Analysing ecological dynamics with relational event models: the case of invasion events

Rūta Juozaitienė

VU Institute of Data Science and Digital Technologies, Vilnius, Lithuania;
Vytautas Magnus University, Kaunas, Lithuania

Hanno Seebens

Senckenberg Biodiversity and Climate Research Centre, Frankfurt, Germany

Guillaume Latombe

Institute of Evolutionary Biology, University of Edinburgh, UK

Franz Essl

Bioinvasions. Global Change. Macroecology Group,
Department of Botany and Biodiversity Research, University of Vienna, Austria

Ernst C. Wit[*]

Institute of Computing, Università della Svizzera italiana, Lugano, Switzerland

---

[*]To whom correspondence should be addressed: wite@usi.ch

Abstract

Spatio-temporal processes play a key role in ecology from genes to large-scale macroecological and biogeographical processes. For analyzing such spatio-temporally structured data timing and time-ordering of events is particularly important. In this paper, we present a generic method - relational event modelling - for studying spatio-temporal patterns of biological invasions at large spatial scales and by including variables that drive these dynamics. Relational event modelling (REM) relies on temporal interaction dynamics, that encode sequences of relational events connecting a sender to a recipient at a specific point in time.

We present a relational event model case study of the spread of alien species, which are species that are introduced accidentally or deliberately into geographical regions outside their native range. In this context, a relational event represents the new occurrence of an alien species given its former distribution. By considering the bipartite network of species and regions, we embed the first records process into a relational event setting in order to detect drivers of invasions in the presence of complex confounding. Thus, taking the temporal sequence of occurrences into account, the relational event model identifies commonalities among species' spread and their relation to underlying variables. Using years of first records of 16,019 established alien species from major taxonomic groups we have shown that the relational event model can be used constructively in environmental science to model time-stamped ecological interactions, such as the invasive species process, without having to resort to more traditional simplifications. Combining the first records data with other spatio-temporal information, enables us to discover which factors have been driving the spread of species across the globe.

Keywords: *Ecological dynamics; First records database; Relational event model; Species invasions.*

1 Introduction

Biological interactions, which we will refer to as relational events, are the fundamental process shaping the diversity of life (Del-Claro and Torezan-Silingardi, 2021a). So-called relational events are important basic relationships in ecology (Del-Claro and Torezan-Silingardi, 2021b). These interactions can be intraspecific, i.e., such as male-male interference competition (Dijkstra and Border, 2018) and mating, interspecific, such as predation and pollination, or form some more generalized interaction, e.g., between a species and its habitat or environment. Often it is possible to associate

particular temporal events to these interactions, where both the event type and the timing of the event are relevant.

An example of a more general interaction is the spread of an alien species among regions, where the relational events describe either the subsequent spread of other species along similar routes or the interaction between a region and the respective species (Seebens et al., 2021). The dynamics of such inter-regional spread can be complex, thereby varying in time, space and among species. In addition, dynamics are usually driven by multiple factors simultaneously, involving exogenous ecological and socio-economic factors, such as transport hubs (Floerl et al., 2009), habitat suitability, climatic conditions, land use (Pyšek et al., 2010; Dyer et al., 2017; Essl et al., 2019), prevailing in a particular geographic region. Given the complex interplay of natural and human spatio-temporal factors, it is crucial to study the effects of many factors simultaneously.

Relational event modelling was first developed in the field of social network analysis (Butts, 2008; Perry and Wolfe, 2013), before being applied in number of other fields, such as interactions among cattle (Patison et al., 2015), transfers of patients between hospitals (Vu et al., 2017) and political interactions (DuBois and Smyth, 2010). A relational event describes a time-stamped interaction, such as the appearance of a species in a new location. The relational event model analyses sequences of relational events in their observed temporal order. Ecological invasion processes have crucial relational event characteristics: an invasion is a temporal event whereby a species gets associated to a certain new region. The methodological advantage of using relational event models is the fact that they are able to describe the underlying general interaction patterns, can deal with time-varying variables and have a suit of diagnostic tools to test the suitability of the models on actual data.

There have been a number of statistical approaches to predict the occurrence and spread of alien species ranging from correlative (e.g. (Elith and Leathwick, 2009)) to dynamic modelling tools (Dullinger et al., 2009; Chapman et al., 2016; Paini and Yemshanov, 2012). Usually, these approaches are constrained to individual species and disregard the timing of an invasion in a particular region; thus, potential temporal changes in the magnitude and even direction of drivers on invasions are not considered. Species Distribution Models (SDMs) usually model one species at a time (Guisan and Thuiller, 2005) and although methods that can simultaneously model more than one species have been developed (Caradima et al., 2019, e.g.), SDMs are static models, which do not account for temporal variation of predictor and response variables.

In this paper, we introduce a dynamic modelling technique based on relational events and showcase its application by simulating dynamics spreading dynamics of alien species and identifying predictors of species introductions through time and space. The model considers multiple species simultaneously and jointly takes into account the effects of potential drivers, while controlling for potential confounders.

By using the data of four taxonomic groups (i.e. mammals, birds, insects, vascular plants) of the Alien Species First Record Database (FirstRecords), the most exhaustive source of first records of species in regions of the world (Seebens et al., 2017), we introduce a relational event model (REM) for studying the factors driving the spread of alien species across large spatial (i.e. globally) and temporal (i.e. 1880-2005) scales. We use a set of spatio-geographic and socio-economic variables that have been shown to be of importance for explaining the current distribution of alien species (Pyšek et al., 2010; Dawson et al., 2017; Essl et al., 2019, e.g.). We analyse for changes in the magnitude and direction of the set of variables on the timing of first records of alien species over time. Based on these results, we provide a synthesis of the advantages and caveats of using REMs for studying ecological invasions.

2 Material and methods

2.1 Data

This study focuses on events of first records of mammals, birds, insects and plants from 1880 until 2005 . We consider a range of explanatory variables to characterise the spatial locations, the climatic conditions, political relationships such as former colonial ties among countries and various socio-economic variables during that period (see Table 1).

The Alien Species First Records database (Seebens et al., 2017) contains the years of the first records in regions (i.e. countries, and sub-national regions such as islands, archipelagos and federal states) of established alien species across taxonomic groups worldwide. Established alien species refer to species introduced, accidentally or intentionally, outside of their natural geographic range and that have established a viable population (Blackburn et al., 2011). The data contains more than 47K records of 16,922 established alien species across 275 regions worldwide. Most data are from plants (48%), followed by insects (26%) and birds (6%), whereas mammals are present in 3% of the records. Regions usually correspond to countries or islands having a particularly large number of samples. In this manuscript, we focus

Variable	Description	Data source
$d_{sc}(t)$	Closest distance between regions	(Hijmans et al., 2017)
$tr_{sc}(t)$	International trade	(Barbieri et al., 2009)
$dt_{sc}(t)$	Near-surface air temperature	(Watanabe et al., 2011)
$l_c(t)$	Sum of cropland and pasture proportions	(Hurtt et al., 2011)
$u_c(t)$	Proportion of urban area	(Hurtt et al., 2011)
$k_{sc}(t)$	Colonial ties	(Becker, 2019)
$pi_c(t)$	The Alien Species First Records database	(Seebens et al., 2018)

Table 1: List of variables and data sources used in analysis

on four major land-based taxonomic groups with the largest number of first records, namely vascular plants, insects, birds and mammals. The selected taxonomic groups cover 81% of the data contained in the Alien Species First Record Database.

Land cover types have been shown to be associated with different levels of alien species presence (Chytrỳ et al., 2009; Polce et al., 2011; Sánchez-Ortiz et al., 2020). Here, information on land cover is used to characterise the potential suitability of regions for the establishment of alien species and the degree of anthropogenic disturbance, which facilitates establishments. Land cover is described by the proportion of cropland, pasture and urban area; data were obtained from the Harmonized Global Land Use data set, V1 (Hurtt et al., 2011). The grid-based information has been averaged for the regions and years (1971-2012) considered in this study. As land cover data were provided on a decadal basis, while the analysis was run on an annual scale, land cover variables were interpolated among decades to obtain the annual resolution.

It is expected that favourable climate conditions facilitate the spread of alien species and harsh environments with stressful conditions often lead to a lower success of an establishment. Near-surface air temperatures were used to represent climate conditions, using simulated annual values covering the period from 1871 to 2012 (Watanabe et al., 2011). Annual near-surface air temperature values were averaged for each region and year between 1871 and 2012.

International trade has been identified as a major driver of alien species introductions (via increased propagule and colonization pressures), and therefore increases the probability of species to become established (Westphal et al., 2008; Chapman et al., 2017; Blackburn et al., 2020). International trade values were extracted from the Correlates of War project (Barbieri

et al., 2009). The dataset includes information on trade flows between countries from 1871 to 2009. The dyadic trade dataset values are provided in current U.S. dollars. However, the dataset has missing values, because trade flow is unknown for some countries within certain time periods, and some countries are not included in the dataset, for a number of reasons. To address this issue, missing values that appear at the beginning of the observational period were replaced with zeros in those time series that start at zero value. In other cases, i.e., for time series for which the first recorded value is not equal to zero, we extrapolated missing values using the linear model $\log(trade + 1) = \beta_0 + \beta_1 t$.

A substantial part of the most important trade routes has formed since colonial times. It is expected that former colonial ties facilitated the spread and establishment of alien species. Colonial ties were therefore included in this analysis using the Colonial Dates Dataset (COLDAT) (Becker, 2019), which provides information on the reach and duration of European colonial empires. Each country is uniquely characterised by the colonial empire it belonged to or classified as independent, in case there were no colonial ties.

Native ranges of species were not provided in an individual database and thus were gathered from various data sources. For mammals and birds, native ranges were provided as range maps provided by the IUCN range maps (`https://www.iucnredlist.org`, accessed 08.07.2016) and BirdLife (`http://datazone.birdlife.org/species/requestdis`, accessed 01.11.2016), respectively. A species was then assigned to a region as listed in the first record database if the range map intersected with the region. For vascular plants, native ranges were provided by (van Kleunen et al., 2018) and for insects ranges were obtained from CABI Invasive Species Compendium (`https://www.cabi.org`, accessed 15.07.2016). Region names were harmonised to match the regions as provided in the first record database.

2.2 Relational event model for first invasion events

A relational event model (REM) is a probabilistic model that describes the temporal interactions between a set of sender nodes and a set of receiver nodes (Butts, 2008). Sender and receiver nodes can either represent elements of the same group (e.g. to express relationships between species - one-mode networks), or elements of different groups (e.g. to express relationships between species and regions - bipartite networks). In our study, sender nodes are species and receiver nodes are regions as defined in the FirstRecords database. We consider four groups of species, $L_i = \{s_1, \ldots, s_n\}$ (i=mammals, birds, plants, insects) and a fixed time-frame $[1880, 2005]$

over which the invasion process of the species occurs into the regional set $C = \{c_1, \ldots, c_{272}\}$ of the 272 pre-specified regions. If at time t a species emerged in a new region c, then this invasion event can be described by a tuple $e = (s, c, t)$, whereby species s spread to a region c at time t. The relational event model describes a general framework for modelling and inferring the stochastic species invasion process $\{(s, c, t)\}_{t \in [1880, 2005]}$ in order to identify the reasons why particular species invade particular regions and why they did so at the time they did.

We focus on a number of possible drivers of invasions that can be roughly distinguished into four groups (see Table 2). The first group consists of, possibly time-varying, factors with constant effects. The second group focuses on, again possibly time-varying, factors whose effect changes over time. Therefore, we take into account two types of dynamic processes. Time-varying factors allow encompassing variables whose values change over time (e.g. reflect changes in the volume of international trade). The parameter value estimating the effect of the time-varying variable can also vary over time. That is, not only can the volume of international trade vary across time, but the effect of trade volume on the rate of invasion events also varies over time. Modelling dynamic changes provide insights on specific time patterns and valuable information that could be missed otherwise. The third group covers random effects of a monadic type, which are variables measured at the level of a region or species, whereas the final group of factors considers dyadic random effects, which denote variables measured at the level of pairs of species, capturing the interaction between species in invasions. The overall list of explanatory variables contains the information on geographic distance, trade volume, climatic data, land cover variables, colonial ties. Furthermore, we consider a number of endogenous drivers, which focuses on the self-exciting nature of the process itself, such as the possibility that certain species tend to co-invade regions, that certain regions have a larger capacity for accommodating new species or that certain species are more invasive. These potential drivers are time-varying variables, as their values are not fixed over time.

For each taxonomic group (plants, birds, insects, mammals), we fit a relational event model representing global invasion dynamics over the period from 1880 till 2005. The overall time-dependent model of species invasion dynamics that we consider for each taxonomic group is given as,

$$
\begin{aligned}
\log \lambda_{sc}(t) &= \log \lambda_0(t) d_{sc}(t) \beta_1(t) + tr_{sc}(t) \beta_2(t) + dt_{sc}(t) \beta_3 + l_c(t) \beta_4(t) \\
&\quad + u_c(t) \beta_5 + k_{sc}(t) \beta_6 + pi_c(t) \beta_7 + b_s + b_c + b_{ss_c(t)},
\end{aligned} \tag{1}
$$

where $s_c(t)$ is defined as the last species to invade c before t. $d_{sc}(t)$ de-

notes the distance to the nearest region c invaded by species s by time t. Long-distance invasion events through natural dispersal are typically rare. Therefore, the distance between two regions that species have to overcome could be an important one of the main factors responsible for alien species invasions. The distance between two regions is defined as the distance between their closest borders, and thus the distance between neighbouring regions is zero. The quantity $tr_{sc}(t)$ is the logarithm of the sum of annual trade flows between region c and other regions invaded by species s by time t (log-transformation is used to comply with assumptions of normality). $tr_{sc}(t)$ is a time-dependent variable as the amount of traded commodities varied over time. This variable representing international trades is a time-dependent variable as alien species are introduced to new areas over time. The variable $dt_{sc}(t)$ represents the absolute value of difference in temperature between region c and the nearest region invaded by species s by time t. It is a time-dependent variable since the list of regions invaded by species s is changing through time. We consider two variables representing the land cover in region c at time t. The first variable $l_c(t)$ is the variable that describes the agricultural land which is defined as the sum of cropland and pasture proportions in the region c at time t. $u_c(t)$ is the proportion of urban area in region c at time t. $k_{sc}(t)$ indicates the presence of species s at time t in a region within the area of a colonial power. The variable $pi_c(t)$ is a weighted number of prior invasions in region c. Assuming that recent invasions are more informative for the current invasion process, this covariate is defined using a temporal discounting factor. The relational event model also includes random effects representing species spread rate (b_s) and region invasion propensity (b_c). It is expected that different species may vary in their spread behaviour, whereas similarly certain regions may be more "popular" destinations than others.

The last group of model explanatory variables considers dyadic random effects that aim to capture species interaction effects in invasions. This effect $b_{ss'}$ describes the affinity of species s and s' in their diffusion patterns. We consider both a general formulation in which the order of invasion by the species is accounted for, as well as the more parsimonious symmetric formulation that constrains $b_{ss'} = b_{s's}$. It turns out that although the symmetric formulation is often sufficient, for some taxonomic groups the general non-symmetric co-invasion random effect offers additional explanatory power. Species co-invasion was analysed for the top thirty most alien species in each taxonomic group due to the high computational complexity.

Furthermore, for a number of explanatory variables we consider time-varying effects. For distance, land cover and trade we found that the ef-

fect changes over time and therefore a piecewise constant coefficient was estimated for each of the five equally sized periods (1880-1905, 1906-1930, 1931-1955, 1956-1980, and 1981-2005).

Analysis was conducted using *coxme* function from the standard `coxme` package within the R statistical software to fit a Cox proportional hazard model. To assess whether a fitted Cox regression models adequately described the data we performed model diagnostics. The proportional hazards assumption was checked using statistical tests and graphical diagnostics based on the scaled Schoenfeld residuals. Thus, we used the *cox.zph* function from the `survival` package to test for independence between residuals and time, the *ggcoxzph* function from the `survminer` package to perform a graphical diagnostics and the *ggcoxdiagnostics* function (`survminer` package] to test influential observations or outliers. A more detailed description of relational event modelling and diagnostics is provided in the online supplementary material.

3 Results

3.1 3.1 The temporal variation in importance of drivers of invasion events

The final REMs provide a parsimonious description of the data as summarized in Table 3. The models explain between 16.9 and 27.4% of the variance, depending on the taxonomic groups as shown in Table 3 (all p-values $< 10^{-4}$). Chi-squared tests show that models with time-varying coefficients including distance, land cover and trade are highly significant, with the sole exception of cropland-pasture effects for insects (Tables Sx-Sy in Supplementary Material).

Distance. The effect of distance on the invasion rate (i.e. recording in anew regions) has grown over time, with the rate of invasion events (i.e. the instantaneous rate at which the invasion event occurs among other possible invasions) decreasing with distance to the closest invaded region (Figure 1a). This relationship became especially strong after 1980 for mammals and insects.

Land cover variables. A larger fraction of urban areas in a region was positively correlated with spread rate for plants, insects and mammals (Table 4, p-values < 0.05). For every increase in urbanization by 1%, the rate of
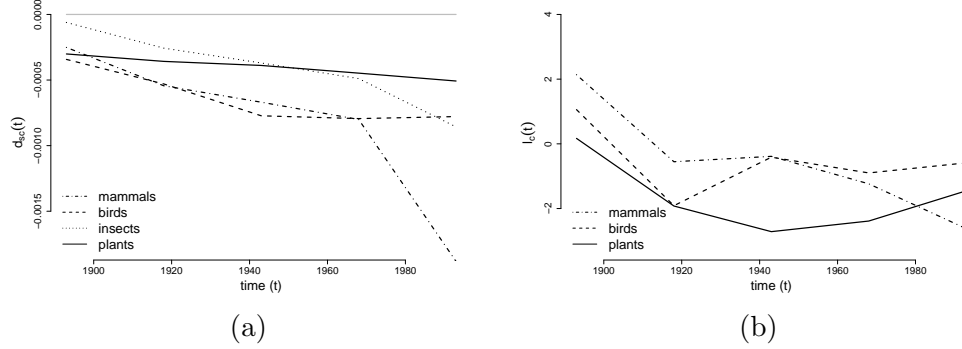
Figure 1: Effect of a) distance and b) cropland and pasture proportion on rates of invasion. a). Negative effect of distance on the rate of invasion events has increased over time. b). Crop land and pasture proportions have flipped their effects on the rate of invasion events from positive to negative over time.

invasion events in that region rose by 4.0% (vascular plants), 14.0% (insects) and 17.0% (mammals), respectively. On the other hand, the effects of cropland and pasture on the rate of invasion events were more complex. This variable was significant for plants, mammals and birds and its time-varying coefficients were positive in the first period of consideration [1880-1905], but became negative thereafter, even though its effect for plants and birds tended towards becoming positive again more recently (Figure 1b). This means that in the first period (1880-1905) larger fractions of cropland and pasture in a region resulted in higher spread rates, while this effect has reversed since the early 20th century for plants, mammals and birds.

The number of alien species found in a particular region. Both plants and insects were significantly affected by the number of species that have already invaded a certain region. However, whereas plants were recorded more frequently in regions already invaded by other plants, insects were less frequently recorded in regions that had already been invaded by other insects. The effects were nonetheless very small in both cases, i.e., affecting the hazard by 0.3% per invasion in the opposite direction.

Climatic similarity. For all taxonomic groups, rates of invasion events were significantly affected by climatic conditions of the receiving region. The temperature differential effect for all taxonomic groups was negative

(Table 4). The parameter values associated with $dt_{sc}(t)$ mean that for each degree of mean annual temperature difference to the nearest invaded region the spread rate decreased by 13.0%, 20.5%, 11.3% and 7.7% for plants, insects, mammals and birds, respectively, keeping everything else constant.

Colonial ties. The effect of colonial ties was not statistically significant for any taxonomic groups.

Trade flow. During the first two periods (1880-1930), the effect of trade was positive, i.e. the spread rate increased with the amount of trade between a region (Figure 2). However, the effect became negative afterwards.
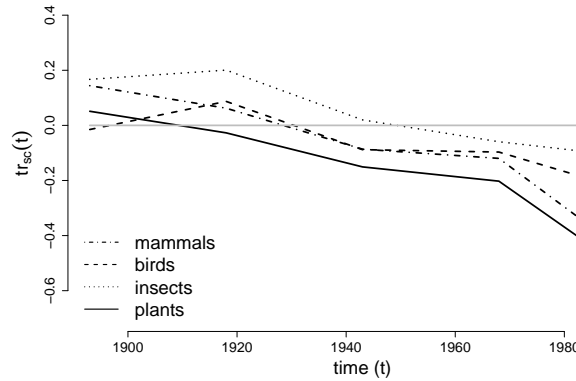

Figure 2: Effect of trade on the diffusion of species. After an initial positive effect, in more recent times the effect of trade seems to have become negative.

Species spread rate. Analysing the random effects, we can identify the species with the highest spread propensity. After accounting for the other effects described above, the random species spread rate captures the tendency of different species to spread. Table 4 presents the top five alien species in each taxonomic group. The results show that Siberian Chipmunk (*Tamias sibiricus*), Rraccoon dogs (*Nyctereutes procyonoides*), and American mink (*Neovison vison*) are the mammals with the highest species spread rate, while for birds, Rock pigeon (*Columba livia*), followed by Alexandrine parakeet (*Psittacula eupatria*) and Egyptian goose (*Alopochen aegyptiaca*). For plants, water plants (Hydrocotyle ranunculoides) and conifers such as

11

Caribbean pine (*Pinus caribaea*) and Douglas fir (*Pseudotsuga menziesii*) rank first. The highest ranking insects are pests, i.e. a fruit fly (*Bactrocera invadens*), pink hibiscus mealybug (*Maconellicoccus hirsutus*), and western flower thrips (*Frankliniella occidentalis*).

Region invasion propensity. When considering the random region effects to identify those regions with the highest invasion propensity, while correcting for all fixed effects in the REM, regions with the greatest number of recorded alien species are the United States, Australia, United Kingdom, South Africa, New Zealand and Argentina (Table 6).

Species co-invasions. Standard deviations of the random effects show that the symmetric formulation of the co-invasion propensity of pairs of species provides similar explanatory power as the non-symmetric formulation in most cases (Table 4). Only for mammals does the non-symmetric co-invasion random effect offer additional explanatory power. The estimated random effects for insects are represented in Figure 3.

Baseline hazard. The baseline hazard $\lambda_0(t)$ is a remainder term of the relational event model. It represents the underlying rate of spread over time that is unexplained by the drivers included in the model. Figure 4 shows the cumulative baseline hazard estimates for the four taxonomic groups covered. The linear shape of the baseline hazard for both plants and birds suggest that they have an approximately constant baseline hazard $\lambda_0(t) = \frac{d\Lambda_0(t)}{dt}$, indicating that the spread rate as explained by the model was mostly constant for these species. However, insects have seen a marked increase of their baseline hazard since the 1980s. This suggests that the spread rate has been accelerating for insects since then in a way that cannot be explained by the current variables in the model. For mammals, the cumulative baseline hazard has flattened since the 1960s, suggesting that their spread rate was decreasing. Earlier, the spread rate of mammals outpaced that of the other taxonomic groups by a factor of two to three.

4 Discussion

4.1 The changing role of drivers in spread dynamics

Alien species invasions are complex, dynamic processes that involve both natural factors, such as climate, and man-made conditions, such as trade and

Figure 3: Symmetric insects interaction effects in invasion events. Each node represents a species and an edge depicts a co-invasion propensity. Solid lines indicate that the spread of two species tends to follow each other, while dashed edges suggest that two species tend to avoid each other in their spatio-temporal spread. Moreover, the edge thickness represents the strength of the measured interaction. The provided graphs include only the subset of edges having the highest weight values, i.e. increasing or decreasing the invasion hazard by a factor of 2. Smaller co-invasion strengths have been removed to make the graphs easier to interpret.

land use (intensity) as represented by anthropogenic land cover types. Its temporal unfolding means that any analysis aggregating invasion events over
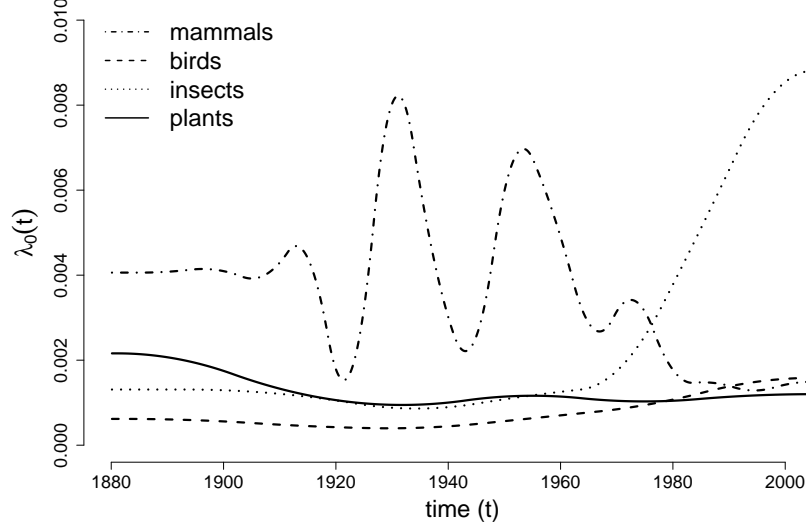
Figure 4: Baseline hazard for the underlying invasion speeds of the different taxonomic groups. Although mammals have shown the quickest pace, it is tapering off recently, whereas the spread of insects is speeding up since 1980.

long time periods would miss important aspects related to timing, temporal changes of relative importance of drivers and ordering of invasion events. The relational event model proposed in this paper has the advantage of being able to capture multiple factors in a temporally ordered analysis of invasion events, and is therefore able to account for much of the complex involved in the spread process of alien species.

Our analysis affirmed that a changing environment affected invasion event patterns. Urbanization - represented by the amount of urban areas in the included regions - has been increasing the speed of invasions, specifically for insects and mammals, and this effect has been consistent over the whole time span we analysed. In contrast, the relationship between agricultural land use-systems (represented by the amount of cropland and pasture) and biological invasion is more complex and has changed over time non-linearly: as shown by the fact that changes in the amount of cropland and pastures varied in seem to be responsible for more complex temporal effects of the amount of cropland and pastures. Whereas initially the fraction of cropland and pastures was positively related to the invasion hazard of birds, plants and mammals, more recently the effect has become neutral for mammals

and negative for plants and birds.

The negative effect of temperature for all taxonomic groups shows that biological invasions are strongly affected by climatic conditions. This is particularly true for insects and to a slightly lesser extent for plants, mammals, and birds. A mean annual temperature difference between source and target regions of one degree was enough to decreases the probability of invasion by 8% for birds and 25% for insects. This finding supports the importance of climate matching between regions for alien species spread and establishment (Abellán et al., 2017; Lovell et al., 2021).

Surprisingly, results revealed that recently regional diffusive processes have become stronger for the spread of invasive behaviour of alien species. Despite substantial evidence on the important role of trade as driver of invasions (Seebens et al., 2015), in recent times more regional drivers have become more important for recent invasion events as can be seen from an increased negative effect of distance. It is well-established that trade is crucial for alien species spread across the globe (Seebens et al., 2015; Dawson et al., 2017). However, this picture seems to be more complicated for invasion events than might be expected. Although trade was certainly of major importance in the early 20th century, the positive effect seems to have largely disappeared more recently. This finding seems counter-intuitive, but could possibly be partially explained by the fact that the number of introduced species did not keep pace with the increase in trade volumes due to a saturating relationship between imported commodities and the number of alien species. Indeed, the number of newly recorded insects distinctly declined with the amount of imported goods (Seebens et al., 2017). As a consequence, the relative importance of trade as a predictor for alien species occurrences declined over time. Further, the proportion of goods with known high relevance for introducing alien species such as agricultural and silvicultural goods (e.g. wood, grain) has declined severely in the last decades (Luppold, 1988). There has also been a structural shift in the commodity composition, bulk commodities were substitute by processed consumer-ready products, as a result of rising incomes. The share of bulk commodities declined from 50% in 1980 to 32% in 1995, while the share of nonbulk products increased from 50% to 68%. Moreover, the share of products with low risks of introducing alien species (e.g. electronics, goods from other technology intensive industries, and labour-intensive products, particularly clothing) have grown most rapidly in world exports during the period 1980-2000 (Mayer et al., 2003). Thus, it is likely that recent trade volumes are less closely reflecting trade-related introduction risks of alien species.

Baseline hazard estimates suggest that the international containment

strategies have disparate effects on the different taxonomic groups. In particular, whereas for birds, mammals and plants we observe a rather constant hazard (with some fluctuations for mammals), suggesting no dramatic increase in the general spread rate, beyond what is explained by the model, the picture for insects is quite different. Particularly since the 1970s, the unexplained spread rate of insects has increased five fold. We hypothesize that increases in velocity of transport of many goods (e.g. via the expansion of trade via cargo planes) has facilitated the spread of insects, as many short-lived insects are not able to survive long ship-voyages (Hulme, 2009).

4.2   Insights offered by alien species- and region-level factors

Our analyses show that several countries being part of the former British Empire (Australia, New Zealand, South-Africa, United Kingdom, USA) are on top of the list of the country invasion propensity list for all four taxonomic groups. This finding is in line with previous findings on hot spots of global invasion levels across eight taxonomic groups, where these countries all show up prominently (Dawson et al., 2017). There are likely a number of reasons that jointly explain this result. Most of these countries are large and have a long history of an advanced ecological-scientific infrastructure that has kept accurate records of biological data including species invasions. Further, these countries have been part of one of the largest European Empires and the integration into a near-global political entity and higher levels of trade and people movements among these regions may have fostered biological invasions as well.

On the other hand, several of these countries such as New Zealand and Australia have implemented strict control measures to stop the introduction of alien species (Turbelin et al., 2017). However, these policies are mostly of rather recent times and may be masked by the time-frame (1880-2005) under observation here.

4.3   Application potential and potential limitations of REMs in ecology

Relational event models offer the capacity to account for the temporality of ecological events such as biological invasions and to investigate how relationships between these events and potential drivers change over time. This approach overcomes both non-dynamic analyses as well as single variable models, which always run the risk of misidentifying the real causal factor. REMs nonetheless require comprehensive and accurate temporal data on

these drivers and events, which are often scarce. In our study, the First Records Database, although the currently most comprehensive one, is biased towards regions in Europe and certain taxonomic groups such as plants and vertebrates. And even in Europe, the availability of first records varies among countries. In addition, information about the year of first record often does not correspond to the year of actual introduction, which might be decades earlier. These uncertainties might explain the comparatively low degree of explained variation. Information on trade, land use, and economic variables was easily available and therefore proxies had to be employed. This means that although our relational event models give a more subtle picture of the time-dependent invasion process of species, the parameter estimates have to be interpreted in the light of these limitations.

Originally, relational event models were designed for small scale networks and their computational costs are quite extensive. In this study, we have shown how they can nonetheless be applied to more complex networks based on high-quality standardized databases of ecological data. Further developments of novel, computationally efficient implementations and user-friendly software will help quantitative ecologists to make the standard application and diffusion of the methodology possible and more widely investigate time-dependent ecological phenomena.

Supplementary Materials

Supplementary material is available at `http://biostatistics.oxfordjournals.org`.

Acknowledgments

References

P. Abellán, J. L. Tella, M. Carrete, L. Cardador, and J. D. Anadón. Climate matching drives spread rate but not establishment success in recent unintentional bird introductions. *Proceedings of the National Academy of Sciences*, 114(35):9385–9390, 2017.

K. Barbieri, O. Keshk, and B. Pollins. Trading data: Evaluating our as-

sumptions and coding rules. *Faculty Publications*, 26, 11 2009. doi: 10.1177/0738894209343887.

B. Becker. Colonial Dates Dataset (COLDAT). 2019. doi: 10.7910/DVN/ T9SDEW. URL `https://doi.org/10.7910/DVN/T9SDEW`.

T. Blackburn, G. Edwards, S. Tulaczyk, M. Scudder, G. Piccione, B. Hallet, N. McLean, J. Zachos, B. Cheney, and J. Babbe. Ice retreat in wilkes basin of east antarctica during a warm interglacial. *Nature*, 583:554–559, 07 2020. doi: 10.1038/s41586-020-2484-5.

T. M. Blackburn, P. Pyšek, S. Bacher, J. T. Carlton, R. P. Duncan, V. Jarošík, J. R. Wilson, and D. M. Richardson. A proposed unified framework for biological invasions. *Trends in ecology & evolution*, 26(7): 333–339, 2011.

N. Breslow. Disussion of regression models and life-tables by cox, dr. *J. Roy. Statist. Assoc., B*, 34:216–217, 1972.

C. T. Butts. A relational event framework for social action. *Sociological Methodology*, 38(1):155–200, 2008.

B. Caradima, N. Schuwirth, and P. Reichert. From individual to joint species distribution models: A comparison of model complexity and predictive performance. *Journal of Biogeography*, 46(10):2260–2274, 2019.

D. Chapman, B. V. Purse, H. E. Roy, and J. M. Bullock. Global trade networks determine the distribution of invasive non-native species. *Global Ecology and Biogeography*, 26(8):907–917, 2017.

D. S. Chapman, L. Makra, R. Albertini, M. Bonini, A. Páldy, V. Rodinkova, B. Šikoparija, E. Weryszko-Chmielewska, and J. M. Bullock. Modelling the introduction and spread of non-native species: International trade and climate change drive ragweed invasion. *Global Change Biology*, 22(9): 3067–3079, 2016.

M. Chytrỳ, P. Pyšek, J. Wild, J. Pino, L. C. Maskell, and M. Vilà. European map of alien plant invasions based on the quantitative assessment across habitats. *Diversity and Distributions*, 15(1):98–107, 2009.

D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B*, 34(2):187–220, 1972.

W. Dawson, D. Moser, M. Van Kleunen, H. Kreft, J. Pergl, P. Pyšek, P. Weigelt, M. Winter, B. Lenzner, T. M. Blackburn, et al. Global hotspots and correlates of alien species richness across taxonomic groups. *Nature Ecology & Evolution*, 1(7):1–7, 2017.

K. Del-Claro and H. Torezan-Silingardi. *Plant-Animal Interactions Source of Biodiversity.* 05 2021a. ISBN 978-3-030-66876-1. doi: 10.1007/978-3-030-66877-8.

K. Del-Claro and H. M. Torezan-Silingardi. An evolutionary perspective on plant-animal interactions. *Plant-Animal Interactions: Source of Biodiversity*, page 1, 2021b.

P. D. Dijkstra and S. E. Border. How does male–male competition generate negative frequency-dependent selection and disruptive selection during speciation? *Current zoology*, 64(1):89–99, 2018.

C. DuBois and P. Smyth. Modeling relational events via latent classes. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 803–812, 2010.

S. Dullinger, I. Kleinbauer, J. Peterseil, M. Smolik, and F. Essl. Niche based distribution modelling of an invasive alien plant: effects of population status, propagule pressure and invasion history. *Biological Invasions*, 11 (10):2401–2414, 2009.

E. E. Dyer, P. Cassey, D. W. Redding, B. Collen, V. Franks, K. J. Gaston, K. E. Jones, S. Kark, C. D. L. Orme, and T. M. Blackburn. The global distribution and drivers of alien bird species richness. *PLoS biology*, 15 (1):e2000942, 2017.

B. Efron. The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, 72(359):557–565, 1977.

J. Elith and J. R. Leathwick. Species distribution models: ecological explanation and prediction across space and time. *Annual review of ecology, evolution, and systematics*, 40:677–697, 2009.

F. Essl, W. Dawson, H. Kreft, J. Pergl, P. Pyšek, M. Van Kleunen, P. Weigelt, T. Mang, S. Dullinger, B. Lenzner, D. Moser, N. Maurel, H. Seebens, A. Stein, E. Weber, C. Chatelain, Inderjit, P. Genovesi, J. Kartesz, O. Morozova, M. Nishino, P. M. Nowak, S. Pagad,

W.-S. Shu, and M. Winter. Drivers of the relative richness of naturalized and invasive plant species on Earth. *AoB PLANTS*, 11(5), 09 2019. ISSN 2041-2851. doi: 10.1093/aobpla/plz051. URL `https://doi.org/10.1093/aobpla/plz051`. plz051.

O. Floerl, G. Inglis, K. Dey, and A. Smith. The importance of transport hubs in stepping-stone invasions. *Journal of Applied Ecology*, 46(1):37–45, 2009.

A. Guisan and W. Thuiller. Predicting species distribution: offering more than simple habitat models. *Ecology letters*, 8(9):993–1009, 2005.

R. J. Hijmans, E. Williams, C. Vennes, and M. R. J. Hijmans. Package 'geosphere'. *Spherical trigonometry*, 1(7), 2017.

P. E. Hulme. Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of applied ecology*, 46(1):10–18, 2009.

G. Hurtt, L. Chini, S. Frolking, R. Betts, J. Feddema, G. Fischer, J. Fisk, K. Hibbard, R. Houghton, A. Janetos, C. Jones, G. Kindermann, T. Kinoshita, K. K. Goldewijk, K. Riahi, and E. Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Climatic Change*, 109(1):117–161, November 2011. doi: 10.1007/s10584-011-0153-2. URL `https://ideas.repec.org/a/spr/climat/v109y2011i1p117-161.html`.

R. S. Lovell, T. M. Blackburn, E. E. Dyer, and A. L. Pigot. Environmental resistance predicts the spread of alien species. *Nature Ecology & Evolution*, 5(3):322–329, 2021.

W. G. Luppold. *Hardwood import trends*, volume 619. US Department of Agriculture, Forest Service, Northeastern Forest Experiment . . . , 1988.

J. Mayer, A. Butkevicius, A. Kadri, and J. Pizarro. Dynamic products in world exports. *Review of World Economics*, 139(4):762–795, 2003.

D. R. Paini and D. Yemshanov. Modelling the arrival of invasive organisms via the international marine shipping network: a khapra beetle study. *PloS one*, 7(9):e44589, 2012.

K. Patison, E. Quintane, D. Swain, G. Robins, and P. Pattison. Time is of the essence: an application of a relational event model for animal social networks. *Behavioral ecology and sociobiology*, 69(5):841–855, 2015.

P. O. Perry and P. J. Wolfe. Point process modelling for directed interaction networks. *Journal of the Royal Statistical Society: SERIES B: Statistical Methodology*, pages 821–849, 2013.

C. Polce, W. E. Kunin, J. C. Biesmeijer, J. Dauber, O. L. Phillips, and A. F. S. Network. Alien and native plants show contrasting responses to climate and land use in europe. *Global Ecology and Biogeography*, 20(3): 367–379, 2011.

R. L. Prentice, Y. Yoshimoto, and M. Mason. Relationship of cigarette smoking and radiation exposure to cancer mortality in Hiroshima and Nagasaki. *Journal of National Cancer Institute*, 70(4):611–622, 1983.

P. Pyšek, V. Jarošík, P. E. Hulme, I. Kühn, J. Wild, M. Arianoutsou, S. Bacher, F. Chiron, V. Didžiulis, F. Essl, et al. Disentangling the role of environmental and human pressures on biological invasions across europe. *Proceedings of the National Academy of Sciences*, 107(27):12157–12162, 2010.

S. Ripatti and J. Palmgren. Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, 56(4):1016–1022, 2000.

K. Sánchez-Ortiz, K. J. Taylor, A. De Palma, F. Essl, W. Dawson, H. Kreft, J. Pergl, P. Pyšek, M. van Kleunen, P. Weigelt, et al. Effects of land-use change and related pressures on alien and native subsets of island communities. *PloS one*, 15(12):e0227169, 2020.

H. Seebens, F. Essl, W. Dawson, N. Fuentes, D. Moser, J. Pergl, P. Pyšek, M. van Kleunen, E. Weber, M. Winter, et al. Global trade will accelerate plant invasions in emerging economies under climate change. *Global change biology*, 21(11):4128–4140, 2015.

H. Seebens, T. Blackburn, E. Dyer, P. Genovesi, P. Hulme, J. Jeschke, S. Pagad, P. Pyšek, M. Winter, M. Arianoutsou, S. Bacher, B. Blasius, G. Brundu, C. Capinha, L. Celesti-Grapow, W. Dawson, S. Dullinger, N. Fuentes, H. Jäger, and F. Essl. No saturation in the accumulation of alien species worldwide. *Nature Communications*, 8, 02 2017. doi: 10.1038/ncomms14435.

H. Seebens, T. Blackburn, E. Dyer, P. Genovesi, P. Hulme, J. Jeschke, S. Pagad, P. Pyšek, M. van Kleunen, M. Winter, M. Ansong, M. Arianoutsou, S. Bacher, B. Blasius, E. Brockerhoff, G. Brundu, C. Capinha, C. Causton, L. Celesti-Grapow, and F. Essl. Global rise in emerging alien species results from increased accessibility of new source pools. *Proceedings of the National Academy of Sciences*, 115:201719429, 02 2018. doi: 10.1073/pnas.1719429115.

H. Seebens, T. M. Blackburn, P. E. Hulme, M. van Kleunen, A. M. Liebhold, M. Orlova-Bienkowskaja, P. Pyšek, S. Schindler, and F. Essl. Around the world in 500 years: Inter-regional spread of alien species over recent centuries. *Global Ecology and Biogeography*, 2021.

T. M. Therneau and M. T. M. Therneau. Package 'coxme'. *Mixed effects cox models. R package version*, 2, 2015.

D. C. Thomas. General relative-risk models for survival time and matched case-control analysis. *Biometrics*, 37(4):673–686, 1981.

A. J. Turbelin, B. D. Malamud, and R. A. Francis. Mapping the global state of invasive alien species: patterns of invasion and policy responses. *Global Ecology and Biogeography*, 26(1):78–92, 2017.

M. van Kleunen, P. Pyšek, W. Dawson, F. Essl, H. Kreft, J. Pergl, P. Weigelt, A. Stein, S. Dullinger, C. König, B. Lenzner, N. Maurel, D. Moser, H. Seebens, J. Kartesz, M. Nishino, A. Aleksanyan, M. Ansong, L. Antonova, and M. Winter. The global naturalized alien flora (glonaf) database. *Ecology*, 100, 10 2018. doi: 10.1002/ecy.2542.

D. Vu, A. Lomi, D. Mascia, and F. Pallotti. Relational event models for longitudinal network data with an application to interhospital patient transfers. *Statistics in medicine*, 36(14):2265–2287, 2017.

S. Watanabe, T. Hajima, K. Sudo, T. Nagashima, T. Takemura, H. Okajima, T. Nozawa, H. Kawase, M. Abe, T. Yokohata, T. Ise, H. Sato, E. Kato, K. Takata, S. Emori, and M. Kawamiya. Miroc-esm 2010: model description and basic results of cmip5-20c3m experiments. *Geoscientific Model Development*, 4(4):845–872, 2011. doi: 10.5194/gmd-4-845-2011. URL https://gmd.copernicus.org/articles/4/845/2011/.

C. Westphal, R. Bommarco, G. Carré, E. Lamborn, N. Morison, T. Petanidou, S. G. Potts, S. P. Roberts, H. Szentgyörgyi, T. Tscheulin,

et al. Measuring bee diversity in different european habitats and biogeographical regions. *Ecological monographs*, 78(4):653–671, 2008.

E. Wit, E. v. d. Heuvel, and J.-W. Romeijn. 'all models are wrong...': an introduction to model uncertainty. *Statistica Neerlandica*, 66(3):217–236, 2012.

A   Relational event modelling (REM)

A relational event model is a probabilistic model that describes the temporal interactions between a set of sender nodes and a set of receiver nodes (Butts, 2008). In our case study, sender nodes are species and receiver nodes are regions - i.e. countries, or subnational units such as federal states, islands or archipelagos. Each relational event represents the time that a particular species appeared in a particular region. The aim of an REM is to describe the pattern of interactions between pairs of nodes and to identify the drivers of the relational events; in other words, what are the underlying reasons that certain invasions did happen at a certain time, whereas others did not?

We consider a group of n species, $L = \{s_1, \ldots, s_n\}$ and a particular timeframe $[t_o, t_f]$, over which the invasion process of the species occurs into the regional set $C = \{c_1, \ldots, c_{272}\}$ of 272 regions. At t_0 each species s has its own *native range* $N_s(t_0)$, which is defined as the collection of regions where the species s has already been established,

$$N_s(t) = \{c \in C \mid s \text{ is present in } c \text{ before time } t\}.$$

The complement of this set is referred to as the *risk set*, $\mathcal{R}(t) = \{(s,c)|c \notin N_s(t)\}$. If at time t a species emerged in a region c, then that species belongs to the risk set at time t and, therefore, not yet at the native species set at that time. An invasion event can now be described by a tuple $e = (s,c,t)$, whereby species s spread to a region c at time t. The relational event model describes a general framework for modelling and inferring the stochastic species invasion process $\{(s_i, c_i, t_i)\}_{t_i \in [t_o, t_f]}$ in order to identify the reasons why particular species invade particular regions and why they did so at the time they did.

In defining a relational event process, the most important quantity is the hazard function that expresses the risk or hazard of an event at some time t. Formally, let T_{sc} be the (absolutely) continuous random variable associated with the invasion time of s into c and let $f_{sc}(t)$ be the corresponding probability density function. The hazard function is defined as $\lambda_{sc}(t) = f_{sc}(t)/\{1-$

$\int_0^t f_{sc}(s)\mathrm{d}s\}$ and specifies the instantaneous rate at which the invasion event occurs among those other invasions that are possible at time t. We suppose that the hazard function can depend on a p-dimensional, possibly time-dependent, vector of fixed effects, denoted by $\mathbf{x}_{sc}(t) = (x_{sc1}(t), \ldots, x_{scp}(t))^\top$ and a q-dimensional, possibly time-dependent, vector of fixed effects, denoted by $\mathbf{z}_{sc}(t) = (z_{sc1}(t), \ldots, z_{scq}(t))^\top$. Proportional hazard models are based on the assumption that $\mathbf{x}_{sc}(t)$ and $\mathbf{z}_{sc}(t)$ influence the hazard function through the following relation

$$\lambda_{sc}(t; \mathbf{x}_{sc}, \mathbf{z}_{sc}) = \lambda_0(t)e^{\mathbf{x}_{sc}^\top(t)\boldsymbol{\beta} + \mathbf{z}_{sc}^\top(t)\boldsymbol{b}}, \tag{2}$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ is a p-dimensional is a vector of fixed parameters, $\boldsymbol{b} \in \mathbb{R}^q$ is a vector of random parameters, $\boldsymbol{b} \sim N(0, \Sigma(\theta))$ whose variance possibly can be written as a function of a lower dimensional parameter θ, and $\lambda_0(t)$ is the baseline hazard function at time t, which is assumed the same for all species-region combinations. This model is known as the mixed-effect Cox proportional hazard model (Cox, 1972), although it is possible to replace $e^{\mathbf{x}_{sc}^\top(t)\boldsymbol{\beta} + \mathbf{z}_{sc}^\top(t)\boldsymbol{b}}$ by a more general relative risk function $\psi\{\mathbf{x}_{sc}(t), \mathbf{z}_{sc}(t); \boldsymbol{\beta}, \boldsymbol{b}\} > 0$, with $\psi : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ a differentiable function and normalized to satisfy $\psi(\mathbf{0}, \mathbf{0}; \boldsymbol{\beta}, \boldsymbol{b}) = 1$ (Thomas, 1981). Possible choices are the linear relative risk (Prentice et al., 1983), the logit relative risk (Efron, 1977), or the the excess relative risk.

Suppose that n events, $\{(s_i), c_i, t_i)\}_{i=1}^n$, are available in $[t_o, t_f]$. In particular, these correspond to actual invasion events or possibly to non-invasions at the end of the observation period t_f. Additionally, non-invasions may be Let $y_{sc}(t)$ be defined as an indicator corresponding to the risk set $\mathcal{R}_s(t)$, i.e.,

$$y_{sc}(t) = \begin{cases} 1 & \text{if } c \in \mathcal{R}_s(t) \\ 0 & \text{otherwise} \end{cases}$$

then we can define the partial likelihood for $\boldsymbol{\beta}$ and a fixed random effect $\boldsymbol{b}$ as

$$\ell(\boldsymbol{\beta}, \boldsymbol{b}) = \sum_{i=1}^n \int_{t_o}^{t_f} y_{s_i c_i}(t)\eta_{s_i c_i}(t) - \log\left(\sum_{s,c} y_{sc}(t)e^{\eta_{sc}(t)}\right) \, dt,$$

where $\eta_{sc}(t) = \mathbf{x}_{sc}^\top(t)\boldsymbol{\beta} + \mathbf{z}_{sc}^\top(t)\boldsymbol{b}$ is the time-dependent linear predictor for species-region combination (s, c). The random effects need to be integrated out, resulting in mixed-effect partial likelihood, sometimes referred to as the *penalized partial likelihood*,

$$\ell(\boldsymbol{\beta}, \boldsymbol{\theta}) = |\Sigma(\theta)|^{-1/2} \int e^{\ell(\boldsymbol{\beta}, \boldsymbol{b})} e^{-\boldsymbol{b}'\Sigma(\theta)\boldsymbol{b}/2} \, db. \tag{3}$$

A partial likelihood estimate is obtained by joint maximization over $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, which can be obtained via a Laplace approximation of (3) as suggested by (Ripatti and Palmgren, 2000). The method is implemented in the R package coxme (Therneau and Therneau, 2015).

Given that the saturated partial loglikelihood is exactly zero, we can define an approximate R^2 to indicate the goodness of fit of the model, as

$$R^2 = 1 - \frac{\ell(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})}{\ell(0,0)},$$

where the numerator indicates the integrated partial log-likelihood and the denominator indicates the null log-likelihood. The cumulative baseline hazard function $\Lambda_0(t)$ is not estimated as part of the model fitting procedure. However, given the fits of both the fixed, $\hat{\boldsymbol{\beta}}$, and random effects, $\hat{\boldsymbol{b}}$, it is possible to derive a non-parametric estimate in the following fashion,

$$\hat{\Lambda}_0(t) = \sum_{t_i < t} \left[\frac{dN(t_i)}{\sum_{s,c} y_{sc}(t_i) e^{\mathbf{x}_{sc}^{\top}(t_i)\hat{\boldsymbol{\beta}} + \mathbf{z}_{sc}^{\top}(t_i)\hat{\boldsymbol{b}}}} \right], \tag{4}$$

where $dN(t_i)$ are the number of invasion events at time t_i. The estimator is an adaptation of the Breslow estimator (Breslow, 1972).

The relational event model makes a number of assumptions and it is important to check whether the three main assumptions are satisfied in practice. First, the *proportional hazard* assumption focuses on the parametric log-linear assumption of the hazard in (2). This assumption can be checked by means of the scaled Schoenfeld residuals. Secondly, *outliers* and potentially influential observations can be identified by means deviance residuals. Finally, we need to test for any *non-linear effects* of the included variables. Any significant trend in the martingale residuals might indicate some non-linear effect. Assumptions can be checked in a number of ways. Graphical diagnostic plots can give visual clues to the correctness of the assumptions, whereas diagnostic tests evaluate the statistical evidence for any departures of the assumptions. Specifically in large data settings, these tests are, however, not very useful: as all models are wrong (Wit et al., 2012), all tests tend to be rejected with a lot of data even though the practical departures from the assumptions might be minimal. The more interesting question is whether the departures from the model affect the interpretation of the results and for this it is more important to evaluate the diagnostic plots.
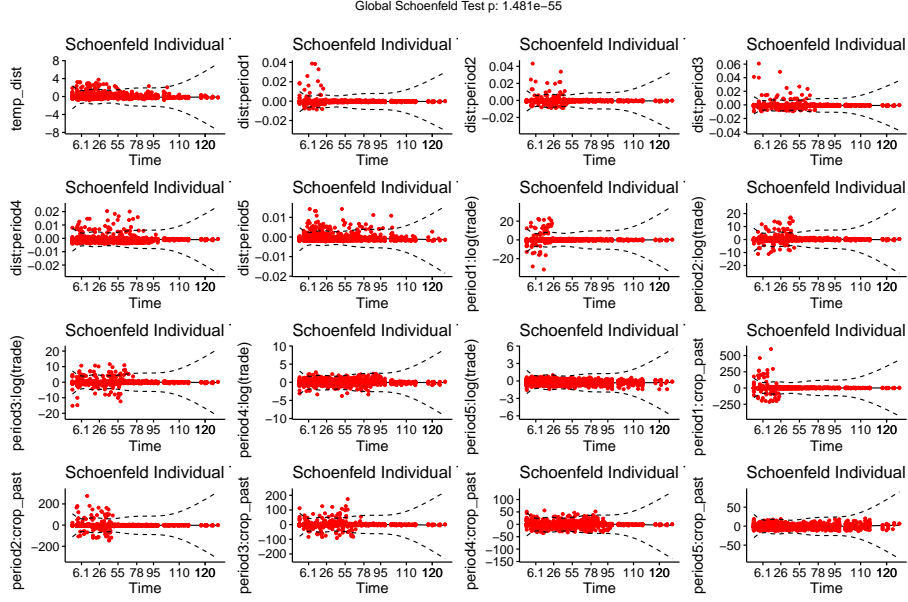
Figure 5: Scaled Schoenfeld residuals for birds to test the proportional hazard assumption.

B   Model diagnostics

It is important that the final model provides a good fit to the data. We provide some selected diagnostic plots for the birds model, whereas other plots and the plots for plants, mammals and insects can be found in the supplementary materials. The proportional hazard assumption is supported by the absence of a relationship between Schoenfeld residuals and time. Figure 5 shows these residuals for the birds relational event model. No serious deviations from proportionality are evident. Secondly, if plots of the deviance residuals are consistent with a mean of zero and a variance of one, then this suggest that there are no outliers or potentially influential observations. As can be seen in Figure 6a, this seems not to be violated in any serious way. In particular, there are no observations with extreme small or large deviance residuals that could have potentially skewed the analysis. Finally, we need to test the functional form of (1), in particular, test for any *non-linear effects* of the included variables. Figure 6b shows the example of the temperature difference variable in the birds model. As can be seen, there is no structural effect of the residual, which suggest that including this variable as the linear effect is sufficient.
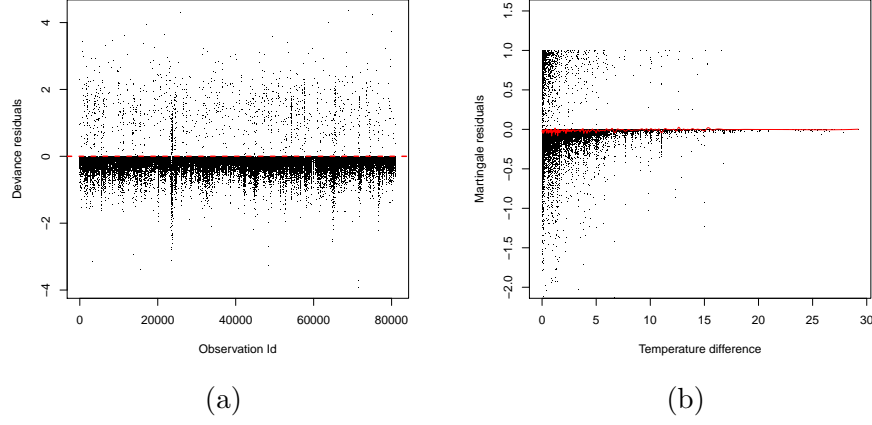
26

Figure 6: (a) Deviance residuals for birds to test for outliers and (b) martingale residuals for birds to test for non-linearity.

C   Time-varying variables

For example, Figure 7 depicts the distance between Galapagos and the nearest region invaded by the red avadavat (*Amandava amandava*). The curve represents the decrease in distance as the red avadavat settles in new regions that are closer to Galapagos.
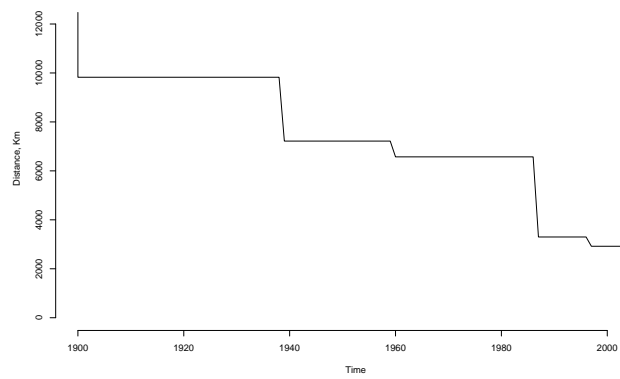
Figure 7: Distance to region nearest to Galapagos that is already invaded by the red avadavat (*Amandava amandava*)

Effect	Definition	Type	Justification
$d_{sc}(t)$	Distance to the nearest region c invaded by species s by time t	Time-varying variable with effect changing through time	Geographic distance may influence migratory behaviours and strategies
$tr_{sc}(t)$	Logarithm of the sum of annual trade flows between region c and other regions invaded by species s by time t	Time-varying variable with effect changing through time	International trade is a key pathway for the global spread of alien species
$dt_{sc}(t)$	Absolute value of difference in temperature between region c and the nearest region invaded by species s by time t	Time-varying variable with constant effect	Climatically suitable ranges ease establishment of alien species
$l_c(t)$	Sum of cropland and pasture proportions in the region c at time t	Time-varying variable with constant effect	Land use changes provide opportunities for the spread and establishment of alien species
$u_c(t)$	Proportion of urban area in region c at time t	Time-varying variable with constant effect	Land use changes provide opportunities for the spread and establishment of alien species
$k_{sc}(t)$	Presence of species s at time t in a region within the area of a colonial power	Time-varying variable with constant effect	In the Colonial era alien species have been intentionally introduced
$pi_c(t)$	Weighted number of prior invasions in region c	Time-varying variable with constant effect	Temporal variable represents the importance of different pathways that shifts over time
b_s	Species spread rate	Monadic random effect	Different species may vary in their spread behaviour
b_c	Region invasion propensity	Monadic random effect	Certain regions may be more "popular" destinations than others
$b_{ss'}$	Affinity of species s and s' in their diffusion patterns	Dyadic random effect	Alien species may take similar invasion pathways

Table 2: List of explanatory variables used for model fitting.

Taxonomic group	R^2	Null loglik	Model loglik	χ^2 stat	df	p-value
Birds	23.6	-13761.6	-10519.8	6483.5	19	$p < 10^{-4}$
Mammals	27.4	-5284.4	-3833.9	2901.0	19	$p < 10^{-4}$
Plants	23.4	-86409.5	-66171.4	40476.2	20	$p < 10^{-4}$
Insects	16.9	-3614.1	-3004.0	1220.3	15	$p < 10^{-4}$

Table 3: Summary table of the final invasion models selected, showing an overall fit between 16.9% and 27.4% for allvtaxonomic groups.

effect	plants	insects	mammals	birds
$pi_c(t)$: prior invasions	0.003***	-0.003***	n.s.	n.s.
$dt_{sc}(t)$: temperature difference	-0.139***	-0.230***	-0.120***	-0.080***
$k_{cs}(t)$: colonial ties	n.s.	n.s.	n.s.	n.s.
$u_c(t)$: urban landscape	3.92*	13.1*	15.74*	n.s.
σ_{spc}: species invasiveness	0.75***	1.47***	0.94***	0.84***
σ_{cnt}: region popularity	2.66***	1.17***	1.12***	1.36***
σ_{int}: species interactions	n.s.	0.71***	**0.22*****	0.41***
$\sigma_{\mathrm{int}}^{\mathrm{sym}}$: symmetric interactions	n.s.	**0.82*****	n.s.	**0.34*****

Table 4: Parameter estimates for the non-temporal fixed effects and random effect standard deviations across the four taxonomic groups (*** significant at $\alpha = 0.001$, * significant at $\alpha = 0.05$, n.s.: non-significant at $\alpha = 0.05$.)

plants	insects
Hydrocotyle ranunculoides	*Bactrocera invadens*
Pseudotsuga menziesii	*Maconellicoccus hirsutus*
Pinus caribaea	*Frankliniella occidentalis*
Sedum rupestre	*Cameraria ohridella*
Sphagneticola trilobata	*Phyllonorycter issikii*

mammals	birds
Tamias sibiricus	*Columba livia*
Nyctereutes procyonoides	*Psittacula eupatria*
Neovison vison	*Alopochen aegyptiaca*
Ondatra zibethicus	*Passer domesticus*
Myocastor coypus	*Lonchura malacca*

Table 5: Species with the highest species spread rate (i.e. random effects) after accounting for all fixed effects in the REM.

plants	insects	mammals	birds
Australia	United States	South Africa	United States
Chile	New Zealand	Australia	South Africa
United Kingdom	Italy	United States	United Kingdom
Belgium	South Africa	New Zealand	Spain
South Africa	Argentina	Argentina	Australia

Table 6: Regions with the highest invasion propensity when accounting for all other variables in the REM.