2005.5.305>

Desgagne-Bouchard, Pandey A, S J, et al (2025) Evovest/ EvoTrees.Jl: V0.18.0

DiRenzo GV, Che-Castaldo C, Saunders SP, et al (2019) Disease-Structured N-mixture Models: A Practical Guide to Model Disease Dynamics Using Count Data. *Ecology and Evolution* 9:899–909. <https://doi.org/10.1002/ece3.4849>

Dobson A (2004) Population Dynamics of Pathogens with Multiple Host Species.. *The American Naturalist* 164:S64–S78. <https://doi.org/10.1086/424681>

Dumelle M, Higham M, Ver Hoef JM, et al (2022) A Comparison of Design-Based and Model-Based Approaches for Finite Population Spatial Sampling and Inference. *Methods in Ecology and Evolution* 13:2018–2029. <https://doi.org/10.1111/2041-210X.13919>

Eby P, Peel AJ, Hoegh A, et al (2023) Pathogen Spillover Driven by Rapid Changes in Bat Ecology. *Nature* 613:340–344. <https://doi.org/10.1038/s41586-022-05506-2>

Elith J, Leathwick JR, Hastie T (2008) A Working Guide to Boosted Regression Trees. *Journal of Animal Ecology* 77:802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>

Elith J, Phillips SJ, Hastie T, et al (2011) A Statistical Explanation of MaxEnt for Ecologists. *Diversity and Distributions* 17:43–57. <https://doi.org/10.1111/j.1472-4642.2010.00725.x>

Elmasri M, Farrell MJ, Davies TJ, Stephens DA (2020) A Hierarchical Bayesian Model for Predicting Ecological Interactions Using Scaled Evolutionary Relationships. *The Annals of Applied Statistics* 14. <https://doi.org/10.1214/19-AOAS1296>

Fairbrother J, Nemeth C, Rischard M, et al (2022) GaussianProcessesJl: A Nonparametric Bayes Package for the Julia Language. *Journal of Statistical Software* 102:1–36. <https://doi.org/10.18637/jss.v102.i01>

Farver TB, Thomas C, Edson RK (1985) An Application of Sampling Theory in Animal Disease Prevalence Survey Design. *Preventive Veterinary Medicine* 3:463–473. [https://doi.org/10.1016/0167-5877\(85\)90007-8](https://doi.org/10.1016/0167-5877(85)90007-8)

Fick SE, Hijmans RJ (2017) WorldClim 2: New 1-Km Spatial Resolution Climate Surfaces for Global Land Areas. *International Journal of Climatology* 37:4302–4315. <https://doi.org/10.1002/joc.5086>

Fourcade Y, Besnard AG, Secondi J (2018) Paintings Predict the Distribution of Species, or the Challenge of Selecting Environmental Predictors and Evaluation Statistics. *Global Ecology and Biogeography* 27:245–256. <https://doi.org/10.1111/geb.12684>

Friant S, Mistrick J, Luis AD, et al (2025) Reducing the Threats of Rodent-Borne Zoonoses Requires an Understanding and Leveraging of Three Key Pillars: Disease Ecology, Synanthropy, and Rodentation. *The Lancet Planetary Health* 101300. <https://doi.org/10.1016/j.lanph.2025.101300>

Gibb R, Albery GF, Mollentze N, et al (2022) Mammal Virus Diversity Estimates Are Unstable Due to Accelerating Discovery Effort. *Biology Letters* 18:20210427. <https://doi.org/10.1098/rsbl.2021.0427>

Gibb R, Franklinos LHV, Redding DW, Jones KE (2020a) Ecosystem Perspectives Are Needed to Manage Zoonotic Risks in a Changing Climate. *BMJ* 371:m3389. <https://doi.org/10.1136/bmj.m3389>

Gibb R, Redding DW, Chin KQ, et al (2020b) Zoonotic Host Diversity Increases in Human-Dominated Ecosystems. *Nature* 584:398–402. <https://doi.org/10.1038/s41586-020-2562-8>

Gibb R, Redding DW, Friant S, Jones KE (2025) Towards a 'People and Nature' Paradigm for Biodiversity and Infectious Disease. *Philosophical Transactions of the Royal Society B: Biological Sciences* 380:20230259. <https://doi.org/10.1098/rstb.2023.0259>

Gibb R, Ryan SJ, Pigott D, et al (2024) The Anthropogenic Fingerprint on Emerging Infectious Diseases. <https://doi.org/10.4236/2024.052224307684>

Gonzalez A, Chase JM, O'Connor MI (2023a) A Framework for the Detection and Attribution of Biodiversity Change. *Philosophical Transactions of the Royal Society B: Biological Sciences* 378:20220182. <https://doi.org/10.1098/rstb.2022.0182>

Gonzalez A, Vihervaara P, Balvanera P, et al (2023b) A Global Biodiversity Observing System to Unite Monitoring and Guide Action. *Nature Ecology & Evolution* 1–6. <https://doi.org/10.1038/s41559-023-02171-0>

Gottdenker NL, Streicker DG, Faust CL, Carroll CR (2014) Anthropogenic Land Use Change and Infectious Diseases: A Review of the Evidence. *EcoHealth* 11:619–632. <https://doi.org/10.1007/s10393-014-0941-z>

Grafström A, Lundström NLP, Schelin L (2012) Spatially Balanced Sampling through the Pivotal Method. *Biometrics* 68:514–520. <https://doi.org/10.1111/j.1541-0420.2011.01699.x>

Gras P, Knuth S, Börner K, et al (2018) Landscape Structures Affect Risk of Canine Distemper in Urban Wildlife. *Frontiers in Ecology and Evolution* 6: <https://doi.org/10.3389/fevo.2018.00136>

Grenié M, Violle C, Munoz F (2020) Is Prediction of Species Richness from Stacked Species Distribution Models Biased by Habitat Saturation?. *Ecological Indicators* 111:105970. <https://doi.org/10.1016/j.ecolind.2019.105970>

Gu SH, Kang HJ, Baek LJ, et al (2011) Genetic Diversity of Imjin Virus in the Ussuri White-Toothed Shrew (*Crocidura Lasiura*) in the Republic of Korea, 2004–2010. *Virology Journal* 8:56. <https://doi.org/10.1186/1743-422X-8-56>

Guisan A, Tingley R, Baumgartner JB, et al (2013) Predicting Species Distributions for Conservation Decisions. *Ecology Letters* 16:1424–1435. <https://doi.org/10.1111/ele.12189>

Halliday FW, Rohr JR, Laine A-L (2020) Biodiversity Loss Underlies the Dilution Effect of Biodiversity. *Ecology Letters* 23:1611–1622. <https://doi.org/10.1111/ele.13590>

Han BA, Schmidt JP, Bowden SE, Drake JM (2015) Rodent Reservoirs of Future Zoonotic Diseases. *Proceedings of the National Academy of Sciences* 112:7039–7044. <https://doi.org/10.1073/pnas.1501598112>

Hosseini PR, Mills JN, Prieur-Richard A-H, et al (2017) Does the Impact of Biodiversity Differ between Emerging and Endemic Pathogens? The Need to Separate the Concepts of Hazard and Risk. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* 372:20160129. <https://doi.org/10.1098/rstb.2016.0129>

Johnson PTJ, de Roode JC, Fenton A (2015) Why Infectious Disease Research Needs Community Ecology. *Science* 349:1259504. <https://doi.org/10.1126/science.1259504>

Karger DN, Conrad O, Böhner J, et al (2017) Climatologies at High Resolution for the Earth's Land Surface Areas. *Scientific Data* 4:170122. <https://doi.org/10.1038/sdata.2017.122>

Keesing F, Holt RD, Ostfeld RS (2006) Effects of Species Diversity on Disease Risk. *Ecology Letters* 9:485–498. <https://doi.org/10.1111/j.1461-0248.2006.00885.x>

Keesing F, Ostfeld RS (2021) Impacts of Biodiversity and Biodiversity Loss on Zoonotic Diseases. *Proceedings of the National Academy of Sciences of the United States of America* 118:e2023540118. <https://doi.org/10.1073/pnas.2023540118>

Kim H-C, Kim W-K, Klein TA, et al (2017) Hantavirus Surveillance and Genetic Diversity Targeting Small Mammals at Camp Humphreys, a US Military Installation and New Expansion Site, Republic of Korea. *PLOS ONE* 12:e176514. <https://doi.org/10.1371/journal.pone.0176514>

Kim H-C, Klein TA, Chong S-T, et al (2007) Seroepidemiological Survey of Rodents Collected at a U.S. Military Installation, Yongsan Garrison, Seoul, Republic of Korea. *Military Medicine* 172:759–764. <https://doi.org/10.7205/milmed.172.7.759>

Kim W-K, No JS, Lee D, et al (2020) Active Targeted Surveillance to Identify Sites of Emergence of Hantavirus. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 70:464–473. <https://doi.org/10.1093/cid/ciz234>

Klein TA, Kim H-C, Chong S-T, et al (2015) Hantaan Virus Surveillance Targeting Small Mammals at Nightmare Range, a High Elevation Military Training Area, Gyeonggi

Province, Republic of Korea. PLOS ONE 10:e118483. <https://doi.org/10.1371/journal.pone.0118483>

Kopsco HL, Smith RL, Halsey SJ (2022) A Scoping Review of Species Distribution Modeling Methods for Tick Vectors. *Frontiers in Ecology and Evolution* 10:. <https://doi.org/10.3389/fevo.2022.893016>

Kéry M, Dorazio RM, Soldaat L, et al (2009) Trend Estimation in Populations with Imperfect Detection. *Journal of Applied Ecology* 46:1163–1172. <https://doi.org/10.1111/j.1365-2664.2009.01724.x>

Lawlor JA, Comte L, Grenouillet G, et al (2024) Mechanisms, Detection and Impacts of Species Redistributions under Climate Change. *Nature Reviews Earth & Environment* 5:351–368. <https://doi.org/10.1038/s43017-024-00527-z>

Lee HW, Lee PW, Johnson KM (1978) Isolation of the Etiologic Agent of Korean Hemorrhagic Fever. *The Journal of Infectious Diseases* 137:298–308. <https://doi.org/10.1093/infdis/137.3.298>

Lee S-H, Kim W-K, No JS, et al (2017) Dynamic Circulation and Genetic Exchange of a Shrew-borne Hantavirus, Imjin Virus, in the Republic of Korea. *Scientific Reports* 7:44369. <https://doi.org/10.1038/srep44369>

Lippi CA, Mundis SJ, Sippy R, et al (2023) Trends in Mosquito Species Distribution Modeling: Insights for Vector Surveillance and Disease Control. *Parasites & Vectors* 16:302. <https://doi.org/10.1186/s13071-023-05912-z>

Merow C, Smith MJ, Silander Jr JA (2013) A Practical Guide to MaxEnt for Modeling Species' Distributions: What It Does, and Why Inputs and Settings Matter. *Ecography* 36:1058–1069. <https://doi.org/10.1111/j.1600-0587.2013.07872.x>

Milanesi P, Della Rocca F, Robinson RA (2020) Integrating Dynamic Environmental Predictors and Species Occurrences: Toward True Dynamic Species Distribution Models. *Ecology and Evolution* 10:1087–1092. <https://doi.org/10.1002/ece3.5938>

Mogensen PK, Riseth AN (2018) Optim: A Mathematical Optimization Package for Julia. *Journal of Open Source Software* 3:615. <https://doi.org/10.21105/joss.00615>

No JS, Kim W-k, Kim J-A, et al (2016) Detection of Hantaan Virus RNA from Anti-Hantaan Virus IgG Seronegative Rodents in an Area of High Endemicity in Republic of Korea. *Microbiology and Immunology* 60:268–271. <https://doi.org/10.1111/1348-0421.12370>

Norman KE, Poisot T (2025) Site Selection Algorithms for Optimal Ecological Monitoring Design. *Ecological Indicators* 178:114104. <https://doi.org/10.1016/j.ecolind.2025.114104>

Nusser SM, Clark WR, Otis DL, Huang L (2008) Sampling Considerations for Disease Surveillance in Wildlife Populations. *The Journal of Wildlife Management* 72:52–60. <https://doi.org/10.2193/2007-317>

Park K, Kim J, Noh J, et al (2025) Epidemiological Surveillance and Phylogenetic Diversity of Orthohantavirus Hantanense Using High-Fidelity Nanopore Sequencing, Republic of Korea. *PLOS Neglected Tropical Diseases* 19:e12859. <https://doi.org/10.1371/journal.pntd.0012859>

Park K, Lee S-H, Kim J, et al (2022) A Portable Diagnostic Assay, Genetic Diversity, and Isolation of Seoul Virus from *Rattus Norvegicus* Collected in Gangwon Province, Republic of Korea. *Pathogens (Basel, Switzerland)* 11:1047. <https://doi.org/10.3390/pathogens11091047>

Plowright RK, Parrish CR, McCallum H, et al (2017) Pathways to Zoonotic Spillover. *Nature Reviews Microbiology* 15:502–510. <https://doi.org/10.1038/nrmicro.2017.45>

Poisot T (2024) Conformal Prediction Quantifies the Uncertainty of Species Distribution Models

Poisot T, Becker DJ, Catchen MD, et al (2025a) Biodiversity Science and Biosurveillance Are Fellow Travelers. *Bioscience* biaf91. <https://doi.org/10.1093/biosci/biaf091>

Poisot T, Bussières-Fournel A, Dansereau G, Catchen MD (2025b) A Julia toolkit for species distribution data. *Peer Community Journal* 5:. <https://doi.org/10.24072/pcjournal.589>

Pollock LJ, Tingley R, Morris WK, et al (2014) Understanding Co-occurrence by Modelling Species Simultaneously with a Joint Species Distribution Model (\text{JSDM}). *Methods in Ecology and Evolution* 5:397–406. <https://doi.org/10.1111/2041-210X.12180>

Ramakrishnan N (2023) Bio-Surveillance as One Health: A Critique of Recent Definitions and Policy Initiatives. *Development (Society for International Development)* 66:215–225. <https://doi.org/10.1057/s41301-023-00396-x>

Redding DW, Gibb R, Jones KE (2024) Ecological Impacts of Climate Change Will Transform Public Health Priorities for Zoonotic and Vector-Borne Disease. <https://doi.org/10.1111/24302575>

Robertson BL, Brown JA, McDonald T, Jaksons P (2013) BAS: Balanced Acceptance Sampling of Natural Resources. *Biometrics* 69:776–784. <https://doi.org/10.1111/biom.12059>

Rodrigues F, Gupta N, Pinto B (1978) Serological Survey for the Detection of Antibodies to Lassa Virus in India. *Journal of the Indian Medical Association* 70:25–28

Royle JA (2004) N-Mixture Models for Estimating Population Size from Spatially Replicated Counts. *Biometrics* 60:108–115. <https://doi.org/10.1111/j.0006-341X.2004.00142.x>

Ryou J, Lee HI, Yoo YJ, et al (2011) Prevalence of Hantavirus Infection in Wild Rodents from Five Provinces in Korea. *Journal of Wildlife Diseases* 47:427–432. <https://doi.org/10.7589/0090-3558-47.2.427>

Sames WJ, Klein TA, Kim HC, et al (2009) Ecology of Hantaan Virus at Twin Bridges Training Area, Gyeonggi Province, Republic of Korea, 2005–2007. *Journal of Vector Ecology: Journal of the Society for Vector Ecology* 34:225–231. <https://doi.org/10.1111/j.1948-7134.2009.00030.x>

Scholes RJ, Walters M, Turak E, et al (2012) Building a Global Observing System for Biodiversity. *Current Opinion in Environmental Sustainability* 4:139–146. <https://doi.org/10.1016/j.cosust.2011.12.005>

Schwantes CJ, Sánchez CA, Stevens T, et al (2025) A Minimum Data Standard for Wildlife Disease Research and Surveillance. *Scientific Data* 12:1054. <https://doi.org/10.1038/s41597-025-05332-x>

Seo M hee, Kim C-M, Kim D-M, et al (2022) Emerging Hantavirus Infection in Wild Rodents Captured in Suburbs of Gwangju Metropolitan City, South Korea. *PLOS Neglected Tropical Diseases* 16:e10526. <https://doi.org/10.1371/journal.pntd.0010526>

Simons D, Rivero R, Guiote AM-C, et al (2025) Protocol to Produce a Systematic Arenavirus and Hantavirus Host-Pathogen Database: Project ArHa.. Wellcome Open Research 10:227. <https://doi.org/10.12688/wellcomeopenres.24037.2>

Smith JR, Levine JM (2025) Linking Relative Suitability to Probability of Occurrence in Presence-Only Species Distribution Models: Implications for Global Change Projections. *Methods in Ecology and Evolution* 16:854–865. <https://doi.org/10.1111/2041-210X.70003>

Song J-W, Kang HJ, Gu SH, et al (2009) Characterization of Imjin Virus, a Newly Isolated Hantavirus from the Ussuri White-Toothed Shrew (*Crocidura Lasiura*). *Journal of Virology* 83:6184–6191. <https://doi.org/10.1128/JVI.00371-09>

Stephens PR, Altizer S, Smith KF, et al (2016) The Macroecology of Infectious Diseases: A New Perspective on Global-Scale Drivers of Pathogen Distributions and Impacts. *Ecology Letters* 19:1159–1171. <https://doi.org/10.1111/ele.12644>

Stevens DL, Olsen AR (2004) Spatially Balanced Sampling of Natural Resources. *Journal of the American Statistical Association* 99:262–278. <https://doi.org/10.1198/016214504000000250>

Thompson SK (2012) Sampling, 1st edn.. Wiley

Thuiller W, Pollock LJ, Gueguen M, Münkemüller T (2015) From Species Distributions to Meta-Communities. *Ecology Letters* 18:1321–1328. <https://doi.org/10.1111/ele.12526>

Tseng KK, Koehler H, Becker DJ, et al (2025) Viral Genomic Features Predict Orthopoxvirus Reservoir Hosts. *Communications Biology* 8:309. <https://doi.org/10.1038/s42003-025-07746-0>

van Dijk JGB, Hoye BJ, Verhagen JH, et al (2014) Juveniles and Migrants as Drivers for Seasonal Epizootics of Avian Influenza Virus. *Journal of Animal Ecology* 83:266–275. <https://doi.org/10.1111/1365-2656.12131>

Vollering J, Halvorsen R, Auestad I, Rydgren K (2019) Bunching up the Background Betters Bias in Species Distribution Models. *Ecography* 42:1717–1727. <https://doi.org/10.1111/ecog.04503>

Wang J-F, Stein A, Gao B-B, Ge Y (2012) A Review of Spatial Sampling. *Spatial Statistics* 2:1–14. <https://doi.org/10.1016/j.spasta.2012.08.001>

WHO (2025) WHO Pandemic Agreement

Wilkinson DP, Golding N, Guillera-Arroita G, et al (2021) Defining and Evaluating Predictions of Joint Species Distribution Models. *Methods in Ecology and Evolution* 12:394–404. <https://doi.org/10.1111/2041-210X.13518>

Zurell D, Franklin J, König C, et al (2020) A Standard Protocol for Reporting Species Distribution Models. *Ecography* 43:1261–1277. <https://doi.org/10.1111/ecog.04960>

Zurell D, Zimmermann NE, Gross H, et al (2020) Testing Species Assemblage Predictions from Stacked and Joint Species Distribution Models. *Journal of Biogeography* 47:101–113. <https://doi.org/10.1111/jbi.13608>