Forestry An International Journal of Forest Research

Institute of Chartered Foresters

Forestry 2021; **94**, 86–101, doi:10.1093/forestry/cpaa014
Advance Access publication 1 June 2020

Application of hierarchical clustering to identify high risk pests to Sitka spruce: Ireland as a case study

Catriona Duffy[1,2,*], Melanie G. Tuffen[3], Rowan Fealy[2] and Christine T. Griffin[1]

[1]Department of Biology, Maynooth University, Maynooth, Co. Kildare, Ireland
[2]Department of Geography, Maynooth University, Maynooth, Co. Kildare, Ireland
[3]Horticulture Development Department, Teagasc, Ashtown Research Centre, Dublin 15 D15 KN3K, Ireland

*Corresponding author E-mail: catriona.duffy@mu.ie

Received 20 December 2018

Invertebrate forest pests and pathogens can cause considerable economic losses and modern patterns of trade have facilitated the international movement of pest species on an unprecedented level. This upsurge in trade has increased the pathways available to high risk species, facilitating entry and potential establishment in nations where they were previously absent. To support policy and pest prioritization, pest risk analyses are conducted to decide 'if' and 'how' pests should be regulated in order to prevent entry or establishment; however, they cannot be carried out for every potential pest. This paper utilizes a hierarchical clustering (HC) approach to analyse distribution data for pests of Sitka spruce (*Picea sitchensis* (Bong.) Carr.) in order to identify species of high risk to Ireland, as well as potential source regions of these pests. The presence and absence of almost a 1000 pests across 386 regions globally are clustered based on their similarity of pest assemblages, to provide an objective examination of the highest risk pests to Irish forestry. Regional clusters were produced for each taxon analysed including the Coleoptera, Diptera, Hemiptera, Hymenoptera, Nematoda, Lepidoptera and the Fungi. The results produced by the HC analysis were interpreted with regard to biological realism and climate. Biologically meaningful clusters were produced for each of the groups, except for the Diptera and Nematoda, and each of the species analysed were ranked within their group by a quantitative risk index specific to the island of Ireland. The impact of uncertainty in the distribution data is also examined, in order to assess its influence over the final groupings produced. The outputs from this analysis suggest that the highest risk pests for Ireland's Sitka spruce plantations will originate from within Europe. Ultimately, Ireland could benefit from seeking regulation for some of the higher ranking pests identified in this analysis. This analysis provides the first of its type for Sitka spruce, as well as its application in Ireland. It also serves to highlight the potential utility of HC as a 'first approach' to assessing the risk posed by alien species to hitherto novel regions.

Introduction

The number of new invertebrate pest and pathogen (hereafter referred to collectively as pests) threats to plant health has been increasing in recent years, as a result of increasing globalization, trade in plants and plant products, and climate change (Desprez-Loustau *et al.*, 2007; Liebhold *et al.*, 2012; Bebber *et al.*, 2013; Santini *et al.*, 2013; Bebber, 2015). In Europe, the number of non-native invertebrate pests and pathogens establishing outside their native regions has rapidly increased over the last century (Desprez-Loustau, 2009; Roques *et al.*, 2009) requiring phytosanitary regulations in order to reduce this trend. Within Europe, the majority of plant health regulation is currently undertaken on an individual pest basis, where specific listed pests are regulated via phytosanitary measures (Regulation (EU) 2016/2031 and its associated legislation). One of the first steps towards achieving regulation of a pest is horizon scanning (the process of gathering and analyzing pest data in order to prioritize certain species). This is then followed by the process of pest risk analysis (PRA), where experts examine available information about a pest to determine if the introduction of phytosanitary measures are justified as per international standards (IPPC, 2017). Governments in most developed countries utilize PRA to aid in the decision-making process regarding what species and pathways to regulate (FAO, 2016; Leung *et al.*, 2012).

The objective assessment of the risk of a pest species entering a region can be achieved using global pest distribution data along with techniques such as hierarchical clustering (HC) (Eschen *et al.*, 2014), k-means clustering (Worner *et al.*, 2013) or self organizing maps (SOMs) (Paini *et al.*, 2010, 2011; Worner *et al.*, 2013). The presence of a species within a specific region indicates that suitable biotic and abiotic conditions exist for that pest.

Therefore, grouping regions based on a species' presence or absence indicates a level of similarity of those same conditions. This is the fundamental assumption of the approach: that a region's pest assemblage is the result of a multitude of complex biotic (e.g. host) and abiotic (e.g. climatic) factors within that area and that these pest profiles act as a proxy for these factors when attempting to group 'like' pest organisms together (Worner and Gevrey, 2006). Species assemblages are widely used as indicators of environmental conditions in various fields, including freshwater ecology (e.g. Chon *et al.,* 1996; Omar, 2010), climatology (e.g. Huppert and Solow, 2004) and extinction events (e.g. Bromham *et al.,* 2012) and have been successfully applied in tandem with multivariate statistical techniques, including clustering to analyse ecological communities (e.g. Legendre and Legendre, 2012). The resultant species associations and spatial patterns from such analyses can form the basis for management as suggested in climate change (e.g. Costanza *et al.,* 2017) and fisheries studies (e.g. Singh *et al.,* 2011). Worner and Gevrey *et al.* (2006) were the first to apply this approach within an invasive species context and since then, others have used species assemblages to infer pest risk for various taxa (e.g. Paini *et al.,* 2011; Morin *et al.,* 2013; Eschen *et al.,* 2014). Pest assemblages also reflect historical trade routes, which may have led to the introduction of exotic species in the first instance (Langor *et al.,* 2009). It follows then that regions with similar pest assemblages are likely to exhibit similar biotic and abiotic conditions and potential trade links, which would likely facilitate the establishment of a species from one region to another (if a pathway for entry was present).

Within the area of plant health risk, the superiority of one clustering technique over another has not yet been confirmed; however, SOMs have been employed more widely (e.g. Worner and Gevrey, 2006; Watts and Worner, 2009; Paini *et al.,* 2011; Roige *et al.,* 2016) than HC (e.g. Worner *et al.,* 2013; Eschen *et al.,* 2014) for identifying high risk pest species. Despite this, the less complex of the two approaches (e.g. HC) has been shown to produce meaningful clusters when utilized within PRA (e.g. Eschen *et al.,* 2014) and other fields (e.g. Laize and Hannah, 2010; Gubareva, 2012). Indeed, Roige (2017) found that HC performed best using a number of statistical metrics when compared with SOMs and k-means for clustering pest assemblages. While this approach does not produce weights (as with SOMs), the frequency of a species within a cluster is employed as a proxy indicator for the likelihood of a species establishing within the target region (after Watts and Worner, 2009). If a pest species is absent from a region of interest, but present in a large number of other regions in the same cluster, then it is assumed that the target region has an environment suitable for establishment of that pest if it were to be introduced. Thus, the frequency metric can be used to illustrate levels of risk, based on the hypothesis that the frequency with which a species appears in a cluster conveys a quantitative level of risk of that species establishing in regions where it is not currently found.

In Ireland, forestry contributes an estimated €2.3 billion to the economy each year (Coillte, 2015), and covers an area of 730 000 ha (10.5 per cent of total) of land (DAFM, 2018) and 112 000 ha of land in Northern Ireland. Sitka spruce is the most economically important forestry species on the island, making up over 50 per cent of the forest estate in Ireland (exact percentage not available for Northern Ireland) (Forest Service, 2013). The species is currently grown in 16 countries worldwide outside its native North American range and constitutes a significant portion of the forested area in the UK and Denmark (Durrant Houston *et al.,* 2016). To date, Irish forestry has remained free from many of the most damaging pests of Sitka spruce that are present on *Picea* Miller in Europe (Grégoire and Evans, 2004; McCracken, 2013). However, in recent years, the introduction and establishment of a number of pests of other tree hosts, such as *Hymenoscyphus fraxineus* (ash dieback) on ash (*Fraxinus* L.) and *Phytophthora ramorum* Werres, De Cock & Man in 't Veld (sudden larch death) on Japanese larch (*Larix kaempferi* (Lamb.) Carr.), have highlighted the potential for introductions to precipitate significant impacts to the national forest estate. The ability to pre-emptively identify potential pests of Sitka spruce is essential to protecting this economically important tree species. Such protection can be afforded by the regulation of pests and the application of phytosanitary measures on international trades. Ireland benefits from its island status on the periphery of Europe (meaning there is limited scope for natural spread); however, pests can still be introduced on infested traded commodities. Ireland maintains a number of 'protected zones', a form of EU plant health regulation that enables it to regulate pests present in the EU that are not already EU quarantine pests, but absent from its own territory. PRAs provide the technical evidence required for such protected zones to be established, as well as providing justification for the introduction of phytosanitary measures on trades at a regional and global level.

The aim of this research is to provide an objective and quantitative assessment of the risk posed by individual species of pests of Sitka spruce in Ireland. This is achieved through the use of HC, which is also assessed for its ability to provide meaningful results within the context of potential Sitka spruce pests. Additionally, the identification of potential regional sources of high-risk pests for Ireland is sought. This paper describes the application of HC to a dataset of more than 1300 pests and potential pests of Sitka spruce globally (Tuffen and Grogan, 2019) to assess the similarity of pest assemblages between regions. Ultimately, pests from regions with similar assemblages to the island of Ireland would be priority candidates for PRA, similar to those identified in the HC of pest assemblages in Europe by Eschen *et al.* (2014). Interpretation of the cluster solutions is provided and species are identified for follow-up PRA based on the maps, risk indices (RIs) calculated and analysis of climate types within the clusters.

Methods

The pest list

The datasets employed for each pest taxon analysed (Table 1) were derived from a list of pest species of the genus *Picea* (described in detail in Tuffen and Grogan, 2019). The Tuffen and Grogan (2019) pest list contains data on the global distribution of pests of spruce (*Picea* spp.), as well as other potential hosts for the reported organisms. Species deemed to be potential pests were also included in this list—that is, phytophagous insects and other organisms associated with spruce, but not recorded as damaging to it. Their inclusion was deemed apt, as some of the species have been recorded as becoming damaging pests within an invaded range. Where data on distribution was lacking, sources utilized included databases with records submitted by amateur recorders; any records far outside the pest's known distribution were dismissed as potential vagrants. To account for potential adaptability of pests to new hosts, a number of pests were included in the pest list that could infest multiple genera of conifers, even if a species of the genus *Picea* was not recorded as a host. During the compilation of the pest list, reports of pest distributions were often limited to larger regions, e.g. 'Central Europe' or 'Southern Canada', as opposed to a specific collection of countries/regions. In order to enable the inclusion of pests within the HC analysis when regional distribution data was not available, a defined list of countries/sub-country regions was created for geographical terms commonly used to describe pest distribution (see Supplementary Material S1). These definitions were then used consistently for pests, but it should be noted that despite widespread use, some regional names remain poorly defined. Therefore, the inclusion of a country or region within a pest's distribution for this analysis should not be used as a reference for the presence of that pest in that region. Instead, references used to determine the distribution of each pest have been retained and queries about specific pest distributions can be sent to the authors. A country or region was not included within the distribution for a pest when it was only recorded within protected cultivation (e.g. glasshouses), since the primary aim of the work was to identify pest threats to Sitka spruce which is not grown under protection. Presence/absence data were recorded for 1374 pest species from a number of taxonomic groups (Table 1) across 386 geographic areas. Large areas were further subdivided into regions, based on the European and Mediterranean Plant Protection Organization (EPPO) delineation of regions (EPPO, 2018), in order to reduce the impact of climatic and ecological variability typical of large countries. These larger regions included Russia (subdivided into six regions), US (50), Canada (13), China (34), India (35), Brazil (28), Japanese islands (5) and Australia (6). It was expected that the division of these geographical areas would produce more ecologically robust clusters due to the reduction of biotic and abiotic variability. Similarly to Worner *et al.* (2013), rare species (defined as those occurring in less than six regions (i.e. <2 per cent of the regions)) were removed from the list in an effort to avoid incorporating overly-influential data in the analysis (Table 1).

Table 1 Summary of the numbers of pest species included in the analysis (rare = occurring in less than six regions).

Order	Number of species in pest list	Number after 'rare' species removed
Coleoptera	395	292
Diptera	19	12
Fungi	305	192
Hemiptera	130	92
Hymenoptera	87	64
Lepidoptera	360	275
Nematoda	26	21
Acari	10	7
Thysanoptera	1	1
Oomycetes	23	7
Phytoplasma	3	2
Other invertebrates	4	3
Other pathogens	11	5

Hierarchical clustering

Using the pest list as the primary database, individual data matrices of presence/absence data across each of the 386 regions utilized were prepared for the following groups: Coleoptera, Diptera, Hemiptera, Hymenoptera, Lepidoptera, Nematoda and the Fungi. The remaining groups were omitted from the analysis owing to low numbers of species records for the groups (Table 1). For each region, pest assemblage data for each group of organisms were defined by the presence or absence of known species for that location. Presence was denoted as '1', while absence was recorded as '0', creating a binary distribution data file for each group analysed. Within the matrix each of the species was listed in the columns, while each of the regions listed by row. Regions were then clustered using HC, employing the Ward linkage method (Ward, 1963). The analysis was performed using Matlab (2013) and Ward's method was utilized as it has been shown to provide interpretable clusters in the areas of pests (e.g. Eschen *et al.,* 2014; Roige, 2017) hydrology (e.g. Ramachandra Rao and Srinivas, 2006; Laize and Hannah, 2010; Chiverton *et al.,* 2015), desertification studies (e.g. Zolfaghari *et al.,* 2019) and bird ecology (e.g. Kati *et al.,* 2009). On the basis of their similarity of pest assemblages, clusters of geographical areas are presumed to exhibit similar biotic and abiotic conditions suitable for species establishment. Those species that occur frequently within a cluster, but are not currently recorded within the region of interest (in this case Ireland) within that cluster, are assumed to be capable of establishing in that region if suitable hosts are available (Watts and Worner, 2009).

Indices

One of the fundamental problems with the application of cluster analyses is identifying an optimal number of clusters (with optimal partitioning being defined as the outcome of a clustering algorithm that best fits the inherent partitions of the data set (Halkidi *et al.,* 2001)). The Davies Bouldin index (Davies and Bouldin, 1979) has been used widely as a measure of cluster separation to identify an appropriate number of clusters for pest data (e.g. Céréghino *et al.,* 2005; Gevrey *et al.,* 2006; Worner *et al.,* 2013; Eschen *et al.,* 2014). However, no one method for identifying the optimal partition of clusters has been identified as superior to all others. For this reason, three different indices were employed to assess the 'optimal' number of clusters for the data belonging to each group. These indices were (1) Davies-Bouldin: based on the ratio of the within-cluster scatter to the between-cluster separation; (2) Calinski Harabasz: based on the average between- and within-cluster sum of squares (Caliński and Harabasz, 1974) and (3) Silhouette: based on the pairwise difference of between- and within-cluster distances (Rousseeuw, 1987).

Risk index

This analysis outlined here follows that of Watts and Worner (2009), wherein the frequency of a species within each cluster is assumed to provide an indication of the potential for that species to establish within a region in the cluster where it is not currently recorded. The RI for each species was calculated by dividing the number of regions in the cluster in which the species was present, by the entire number of regions in the cluster. Where Ireland clustered with the rest of Europe, species in that cluster that had not been recorded in Ireland were ranked by their RI from high to low, in order to identify those species with the highest perceived risk of establishing.

Impacts of uncertainty regarding pest presence or absence

It has been acknowledged that the biodiversity of Ireland is under-recorded, with some taxa such as the nematodes lacking even a basic list of the species present (Fitzpatrick *et al.,* 2010). In addition, there is a lack of expertise in the fields of forest entomology and pathology (Copeland and Dowley, 2010; Hanlon *et al.,* 2016). Therefore, it is likely that there are several species that are present but unrecorded in Ireland. This will be especially true for those species that have cryptic lifecycles, or closely resemble species that are already present and widespread in Ireland. For example, Ireland is currently developing dedicated seed orchards for Sitka spruce (DAFM, 2018), and since the majority of seed and cone pests do not affect other aspects of stand health, many of these species may have already entered Ireland but remain unrecorded due to their current negligible economic importance. This is important in the context of the current analysis, as under-recording of species which are already present in Ireland could impact the way the country clusters within the analysis (due to underrepresentation within the pest assemblage).

Pests whose presence in Ireland was uncertain were marked within the pest list as ‘uncertain’ rather than present/absent. Pests were deemed to have an uncertain status if they met at least one of following criteria:

- The pest was (1) widespread in the rest of the European Union and associated with traded commodities that are not subject to phytosanitary regulations that would reduce the likelihood of the pest being associated and (2) difficult to identify.
- There were several older records of the pest (at least more than 30-years old), but no recent records.
- The pest had been intercepted on multiple occasions at ports of entry, and there have been no official surveys to establish if the pest has successfully entered and established on the island of Ireland.

To investigate if the potential under-recording of species had a significant impact on the clustering analysis, the analysis was run twice: with pests whose status was considered uncertain marked as ‘absent’ in the first run, and marked as ‘present’ in the second run.

Climate classification

The hierarchical approach used here inherently incorporates the climate of an area into the clusters produced via the pest assemblages. In an effort to further investigate the relationship between the clusters produced and typical climate for a region, the climate types for each of the clusters were assessed using the Köppen classification scheme (Beck *et al.,* 2018). The Köppen-Geiger climate classification is a system of global climate classification originally created by Wladimir Köppen and subsequently updated in collaboration with Rudolf Geiger (Köppen and Geiger, 1936). This type of scheme facilitates the description of the climatic conditions of a region by combining multiple meteorological variables into a single metric (Chen and Chen, 2013). It is one of the most widely applied climate classification schemes due to the generally accepted ecological relevance of the classifications (MacLeod and Korycinska, 2019). This is due to the inherent relationship that exists between plants, animals and their abiotic environment. This relationship is especially important for those species that cannot regulate their own temperature (Andrewartha and Birch, 1984), which includes all the pest species in the current analyses. The Köppen scheme uses seasonal temperature and precipitation regimes to distinguish 31 different categories of climate (listed in Appendix 1). Here, an updated version of the classification scheme is used which provides the data at a 1 km resolution using an ensemble of four high resolution topographically-corrected climatic maps (available at www.gloh2o.org/koppen). Each region within each cluster, for all the taxa analysed, was subsequently mapped and overlaid on the updated Köppen-Geiger scheme using ArcGIS 10.2.2. For each of the clusters, the number of grid cells per Köppen classification was calculated for each of the taxa studied in order to determine the dominant climate zones/types for each cluster.

Results

Indices

The three indices utilized were examined to identify if an ‘optimal’ clustering solution existed for each of the groups analysed, initially with uncertain species marked as present. An example of the output (for the Hemiptera) is shown in Figure 1. Figure 1a shows the output from the Calinski-Harabasz index, which evaluates the cluster solution by assessing both the similarity of objects within cluster and the dissimilarity between clusters. The maximum value indicates the optimal number of clusters, which in this case, is three. Conversely, a low value indicates a good cluster structure for the Davies-Bouldin criterion, which in this case is three (Figure 1b). Finally, a large value for the Silhouette index indicates an optimal number of clusters, which is also three (Figure 1c). Table 2 displays all of the cluster solutions identified by the three indices for each group analysed.

Where all of the indices agreed on a cluster number, this was adopted. Where the cluster solutions were not consistent across the indices, each index output was assessed based on whether it provided biologically coherent, distinct and interpretable pest groupings. Using this approach, the indices acted as a guide for prioritizing the cluster solutions to be examined for ecological

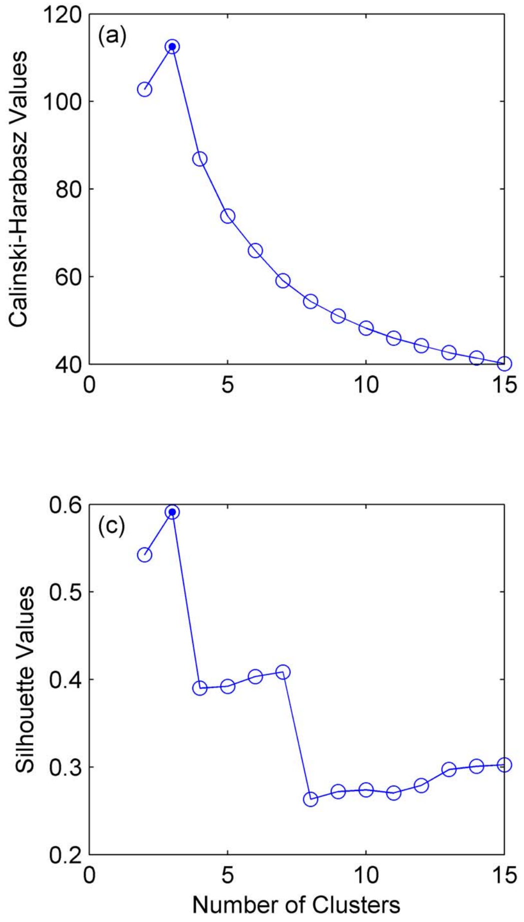

Figure 1 The optimal number of clusters (solid blue dot) based on the analysis of the Hemiptera dataset using (a) the Calinski-Harabasz index, (b) the Davies-Bouldin index and (c) the Silhouette index.

Table 2 The optimal number of clusters identified for each pest group using three different indices.

Organism group	Calinski	Davies	Silhouette	Final cluster	Difference
Coleoptera	3	2	2	5	No
Diptera	15	2	15	N/A	No
Fungi	2	2	10	6	Yes
Hemiptera	3	3	3	3	No
Hymenoptera	2	4	2	3	No
Lepidoptera	3	2	2	3	No
Nematoda	2	3	3/**2***	3	Yes

‘Final cluster’ denotes the final number of clusters used in the analysis. N/A is used when a sensible classification was not attainable. ‘Difference’ indicates that there was a difference in the clustering solutions returned depending on whether ‘uncertain’ species were coded as present or absent.
*Value differed depending on whether uncertain were marked as present or absent (present in bold).

realism. The resulting cluster maps (Figure 2) were examined to determine if the cluster numbers identified by the indices made biological sense, or if a different number of clusters may be more appropriate. Following analysis of the maps using the index-proposed cluster numbers, it was determined that the cluster solution selected for the Hemiptera and Lepidoptera produced biologically meaningful groups. Cluster solutions outside those proposed by the indices were also assessed in order to ascertain if other cluster numbers could provide ecologically valid groupings for the remaining groups of organisms. Biologically meaningful clusters were identified for the Coleoptera (five clusters), Hymenoptera (three clusters) and Fungi (six clusters). The three cluster solution identified by the indices produced meaningful groupings for the Nematoda; however, the groups formed tended

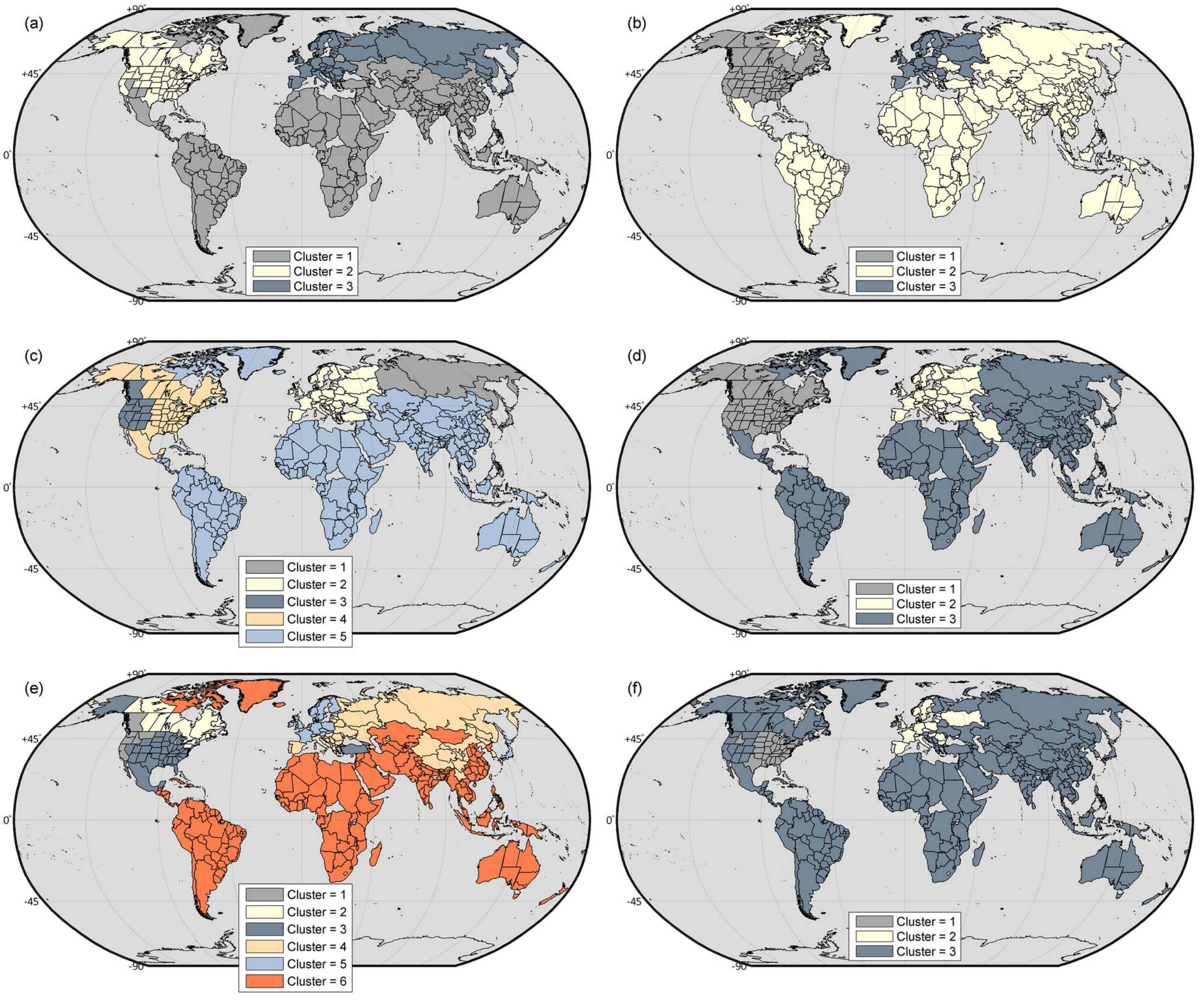

Figure 2 Cluster partitions for the (a) Hymenoptera, (b) Lepidoptera, (c) Coleoptera, (d) Hemiptera, (e) Fungi and (f) Nematoda.

to be highly influenced by the presence of a very small number of species due to the small size of the input data ($n = 21$). For the Diptera, the cluster solutions were not biologically interpretable in the context of pest assemblages, potentially due to the relatively low number of species ($n = 12$) in the dataset. The final accepted number of clusters per taxonomic group ('Final Cluster') is listed in Table 2 along with the index derived cluster numbers. The RIs were calculated to rank species according to their quantified level of risk and the top 10 potential pests of Sitka spruce from each of the groups analysed are presented in Table 3.

Uncertain marked as present

In order to investigate the potential for under-recording of species' distributions to influence the cluster outcomes, the analysis was re-run with all instances of a species marked as 'uncertain' recorded in the distribution, recoded to 'present'. The same approach used in the section above was adopted, with indices being generated for each group (Table 2) and the biological validity of the final cluster solutions assessed. In this run, the Silhouette method identified two clusters for Nematoda (rather than the three identified when 'uncertain' were coded as absent), but all other indices for other species groups remained the same (Table 2). In addition, no differences were noted in the resulting maps for the majority of the groups analysed (Table 2). The only two changes to the cluster solutions were recorded for the (1) Nematoda, where Ireland was moved into the European cluster and (2) Fungi, where the Netherlands and Denmark moved out of the European cluster containing Ireland.

Climate classification

Analysis of the Köppen types found in the native Pacific Northwest range for Sitka spruce displayed 11 different climate classifications, with *Dfc* (cold, no dry season, cold summer) and *Cfb* (temperate, no dry season, warm summer) accounting for the top two climate types. When the European distribution for Sitka spruce was assessed, four dominant climate types were

Table 3 Ranked indices for the 10 species with the highest RI across each of the groups for which meaningful cluster solutions were derived (protected zone species marked with an asterisk).

Species	Risk index	Species	Risk index
Hemiptera		**Fungi**	
Cinara cedri	0.95	*Phacidium abietinum*	0.88
Orsillus depressus	0.70	*Heterobasidion parviporum*	0.81
Gastrodes abietum	0.68	*Lirula macrospora*	0.81
Carulaspis juniperi	0.61	*Veluticeps abietina*	0.81
Dynaspidiotus abietis	0.59	*Pestalotia hartigii*	0.81
Physokermes Piceae	0.55	*Herpotrichia juniperi*	0.75
Physokermes hemicryphus	0.52	*Lachnellula suecica*	0.75
Dichrooscytus intermedius	0.50	*Ophiostoma Piceae*	0.75
Cinara piceicola	0.48	*Cytospora abietis*	0.63
Heliococcus bohemicus	0.45	*Tympanis hypopodia*	0.63
Coleoptera		**Hymenoptera**	
Arhopalus rusticus	0.97	*Xiphydria camelus*	0.57
Pityphilus fasciculatus	0.97	*Gilpinia hercyniae*∗	0.53
Ips acuminatus	0.95	*Cephalcia arvensis*	0.51
Ips sexdentatus∗	0.95	*Pristiphora saxesenii*	0.47
Prionus coriarius	0.95	*Cephalcia abietis*	0.43
Orthotomicus suturalis	0.92	*Megastigmus strobilobius*	0.39
Trachyphloeus bifoveolatus	0.92	*Cephalcia erythrogaster*	0.37
Xyleborus dispar	0.90	*Megastigmus suspectus*	0.33
Arhopalus tristis	0.87	*Pristiphora compressa*	0.33
Callidium violaceum	0.87	*Gilpinia abieticola*	0.29
Lepidoptera		**Nematoda**	
Archips oporana	0.96	*Helicotylenchus dihystera*	0.44
Pandemis cinnamomeana	0.93	*Paratrichodorus minor*	0.24
Thera variata	0.93	*Pratylenchus penetrans*	0.22
Sphinx pinastri	0.89	*Bursaphelenchus xylophilus*	0.12
Assara terebrella	0.89	*Bursaphelenchus mucronatus*	0.12
Peribatodes secundaria	0.89	*Hoplolaimus galeatus*	0.10
Cymolomia hartigiana	0.89	*Pratylenchus pratensis*	0.08
Cydia pactolana	0.86	*Pratylenchus crenatus*	0.06
Chionodes electella	0.86	*Paratylenchus projectus*	0.05
Choristoneura diversana	0.86	*Tylenchorhynchus claytoni*	0.05

identified (Cfb, Cfc,Dfb and Dfc) (for a full description of climate types, see Appendix 1). Finally, for each of the pest clusters produced, the number of grid cells for each climate type was converted to percentages and the results are displayed as a heatmap in Figure 3.

Discussion

Indices and cluster maps

All of the groups analysed include a European or Eurasian (in the case of the Hymenoptera) cluster (Figure 2), highlighting the similarity of pest assemblages within Europe (and their corollary biotic/abiotic conditions). A clustering solution of six was identified as the most biologically meaningful solution for the Fungi (as opposed to the two-cluster solution identified by the indices (Table 2)). This split North America into three clusters consisting of the Pacific Northwest (excluding Alaska), Canada and the north-eastern US and the rest of the US—anomalously Turkey also clusters into this latter cluster. Two Eurasian clusters are formed consisting of Northern and Central Europe and the Japanese islands of Hokkaido and Honshu, and then the rest of Eurasia. A three cluster grouping was identified as optimal for the Hymenoptera, with two clusters covering the Holarctic (the natural distribution of spruce), which are split between Eurasia and Europe due to the geographic barriers that lead to different pest assemblages in these two regions. The three-cluster partition resulting from analysis of the indices for the Lepidoptera constituted interpretable groups, including a group forming a North American cluster, one that contains much of Europe, and finally, the rest of the world. Upon examination of the proposed cluster partitions for the Coleoptera, the three cluster solution (identified using the indices) displayed biological groups that made ecological and biological sense, with a Eurasian, North American and ‘rest of the world’ cluster. The four-cluster solution furthered the explanatory power of the partitions, including a

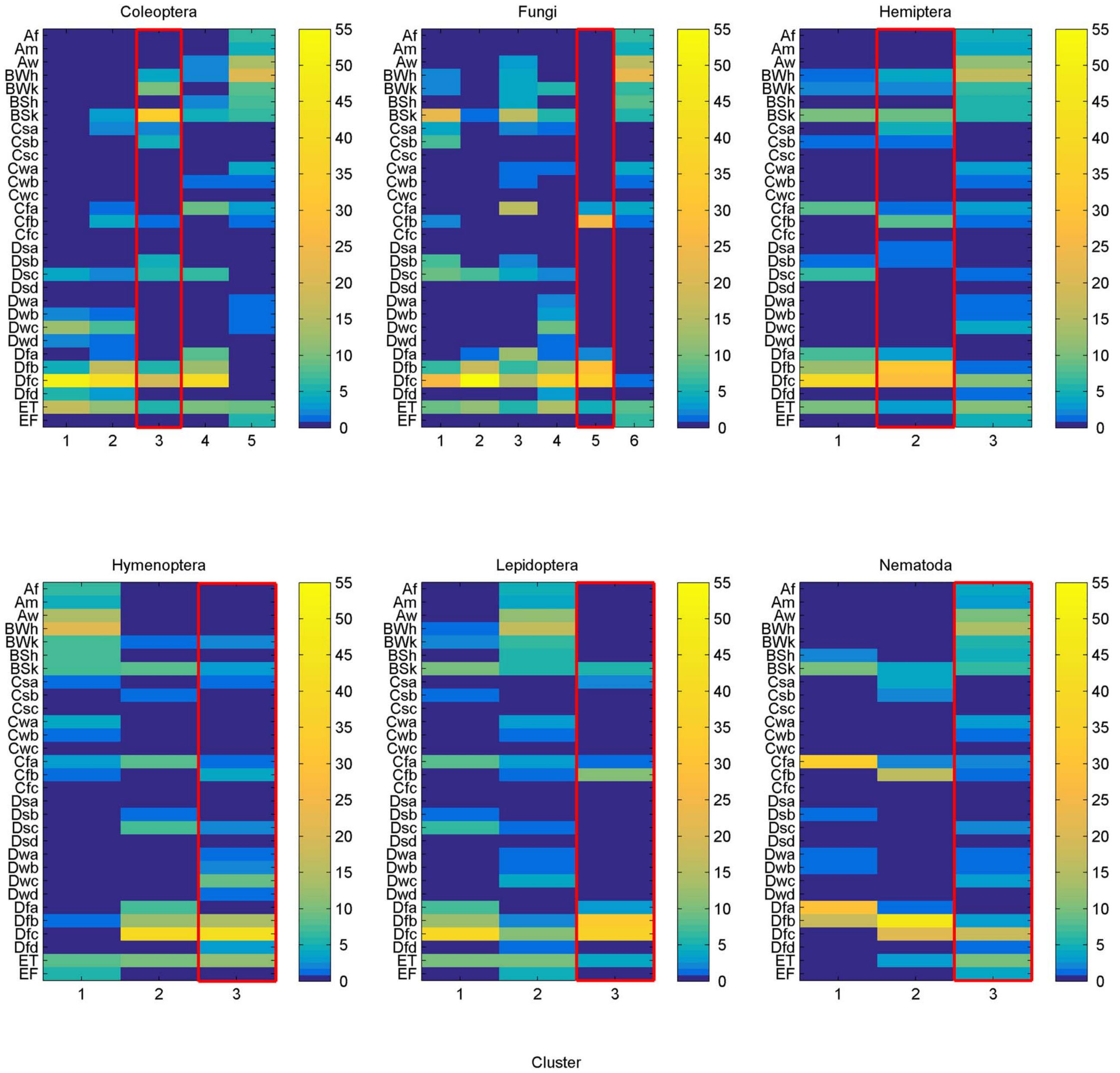

Figure 3 Heatmaps displaying the percentage of Köppen classifications per cluster per taxon. Legend for climate types is in Appendix 1. Red rectangle denotes the cluster containing Ireland for each of the groups.

Western and Eastern North American cluster (split along the Rocky Mountains), Eurasia and the rest of the world. Upon further examination, the five cluster solution was identified as optimal in biological terms, providing additional physical meaning to the group demarcations, with Asia and Europe being split along the Ural Mountains. Finally, the three groupings identified for the Hemiptera show distinct European, North American and 'rest-of-the-world' clusters (Figure 2d).

For five out of the six groups interpreted, Ireland was included within the European cluster, indicating the potential for regions within this cluster to be 'donor' regions of pests to Ireland. The study by Eschen *et al.* (2014) revealed similar grouping of European countries for the pests analysed. For each grouping produced (Figure 2), a counterintuitive cluster exists which includes countries such as Australia, Brazil and Greenland. This 'rest of the world' cluster exhibited by all of the groups are clusters that are defined by their absences of pest species of *Picea*, i.e. the majority of the species in the analysis do not occur in these regions and they form a cluster as they are similar in their dissimilarity to the other clusters whose production is based on similar presences within the assemblages. The presence/absence data for these 'absent clusters' predominantly display zeroes, highlighting the fact that there are very few pests of *Picea* found in these regions. This makes sense considering the distribution of Spruce worldwide, however, it raises the question whether countries such as these should be included in the analysis. Their inclusion however is important for two reasons (1) they provide a benchmark-assemblage against which assemblages characterized by many presences can be compared (i.e. if the 'absent clusters' were not included, the dissimilarity between the remaining countries would be altered, as would the resultant clusters). (2) *A priori* removal of countries based on an any number of presences would be entirely arbitrary and potentially confound the production of scientifically robust outputs. It is important to still include this data in case it has something to offer the clustering (a mostly cosmopolitan pest not currently found in Ireland may well pose a risk to Ireland, even if it is also present in Africa (perhaps reflecting a trade route with Europe)).

Maps: uncertain marked as absent

For the clustering procedure where the presence of a species was indicated as uncertain in the original data, and subsequently recoded as absent, the nematode cluster solution was difficult to interpret initially. Further examination revealed that the split between cluster 1 and 2 was heavily influenced by the presence of the species *Paratylenchus projectus* (Jenkins), *Paratrichodorus minor* ((Colbran) Siddiqi) and *Hoplolaimus galeatus* ((Cobb) Thorne), all of which display the highest RI achievable of one for cluster 1 within the nematode analysis. The presence of these three species together appears to be the primary driving force behind the grouping of US states in cluster 1. Their presence also explains the inclusion of Idaho, the only western state to be incorporated into this cluster (also the only western state with these three species present) (Figure 3). Only a single nematode from the pest distribution list could be confirmed as present in Ireland, explaining why Ireland fails to cluster with the rest of Europe.

For Fungi, some aspects of the clustering analysis are difficult to interpret. Two clusters were suggested by the indices, the first cluster broadly covered the Holarctic with the second cluster covering the rest of the world. This is likely a reflection of the natural distribution of spruce, and thus the fungal species associated with it. Upon further examination of all the potential cluster solutions for this group, six clusters provided groupings with ecological and geographical meaning when the US and Canada are split into three clusters. Cluster 1 forms in the Western US, reflecting the fact the pest list had concentrated on identifying pests to Sitka spruce, which is native to the Pacific Northwest (Durrant Houston *et al.,* 2016). The Cfb climate classification is shared by this native region and Ireland (Figure 3b). Cluster 2 incorporates much of Canada and New England, whereas cluster 3 takes in the rest of the US. This is likely representative of the distribution of spruce in North America. Pine is the dominant species across much of the US, except for New England where there is a spruce/pine mix similar to Canada (Forest Service, 1965). It is uncertain why Alaska is in cluster 3 rather than clustering with the Pacific Northwest in cluster 1; this is most likely due to under-recording of fungi in Alaska due to it being a less densely populated state with vast areas of wilderness and therefore potentially subject to fewer fungi surveys. It is also highly anomalous that Turkey clusters with cluster 3 and the reason for this is not known. In cluster 5, Hokkaido and Honshu in Japan cluster with the European regions of Great Britain, Ireland, Norway, Sweden, Finland, France, Switzerland, Czech Republic, Slovakia, Poland and Austria. Japan clustering with Europe is likely to be related to the availability of the 'Index of Fungi on Woody Hosts in Japan' (Kobayashi, 2007). Access to this document meant that Japan was well documented and contributed a largely representative species sample towards the HC, whereas Fungi are less well documented for Russia (or the literature is less accessible). Many of the fungi within this cluster are potentially present across the boreal forests of Eurasia, but a lack of distribution data for cryptic fungi may have prevented such a cluster being formed in the analysis—the rest of the fungi widespread in Eurasia are likely represented in cluster 4. In a pest risk context, the clustering of Japan with Europe could also serve to highlight that donor regions of potential pests for the group are not limited to European nations (like many of the other groups) but also parts of Asia displaying similar pest assemblages from this group.

Upon examination of the proposed cluster partitions for the Coleoptera, both 4 and 5 cluster solutions were ecologically interpretable—the four-cluster partition included a western and eastern North American cluster (with the Rocky Mountains providing a physical boundary for the split), Eurasia and the rest of the world, while at five clusters, Asia and Europe split along the Ural Mountains. For Lepidoptera, it is uncertain why countries along the Adriatic and Portugal do not cluster with the rest of Europe, but it may be to do with the native range of Norway spruce (the dominant spruce species in Europe), which reaches its limits in this area.

Maps: uncertain marked as present

The inclusion of all 'uncertain' nematode species as 'present' within the input files to the HC serves to highlight how sensitive this type of analysis is to underrepresentation of a group. In this iteration, the inclusion of the uncertain species produces a European cluster which includes Ireland. This may be partly explained by the poor records of nematode species within the country, which has no official list of nematode species present (Fitzpatrick *et al.,* 2010) (which led to the recording of many of the species as 'uncertain'). While Ireland is generally considered depauperate in its species in comparison with our mainland Europe counterparts, it is possible that a number of the nematode species on the list are in fact present and that Ireland's omission from the European cluster in the first part of the analysis is a direct result of their under-recording. The second change resulting from analyzing the uncertain species as absent is found within the Fungi and is relatively minor. Denmark and Netherlands move out of the European cluster, which could suggest that the approach provides more nuanced results, where uncertainty exists (in the case of the Fungi, 13 per cent of the data was recorded as 'uncertain'). The removal of the Netherlands moves it into the same cluster as its neighbour, Belgium, to which it is ecologically similar. The removal of Denmark however is a little more difficult to explain. Suffice to say, that the alteration of the uncertain species associated with Ireland changed the pest profile sufficiently to render Ireland more 'similar' to the countries within the European cluster than they were to Denmark or the Netherlands. Paini *et al.* (2010) investigated the sensitivity of SOMs to errors in input data (absences recorded as presences and vice versa) and found that the method was insensitive to errors of up to 20 per cent. This finding suggests that perhaps SOM is less sensitive to errors than the HC approach, considering that a 13 per cent change in input data for the Fungi precipitated a change in the map output of this analysis.

The authors also considered whether or not a potential cutoff or limit existed within a dataset size, wherein the clusters would remain stable once a critical number of species was incorporated. This was a different way of looking at the 'uncertain marked as present' analysis wherein it was not the effect of uncertainty around a species' presence, but rather the number of species incorporated in the analysis being examined. This was informally tested for each of the groups, by running the cluster analysis for the cluster solutions identified for each group, wherein a single species was removed at each iteration (the premise being that a stable cluster solution may exist once a specific number of species was incorporated). A single critical size of input data was not identified, as for each group, the amount of species data required for the clusters to 'stabilize' (remain relatively unchanged) changed depending on which group was examined. This suggests that the size of the dataset may be less important than the data that it contains, highlighting once again the importance of reliable, available data if this sort of approach is to be truly useful.

Risk index

The RIs calculated here apply only to the species recorded within the EU clusters produced. This is not perceived as an issue however, as the entities within the cluster represent the climatic conditions, ecological state and history of trade of the cluster regions. It follows that regions which do not share the cluster will display different biotic and abiotic conditions and constitute a lower risk of being a 'donor' of alien species. Additionally, as the focus here is on Ireland, only those species that share a cluster with the island are of interest in the context of 'high risk of establishment' pests. The pests identified within the current analysis using the calculated RI's (Table 3) merit further examination and analysis in the form of a PRA in order to assess their potential risk to Sitka spruce in Ireland (and potentially start the process of regulation of some of these species).

The spruce pest list described by Tuffen and Grogan (2019) includes information on the pathways the pests could be associated with, as well as aspects of their biology (such as a cryptic lifecycle which would make their detection during inspections more difficult). Of the 60 species listed in the present paper as having the highest RIs for their taxa, 52 have 'plants for planting' as a potential pathway of entry; 50 had a pathway of entry for which there were no pest-specific phytosanitary measures in place (i.e. an 'unregulated pathway'), and with evidence of trade along that pathway, while 17 of the 60 also have an 'uncertain' distribution status on the island of Ireland. Many of the high risk pests identified have a cryptic lifecycle (e.g. *Ips sexdentatus* (Börner), *Arhopalus rusticus* (Linnaeus), *Xiphydria camelus* (Linnaeus), *Archips oporana* (Linnaeus)) meaning that those entering on an unregulated path would be difficult to detect during inspections. Perhaps surprisingly, only 19 of these 60 high-risk species have been recorded as invading new regions. However, it is important to consider this in the context of the Irish landscape: spruce is not native to Ireland and for many years Ireland had exceptionally low forest cover, with the current levels of 10.5 per cent the highest for over 350 years. Thus, species naturally associated with spruce in Europe will increasingly have the opportunity to be introduced to Ireland. In contrast, for the lowest ranking 60 species across taxa (Appendix 2), plants for planting was a pathway for 41 of the pests; however, only 26 of the 60 low-risk species had an 'unregulated pathway' of entry, with 18 of those species having evidence of trade along potential pathways of entry. This indicates that pest rankings also potentially reflect current trading patterns—lower ranking pests tended to be limited in their distribution to North America from which import of many coniferous plants for planting is prohibited, and coniferous wood imports must undergo phytosanitary treatments. It is, however, important that the risk posed by lower ranking pests is still considered—as the clusters reflect historical trade routes, and thus new trades initiated could provide the pests with a potential pathway of entry.

Pests identified as high risk using the RIs include those for which Ireland already has legislation in place (in the form of a protected zone); namely *Ips sexdentatus* (Börner) within the Coleoptera and *Gilpinia hercyniae* (Hartig) within the Hymenoptera. Interestingly, the bark beetle *Ips typographus*, L. ranked number 20 in the overall Coleoptera RIs and this particular species was recently discovered in Kent in the UK (Gov.UK, 2018). The discovery was of a breeding population and the species is one of the most significant pests of European spruce forests (Wermelinger *et al.*, 2012).

As a consequence of the current analysis, *Physokermes* (spruce bud scale) species which ranked sixth and seventh within the Hemiptera cluster, were subsequently prioritized for PRA (Tuffen and Grogan, 2019). In Europe, populations of *Physokermes* species build up occasionally to outbreak levels—causing severe damage to trees by producing excessive amounts of honeydew that lead to sooty mould growth and defoliation (Malumphy, 2009; Miezite *et al.*, 2013). Another pest of interest is *Cydia pactolana* (Zeller), ranked eighth in the Lepidoptera cluster, which was recently reported causing cumulative impacts with the fungus *Neonectria fuckeliana* ((C. Booth) C. González & P. Chaverri) on *Picea abies* in Finland (Uimari *et al.*, 2018). And within the fungi, *Heterobasidion parviporum* (Niemelä & Korhonen) which ranks second is an economically important root and butt rot of *Picea abies* which is widespread in Europe but not recorded in Ireland or the UK.

Climate classification

Across the five invertebrate groups where Ireland is clustered within Europe (Figure 3), the predominant climate types are from the 'temperate' and 'continental' or 'cold' groups (C and D respectively). Ireland's climate is defined as a temperate oceanic climate (or Cfb on the Köppen scheme), which is one of the primary climate types from the native range of *Picea sitchensis*. For each of the groups, the European cluster contains many of the climate classifications identified from the native range. For example, for the fungal group, the dominant Köppen classifications are those where Sitka spruce is found (Cfb: 25 per cent, Dfb: 29 per cent, Dfc: 34 per cent), suggesting a strong relationship

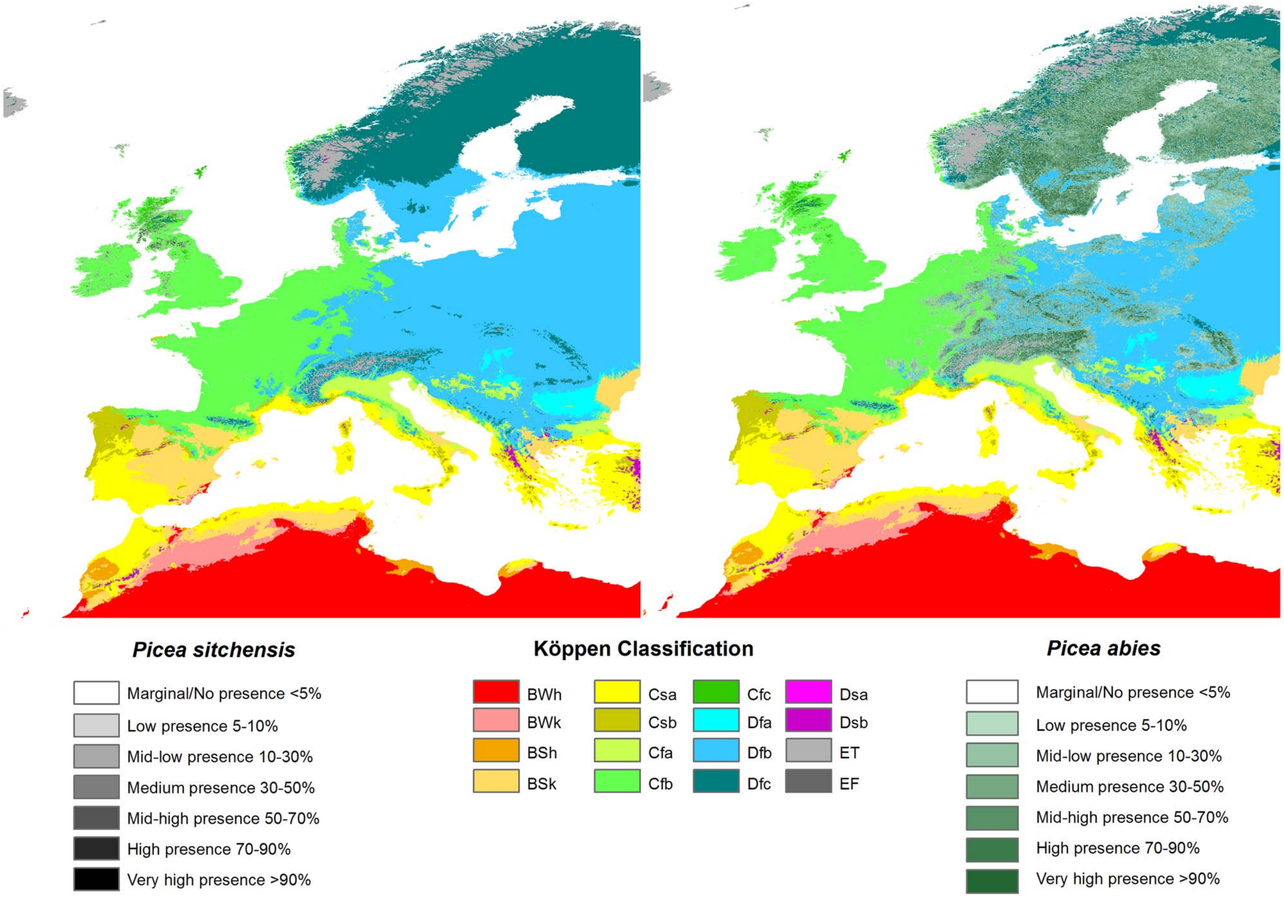

Figure 4 Relative probability of presence of *P. sitchensis* (left) (Durrant Houston *et al.*, 2016) and *P. abies* (right) (Caudullo *et al.*, 2016) in Europe overlaid over the Köppen classifications displayed (Beck *et al.*, 2018).

between the species included in the analysis and their *Picea* hosts (see Figure 4) and indicating favourable climatic conditions for the fungi. Within the other groups analysed (for which Ireland clustered within Europe), the temperate and continental classifications again featured strongly (alongside some instances of dry classifications in the cases of the Coleoptera, Lepidoptera, Hymenoptera and Hemiptera (BSk and BWk)) (Figure 3). The high percentages of the Dfb and Dfc climate classifications across all the Irish clusters are also indicative of the native range for Norway spruce, another host for many of the Sitka spruce pests analysed here. Indeed, the spatial analysis of the Köppen classifications across the native range for Norway spruce described in Caudullo *et al.* (2016) indicated that the two 'cold' climate types account for 98 per cent of the climate types in that range (Figure 4).

Each of the taxa analysed display a cluster, which is clearly climatically different from the rest (i.e. it contains climate types across the spectrum of classifications and includes polar and tropical rainforest classifications, e.g. cluster 1 in the Hymenoptera, cluster 2 in the Lepidoptera, cluster 3 for the Hemiptera and Nematoda, cluster 5 in the Coleoptera and Cluster 6 in the Fungi). These clusters are referred to in the 'indices and cluster maps' section and contain all those species whose pest assemblages are defined by their absence from more temperate climates. Figure 3 indicates the heterogeneity in climate across all the regions included in these clusters, signifying the diversity of environments present in these groups, many of which are unsuitable for *Picea* spp.

Conclusion

The outputs from the HC analysis presented here indicate that the approach is useful at identifying biologically coherent groups based on pest assemblage data. As a result, HC can be used for prioritizing species for evaluation as potential alien pests to Sitka spruce in Ireland or other target regions. The original aim of the research, to provide an objective and quantitative assessment of the risk posed by individual pest species to Sitka spruce in Ireland was achieved by: (1) providing credible map outputs from the HC analysis along with RIs for each species analysed, and (2) identifying Europe as the primary source for new pests. Despite the fact that HC is a less complex method than SOM, the results presented here have real biological meaning. The HC analysis, along with the indices analysed to identify the optimal number of clusters all served to refine a considerably large presence/absence dataset, into a smaller set of maps for analysis. For each of the six groups studied, interpretable European or Eurasian clusters were formed, however, in some cases (like with the Nematoda), small countries (or those with very few records) were placed in different clusters to the rest of the European clusters. The recurring production of a predominantly European cluster across all of the groups indicates that the highest risk pests for Ireland's Sitka spruce plantations will originate from within Europe. As an island, there is more limited potential for pests to arrive via natural spread and pest arrival will most likely be with traded goods—the majority of which (plants and plant products traded into Ireland) originate from within the EU.

More generally, this work highlights a recurring problem in ecological modelling studies where data are frequently limited and of doubtful quality. Despite the use of a novel pest-dataset here, there is continued uncertainty regarding species distribution due to several factors: lack of historical surveys; incidences reported in un-indexed foreign literature; or information that is unreliable (misidentification), imprecise (reported as present within a national boundary) and unrepresentative (only from a portion of a species range) (Venette, 2017). The clustering of Japan with Europe within the fungal group highlights how well this approach works in grouping similar pest assemblages when representative data are available (as is the case with the *Index of Fungi on Woody Hosts in Japan* used in the construction of the pest list (Kobayashi, 2007)). Simultaneously however, it emphasizes that a hierarchy of 'representativeness' exists across geographic ecological data and that those areas that are less well-documented will not be fully represented in presence/absence analysis. Despite these uncertainties, the analysis presented here indicates that ecologically realistic outputs can still be produced.

This application of HC is novel in its application to pests of Sitka spruce and its focus on the island of Ireland to produce biologically meaningful results. HC has been shown to constitute a robust approach in a pest risk context (Eschen *et al.*, 2014; Roige, 2017) and the work presented here supports its applicability in this area. Considering the issue raised in this paper regarding the potential for under-recording to impact the final maps produced, a SOM analysis should be carried out on the current dataset, in order to provide a direct comparison between the resultant maps. This work is currently underway, however, even without it, it is fair to say that the current approach using HC has provided preliminary information for the evaluation and prioritization of pests of Sitka spruce for Ireland. It has also served to highlight potential geographic areas of origin for future pests of this tree species. The successful application of this approach incorporates both the refinement of a large dataset using the HC analysis and indices, as well analysis of resultant cluster maps and cluster climatologies, to derive useful information for PRA. This approach could equally be applied for other target regions to assess vulnerability to pests of *Picea* species and identify high risk species that should be subjected to PRA. This type of approach could be extremely useful for any National Plant Protection Organization seeking to prioritize pests for a full PRA. Indeed, MacLeod and Korycinska (2019) discuss how screening pests prior to in-depth analyses avoids the unnecessary use of limited resources within these organizations. The findings presented here suggest that many of the most significant regions for introduction of new pests to Sitka spruce in Ireland will be from within Europe and a number of species identified within this work have already become the focus of PRA for Ireland (as outlined above). Finally, the importance of high quality, reliable data to this type of analysis has been highlighted in this study, emphasizing how critical data collection and long term targeted monitoring are to plant health on both a national and international scale.

Supplementary data

Supplementary data are available at *Forestry* online.

Acknowledgements

The authors would like to thank the plant health laboratories, DAFM and Forest Service, DAFM for their cooperation on the project.

Conflict of interest statement

None declared.

Funding

Department of Agriculture Food and the Marine (DAFM), Ireland (Project reference: 14/C/889).

Data availability statement

The data underlying this article were provided by Teagasc (The agriculture and food development authority in Ireland) by permission. Data will be shared on request to the corresponding author with permission of Teagasc, including a signed data sharing agreement.

References

Andrewartha, H.G. and Birch, L.C. 1984 *The Ecological Web: More on the Distribution and Abundance of Animals*. University of Chicago Press.

Bebber, D.P. 2015 Range-expanding pests and pathogens in a warming world. *Annu. Rev. Phytopathol.* **53**, 335–356. doi: 10.1146/annurev-phyto-080614-120207.

Bebber, D.P., Ramotowski, M.A.T. and Gurr, S.J. 2013 Crop pests and pathogens move polewards in a warming world. *Nat. Clim. Chang.* **3**, 985–988. doi: 10.1691/ph.2017.6747.

Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A. and Wood, E.F. 2018 Present and future köppen-Geiger climate classification maps at 1-km resolution. *Sci. Data* **5**, 180214. doi: 10.1038/sdata.2018.214.

Bromham, L., Lanfear, R., Cassey, P., Gibb, G. and Cardillo, M. 2012 Reconstructing past species assemblages reveals the changing patterns and drivers of extinction through time. *Proc. R. Soc. B Biol. Sci.* **279**, 4024–4032. doi: 10.1098/rspb.2012.1437.

Caliński, T. and Harabasz, J. 1974 A dendrite method for cluster analysis. *Commun. Stat. Methods* **3**, 1–27. doi: 10.1080/03610917408548446.

Caudullo, G., Tinner, W. and de Rigo, D. 2016 Picea abies in Europe: distribution, habitat, usage and threats. In *European Atlas of Forest Tree Species*. J., San-Miguel-Ayanz, D.D., Rigo, G., Caudullo, A., Mauri (eds.), pp. 114–116.

Céréghino, R., Santoul, F., Compin, A. and Mastrorillo, S. 2005 Using self-organizing maps to investigate spatial patterns of non-native species. *Biol. Conserv.* **125**, 459–465. doi: 10.1016/j.biocon.2005.04.018.

Chen, D. and Chen, H.W. 2013 Using the Köppen classification to quantify climate variation and change: An example for 1901–2010. *Environ. Dev.* **6**, 69–79. doi: 10.1016/j.envdev.2013.03.007.

Chiverton, A., Hannaford, J., Holman, I., Corstanje, R., Prudhomme, C., Bloomfield, J. *et al.* 2015 Which catchment characteristics control the temporal dependence structure of daily river flows? *Hydrol. Process.* **29**, 1353–1369. doi: 10.1002/hyp.10252.

Chon, T.S., Park, Y.S., Moon, K.H. and Cha, E.Y. 1996 Patternizing communities by using an artificial neural network. *Ecol. Model.* **90**, 69–78. doi: 10.1016/0304-3800(95)00148-4.

Coillte Annual Report 2015. 1–104. https://www.coillte.ie/media/2017/02/2015CoillteAnnualReport.pdf.

Copeland, R. and Dowley, L. 2010 Plant pathology: A disappearing science. *Soc. Irish Plant Pathol. Newsl.* **4**. doi: https://dce37fd4-d0cf-4298-b57f-792da77c03ee.filesusr.com/ugd/f1c413_76226e050ffb4b87b35dc728aa048263.pdf

Costanza, J.K., Coulston, J.W. and Wear, D.N. 2017 An empirical, hierarchical typology of tree species assemblages for assessing forest dynamics under global change scenarios. *PLoS One* **12**(9), e0184062. doi: 10.1371/journal.pone.0184062.

DAFM, 2018 Forest statistics Ireland 2017. Johnstown Castle Estate, Co. Wexford. https://www.agriculture.gov.ie/media/migration/forestry/forestservicegeneralinformation/ForestStatisticsIreland2017090318.pdf.

Davies, D.L. and Bouldin, D.W. 1979 A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1**, 224–227.

Desprez-Loustau, M.L. 2009 The alien fungi of Europe. In *DAISIE Handbook of Alien Species in Europe*. Springer, pp. 15–28. https://link.springer.com/chapter/10.1007/978-1-4020-8280-1_2.

Desprez-Loustau, M.L., Robin, C., Buée, M., Courtecuisse, R., Garbaye, J., Suffert, F. *et al.* 2007 The fungal dimension of biological invasions. *Trends Ecol. Evol.* **22**, 472–480. doi: 10.1016/j.tree.2007.04.005.

Durrant Houston, T., Mauri, A., Rigo, D.,.D. and Caudullo, G. 2016 *Picea sitchensis* in Europe: distribution, habitat, usage and threats. In *European Atlas of Forest Tree Species*. J., San-Miguel-Ayanz, D.D., Rigo, G., Caudullo, A., Mauri (eds.). European Commission, pp. 118–119.

EPPO. 2018 *European and Mediterranean Plant Protection Organization Global Database*. https://gd.eppo.int.

Eschen, R., Holmes, T., Smith, D., Roques, A., Santini, A. and Kenis, M. 2014 Likelihood of establishment of tree pests and diseases based on their worldwide occurrence as determined by hierarchical cluster analysis. *For. Ecol. Manag.* **315**, 103–111. doi: 10.1016/j.foreco.2013.12.021.

FAO. 2016 ISPM 2: Framework for pest risk analysis. International standards for phytosanitary measures. https://doi.org/ http://www.fao.org/3/a-k0125e.pdf.

Fitzpatrick, U., Regan, E. and Lysaght, L. 2010 *Ireland's Biodiversity in 2010: State of Knowledge*. Waterford. https://doi.org/www.biodiversityireland.ie/wordpress/wp-content/uploads/Irelands-Biodiversity-20101.pdf.

Forest Service. 1965 *Silvics of North America: Agriculture handbook, Agricultur*. ed. United States Department of Agriculture, Washington DC. https://doi.org/www.srs.fs.usda.gov/pubs/misc/ag_654_vol1.pdf.

Gevrey, M., Worner, S., Kasabov, N., Pitt, J. and Giraudel, J.L. 2006 Estimating risk of events using SOM models: A case study on invasive species establishment. *Ecol. Model.* **197**, 361–372. doi: 10.1016/j.ecolmodel.2006.03.032.

Gov. UK. 2018. *Forestry Commission act on tree pest detected in Kent*. https://www.gov.uk/government/news/forestry-commission-act-on-tree-pest-detected-in-kent (accessed on 12 December, 2018).

Grégoire, J.C. and Evans, H.F. 2004. Damage and control of BAWBILT organisms: an overview. In *Bark and Wood Boring Insects in Living Trees in Europe, a Synthesis*. F. Lieutier, K.R. Day, A. Battisti, J.C. Grégoire and H.F. Evans (Eds). Springer, pp. 19–37. doi: 10.1007/978-1-4020-2241-8

Gubareva, T.S. 2012 Classification of river basins and hydrological regionalization (as exemplified by Japan). *Geogr. Nat. Resour.* **33**, 74–82. doi: 10.1134/S187537281201012X.

Halkidi, M., Batistakis, Y. and Vazirgiannis, M. 2001 On clustering validation techniques. *J. Intell. Inf. Syst.* **17**, 107–145. doi: 10.1023/A:1012801612483.

Huppert, A. and Solow, A.R. 2004 A method for reconstructing climate from fossil beetle assemblages. *Proc. R. Soc. B Biol. Sci.* **271**, 1125–1128. https://doi.org/10.1098/rspb.2004.2706.

IPPC. 2017. *Pest Risk Analysis for Quarantine Pests*. https://www.ippc.int/en/publications/639/.

Kati, V., Dimopoulos, P., Papaioannou, H. and Poirazidis, K. 2009 Ecological management of a Mediterranean mountainous reserve (Pindos National Park, Greece) using the bird community as an indicator. *J. Nat. Conserv.* **17**, 47–59. doi: 10.1016/j.jnc.2008.12.002.

Kobayashi, T. 2007 *Index of fungi inhabiting woody plants in Japan. Host, Distribution and Literature*. Zenkoku-Noson-Kyiku Kyokai Publishing Co.

Köppen, W., Geiger, R., 1936. *Handbuch der Klimatologie: Das geographische System der Klimate*. Verlag von Gebrüder Borntraeger. doi: 10.3354/cr01204

Laize, C.L.R. and Hannah, D.M. 2010 Modification of climate-river flow associations by basin properties. *J. Hydrol.* **389**, 186–204. doi: 10.1016/j.jhydrol.2010.05.048.

Langor, D.W., DeHaas, L.J. and Foottit, R.G. 2009 Diversity of non-native terrestrial arthropods on woody plants in Canada. *Biol. Invasions* **11**, 5–19. doi: 10.1007/s10530-008-9327-x.

Legendre, P. and Legendre, L., 2012. Developments in environmental modelling. In *Developments in Environmental Modelling*. Elsevier, pp. 337–424. doi: 10.1016/B978-0-444-53868-0.50007-1

Leung, B., Roura-Pascual, N., Bacher, S., Heikkilä, J., Brotons, L., Burgman, M.A. *et al.* 2012 TEASIng apart alien species risk assessments: A framework for best practices. *Ecol. Lett.* . doi: 10.1111/ele.12003.

Liebhold, A.M., Brockerhoff, E.G., Garrett, L.J., Parke, J.L. and Britton, K.O. 2012 Live plant imports: The major pathway for forest insect and pathogen invasions of the US. *Front. Ecol. Environ.* 10, 135–143. https://doi.org/10.1890/110198.

MacLeod, A. and Korycinska, A. 2019 Detailing Köppen–Geiger climate zones at sub-national to continental scale: a resource for pest risk analysis. *EPPO Bull.* 49, 73–82. https://doi.org/10.1111/epp.12549.

Malumphy, C.P. 2009 Confirmation of the presence of small spruce bud scale, Physokermes hemicryphus (Hemiptera: Coccoidea: Coccidae), in Britain. *Br. J. Entomol. Nat. Hist.* **22**, 15–22.

Matlab 2013 *Matlab R2013b (8.2.0.701)*. The Mathworks, Inc.

McCracken, A. 2013 Current and emerging threats to Ireland's trees from diseases and pests. *Irish For.* **70**, 36–60.

Miezite, O., Okmanis, M. and Indriksons, A. 2013 Assessment of sanitary conditions in stands of Norway spruce (*Picea abies* karst.) damaged by spruce bud scale (*Physokermes piceae* Schrnk.). *iForest* 6, 73–78. doi: 10.3832/ifor0703-006.

Morin, L., Paini, D.R. and Randall, R.P. 2013 Can global weed assemblages be used to predict future weeds? *PLoS One* **8**, e55547. doi: 10.1371/journal.pone.0055547.

Hanlon, R.O., McCracken, A.R. and Cooke, L.R. 2016 Diversity and ecology of Phytophtora species on the island of Ireland. *Biol. Envirn. Proc. R. Irish Acad.* **116**, 1–25.

Omar, W.M.W. 2010 Perspectives on the use of algae as biological indicators for monitoring and protecting aquatic environments, with special reference to Malaysian freshwater ecosystems. *Trop. Life Sci. Res.* **21**, 51–67.

Paini, D.R., Bianchi, F.J.J.A., Northfield, T.D. and de Barro, P.J. 2011 Predicting invasive fungal pathogens using invasive pest assemblages: Testing model predictions in a virtual world. *PLoS One* **6**, e25695. doi: 10.1371/journal.pone.0025695.

Paini, D.R., Worner, S.P., Cook, D.C., De Barro, P.J. and Thomas, M.B. 2010 Using a self-organizing map to predict invasive species: Sensitivity to data errors and a comparison with expert opinion. *J. Appl. Ecol.* **47**, 290–298. doi: 10.1111/j.1365-2664.2010.01782.x.

Ramachandra Rao, A. and Srinivas, V.V. 2006 Regionalization of watersheds by hybrid-cluster analysis. *J. Hydrol.* **318**, 37–56. doi: 10.1016/j.jhydrol.2005.06.003.

Roige, M. 2017 *The Structure of Global Invasive Species Assemblages and their Relationship to Regional Habitat Variables: Converting Scientifically Relevant Data into Decision Relevant Information*. Lincoln University.

Roige, M., Parry, M., Phillips, C. and Worner, S. 2016 Self-organizing maps for analysing pest profiles: Sensitivity analysis of weights and ranks. *Ecol. Model.* **342**, 113–122. doi: 10.1016/j.ecolmodel.2016.10.003.

Roques, A., Rabitsch, W., Rasplus, J.-Y., Lopez-Vaamonde, C., Nentwig, W. and Kenis, M. 2009 Alien terrestrial invertebrates of Europe. In *DAISIE Handbook of Alien Species in Europe*. Springer, pp. 63–79.

Rousseeuw, P.J. 1987 Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65.

Santini, A., Ghelardini, L., De Pace, C., Desprez-Loustau, M.L., Capretti, P., Chandelier, A. *et al.* 2013 Biogeographical patterns and determinants of invasion by forest pathogens in Europe. *New Phytol.* **197**, 238–250. doi: 10.1111/j.1469-8137.2012.04364.x.

National Forest Service. 2013 *National Forest Inventory Results Data (2012). Chapter 4: Species*. https://www.agriculture.gov.ie/nfi/nfisecondcycle2012/nationalforestinventoryresultsdata2012/

Singh, W., Hjorleifsson, E. and Stefansson, G. 2011 Robustness of fish assemblages derived from three hierarchical agglomerative clustering algorithms performed on Icelandic groundfish survey data. *ICES J. Mar. Sci.* **68**, 189–200. doi: 10.1093/icesjms/fsq144.

Tuffen, M.G. and Grogan, H.M. 2019 Current, emerging and potential pest threats to Sitka spruce plantations and the role of pest risk analysis in preventing new pest introductions to Ireland. *Forestry* **92**, 26–41. doi: 10.1093/forestry/cpy036.

Uimari, A., Heliövaara, K., Tuba, K., Poteri, M. and Vuorinen, M. 2018 Occurrence of the moth Cydia pactolana is associated with the spruce canker fungus *Neonectria fuckeliana*. *Scand. J. For. Res.* **33**, 529–534. doi: 10.1080/02827581.2018.1427791.

Venette, R.C. 2017 Climate analyses to assess risks from invasive Forest insects: Simple matching to advanced models. *Curr. For. Rep.* **3**, 255–268. doi: 10.1007/s40725-017-0061-4.

Ward, J.H. 1963 Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58**, 236–244. https://doi.org/10.1080/01621459.1963.10500845.

Watts, M.J. and Worner, S.P. 2009 Estimating the risk of insect species invasion: Kohonen self-organising maps versus k-means clustering. *Ecol. Model.* **220**, 821–829. doi: 10.1016/j.ecolmodel.2008.12.016.

Wermelinger, B., Epper, C., Kenis, M., Ghosh, S. and Holdenrieder, O. 2012 Emergence patterns of univoltine and bivoltine *Ips typographus* (L.) populations and associated natural enemies. *J. Appl. Entomol.* **136**, 212–224. doi: 10.1111/j.1439-0418.2011.01629.x.

Worner, S., Gevrey, M., Eschen, R., Kenis, M., Paini, D., Singh, S. *et al.* 2013 Prioritizing the risk of plant pests by clustering methods; self-organising maps, k-means and hierarchical clustering. *NeoBiota* **18**, 83–102. doi: 10.3897/neobiota.18.4042.

Worner, S.P. and Gevrey, M. 2006 Modelling global insect pest species assemblages to determine risk of invasion. *J. Appl. Ecol.* **43**, 858–867. doi: 10.1111/j.1365-2664.2006.01202.x.

Zolfaghari, F., Khosravi, H., Shahriyari, A., Jabbari, J. and Abolhasani, A. 2019 Hierarchical cluster analysis to identify the homogeneous desertification management units. *PLoS One* **14**, e0226355. https://doi.org/10.1371/journal.pone.0226355.

Appendix 1: Köppen classification legend (after Beck *et al.*, 2018).

Number	Letters	Climate description		
1	Af	Tropical	Rainforest	
2	Am	Tropical	Monsoon	
3	Aw	Tropical	Savannah	
4	BWh	Arid	Desert	Hot
5	BWk	Arid	Desert	Cold
6	BSh	Arid	STEPPE	Hot
7	BSk	Arid	Steppe	Cold
8	Csa	Temperate	Dry summer	Hot summer
9	Csb	Temperate	Dry summer	Warm summer
10	Csc	Temperate	Dry summer	Cold summer
11	Cwa	Temperate	Dry winter	Hot summer
12	Cwb	Temperate	Dry winter	Warm summer
13	Cwc	Temperate	Dry winter	Cold summer
14	Cfa	Temperate	No dry season	Hot summer
15	Cfb	Temperate	No dry season	Warm summer
16	Cfc	Temperate	No dry season	Cold summer
17	Dsa	Cold	Dry summer	Hot summer
18	Dsb	Cold	Dry summer	Warm summer
19	Dsc	Cold	Dry summer	Cold summer
20	Dsd	Cold	Dry summer	Very cold winter
21	Dwa	Cold	Dry winter	Hot summer
22	Dwb	Cold	Dry winter	Warm summer
23	Dwc	Cold	Dry winter	Cold summer
24	Dwd	Cold	Dry winter	Very cold winter
25	Dfa	Cold	No dry season	Hot summer
26	Dfb	Cold	No dry season	Warm summer
27	Dfc	Cold	No dry season	Cold summer
28	Dfd	Cold	No dry season	Very cold winter
29	ET	Polar	Tundra	
30	EF	Polar	Frost	

Appendix 2: Ten species with the lowest RI across each of the groups for which meaningful cluster solutions were derived.

Species	Species
Hemiptera	**Fungi**
Chionaspis heterophyllae	*Diaporthe lokoyae*
Phenacaspis pinifoliae	*Isthmiella crepidiformis*
Dynaspidiotus tsugae	*Pucciniastrum americanum*
Fiorinia externa	*Sparassis spathulata*
Hemiberlesia ithacae	*Peniophora pseudopini*
Lepidosaphes japonica	*Ceriporiopsis rivulosa*
Phenacoccus minimus	*Dasyscyphus ellisiana*
Aphrophora parallela	*Ganoderma oregonense*
Clastoptera juniperina	*Stigmina lautii*
Dynaspidiotus californicus	*Meruliporia incrassata*
Coleoptera	**Hymenoptera**
Stictoleptura canadensis	*Sirex abietinus*
Arhopalus productus	*Sirex nitidus*
Neacanthocinus pusillus	*Urocerus flavicornis*
Neacanthocinus obsoletus	*Urocerus californicus*
Dicerca tenebrosa	*Urocerus cressoni*
Scolytus fiskei	*Urocerus sah*
Scolytus opacus	*Xeris indecisus*
Scolytus tsugae	*Xeris caudatus*
Scolytus unispinosus	*Xeris melancholicus*
Hylastes porculus	*Cephalcia provancheri*
Lepidoptera	**Nematoda**
Spiramater lutra	*Hemicycliophora similis*
Palthis angulalis	*Bitylenchus dubius*
Melanchra pulverulenta	*Tylenchorhynchus maximus*
Melanchra picta	*Xiphinema bakeri*
Mamestra curialis	*Rotylenchus robustus*
Feralia major	*Crossonema menzeli*
Feralia jocosa	*Hoplotylus femina*
Feralia comstocki	*Paratrichodorus pachydermus*
Korscheltellus gracilis	*Paralongidorus maximus*
Apotomops wellingtoniana	*Trichodorus velatus*