The LINNEAN SOCIETY of London BIOLOGICAL Journal of the Linnean Society

Biological Journal of the Linnean Society, 2015, **115**, 664–677. With 2 figures.

Tracking the distribution and impacts of diseases with biological records and distribution modelling

BETHAN V. PURSE[1]* and NICK GOLDING[2]

[1]*NERC Centre for Ecology and Hydrology, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, UK*
[2]*Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford, Oxford, Oxfordshire, OX1 3PS, UK*

Received 6 January 2015; revised 15 March 2015; accepted for publication 17 March 2015

Species distribution modelling is widely used in epidemiology for mapping spatial patterns and the risk of introduction of diseases and vectors and also for predicting how exposure may alter given future environmental change, motivated by the high societal impact and the multiple environmental drivers of disease outbreaks. Although pathogens and vectors have historically been sparsely recorded, monitoring systems and media sources are generating novel, online data sources on occurrence. Moreover, increasing ecological realism is being incorporated into distribution modelling techniques, focussing on dispersal, biotic interactions and evolutionary constraints that shape species distributions alongside abiotic factors and biases in recording effort, common to pathogens and vectors and wildlife species. Considering pathogens and arthropod vector systems with high impact on plant, animal and human health, the present review describes how biological records for vectors and pathogens arise, introduces the concepts behind distribution models and illustrates the potential for ecologically realistic distribution models to yield insight into the establishment and spread of pathogens. Because distribution modellers aim to provide policy makers with evidence and maps for planning and evaluation of disease mitigation measures, we highlight factors that currently constrain direct translation of models to policy. Disease distributions will be better understood and mapped in the future given improved occurrence data access and integration and combined (correlative and mechanistic) modelling approaches that are developed iteratively in concert with stakeholders. © 2015 The Linnean Society of London, © 2015 The Linnean Society of London, *Biological Journal of the Linnean Society*, 2015, **115**, 664–677.

ADDITIONAL KEYWORDS: biotic interactions – environmental change – niche concept – risk maps – species.

INTRODUCTION

Species distribution models are empirical models relating field observations of the distributions of species to environmental predictor variables, based on statistically or theoretically derived response surfaces (Guisan & Zimmermann, 2000). Once such a model has been parameterized, it can be combined with continuous spatially gridded data on environmental variables to make predictions to all locations for which environmental data are available, enabling the user to create a predictive map of the likely distribution of the species across an entire landscape (Elith & Leathwick, 2009). From this basic framework, distribution modelling has been expanded to incorporate a broader range of ecological theory and applied to answer fundamental questions in applied ecology.

Distribution modelling has become particularly widely used in epidemiology; for mapping current spatial patterns in incidence of diseases and their vectors or reservoirs, for understanding how environmental factors may underpin transmission and for predicting how risk of exposure may change in the future under environmental and socioeconomic change (Rogers & Randolph, 2003; Eisen & Eisen, 2010; Hay *et al.*, 2013). Generally, the ultimate goal of epidemiological distribution modelling is to reduce

*Corresponding author. E-mail: beth@ceh.ac.uk

disease impacts (in terms of burden on human or animal health, or negative economic consequences) by providing cartographic information that enables policy makers to make evidence-based decisions; for example, concerning the risk of introduction of diseases or vectors (Benedict *et al.*, 2007; Lindsay *et al.*, 2010) or the planning and targeting of surveillance and interventions (Dicko *et al.*, 2014). The growth in use of these tools has been facilitated by the increasing availability of ecologically relevant, spatial environmental datasets, user-friendly statistical software for fitting distribution models (Phillips, Anderson & Schapire, 2006) and data on where diseases occur. Additional motivation comes from high societal impacts of pathogens on human, animal and plant health (Jones *et al.*, 2008; Pautasso *et al.*, 2012), as well as the increasing realization that disease emergence is often driven by multiple, global, interacting environmental changes, including landscape, climate and social factors (Chaves *et al.*, 2008; Jones *et al.*, 2008; Pautasso *et al.*, 2012). In parallel, ecologists and epidemiologists have begun to incorporate increasing ecological realism into distribution modelling techniques, including dispersal, biotic interactions and evolutionary constraints that shape species distributions alongside abiotic factors (Elith & Leathwick, 2009). New developments have often been inspired by and evaluated on well recorded taxa such as birds, butterflies and ladybirds (Araújo & Luoto, 2007; Powney & Isaac, 2015), comprising species whose ecological and environmental requirements are usually well described. By contrast, pathogens and vector species are often sparsely recorded and their life histories poorly known (Hay *et al.*, 2013).

The present review aims to understand the potential of new distribution modelling techniques that incorporate more ecological realism to enhance our understanding of processes underpinning the arrival, establishment and spread of pathogens and to yield mapped information of significance to stakeholders. Having contrasted the way that biological records are collected for vectors and pathogens vs. other wildlife taxa and having introduced the concepts behind distribution models, we review how (1) dispersal constraints and (2) biotic interactions have been incorporated into distribution modelling and identify epidemiological situations in which these techniques may generate most insight into disease patterns. We highlight situations in which the resulting risk or occurrence maps have been of particular value to disease managers and policy makers and delineate some factors that may lead to some maps and predictions being used 'in anger' by stakeholders, whereas others remain unused in reports or scientific literature (Leach & Scoones, 2013). We consider future prospects for improved disease mapping resulting from new pathways and technologies for gathering and integrating distributional data for pathogens, hosts and vectors and highlight key future challenges, such as accounting for spatial biases in disease reporting or vector and host recording effort.

BIOLOGICAL RECORDS FOR DISEASE, VECTORS AND RESERVOIRS

How are records generated?

The taxonomy and field ecology of pathogen and arthropod vector species are often poorly resolved. Many taxa can only be recorded using specialized field sampling techniques (e.g. traps for arthropod vectors baited with host material or semio-chemicals, or collection of blood samples to screen for human pathogens) and identified to species using expensive morphological and molecular methods. Rarely, for pathogens with symptoms that are easy to identify and distinguish from co-circulating pathogens (e.g. ash dieback), there is the potential for occurrence information to be generated by citizens and expert volunteers. This is also true of tick vector species that are large-bodied, slow moving and can be found feeding on people for several days but not of smaller, dispersive, cryptic insect vectors such as mosquitoes and *Culicoides* biting midges. For example, a scheme run by Public Health England since 2005 in which tick specimens collected by members of public, veterinarians, clinicians, wildlife charities and academics are identified and mapped has substantially increased the known geographical extent of *Ixodes ricinus*, the main tick vector of Lyme disease pathogen (*Borrelia burgdorferi s.l.*) in Britain (Jameson & Medlock, 2011).

More commonly, georeferenced data on the occurrence of diseases and arthropod vectors originates from organized surveys, either as a part of scientific research projects or surveillance carried out by ministries of health or disease or vector control programmes. Epidemiological studies may report a number of different epidemiological metrics of occurrence, ranging from detailed information on the proportion of individuals infected at a point in space and time (disease prevalence) to poorer quality information, such as disease diagnoses made based on symptoms (which may confound pathogens) or serological evidence that a disease had infected individuals at some point in the past. Similarly, studies of arthropod vectors may report simply a snapshot of the presence or absence of a species, or detailed information on abundance levels, seasonality and infection rates. The small geographical extent (ranging from small subnational regions up to a few countries) and often sparse spatial sampling of

organised surveys means that the results of multiple studies must be compiled to produce distribution maps spanning the broad spatial scales required by public and animal health policy makers. Rarely, sufficient data are available for high-quality epidemiological or ecological metrics of maximal use to policy makers to be mapped, such as spatial variation in disease prevalence (Gething *et al.*, 2011a). More often, the information in the high-quality data must be down-graded to match the more frequent poor-quality data, generating simpler maps of disease or vector occurrence (Hay *et al.*, 2013).

Susceptible hosts particularly plants and vertebrates tend to be more widely surveyed than pathogens and arthropod vectors but only a few taxa and regions are subject to long-term standardized monitoring schemes that yield detailed information on population sizes and range changes (Isaac & Pocock, 2015; Maes *et al.*, 2015). Some studies use available opportunistic data, sometimes from networks of volunteers (Purse *et al.*, 2013; Maes *et al.*, 2015) to map hosts; for example, the fruit bat reservoirs of *Ebolavirus* and monkey reservoirs of *Plasmodium knowlesi* malaria (Pigott *et al.*, 2014b). Unfortunately, the latter are the exception, rather than rule among human diseases, in having their wild animal reservoirs mapped to some degree (Hay *et al.*, 2013). Despite arising by different methodologies and processes, datasets of disease or vector occurrence share similarities with the species occurrence datasets more commonly used in ecological applications of distribution modelling such as strong spatial variability in reporting rates. In species occurrence datasets, recording effort can be biased towards more populous areas or areas visited regularly by specialist collectors. For diseases and vectors, occurrence is more likely to be reported in countries, and susceptible livestock or plant hosts are more likely to be intensively surveyed or censused in regions, with strong healthcare systems and the necessary funds for surveillance and control programmes. Recording effort is further biased towards those habitats or hosts (Cumming, 2002) in which primary disease impacts are felt (e.g. livestock, crops, people). These biases in recording rates represent a serious impediment to accurate mapping of species and disease distributions (Phillips *et al.*, 2009) and to understanding the ecological and environmental interactions underpinning transmission.

Databases of epidemiological and vector records

General biological records databases such as the Global Biodiversity Information Facility (http://www.gbif.org) compile occurrence data for a large number of biological species, including some vectors and reservoirs of human diseases. However, vector species tend to be under-represented in such databases compared to charismatic arthropod taxa, both with regard to the total number of records and the spatial sampling effort across countries (Table 1). Consequently, these resources have largely been used to supplement other sources of data when constructing distribution models for vector species (Porretta *et al.*, 2012). To compensate for this under-representation, specific databases for epidemiological records or vector groups have also been developed, most notably in the VectorMap project (http://www.vectormap.org), encompassing the MosquitoMap, SandflyMap and TickMap projects (Foley *et al.*, 2010, 2012). Alongside the vector occurrence data compiled from wide ranging survey reports and literature, VectorMap provides some prediction maps of vectors, pathogen and ecto-parasite data from vertebrate hosts and expert opinion maps of disease occurrence (Foley *et al.*, 2012), fostering broadscale understanding of the ecological context of transmission.

Considering global/continental databases of disease occurrence, these tend to cover only those pathogens that are notifiable because of their high impacts on trade and animal or plant products. The World Animal Health Information System of the Office International des Epizooties (OIE, 2014) provides standardized geographical data on emerging animal pathogens based on official country reports. The European and Mediterranean Plant Protection Organization (EPPO) provides global occurrence data as part of the Plant Quarantine Data Retrieval System (PQR) for plant pathogens (EPPO, 2014). In both databases, occurrence data are accompanied by the ecological and epidemiological trait data which will be critical for generalizing models between groups of pathogens but, as is the case with vectors, the occurrence data are necessarily skewed towards crops and livestock rather than wild species.

It is notable that no centralized (global or continental) database of this kind exists for notifiable human diseases. Epidemiological data provided on the World Health Organization (WHO), Center for Disease Control and European Centre for Disease Control (ECDC) websites tends to be at national scale that is insufficient for distribution modelling. As a consequence, collation of human disease occurrence records for epidemiological distribution modelling typically requires manual searches for suitable disease occurrence information from various literature sources encompassing repositories of published scientific articles such as PubMed (http://www.ncbi.nlm.nih.gov/pubmed) and Web of Knowledge (http://wokinfo.com/), repositories of surveys containing disease incidence data such as the Global Health Data Exchange (http://

Table 1. Availability of records for vector groups (mosquitoes and hard ticks) in national (BRC) and international (GBIF) databases compared to selected charismatic arthropod taxa (butterflies and ladybirds)

Taxa	Family name	Number of records		Coverage of UK (10 km^2)
		BRC	GBIF	BRC
Mosquitoes	Culicidae	3000	186 400	800
Hard ticks	Ixodidae	4500	29 900	970
Butterflies	Nymphalidae	2 000 000	2 335 900	3500
Ladybirds	Coccinellidae	150 000	373 400	2500

BRC, UK Biological Records Centre; GBIF, Global Biodiversity Information Facility (both accessed June 2014).

ghdx.healthdata.org/), as well as less formal records of disease occurrence such as the ProMed mail archives (www.promedmail.org), which provide short ad-hoc reports on emerging and re-emerging diseases. Only a subset of these reports contain disease occurrence data that can be used in a distribution modelling study. Consequently, this approach to data collection is very time consuming and costly, limiting progress in mapping infectious diseases (Hay *et al.*, 2013). It is helpful, where this process has been carried out, for the resulting disease- or vector-specific database to subsequently be made freely available to other researchers (Pigott *et al.*, 2014a), preventing duplication of this significant effort and fostering development of multiple competing models to explain disease patterns. In the absence of a centralized disease occurrence data repository, these datasets are scattered across different online repositories and institutional data archives, reducing their accessibility to researchers.

Even though occurrence data may be more widely available for susceptible hosts than for pathogens or vectors, there is again a lack of centralized, accessible databases. Global atlases have been developed for the distributions of some wild and domestic reservoirs of human and animal pathogens (Jenkins, Pimm & Joppa, 2013; Robinson *et al.*, 2014). In both ecological and epidemiological distribution modelling, the availability of occurrence data is hindered by a lack of transparency from official bodies (Hay *et al.*, 2013), partly as a result of the disparate regulations surrounding the protection of personal information. Where differences in official reporting and regulations follow national or regional boundaries, these issues only add to the aforementioned spatial bias in reporting rates. By developing multiple models for diseases and vectors, it is more likely that the various epidemiological, ecological and social processes underpinning risk will be determined (Leach & Scoones, 2013) and, thus, resolving the issues of accessibility of data to the wider scientific community is critical.

WHAT ARE DISTRIBUTION MODELS AND HOW DO THEY RELATE TO NICHES?

As described above, distribution modelling involves matching geographical patterns in species (or community) occurrence with patterns in environmental factors to understand the mechanisms governing species distributions. A wide range of statistical methods are available to fit such models and their relative performance has been extensively reviewed elsewhere (Elith & Graham, 2009). These correlative or top-down approaches contrast with mechanistic, bottom-up approaches that mathematically describe the biological processes underpinning population performance (or transmission) and require detailed knowledge of species life cycles and ecological requirements (Rogers, 2006; Dormann *et al.*, 2012). For some of the more intensively studied diseases, sufficient knowledge is available to mechanistically model parts of the disease transmission process, such as the link between temperature and vector-borne disease transmission (Gething *et al.*, 2011b). However, even these detailed models are unable to capture all environmental drivers of disease transmission, let alone more complex socioeconomic factors, and consequently are best used in conjunction with correlative approaches (Hartemink *et al.*, 2011). Because such knowledge is generally lacking for most pathogen systems, correlative distribution models are often the most viable option for predicting and understanding patterns in transmission (Rogers, 2006; Hay *et al.*, 2013).

Although correlative distribution models have been widely used to make predictions of species' (and diseases') distributions, there is ongoing debate in the ecological literature regarding exactly what these distributions represent (Warren, 2012, 2013; McInerny & Etienne, 2013). The lack of consensus on how distribution modelling relates to niche concepts is probably caused not only by inconsistency of niche definitions, but also the variability in data, methods and scale across studies (Araújo & Guisan, 2006;

Soberon, 2007). Some studies have suggested that, if the realized niche is the subset of abiotic environmental space to which a species is restricted by biotic interactions (Hutchinson, 1957), then, by definition, known occurrence points used to generate distribution models represent the realized niche (Phillips *et al.*, 2006). Fundamental or environmental niches are only considered to be approximated by distribution models when occurrence data are drawn from a broad geographical extent (relative to the total range of the species in question) (Phillips *et al.*, 2006). Other studies caution against such generalizations (Elith & Leathwick, 2009), arguing that the different niches quantified using observed occurrences of species reflect an unknown conjunction of the environmental niches of the species, the biotic interactions they experience and the habitats available to species and colonized by them (Soberon, 2007).

These concerns are likely to be all the more relevant to pathogens, which are (1) intimately affected by biotic interactions, requiring hosts and sometimes vectors for persistence; (2) often under-recorded (infections of many pathogens occur subclinically in reservoir populations); and (3) increasingly being introduced into new geographical areas where they are not in equilibrium with their environments. If we want to infer niche components from distribution models for pathogens, then it may be particularly important to carefully consider adequacy of input data, sources of prediction uncertainty and test out ways of incorporating dispersal limitation and trophic and competitive interactions into models (Elith & Leathwick, 2009). Some of these aspects are dealt with below.

It is important to note that distribution modelling is unlikely to yield insights into environmental factors and processes governing distribution for all pathogens and reservoirs or vectors. Minimum prerequisites for the success of these techniques are that the organisms involved are sensitive to external environmental factors (e.g. poikilothermic arthropod vectors that are highly sensitive to temperature and moisture variability) and must themselves display spatial variation in occurrence. For example, amongst infectious diseases of humans, Hay *et al.* (2013) highlighted that endogenous infections caused by previously inapparent or dormant pathogens arising from the typical commensal microbial flora of humans showed little sustained global spatial variation in occurrence compared to pathogens in transmission categories that were inherently linked to the environment such as vector-borne pathogens, or soil or water contact pathogens. Ideally, basic information on life history should be available (vectors, reservoirs, hosts and routes of transmission) to enable identification of appropriate environmental drivers, and sufficient quantity and quality of occurrence data to enable robust inference.

RECENT IMPROVEMENTS IN ECOLOGICAL REALISM OF DISTRIBUTION MODELS

Accounting for dispersal and colonization processes during invasion

Correlative species distribution models (SDMs) assume that a species is in equilibrium with its environment (i.e. that the species has had sufficient time to sample all potentially suitable habitat in the region used to train the model) (Elith, Kearney & Phillips, 2010). This assumption clearly does not hold in the case of newly invading pathogens or vectors that are still in the process of expanding their distribution into all environmentally suitable locations. In such situations, the distribution of the invader at any point during the invasion will be strongly driven by dispersal and colonization processes in addition to environmental suitability (Vaclavik & Meentemeyer, 2012). In these cases, absence of the species may simply reflect that the species has yet to arrive, rather than environmental exclusion (Meentemeyer *et al.*, 2008; Jimenez-Valverde *et al.*, 2011).

Regardless of whether the aim of the distribution modelling study is to evaluate the potential distribution of an invading pathogen or vector (all sites at risk of invasion) (Benedict *et al.*, 2007) or the species' current distribution, it is crucial to understand and account for these dispersal processes. Both forms of modelling can provide crucial data to guide surveillance for invasions and to target mitigation options.

Very rarely, researchers have been able to conduct intensive fieldwork to track and model the actual distribution at a fine scale during invasion of a pathogen or vector (Rouget & Richarson, 2003; Meentemeyer *et al.*, 2008; Vaclavik & Meentemeyer, 2009, 2012; Václavík *et al.*, 2010). Studies of this kind on the sudden oak death pathogen, *Phytophthora ramorum*, in the USA have revealed that, if true absence data and dispersal parameters are not incorporated into the model, predictions of the actual distribution are less accurate and tend to overpredict the actual range of invasion (Václavík *et al.*, 2010). The spatial autocorrelation caused by colonized sites tending to cluster around initial invasion foci will inflate not only model accuracy, but also the estimated explanatory power of environmental predictors (particularly when these are distally related to the requirements of the focal species) and underestimate uncertainty in model parameters (Dormann *et al.*, 2007).

Various methods have been proposed for incorporating dispersal processes into distribution models.

These include simple dispersal kernels and distance decay formula where propagule pressure is assumed to decline with distance from already infected sites (Allouche *et al.*, 2008; Meentemeyer *et al.*, 2008; Leighton *et al.*, 2012) and more complex mechanistic simulation models of species dispersal to new locations (Sedda *et al.*, 2012). Often, the knowledge of species-specific dispersal processes and location of already infected sites required to implement these methods is unavailable. This is typical for invading pathogens (where undetected circulation in reservoir hosts may be common) and vectors. In these cases, dispersal limitations have often been accounted for in correlative distribution models, albeit in a less explicit way, by modelling residual spatial autocorrelation. The application of spatially-explicit models to invading pathogens has been reviewed and evaluated on the sudden oak death system at multiple scales (Vaclavik, Kupfer & Meentemeyer, 2012).

Inference about the dispersal mechanisms constraining the distributions of invasive species and predictions of those distributions will both be improved when models incorporate connectivity measures that are related to ecological knowledge. For example, models accounting for landscape or habitat structure (Ellis, Václavík & Meentemeyer, 2010) or pathways of dispersal such as wind (Sedda *et al.*, 2012), host migration (Vaclavik *et al.*, 2012) and human-mediated movement (Gilbert *et al.*, 2005) are likely to have greater predictive power than those which assume connectivity to simply be a function of distance.

When the goal is to predict the potential distribution of pathogens and vectors introduced into new areas, it should be noted that the full environmental niche will not be effectively captured with occurrence data from early in the invasion because full occupancy of suitable habitats is prevented by dispersal constraints (Vaclavik & Meentemeyer, 2012). Consequently, when SDMs are developed in the early stages of invasion, they tend to under-predict the potential range (Vaclavik & Meentemeyer, 2012). The same is likely to be true when surveillance of pathogens or vectors is sparse or geographically biased across the species' true distribution (Elith *et al.*, 2010). Both of these problems may be alleviated to some degree if the available occurrence data are distributed across heterogeneous environments and models employ environmental predictors that are proximally related to species' requirements. The stage of invasion at which occurrence data are drawn not only impacts on model performance, but also on the reliability of accuracy statistics. Metrics of predictive accuracy, such as the very widely used area under the curve, are strongly influenced by variation in the relative occurrence area (i.e. the ratio between the extent of the species' true occurrence and the extent of study region) (Lobo, Jimenez-Valverde & Real, 2008; Elith *et al.*, 2010; Hijmans, 2012). A number of practical, heuristic approaches have been proposed to account for these issues, including the upweighting of occurrence records with few neighbours in geographical space (to compensate for records being relatively dense in areas of first appearance in a region and sparse in newly invaded areas) and testing out different options for representing 'background' locations (for presence-only methods) (Elith *et al.*, 2010), although these have yet to be widely adopted in distribution models of invading vectors or pathogens.

Alternatives to parameterizing models with occurrence data from early in an invasion, include building correlative models based on data from the species' native range and projecting them to potential introduction sites (Benedict *et al.*, 2007) or developing mechanistic models based on physiological responses of organisms to their environment (Elith *et al.*, 2010; Caminade *et al.*, 2012). The former requires careful consideration of whether and where conditions in the extrapolation area are novel either with respect to their combinations of abiotic conditions (e.g. through multivariate environmental distance measures such as the mahalanobis distance) or biotic interactions in terms of natural enemies or competitors and whether the species may have shifted its environmental niche during invasion (Medley, 2010). The approaches are also likely to be of little use in the case of novel pathogens, such as emerging viruses for which there is no clear 'native range' to model. Using reciprocal distribution modelling, Medley (2010) found that the distribution of *Aedes albopictus*, a key mosquito vector of Dengue, was underpredicted in introduced regions of north and south America and Europe based on the environmental niche quantified from the native range, suggestive of a niche shift. Subsequently, methods of quantifying niche overlap have been developed to account for the potential difference in representation of environmental conditions (between native and introduced ranges or between species) (Broennimann *et al.*, 2012). Peterson (2011) also points out that, when niches are parameterized in highly dimensional environmental space, a niche shift is more likely to be detected and it is suggested that this is a likely reason why Medley (2010) found a niche shift during *Aedes albopictus* invasion, whereas Benedict *et al.* (2007) found niche stasis for the same system. This again highlights the importance of using predictors that are tightly related a priori to the known ecological requirements of the focal species and also the need to avoid over-fitting when using species distribution modelling to compare environmental niches.

Understanding biotic interactions influencing transmission

Despite their importance in driving species' distributions, very few species distribution modelling studies have explicitly incorporated biotic interactions (Araújo & Luoto, 2007; Elith & Leathwick, 2009; Wisz *et al.*, 2013). This is particularly problematic when extrapolating models to predict potential distributions under invasion or climate change because natural enemies or competitors may have far reaching effects, particularly if novel combinations of species occur (Elith & Leathwick, 2009). Predictions of actual distributions will be inaccurate, especially for species where hosts are critical for persistence (Elith & Leathwick, 2009).

Possibly because vector-borne pathogens are known to be so intimately constrained by biotic interactions, it has become relatively common in correlative distribution models of vector-borne pathogens or blood-sucking arthropod vectors to add the occurrence or abundance of host or vector species as predictors alongside abiotic variables, resulting in significant improvements in model accuracy (Conte *et al.*, 2007; Purse *et al.*, 2012), insights into the role of hosts as regulators of vector population abundance (Cumming, 1999) and more refined estimates of potential transmission extent under future environmental change (Daszak *et al.*, 2013).

Other studies have developed individual species distribution models for all interacting species potentially involved in transmission in a focal region and then visualized (Peterson & Shaw, 2003; Purse *et al.*, 2007) or explicitly quantified (Rödder, Schulte & Toledo, 2013) overlap of environmental niches in environmental space defined by key abiotic variables. Sometimes, these approaches can allow inferences to be made about the relative role of particular vectors or reservoir species in current transmission when laboratory or field data on infection rates are difficult to obtain. For example, during the emergence of a midge-borne sheep disease, bluetongue, in southern Europe in the last century, it was shown that transmission had spread outside the environmental niche of the historical midge vector, *Culicoides imicola*, into cooler and wetter areas as a result of transmission by palearctic vector groups (Purse *et al.*, 2007, 2008). In the Americas, Rödder *et al.* (2013) found high niche overlap between the fungus, *Batrachochytrium dendrobatidis*, a major cause of amphibian declines, and an invasive alien bullfrog, *Lithobates catesbeianus*, strengthening hypotheses that this frog species is a major carrier. Again, the fungus was found in environmental space in which this bullfrog was not present, implicating native wild frogs also in transmission (Fig. 1).

Other studies have projected environmental niches of the different species involved in the 'epidemiological triangle' of a pathogen system onto future climate conditions aiming to understand the relative sensitivity of member species to environmental change and how and where overlap of hosts and/or vectors critical to transmission may be maintained (Daszak *et al.*, 2013; Pickles *et al.*, 2013). For example, Pickles *et al.* (2013) found a geographical mismatch in areas of habitat predicted to be suitable in the future for a free-living nematode meningeal parasite (*Parelaphostrongylosis tenuis*) of deer vs. those predicted to be suitable for its gastropod intermediate host and were consequently able to identify areas of North America where the resulting disease (parelaphostrongylosis) would be expected to expand or contract.

The approaches thus far described are most feasible for situations where very few species are involved in host-pathogen–vector interactions. In scenarios where multiple hosts and vectors are involved in transmission, with unknown roles, community modelling approaches may hold more promise for understanding biotic interactions (Kissling *et al.*, 2012). Using biological records, Stephens *et al.* (2009) analyzed the co-distribution across Mexico of over 400 mammal species and sand fly species that might be involved in transmission of *Leishmaniasis* to construct inter-species interaction networks, by identifying and ranking the extent to which pairs of vectors and reservoir mammals were positively associated in geographical space. For example, four particular rodent species, previously found to be infected with *Leishmania* in the field, were identified as being associated geographically (in the Yucatan peninsula) with a wide spectrum of vector species, offering high potential for parasite exchange. It is acknowledged that sampling bias arising from use of clustered, presence-only data may have influenced their results. This approach was later extended to examine co-distribution of vectors and mammal reservoirs with particular land cover types aiming to identify species assemblages that (1) posed a high risk of transmission to humans by virtue of being present in both natural habitats and human settlements or (2) posed an additional risk of dispersing the disease by virtue of also being found in cropland habitats that form corridors between natural habitats and human habitation (González-Salazar & Stephens, 2012).

However, by modelling species interactions solely from geographical co-occurence or combining abiotic and biotic predictors in single species distribution models, the influence of species interactions and environmental covariates may be confounded

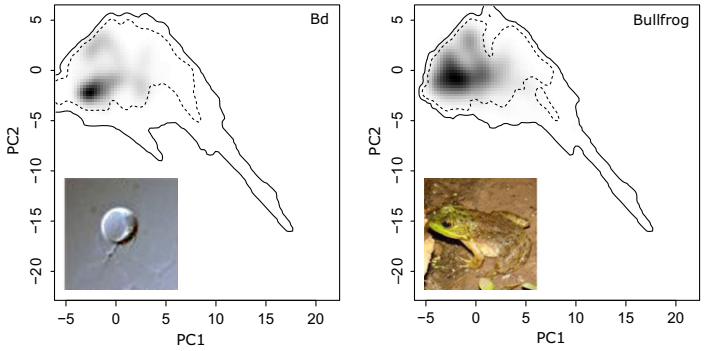

Figure 1. Comparisons of the realized climatic niches of *Batrachochytrium dendrobatidis* (Bd) and the invasive bullfrog (*Lithobates catesbeianus*). The climatic niches are defined by axes from a principal components analysis (PCA) of bioclimatic variables across the study area. Darker areas in the PCA plots indicate a higher density of species records, for which the available climate spaces within the general area of occurrence are indicated. The high niche overlap between the fungus and the invasive bullfrog strengthens the hypotheses that *L. catesbeianus* is a major carrier. Again, the fungus was found in environmental space in which this bullfrog was not present, implicating native wild frogs also in transmission. Modified with permission from Rödder *et al.* (2013).

(Kissling *et al.*, 2012; Pollock *et al.*, 2014). Explanatory power might be spuriously ascribed to a host species simply because the pathogen and host species are independently affected by a missing but unknown environmental predictor. Such similarities in environmental responses of species can be accommodated in multispecies SDMs (and the unobserved environmental gradients driving species co-occurrence may even be inferred (Harris, 2015; Ovaskainen, Hottola & Siitonen, 2010) but, where biotic interactions are the primary focus of investigation, new methods that model interactions between multiple species explicitly using error matrices in multivariate regression models of spatial co-occurrence data may hold more promise (Kissling *et al.*, 2012; Pollock *et al.*, 2014). When such techniques were applied to UK mosquito communities on grazed wetlands, key predatory taxa constraining mosquito distributions alongside temperature and water level variability were revealed (N. Golding, M.A. Nunn & B.V. Purse, in review).

For the most complete understanding of the relative role of vector and reservoir species in transmission from distribution or community models, it is necessary to include data on the occurrence of all potential interacting species in the focal region (Daszak *et al.*, 2013; Sedda *et al.*, 2014). Daszak *et al.* (2013) note that the distribution of wildlife reservoir hosts has been rarely incorporated into distribution models. Indeed, for vector-borne pathogens affecting humans and animals, densities of people and livestock may have been more commonly considered as explanatory predictors of distribution than wildlife hosts, possibly because they tend to be mapped more accurately (Robinson *et al.*, 2014).

VALUE OF MAPPING OF VECTORS, RESERVOIRS, AND DISEASES TO DISEASE MANAGERS AND POLICY MAKERS

Although epidemiological applications of species distribution models are increasingly reported in the scientific literature, particularly over the last 5 years (Fig. 2), direct impacts of predictions and risk maps on policy and mitigation measures are documented only rarely (Leach & Scoones, 2013). Here, with reference to some recent case study models with policy impacts, we highlight some of the factors that may constrain or promote use of outputs of correlative disease distribution models by stakeholders and policy makers, with a view to improving these linkages.

One conceptual problem limiting application of distribution models may be that vector or pathogen occurrence may only be indirectly related to intensity of transmission or disease impacts. Additional factors such as the relative abundance and composition of vectors and hosts, host recovery rates and infectious periods, and opposing impacts of temperature on the transmission cycle also come into play. This complexity and nonlinearity is arguably better captured by mechanistic modelling frameworks that explicitly consider mathematical relationships between demographic rates and environmental factors (e.g. the basic reproduction number framework) (Rogers, 2006). However, the epidemiological data necessary to construct such a model are simply unavailable for a wide range of diseases. Although correlative approaches to estimating epidemiologically relevant metrics are more widely applicable, the fact that underlying processes are not explicitly identified

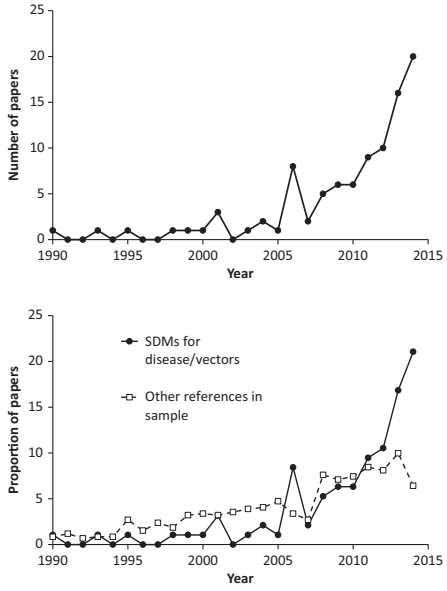

Figure 2. The increasing frequency of papers on species distribution models (SDMs) and vectors or disease over the last 15 years ($N = 95$). A, the number of SDM and disease/vector papers each year in our sample. B, the proportion of papers from the last 15 years appearing in each year for papers on SDMs and disease/vectors vs. the papers that were not on this topic ($N = 592$) from our search sample. Methodology: the search term 'species distribution model disease' was entered into PubMed (in October 2014) and yielded 687 references dated between 1990 and 2014 inclusive. The abstracts of the references were searched to confirm whether species distribution modelling had been used to predict the distribution of a vector or pathogen. Ninety-five of these references conformed to this criteria. It should be noted that this search probably encompasses only small subset of the total SDM and disease/vector papers that would have been yielded had other databases and other terms also been employed (e.g. ecological niche model + disease or vector).

may hamper the design of mitigation measures (Braks *et al.*, 2014).

As a result of flexibility in the number and type of predictors that can be accommodated, correlative models of disease occurrence have sometimes performed better than mechanistic models at capturing the social and ecological context of transmission, resulting in high impacts on policy decisions (Leach & Scoones, 2013). For example, Gilbert *et al.* (2005) found, amongst the environmental, topographic and social predictors considered, that frequency of cattle movements was the most important explanatory predictor of occurrence patterns of bovine tuberculosis in Great Britain (in 2002 and 2003). The study under consideration formed part of the evidence base for introducing a policy of cattle movement controls to control disease spread (Godfray *et al.*, 2013).

For highly pathogenic avian influenza (H5N1) outbreaks in Thailand (Gilbert *et al.*, 2006), correlative models were used to link disease prevalence with the density of free-ranging ducks and, in turn, to a particular rice-paddy duck farming system, common in low elevation farming areas. Consequently, surveillance and control were directed towards domestic ducks, contributing to the eradication of H5N1 from Thailand without vaccination. In this case, correlative models translated local epidemiological knowledge of risk factors in a quantitative way (i.e. risk maps) that could be communicated easily to decision makers to influence policy (M. Gilbert, pers. comm.).

Decisions on mitigation measures often need to be made when the threat of disease is still only modest or before definitive information is available (Leach & Scoones, 2013). A key advantage to policy makers of correlative distribution models vs. mechanistic models may be the potential to provide an explanation of disease patterns early on in an epidemic of an exotic pathogen because the former may be parameterized from initial or historical outbreak data (with care; see the earlier section on Recent improvements in ecological realism of distribution models) before detailed ecological knowledge is available from the invaded region. From a relatively small number of previous outbreak events, a correlative distribution model was recently used to map locations in sub-Saharan Africa where Ebola virus is most likely to spill over from its zoonotic reservoir in and trigger outbreaks in human populations (Pigott *et al.*, 2014b) and to identify regions where outbreaks have so far not occurred but may do in the future. These quantitative maps will facilitate the development and targeting of diagnostic resources to ensure that future outbreaks of the virus are detected and controlled in their early stages before they can accelerate into large-scale public health emergencies such as that seen in West Africa in 2014 (Farrar & Piot, 2014).

Developing mechanistic models for emerging pathogens in new ecological contexts in a timely fashion can be much harder; for example, the midge-borne bluetongue virus, which causes devastating disease in livestock, was circulating in Europe for approximately 10 years before sufficient information was available to parameterize simplified basic reproduction number maps, indicating the geographically variable risk of establishment, at a national (Hartemink *et al.*, 2009) or continental scale (Guis *et al.*, 2012).

One of the most robust uses of correlative distribution modelling has been in motivating and spatial targeting of surveillance for vectors or pathogens, particularly early in invasion when information on the geographical extent of the problematic species

may be very poor. For example, in West Africa, an exotic cattle tick species, *Rhipicephalus microplus*, that transmits *Babesia* and *Anaplasma* pathogens was introduced from Brazil (probably on imported cattle) approximately 8 years ago and has established viable populations in Benin and Cote D'Ivoire. A research project (TICKRISK) was established linking field estimates of occurrence of *R. microplus* with modelling of the environmental niche (De Clercq *et al.*, 2012, 2013). The resulting prediction maps were used by the stakeholders to lobby the Ministry of Agriculture successfully for (1) an increased budget to tackle problems related to ticks in general and the invasive tick in particular and (2) a moratorium on import of cattle from Brazil. In Benin, it is not clear whether it was the research project as a whole or specifically the use of prediction maps that was responsible for these decisions, although the maps were considered as strong images. In Burkina Faso, the prediction maps resulted in a programme of directed sampling in the most suitable areas for this tick, where it was indeed found. After confirmation of the presence of the invasive tick, a nationwide survey was undertaken (E. De Clercq pers. comm.).

It is rarer at present for species distribution model outputs and maps to be used for detailed spatial targeting of control measures at local scales; but see Dicko *et al.* (2014). This is probably because their spatial resolution tends to be coarse relative to the scale at which control or surveillance is conducted. Local scale abundance or prevalence patterns may be harder to predict than regional scale occurrence patterns as a result of the interplay of wide ranging climate, landscape, social and ecological factors for which it may be difficult to identify fine resolution spatial proxies (Purse *et al.*, 2012).

FUTURE CHALLENGES AND OPPORTUNITIES FOR TRACKING THE SPREAD AND IMPACTS OF DISEASES WITH BIOLOGICAL RECORDS AND DISTRIBUTION MODELLING

The high value and lack of centralised global databases for human pathogens and vectors has been highlighted. Considering how records for pathogens and vectors are generated, early warning systems and media sources such as Twitter are providing novel, online, often real-time data sources on disease occurrence that may prove invaluable for disease modelling (Sedda *et al.*, 2014).

A key example is HealthMap (http://healthmap.org/en), as produced by the Computational Epidemiology Group of the Children's Hospital, Boston, MA, USA), which monitors and maps health alerts worldwide from eleven sources of news and reports, including ProMed, Google News, WHO, OIE and ECDC (Brownstein *et al.*, 2008). To foster the development of multiple models that offer different perspectives on epidemiological, ecological and social processes (Leach & Scoones, 2013), or predictions at different geographical/temporal scales, improvements in data accessibility and integration are still required. An understanding of the evolutionary processes involved in disease and vector spread will be facilitated if databases can link molecular and genomics data for pathogens and vectors with occurrence recording (Porretta *et al.*, 2012; Pybus *et al.*, 2012).

Occurrence data from such sources are likely to be subject to significant spatial biases in reporting rates (Isaac & Pocock, 2015). Statistical methodologies to account for such biases in the recording of vectors and pathogens are beginning to be developed. Other key challenges for translating species distribution models built using such data into risk maps of value for risk governance are: (1) understanding how vector population abundance or pathogen transmission intensity scale with predicted occurrence (a problem currently being addressed in species distribution models in ecology (Yañez-Arenas *et al.*, 2014); (2) understanding how human behaviour influences the recorded occurrence of pathogens; and (3) quantifying the uncertainty in risk map outputs in a manner that is comparable between modelling approaches (Sedda *et al.*, 2014). We argue that use of ecologically realistic species distribution modelling methods that incorporate dispersal constraints, biotic interactions and social factors will help to meet all of these challenges.

Where there is a lack of detailed ecological or epidemiological information in an invaded region or where local ecosystem and social processes underpin transmission, correlative models of disease or vector occurrence may offer considerable advantages over mechanistic modelling approaches. Where local or regional epidemiology is better understood, insights from correlative species distribution models on spatial patterns in vectors or hosts may be incorporated into mechanistic models of transmission to better understand processes of establishment and spread (Hartemink *et al.*, 2011; Meentemeyer *et al.*, 2011).

Overall, Leach & Scoones (2013) caution against thinking of modelling as 'an objective, neutral scientific exercise that linearly informs policy' and note that models themselves are shaped by social, political and cultural norms and by prevailing policy narrative. Modellers must interact with stakeholders and policy makers to understand trade-offs in health priorities, key field-level risk factors and the optimal scale and format for communicating maps and

results. The impacts of this approach are clear in several of the case studies reported above where correlative models were used to translate local knowledge of ecosystem and social risk factors into maps and other formats of benefit to policy makers. Ultimately, the challenge will be to combine biological records and ecological modelling with the aim of analyzing the impacts of disease and disease management at a whole ecosystem level to allow decision makers to evaluate potential trade-offs between ecosystem services (Cheatham *et al.*, 2009; Boyd *et al.*, 2013).

ACKNOWLEDGEMENTS

BVP was funded by the Biotechnology and Biological Sciences Research Council (BBSRC), Scottish Government and the Department for International Development under grant number BB/H009167/1, with additional support provided by the Natural Environment Research Council. NG was funded by grants from the Bill & Melinda Gates Foundation (OPP1053338 & OPP1093011). We thank Eva De Clercq, Marius Gilbert, Nienke Hartemink, David Rogers, Sophie Vanwambeke and William Wint for sharing their experiences of translating risk maps into policy, as well as Colin Harrower for preparing the information from the Biological Records Centre for use in Table 1 and assistance with figures.

REFERENCES

Allouche O, Steinitz O, Rotem D, Rosenfeld A, Kadmon R. 2008. Incorporating distance constraints into species distribution models. *Journal of Applied Ecology* **45:** 599–609.

Araújo MB, Guisan A. 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography* **33:** 1677–1688.

Araújo MB, Luoto M. 2007. The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography* **16:** 743–753.

Benedict MQ, Levine RS, Hawley WA, Lounibos LP. 2007. Spread of the tiger: global risk of invasion by the mosquito *Aedes albopictus*. *Vector Borne and Zoonotic Diseases* **7:** 76–85.

Boyd IL, Freer-Smith PH, Gilligan CA, Godfray HCJ. 2013. The consequence of tree pests and diseases for ecosystem services. *Science* **342:** 1235773.

Braks M, van Ginkel R, Wint W, Sedda L, Sprong H. 2014. Climate change and public health policy: translating the science. *International Journal of Environmental Research and Public Health* **11:** 13–29.

Broennimann O, Fitzpatrick MC, Pearman PB, Petitpierre B, Pellissier L, Yoccoz NG, Thuiller W, Fortin M, Randin C, Zimmermann NE, Graham CH, Guisan A. 2012. Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography* **21:** 481–497.

Brownstein JS, Freifeld CC, Reis BY, Mandl KD. 2008. Surveillance Sans Frontières: Internet-based emerging infectious disease intelligence and the HealthMap Project. *PLoS Medicine* **5:** e151.

Caminade C, Medlock JM, Ducheyne E, McIntyre KM, Leach S, Baylis M, Morse AP. 2012. Suitability of European climate for the Asian tiger mosquito *Aedes albopictus*: recent trends and future scenarios. *Journal of the Royal Society Interface* **9:** 2708–2717.

Chaves LF, Cohen JM, Pascual M, Wilson ML. 2008. Social exclusion modifies climate and deforestation impacts on a vector-borne disease. *PLoS Neglected Tropical Diseases* **2:** e176.

Cheatham MR, Rouse MN, Esker PD, Ignacio S, Pradel W, Raymundo R, Sparks AH, Forbes GA, Gordon TR, Garrett KA. 2009. Beyond yield: plant disease in the context of ecosystem services. *Phytopathology* **99:** 1228–1236.

Conte A, Goffredo M, Ippoliti C, Meiswinkel R. 2007. Influence of biotic and abiotic factors on the distribution and abundance of *Culicoides imicola* and the *Obsoletus* Complex in Italy. *Veterinary Parasitology* **150:** 333–344.

Cumming GS. 1999. Host distributions do not limit the species ranges of most African ticks (Acari: Ixodidae). *Bulletin of Entomological Research* **89:** 303–327.

Cumming GS. 2002. Comparing climate and vegetation as limiting factors for species ranges of African ticjs. *Ecology* **83:** 255–268.

Daszak P, Zambrana-Torrelio C, Bogich TL, Fernandez M, Epstein JH, Murray KA, Hamilton H. 2012. Interdisciplinary approaches to understanding disease emergence: the past, present, and future drivers of Nipah virus emergence. *Proceedings of the National Academy of Sciences of the United States of America* **110:** 3681–3688.

De Clercq EM, Vanwambeke SO, Sungirai M, Adehan S, Lokossou R, Madder M. 2012. Geographic distribution of the invasive cattle tick *Rhipicephalus microplus,* a country-wide survey in Benin. *Experimental and Applied Acarology* **58:** 441–452.

De Clercq EM, Estrada-Pena A, Adehan S, Madder M, Vanwambeke SO. 2013. An update on distribution models for *Rhipicephalus microplus* in West Africa. *Geospatial Health* **8:** 301–308.

Dicko AH, Lancelot R, Seck MT, Guerrini L, Sall B, Lo M, Vreysen MJ, Lefrancois T, Fonta WM, Peck SL, Bouyer J. 2014. Using species distribution models to optimize vector control in the framework of the tsetse eradication campaign in Senegal. *Proceedings of the National Academy of Sciences of the United States of America* **111:** 10149–10154.

Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G, Davies RG, Hirzel A, Jetz W, Kissling WD, Kühn I, Ohlemüller R, Peres-Neto PR, Reineking B, Schröder B, Schurr FM, Wilson R. 2007. Methods to account for spatial autocorrelation in the

analysis of species distributional data: a review. *Ecography* **30:** 609–628.

Dormann CF, Schymanski SJ, Cabral J, Chuine I, Graham C, Hartig F, Kearney M, Morin X, Römermann C, Schröder B, Singer A. 2012. Correlation and process in species distribution models: bridging a dichotomy. *Journal of Biogeography* **39:** 2119–2131.

Eisen L, Eisen RJ. 2010. Using geographic information systems and decision support systems for the prevention and control of vector-borne diseases. *Annual Review of Entomology* **56:** 41–61.

Elith J, Graham CH. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography* **32:** 66–77.

Elith J, Leathwick JR. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution and Systematics* **40:** 677–697.

Elith J, Kearney M, Phillips S. 2010. The art of modelling range-shifting species. *Methods in Ecology and Evolution* **1:** 330–342.

Ellis AM, Václavík T, Meentemeyer RK. 2010. When is connectivity important? A case study of the spatial pattern of sudden oak death. *Oikos* **119:** 485–493.

EPPO. 2014. *PQR – EPPO plant quarantine data retrieval system, Version 5.3.1*. 2014-09-08. Availabe at: http://www.eppo.int.

Farrar JJ, Piot P. 2014. The ebola emergency – immediate action, ongoing strategy. *New England Journal of Medicine* **371:** 1545–1546.

Foley D, Wilkerson R, Birney I, Harrison S, Christensen J, Rueda L. 2010. MosquitoMap and the Mal-area calculator: new web tools to relate mosquito species distribution with vector borne disease. *International Journal of Health Geographics* **9:** 11.

Foley DH, Wilkerson RC, Dornak LL, Pecor DB, Nyari AS, Rueda LM, Long LS, Richardson JH. 2012. SandflyMap: leveraging spatial data on sand fly vector distribution for disease risk assessments. *Geospatial Health* **6:** 25–30.

Gething PW, Patil AP, Smith DL, Guerra CA, Elyazar IRF, Johnston GL, Hay SI. 2011a. A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *Malaria Journal* **10:** 378.

Gething PW, Van Boeckel TP, Smith DL, Guerra CA, Patil AP, Snow RW, Hay SI. 2011b. Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax*. *Parasites and Vectors* **4:** 92.

Gilbert M, Mitchell A, Bourn D, Mawdsley J, Clifton-Hadley R, Wint GRW. 2005. Cattle movements and bovine tuberculosis in Great Britain. *Nature* **435:** 491–496.

Gilbert M, Chaitaweesub P, Parakamawongsa T, Premashthira S, Tiensin T, Kalpravidh W, Wagner H, Slingenbergh J. 2006. Free-grazing ducks and highly pathogenic avian influenza, Thailand. *Emerging Infectious Diseases* **12:** 227–234.

Godfray HCJ, Donnelly CA, Kao RR, Macdonald DW, McDonald RA, Petrokofsky G, Wood JLN, Woodroffe R, Young DB, McLean AR. 2013. A restatement of the natural science evidence base relevant to the control of bovine tuberculosis in Great Britain. *Proceedings of the Royal Society of London Series B, Biological Sciences* **280:** doi:10.1098/rspb.2013.1634. Print 2013 Oct 7.

González-Salazar C, Stephens CR. 2012. Constructing ecological networks: a tool to infer risk of transmission and dispersal of leishmaniasis. *Zoonoses and Public Health* **59:** 179–193.

Guis H, Caminade C, Calvete C, Morse AP, Tran A, Baylis M. 2012. Modelling the effects of past and future climate on the risk of bluetongue emergence in Europe. *Journal of the Royal Society, Interface* **9:** 339–350.

Guisan A, Zimmermann NE. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* **135:** 147–186.

Harris DJ. 2015. Generating realistic assemblages with a joint species distribution model. *Methods in Ecology and Evolution* **6:** 465–473.

Hartemink NA, Purse BV, Meiswinkel R, Brown HE, De Koeijer AA, Elbers ARW, Boenderd GJ, Rogers DJ, Heesterbeek JAP. 2009. Mapping the basic reproduction number (R_0) for vector-borne diseases:a case study on bluetongue virus. *Epidemics* **1:** 153–161.

Hartemink NA, Vanwambeke SO, Heesterbeek H, Rogers DJ, Morley DW, Pesson B, Davies C, Mahamdallie S, Ready P. 2011. Integrated mapping of establishment risk for emerging vector-borne infections: a case study of canine leishmaniasis in southwest France. *PLoS ONE* **6:** e20817.

Hay SI, Battle KE, Pigott DM, Smith DL, Moyes CL, Bhatt S, Brownstein JS, Collier N, Myers MF, George DB, Gething PW. 2013. Global mapping of infectious disease. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* **368:** 20120250.

Hijmans RJ. 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology* **93:** 679–688.

Hutchinson GE. 1957. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology* **22:** 415–427.

Isaac NJB, Pocock M. 2015. Bias and information in biological records. *Biological Journal of the Linnean Society* **115:** 522–531.

Jameson LJ, Medlock JM. 2011. Tick surveillance in Great Britain. *Vector-Borne and Zoonotic Diseases* **11:** 403–412.

Jenkins CN, Pimm SL, Joppa LN. 2013. Global patterns of terrestrial vertebrate diversity and conservation. *Proceedings of the National Academy of Sciences of the United States of America* **110:** E2602–E2610.

Jimenez-Valverde A, Peterson AT, Soberon J, Overton JM, Aragon P, Lobo JM. 2011. Use of niche models in invasive species risk assessments. *Biological Invasions* **12:** 2785–2797.

Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P. 2008. Global trends in emerging infectious diseases. *Nature* **451:** 990–993.

Kissling WD, Dormann CF, Groeneveld J, Hickler T, Kühn I, McInerny GJ, Montoya JM, Römermann C, Schiffers K, Schurr FM, Singer A, Svenning J-C,

Zimmermann NE, O'Hara RB. 2012. Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography* **39:** 2163–2178.

Leach M, Scoones I. 2013. The social and political lives of zoonotic disease models: narratives, science and policy. *Social Science & Medicine* **88:** 10–17.

Leighton PA, Koffi JK, Pelcat Y, Robbin Lindsay L, Ogden NH. 2012. Predicting the speed of tick invasion: an empirical model of range expansion for the Lyme disease vector *Ixodes scapularis* in Canada. *Journal of Applied Ecology* **49:** 457–464.

Lindsay SW, Hole DG, Hutchinson RA, Richards SA, Willis SG. 2010. Assessing the future threat from vivax malaria in the United Kingdom using two markedly different modelling approaches. *Malaria Journal* **9:** 70–70.

Lobo JM, Jimenez-Valverde A, Real R. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* **17:** 145–151.

Maes D, Isaac NJB, Harrower C, Collen B, Roy DB, van Strien AJ. 2015. The use of opportunistic data for IUCN Red List assessments. *Biological Journal of the Linnean Society* **115:** 690–706.

McInerny GJ, Etienne RS. 2013. 'Niche' or 'distribution' modelling? A response to Warren. *Trends in Ecology & Evolution* **28:** 191–192.

Medley KA. 2010. Niche shifts during the global invasion of the Asian tiger mosquito, *Aedes albopictus* Skuse (Culicidea), revealed by reciprocal distribution models. *Global Ecology and Biogeography* **19:** 122–133.

Meentemeyer RK, Anacker BL, Mark W, Rizzo DM. 2008. Early detection of emerging forest disease using dispersal estimation and ecological niche modelling. *Ecological Applications* **18:** 377–390.

Meentemeyer RK, Cunniffe NJ, Cook AR, Filipe JAN, Hunter RD, Rizzo DM, Gilligan CA. 2011. Epidemiological modeling of invasion in heterogeneous landscapes: spread of sudden oak death in California (1990–2030). *Ecosphere* **2:** art. 17. Available at: http://dx.doi.org/10.1890/ES10-00192.1

OIE. 2014. *World Animal Health Information System*. Office International des Epizooties. Paris, France.

Ovaskainen O, Hottola J, Siitonen J. 2010. Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology* **91:** 2514–2521.

Pautasso M, Doring TF, Garbelotto M, Pellis L, Jeger MJ. 2012. Impacts of climate change on plant diseases – opinions and trends. *European Journal of Plant Pathology* **133:** 295–313.

Peterson AT. 2011. Ecological niche conservatism: a time-structured review of evidence. *Journal of Biogeography* **38:** 817–827.

Peterson AT, Shaw J. 2003. Lutzomyia vectors for cutaneous leishmaniasis in Southern Brazil: ecological niche models, predicted geographic distributions, and climate change effects. *International Journal for Parasitology* **33:** 919–931.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190:** 231–259.

Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick JR, Ferrier S. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* **19:** 181–197.

Pickles RS, Thornton D, Feldman R, Marques A, Murray DL. 2013. Predicting shifts in parasite distribution with climate change: a multitrophic level approach. *Global Change Biology* **19:** 2645–2654.

Pigott DM, Bhatt S, Golding N, Duda KA, Battle KE, Brady OJ, Messina JP, Balard Y, Bastien P, Pratlong F, Brownstein JS, Freifeld C, Mekaru SR, Gething PW, George DB, Myers MF, Reithinger R, Hay SI. 2014a. Global distribution maps of the leishmaniases. *eLife* doi.org/10.7554/eLife.02851: .

Pigott DM, Golding N, Mylne A, Huang Z, Henry AJ, Weiss DJ, Brady OJ, Kraemer MU, Smith DL, Moyes CL, Bhatt S, Gething PW, Horby PW, Bogoch II, Brownstein JS, Mekaru SR, Tatem AJ, Khan K, Hay SI. 2014b. Mapping the zoonotic niche of Ebola virus disease in Africa. *eLife* doi.org/10.7554/eLife.04395:.

Pollock LJ, Tingley R, Morris WK, Golding N, O'Hara RB, Parris KM, Vesk PA, McCarthy MA. 2014. Understanding co-occurrence by modelling species simultaneously with a joint species distribution model (JSDM). *Methods in Ecology and Evolution* **5:** 397–406.

Powney GD, Isaac NJB. 2015. Beyond maps: a review of the applications of biological records. *Biological Journal of the Linnean Society* **115:** 532–542.

Porretta D, Mastrantonio V, Bellini R, Somboon P, Urbanelli S. 2012. Glacial history of a modern invader: phylogeography and species distribution: modelling of the Asian tiger mosquito *Aedes albopictus*. *PLoS ONE* **7:** e44515.

Purse BV, McCormick BJJ, Mellor PS, Baylis M, Boorman J, Borras D, Burgu I, Capela R, Caracappa S, Collantes F, De Liberato C, Delgado JA, Denison E, Georgiev G, Harak ME, De La Rocque S, Lhor Y, Lucientes J, Mangana O, Miranda MA, Nedelchev N, Nomikou K, Ozkul A, Patakakis M, Pena I, Scaramozzino P, Torina A, Rogers DJ. 2007. Incriminating bluetongue virus vectors with climate envelope models. *Journal of Applied Ecology* **44:** 1231–1242.

Purse BV, Brown HE, Harrup L, Mertens PPC, Rogers DJ. 2008. Invasion of bluetongue and other orbivirus infections into Europe: the role of biological and climatic processes. *Revue Scientifique Et Technique-Office International Des Epizooties* **27:** 427.

Purse BV, Falconer D, Sullivan MJ, Carpenter S, Mellor PS, Piertney SB, Mordue Luntz AJ, Albon S, Gunn GJ, Blackwell A. 2012. Impacts of climate, host and landscape factors on *Culicoides* species in Scotland. *Medical and Veterinary Entomology* **26:** 168–177.

Purse BV, Graeser P, Searle K, Edwards C, Harris C. 2013. Challenges in predicting invasive reservoir hosts of emerging pathogens: mapping *Rhododendron ponticum* as a

foliar host for *Phytophthora ramorum* and *Phytophthora kernoviae* in the UK. *Biological Invasions* **15:** 529–545.

Pybus OG, Suchard MA, Lemey P, Bernardin FJ, Rambaut A, Crawford FW, Gray RR, Arinaminpathy N, Stramer SL, Busch MP, Delwart EL. 2012. Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proceedings of the National Academy of Sciences of the United States of America* **109:** 15066–15071.

Robinson TP, Wint GRW, Conchedda G, Van Boeckel TP, Ercoli V, Palamara E, Gilbert M. 2014. Mapping the global distribution of livestock. *PLoS ONE* **9:** e96084.

Rödder D, Schulte U, Toledo LF. 2013. High environmental niche overlap between the fungus Batrachochytrium dendrobatidis and invasive bullfrogs (*Lithobates catesbeianus*) enhance the potential of disease transmission in the Americas. *North-Western Journal of Zoology* **9:** art. 131509.

Rogers DJ. 2006. Models for vectors and vector-borne diseases. *Advances in Parasitology* **62:** 1–35.

Rogers DJ, Randolph SE. 2003. Studying the global distribution of infectious diseases using GIS and RS. *Nature Reviews Microbiology* **1:** 231–237.

Rouget M, Richarson DM. 2003. Inferring process from pattern in plant invasions: a semimechanistic model incorporating propagule pressure and environmental factors. *American Naturalist* **162:** 713–724.

Sedda L, Brown HE, Purse BV, Burgin L, Gloster J, Rogers DJ. 2012. A new algorithm quantifies the roles of wind and midge flight activity in the bluetongue epizootic in northwest Europe. *Proceedings of the Royal Society of London B Series B, Biological Sciences* **279:** 2354–2362.

Sedda L, Morley DW, Braks MAH, De Simone L, Benz D, Rogers DJ. 2014. Risk assessment of vector-borne diseases for public health governance. *Public Health* **128:** 1049–1058.

Soberon J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters* **10:** 1115–1123.

Stephens CR, Gimenez Heau J, Gonza′lez C, Ibarra-Cerden CN, Sanchez-Cordero V, Gonzalez-Salazar C. 2009. Using biotic interaction networks for prediction in biodiversity and emerging diseases. *PLoS ONE* **4:** e5725.

Vaclavik T, Meentemeyer RK. 2009. Invasive species distribution modeling (iSDM): are absence data and dispersal constraints needed to predict actual distributions? *Ecological Modelling* **220:** 3248–3258.

Vaclavik T, Meentemeyer RK. 2012. Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. *Diversity and Distributions* **18:** 73–83.

Vaclavik T, Kupfer JA, Meentemeyer RK. 2012. Accounting for multi-scale spatial autocorrelation improves performance of invasive species distribution modelling (iSDM). *Journal of Biogeography* **39:** 42–55.

Václavík T, Kanaskie A, Hansen EM, Ohmann JL, Meentemeyer RK. 2010. Predicting potential and actual distribution of sudden oak death in Oregon: prioritizing landscape contexts for early detection and eradication of disease outbreaks. *Forest Ecology and Management* **260:** 1026–1035.

Warren DL. 2012. In defence of 'niche modelling. *Trends in Ecology and Evolution* **27:** 497–500.

Warren DL. 2013. 'Niche modeling': that uncomfortable sensation means it's working. *Trends in Ecology and Evolution* **28:** 193–194.

Wisz MS, Pottier J, Kissling WD, Pellissier L, Lenoir J, Damgaard CF, Dormann CF, Forchhammer MC, Grytnes J-A, Guisan A, Heikkinen RK, Høye TT, Kühn I, Luoto M, Maiorano L, Nilsson M-C, Normand S, Öckinger E, Schmidt NM, Termansen M, Timmermann A, Wardle DA, Aastrup P, Svenning J-C. 2013. The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological Reviews* **88:** 15–30.

Yañez-Arenas C, Guevara R, Martínez-Meyer E, Mandujano S, Lobo JM. 2014. Predicting species' abundances from occurrence data: effects of sample size and bias. *Ecological Modelling* **294:** 36–41.