



Novel spatial analysis methods reveal scale-dependent spread and infer limiting factors of invasion by Sahara mustard

Yue M. Li, Katrina M. Dlugosch and Brian J. Enquist

Y. M. Li (liyue@email.arizona.edu), K. M. Dlugosch and B. J. Enquist, Dept of Ecology and Evolutionary Biology, Univ. of Arizona, Tucson, AZ 85721, USA.

Multiple scale-dependent ecological processes influence species distributions. Uncovering these drivers of dynamic range boundaries can provide fundamental ecological insights and vital knowledge for species management. We develop a transferable methodology that uses widely available data and tools to determine critical scales in range expansion and to infer dominating scale-dependent forces that influence spread. We divide a focal geographic region into different sized square cells, representing different spatial scales. We then used herbarium records to determine the species' occupancy of cells at each spatial scale. We calculated the growth in cell occupancy across scales to infer the scale dependent expansion rate. This is the first time such a 'box-counting' method is used to study range expansion. We coupled this multi-scale analysis with species distribution models to determine the range and spatial scales where suitable climate allows the species to spread, and where other factors may be influencing the expansion. We demonstrate our methodology by assessing the spread of invasive Sahara mustard in North America. We detect critical scales where its spread is limited (100–500 km) or unconstrained (5–50 km) by climatic variables. Using climate-based models to assess the similarity of climate envelopes in its native and invaded range, we find that the climate in the invaded range generally predicts the native distribution, suggesting that either there has been little local adaptation to climate occurring since introduction or the biological interaction experienced in the invaded range has not driven the species to occupy climatic conditions much different from its native range. Our novel method can be broadly utilized in other studies to generate critical insights into the scale dependency of different ecological drivers that influence the spread and distribution limits, as well as to help parameterizing predictions of future spread, and thus inform management decisions.

Species' distributions are shaped by a myriad of ecological factors that fall into three general categories: physical environment, biotic interactions and dispersal (Brown et al. 1996, Kirkpatrick and Barton 1997, McGill 2010). Disentangling these multiple drivers of range boundaries is difficult, and can be confounded by their inherent scale-dependency. Multiple ecological and evolutionary processes that operate on differing spatial and temporal scales work together to set the range boundaries of species (MacArthur 1972, Holt 2003, Gaston 2009, Holt and Barfield 2011), and consequently the distribution of species can show strong correlation with different ecological factors when analyzed at different scales (Wiens 1989, McGill 2010, Estrada-Villegas et al. 2011).

For example, climatic variables such as annual precipitation show the strongest variation at large spatial scales, whereas habitat variables such as soil pH vary more strongly at finer scales. Consequently, climatic variables often appear to be the strongest factor explaining species ranges at the coarsest scale in space (Caughley et al. 1987, Root 1988, Davis and Shaw 2001). Within a coarse scale, a distribution can be patchy given finer scale variation in habitat structure, biotic interactions and dispersal limitation (Greig-Smith 1979, Wethey 2002). Approaches that utilize multi-

scale analyses are essential for revealing how the influence of specific ecological factors changes with scale (Wiens 1989).

In this study, we present a novel methodology that can be broadly-applied to assess the scale-dependency of species' distributions. Distribution changes – such as those that occur in response to species introductions or climate change – provide ecological experiments that can reveal the factors that set range boundaries at specific scales (Holt et al. 2005, Kinlan and Hastings 2005, Pysek and Hulme 2005, Ricklefs 2005, Sexton et al. 2009). Increasing number of studies have assessed the distribution of invasive species across multiple spatial scales. Most of them emphasize on how current extent of invasion changes with spatial scales (Foxcroft et al. 2009) and what ecological factors are associated with this scale dependency of existing invasion extent (Catford et al. 2009, Akasaka et al. 2012). But very few (Lonsdale 1993, Pysek et al. 2008) have examined how historical and current rate of range expansion change with spatial scales. It is important to reveal the temporal dynamics of scale dependent range expansion in order to infer historical and existing ecological factors that may influence distribution changes at specific scales.

We suspect the lack of such study is largely due to methodological barrier. Here we illustrate the use of the

'box-counting' method (Mandelbrot 1983 p. 19) to assess the expansion rate of an invading species along a continuum of spatial scales and across multiple decades. In particular, we ask how the rate of decadal expansion varies across those scales and whether there are critical scales at which the expansion rate changes dramatically. The box-counting method is a simple and powerful tool normally used in scaling analysis for detecting the fractal dimension of a system (Mandelbrot 1983, Morse et al. 1985, Ritchie 1998). Veldtman et al. (2010) used this method to assess how environmental factors affect the occupancy pattern of an invasive *Acacia* species. In this study, instead of taking a static view of invasive distribution, we extend its use to multi spatial scale analysis of the temporal dynamics of range expansion.

To infer specific ecological factors driving scale-dependent expansion, we demonstrate a focal analysis of the importance of climate variables. We use climate based species distribution models to predict the expansion potential of a focal species across spatial scales. Climatic variables are expected to restrict species distribution particularly at large spatial scales (Whittaker et al. 2001, Pearson and Dawson 2003, McGill 2010). Here we not only evaluate this hypothesis but also attempt to quantitatively determine the range of spatial scales where climate has the strongest influence. Next, we infer the influence of dispersal by evaluating the temporal change in the scale dependency of range expansion. One can infer strong contribution by long distance dispersal if high expansion rates prevail at large scales. Moreover, one can infer when this contribution matters the most to range expansion by finding periods during which large scale expansion is most prevalent.

Our focal species is a winter annual plant *Brassica tournefortii* (Brassicaceae), commonly known as Sahara mustard. It is native to Eurasia and Africa and was unintentionally introduced to southern California in the 1920s. It has since invaded vast areas of the southwest (Minnich and Sanders 2000), mainly in the Sonoran and Mojave Deserts. This species grows not only on sand, sandy loam soil and rocky hillside in the Mojave and Sonoran Deserts but also in more mesic environments in California coasts. The distribution of Sahara mustard in its native range covers 0–2400 m elevations (Miller and Cope 1996) and a variety of landscapes such as desert floors, dunes, oases, desert mountains and steppes (Maire 1965, Townsend and Guest 1980, Zohary et al. 1980). Individual seeds germinate after major winter storms. Survived germinants have rapid phenology and can produce seed sets as quick as in less than two months (Marushia et al. 2012). Under favorable conditions, an individual plant can develop basal rosette leaves ~1 m in diameter and stem more than 80 cm tall (unpubl.), and produce > 16 000 seeds (Trader et al. 2006). Its seed coat becomes mucilaginous in contact with water, which may aid dispersal by having wet seeds stick to objects capable of traveling long distance (Bangle et al. 2008). Sahara mustard is one of the very few self-compatible species in the *Brassica* genus (Hinata and Nishio 1980). Compared to congenic self-incompatible species, Sahara mustard produces smaller flowers with a smaller amount of pollen grains. Its anthers face the pistils and pollinate the protruding stigma.

Its successful invasion may be a result of many of the aforementioned attributes: tolerance of a wide range of

environments, rapid phenology, high fecundity, aided seed dispersal, and self-fertilization. Its decades of expansion recorded in herbarium collections provide an opportunity to conduct a multi-scale analysis of its spread. Moreover, its threat to native species (Barrows et al. 2009) begs effective management based on sound understanding of the processes that limit its distribution. We demonstrate the use of our novel method by asking 1) how the expansion of Sahara mustard has varied temporally across spatial scales, 2) whether climate limits the expansion of Sahara mustard at specific spatial scales, and 3) how long distance dispersal affects the scale dependency of its expansion.

Finally, we explore the potential for evolution and biological interaction to limit invasive spread. For introduced species in particular, whether or not a species can occupy novel environments has important implications for predicting spread (Wiens et al. 2010). We use species distribution models constructed from the climate of non-native range to predict the native range of invasive Sahara mustard. If the predicted native range based on non-native distributions consists new territories not observed in its native range, the difference may suggest that either an invasive species has adapted to novel climates during invasion or novel biological interactions in the invaded range has driven the species to occupy climatic conditions that it normally cannot occupy in the native range. If the predicted native range is only a subset of the observed native range, the result may suggest that either the species has not reached the climatic limits of invasion or novel biological interactions in the invaded range prevents the species from occupying all of its native climatic range.

Methods

Estimating the expansion rate of Sahara mustard across multiple spatial scales

We utilized herbarium collection records of Sahara mustard to assess the species' expansion over North America. Like many other species, there is no specific long term study to track the change in distribution of Sahara mustard. We show that herbarium collections are useful for revealing the distribution dynamics of poorly studied species despite the potential sampling biases associated with such collections.

We searched a large number of online databases covering North America (Supplementary material Appendix 3, Table A1). Only the following four herbarium databases had records of Sahara mustard: the Southwest Environmental Information Network (SEINet), Consortium of California Herbaria (CCH), Global Biodiversity Information Facility (GBIF) and New Mexico Biodiversity Collections Consortium. Aggregating the herbarium records by decades, we reconstructed the distribution of the species in each decade since its first discovery in North America. We then used this information to study the scale dependency of range expansion.

Next we assessed the robustness of the box-counting method on a simulated diffusion growth model (Supplementary material Appendix 1) and applied the method to estimate the decadal expansion rate of Sahara mustard across 7 spatial scales (5, 10, 20, 50, 100, 200, and

500 km). We selected the rectangular region that contained all the herbarium records of Sahara mustard in North America (25° to 38.25°N, 99.33° to 122.79°W) and evenly divided this region into square cells of 5 × 5 km² to 500 × 500 km² size, corresponding to the 7 scales. We avoided scales smaller than 5 km because imprecise or inaccurate geographic coordinates of herbarium specimens can introduce large errors to small scale estimates. We summed the number of cells that had herbarium records in each decade. Where the number of occupied cells of size *s* in decade *d* is $N_{s,d}$, the expansion rate at spatial scale of *s* from decade *d* to *d* + 1 was calculated as $r_{d+1,s} = \ln(N_{d+1,s}/N_{d,s})$. Because only five records were reported prior to 1950 (Supplementary material Appendix 3, Table A2), we applied the method to the decadal intervals from year 1950 to 2009.

Our simulated diffusion growth model suggests that as the spatial scale increases, the expansion rate calculated by the box-counting method becomes more sensitive to the placement of the grid. Fortunately, this sensitivity can be largely eased by averaging estimates from multiple placements (Supplementary material Appendix 1). In this analysis, we shifted the placement of the grid in 16 directions ($\pi/8$ radians apart) and three distances (600 m, 26 km, and 108 km) in each direction. We graphically report estimates from all those placements and averaged them to determine the expansion rate at each scale.

To assess how small scale imprecision in geographic coordinates affects our estimations at large scales, we perturbed the original collection data by a normally distributed random error. We used errors with a mean of zero and a standard deviation of 1 km and applied the box-counting method to this perturbed data.

Lastly, we assessed how variation in sampling efforts may affect our analysis. The intensity of floristic surveys at different spatial scales could vary in time, confounding our analysis (e.g. fewer herbarium records within a decade could be a result of less sampling effort). We quantified the variation in sampling efforts by performing the same expansion analysis on native species with presumably stable range (Delisle et al. 2003). We then subtracted this native ‘expansion’ rate (averaged over all grid placements) from the invasive expansion rate to correct for sampling bias. We obtained the herbarium records of three common and widely distributed native winter annual species (*Plantago patagonica*, *Chaenactis stevioides* and *Lepidium lasiocarpum*) from the same databases and within the same geographic boundary as Sahara mustard. Using their combined records, we reconstructed the decadal variation in herbarium sampling efforts (Supplementary material Appendix 2) and used it as a baseline to adjust the expansion rate of Sahara mustard.

Using MaxEnt species distribution models to evaluate the range of Sahara mustard under suitable climatic conditions

We predicted the range of Sahara mustard in North America using species distribution models (SDMs) based on the climatic conditions under which this species has been recorded. We built our SDMs using MaxEnt (ver. 3.3.3k), a program specifically developed for treating presence-only records (Phillips et al. 2006, Elith et al. 2010b). Our MaxEnt

SDMs attempted to construct the species’ distribution under the maximum influence of climatic variables we chose and minimum influence of any other factors.

The expanding range of a non-native species does not reflect a stable relationship with the invaded environments (Elith et al. 2010a). This lack of equilibrium presents challenges for modeling a potential distribution using data from the current distribution. To address this problem, we reduced the complexity of our model by choosing only a few climatic variables that are most biologically relevant and by using features and regularization parameters in MaxEnt that are more appropriate for modeling species with an unstable range (Supplementary material Appendix 4).

Our models used four climatic variables from the Global Climate Database (WorldClim.org): mean temperature of the coldest quarter (TEMPCOLDQ), precipitation of the coldest quarter (RAINCOLDQ), annual precipitation (RAINYEAR), and mean temperature of the warmest quarter (TEMPWARMQ), all of which are biologically relevant to the life history of Sahara mustard (Table 1). Those variables were derived from climatic records between 1950 and 2000 (Hijmans et al. 2005), consistent with the period when most of the occurrences of Sahara mustard were recorded (1940–2010). The variables have a spatial resolution of 30 arcsecond (~1 km²), which was achieved by interpolating data from a global network of weather stations. The interpolation took into account the elevational difference in space but ignored other local scale factors such as aspect (Hijmans et al. 2005).

The background of our models was a polygon that consists of the majority of southwestern North America, from which we drew 10 000 random samples. We also allowed for further expansion by using an enlarged background containing all lower 48 states of the U.S. and the entire territory of Mexico. Models trained under this enlarged background did not give qualitatively different results (Supplementary material Appendix 4, Fig. A3 and A4). Therefore we only report results from models based on the standard background.

Table 1. The four climatic variables used in the species distribution models. In line with modeling a range shifting species, we reduced the complexity of our models by choosing a very few climatic variables that are most biologically relevant to Sahara mustard.

Climatic variable	Abbreviation	Biological relevance
Mean temperature of the coldest quarter	TEMPCOLDQ	The average temperature condition in the growing season of a winter annual plant
Precipitation of the coldest quarter	RAINCOLDQ	The average resource (water) availability in the growing season of a winter annual plant
Annual precipitation	RAINYEAR	The average of maximal potential resource (water) availability for an annual plant
Mean temperature of the warmest quarter	TEMPWARMQ	The high summer temperature required to break seed dormancy in desert winter annual plants (Baskin and Baskin 2006)

None of the climatic variables used were overly correlated with each other (Supplementary material Appendix 4, Table A4); therefore we included all four variables in each model.

We used two datasets of spatially distinctive records to train our models. The first dataset included 414 occurrences drawn from the herbarium records. The second dataset included 2662 records drawn from both the herbaria and invasive plant surveys (hereafter combined records). Those surveys (Supplementary material Appendix 3, Table A1) were conducted in the core region of its invasion (southern California and Arizona) in the 2000s.

The first dataset reflects records acquired by relatively equal sampling efforts across North America. The second dataset reflects records influenced by intensive sampling efforts in its core invaded area. Points close to the species' continental range limit are expected to have a stronger influence on models based on the herbarium records than those on the combined records. We derived ensemble predictions from models based on both datasets to balance the influence of occurrences at range limit and in the range center.

We tested all of our models through 10-fold cross validation and made predictions of presence or absence based on the logistic output of the models (Supplementary material Appendix 4).

Comparing the climatic conditions of Sahara mustard in its native and invaded range to infer evolutionary constraints

We used our developed SDMs to project the range of Sahara mustard over the continents to which it is native (Eurasia and northern Africa). We also surveyed the existing literature and searched the Global Biodiversity Information Facility (GBIF) database to estimate the recorded native range of Sahara mustard (Supplementary material Appendix 3, Table A3). We then compared the model-projected and the recorded native range to identify any key differences.

The literature describes detailed regional distribution and habitat type within each country in the Middle East and Europe, but gives very coarse scale description in northern Africa and central Asia, often mentioning an area that covers multiple countries (Supplementary material Appendix 3, Table A3). We didn't find any flora describing its distribution in Morocco, Tunisia, and Burkina Faso even though the GBIF database shows collections from those countries. We therefore highlighted the entire territory of these countries as potential native range of the species.

Results

Scale-dependent expansion of Sahara mustard

After corrected the bias in sampling efforts, we found that in the 1960s the species experienced rapid expansion across all scales (Fig. 1c). However, estimates beyond the scale of 50 km (inclusive) should be taken with caution given their high sensitivity to grid placement (Fig. 1a). In contrast, range expansion in the 1970s saw a substantial slowdown at

5–200 km, but a dramatic surge at the 500 km scale reflecting range boundary expansion. This is the decade during which the species was first recorded in western Texas, central California and southern Utah. In the 1980s, the expansion rate increased at the 5–20 km scale but decreased at larger scales. Note, the increase at smaller scales is a result of correction for sampling efforts. Without the correction, the expansion rate was indicated to rather decrease at those smaller scales (compare Fig. 1a and c). In the 1990s, the species' range experienced a contraction across most spatial scales.

In the 2000s, the species almost stopped its expansion at the 100–500 km scales but sustained a high rate of expansion at the smaller scales. This is the decade in which the species rose to prominence as a significant concern in the southwest. The herbarium collections nearly quadrupled from 76 records in the 1990s to 279 in the 2000s (Supplementary material Appendix 3, Table A2).

With the exception of the 1960s, estimations of expansion rate of Sahara mustard were generally insensitive to the placements of grids (Fig. 1a). As expected, sensitivity increases with the spatial scale of focus. But even at the largest scales (200 and 500 km), there is no severe fluctuation in estimations as those seen in our simulated model (Supplementary material Appendix 1, Fig. A1). Therefore, the box-counting method produced reliable scale-dependent expansion rate of Sahara mustard for the period between the 1970s and the 2000s. Estimations in the 1960s have wider fluctuation especially at scales larger than 50 km (inclusive), and thus are less reliable. This is probably due to lower number of sampling points.

Small scale error introduced to the geographic coordinates has minor but noticeable effect on the estimated expansion rates across all spatial scales (Fig. 1b). However, its effect does not qualitatively change any pattern of scale dependency. Estimates of the expansion rate are also greatly affected by the variation in sampling efforts (comparing Fig. 1a and 1c).

Predicted range of Sahara mustard in North America under its suitable climate

The range boundaries of Sahara mustard predicted by our climate-based SDMs generally agreed with the recorded boundaries at the regional scale (Fig. 2), indicating little room for further regional scale expansion of this species except in the large area of the Central Valley, California. According to our models, Sahara mustard is likely to occur in areas with mean temperature in the coldest quarter between 3.9 and 17°C, mean temperature of the warmest quarter between 15.7 and 36°C, precipitation of the coldest quarter between 16 and 302 mm, and annual precipitation between 46 and 657 mm (Supplementary material Appendix 4, Fig. A5).

All SDMs performed well according to their cross-validation test scores (Supplementary material Appendix 4, Table A6). Models trained by herbarium records (more influenced by points at the range limit) predicted a more extended range (green areas in Fig. 2) that includes more areas in northern Mexico, Baja California, central and coastal California and a small region in northwestern Florida (not shown in figure). In comparison, models trained by

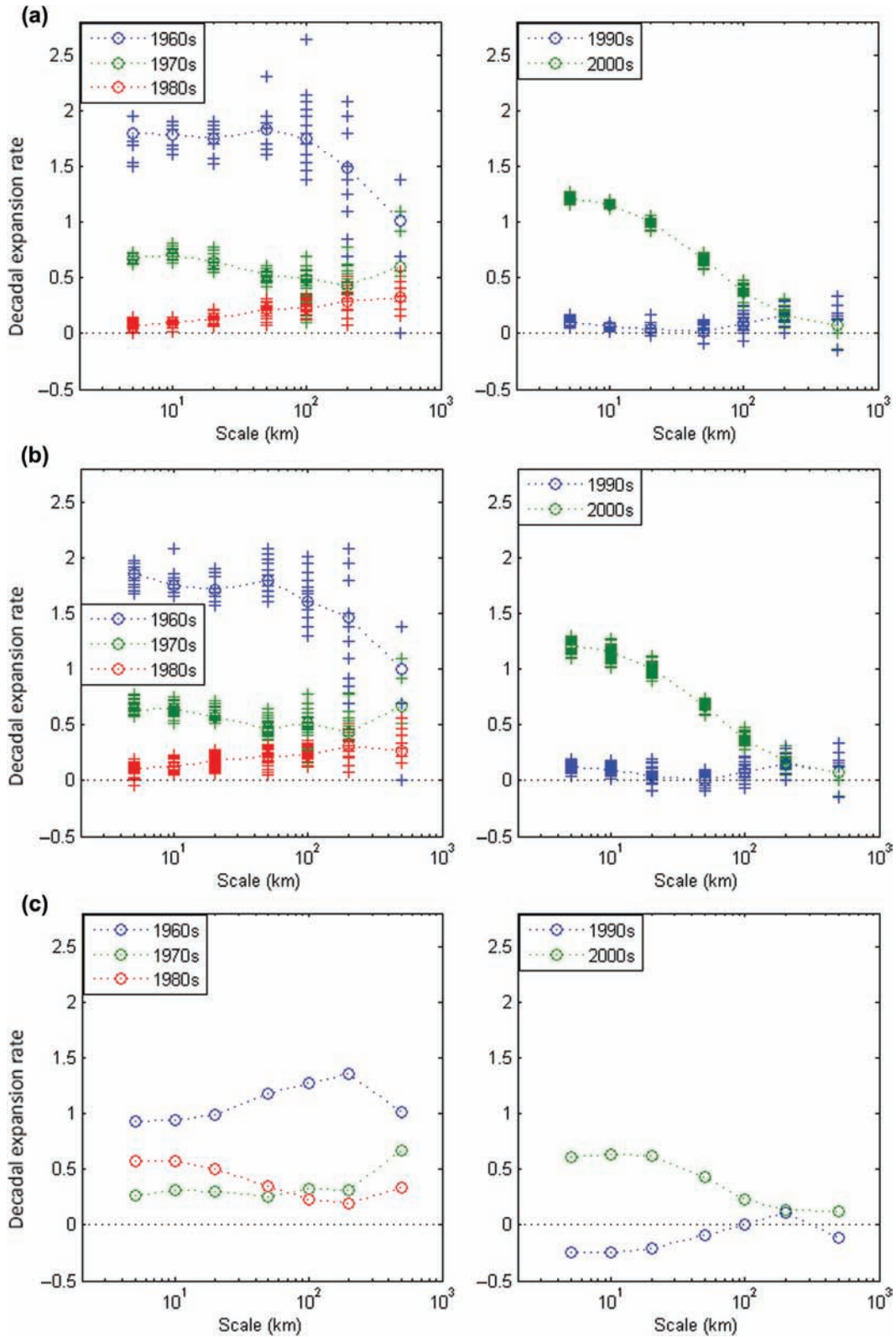


Figure 1. Decadal expansion rates of Sahara mustard from the 1960s to the 2000s at spatial scales of 5–500 km, as estimated from (a) original herbarium data and (b) original data perturbed by a normally distributed random error with a standard deviation of 1 km. '+' indicates each estimate from a single grid placement. Circles indicate a mean value averaged over estimates from all grid placements. (c) Expansion rate adjusted for variation in sampling efforts by subtracting the mean value by an estimate for native species.

combined records (more influenced by points in the core invaded area) predicted a more thorough coverage by the species over the Mojave Desert (yellow areas in Fig. 2).

Agreement of climatic conditions between native and invaded range

The recorded native range of Sahara mustard generally agrees with the projection based on the climatic conditions in its North American range (Fig. 3), suggesting that either this species has not adapted to novel climatic conditions during invasion or biological interactions in the invaded range has not driven the species to occupy new climatic conditions. The SDMs trained by herbarium records predicted a native range that generally extends more southward into the drier regions in Sahara, on the Arabian Peninsula and in central Asia (green areas in Fig. 3), whereas models trained by combined records in North America predicted more coverage in Turkmenistan (yellow areas in Fig. 3), which has drier and colder winters than regions further south. The models predict that Sahara mustard is likely to occur on its native continents in areas with mean temperature of the coldest quarter between -2.3 and 20.6°C , mean temperature of the warmest quarter between 14.8 and 36.7°C , precipitation of the coldest quarter between 0 and 449 mm, and annual precipitation between 0 and 782 mm (Supplementary material Appendix 4, Fig. A5).

There are some mismatches between the model projected and the recorded native range. No records suggest the species' presence in the Nejd region of Saudi Arabia and in Turkmenistan. But our models, even those more influenced by points in the core invaded range predicted its presence in

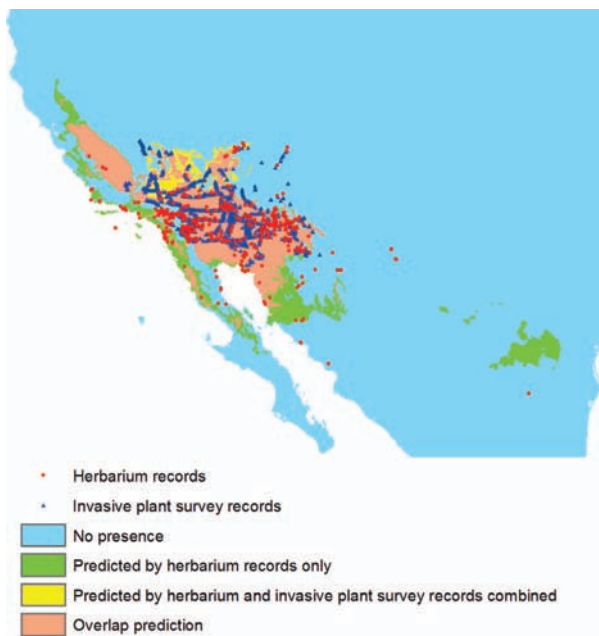


Figure 2. Distribution of Sahara mustard in North America predicted by climate-based SDMs. One model used only herbarium records (red dots) and the other, herbarium and invasive plant survey records (blue triangles) combined. Each ensemble shows the area predicted by both models (peach) and by each model alone (green or yellow). An additional small area predicted by the model in Dixon County, Florida is not shown.

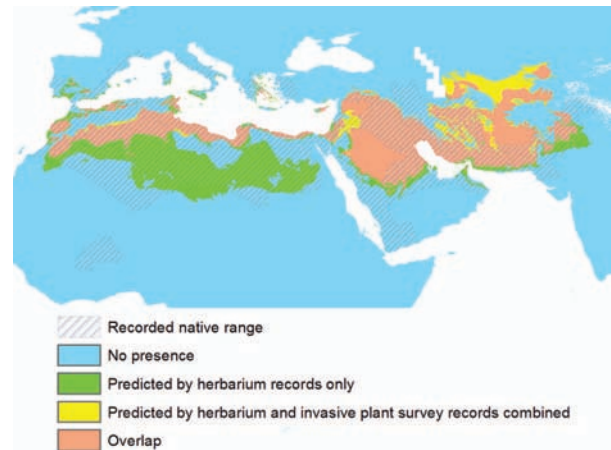


Figure 3. Projected distribution of Sahara mustard in its native range using SDMs based on the climate in its invaded range in North America. Striped areas represent its native range recorded in the literature and the GBIF database.

those regions. None of the models predicted its presence in Burkina Faso, but the species has been found in this country according to GBIF.

Discussion

Climatic constraints on the spread of Sahara mustard and their spatial extent

Our results indicate that the spread of Sahara mustard in North America is indeed spatially scale-dependent and that climate is limiting its spread at regional scales. The expansion of the species at scales beyond 100 km declined to a negligible level in the 2000s (Fig. 1), suggesting a cessation of further expansion of its range boundaries at those scales. The current boundaries of its regional distribution largely agree with those predicted by climate based SDMs (Fig. 2), supporting the role of climate in restricting the spread of the species at those large scales (100 – 500 km). This scale-dependent pattern in our study supports the common view that long-term climate shapes species distributions at relatively large scales. Although this is not a surprising finding, our method is unique in its ability of defining a quantitative extent of the scales of influence. Of course, we cannot rule out the possibility that the expansion is also limited by other unexamined factors that vary strongly at the 100 – 500 km scales.

Our results also reveal that any constraints imposed by long-term climate vanish at smaller scales. Rapid expansion continued at 5 – 50 km scales in the 2000s with a rate generally higher than in the 1970s–1990s (Fig. 1). Our climate based SDMs indicate ample space for the species to further occupy inside its regional climatic envelope, suggesting that long-term climate is not limiting the species' expansion at local to intermediate scales.

The influence of dispersal on the spread of Sahara mustard

Dispersal is generally expected to drive rates of expansion in invasive species (Kot et al. 1996, Phillips et al. 2010),

and Sahara mustard in particular appears to have reached its climatic limits rapidly as a result of long-distance movement. At the 500 km scale, aside from the less reliable estimate in the 1960s, the expansion rate was consistently high in both the 1970s and the 1980s, supporting the view that long distance dispersal at the scale of hundreds of kilometers allowed the species to quickly reach its regional climatic envelope in merely a few decades. Moreover, expansion rates at larger scales were often greater than or equivalent to those at smaller scales between the 1960s and 1980s, suggesting long distance movement that did not rely on sequential population establishment along a shorter dispersal path. Those dispersal leaps could allow Sahara mustard to rapidly extend its range over spatially heterogeneous environments, skipping hostile territories and paving the way for future space filling within its range limit. Its ongoing rapid expansion at the 5–50 km scales also suggests low dispersal limitation at those smaller scales.

Processes that drive the spatiotemporal dynamics of Sahara mustard expansion

The expansion of an invasive species has three phases: a quiescent period between initial colonization and rapid range expansion (the lag phase), a period of rapid range expansion (the spreading phase), and a stage of little further expansion (the ceased phase) (Pysek and Hulme 2005). Scale dependency of invasive spread means that any of the three phases may occur simultaneously when observed at different spatial scales (Pysek and Hulme 2005, Arim et al. 2006). In our case, the expansion of Sahara mustard in the 1960s–1980s was in the spreading phase across all scales. In the 2000s, the species was in the spreading phase at the 5–50 km scales but in the ceased phase at larger scales. Climatic constraints can explain the ceased phase, while the availability of fine scale niche space allowed the species to further expand at smaller scales.

Our results also showed that Sahara mustard had experienced two distinctive lag phases. The first lag phase occurred from the 1920s to the 1940s, during which there was a 14 yr gap between its first and second recorded occurrence (1927 and 1941). The second lag phase occurred in the 1990s, during which expansion was limited across all scales.

An invasive species can experience multiple lag phases due to different causes (Wangen and Webster 2006). In our case, the first lag phase may reflect the time required by Sahara mustard to overcome the problems commonly faced by small founding populations such as loss of genetic diversity, Allee effects, and demographic stochasticity (Crooks and Soulé 1999, Sakai et al. 2001). But it might also reflect the time between the first introduction and a later introduction of a more invasive form of the species. There is some anecdotal evidence that the first introduced form of Sahara mustard, which can still be found near Mecca, California, is smaller and more gracile than those spreading in North America and is more specialized in disturbed soils (M. Dimmit pers. comm.).

The second lag phase, during which the species was well established in North America, was most likely caused by factors other than those that hindered its early spread. A

likely candidate is a period of less-favorable environmental conditions, such as low water availability (a key resource for desert annuals). We examined whether reduced cold (growing) season precipitation was associated with this lag period. Our analysis shows that the cold season precipitation averaged over the 1990s was in fact above the long-term mean in areas where Sahara mustard achieved low (but positive) local-scale expansion (Supplementary material Appendix 5), contradicting a simple explanation of low resource availability.

Plasticity vs adaptation as a process for successful invasion

An introduced species can become a widespread invader by either having low genetic diversity and a ‘general purpose’ genotype or rapidly adapting to its various new environments (Parker et al. 2003). A species with a general purpose genotype tolerates a wide range of environments and can grow in a multitude of climates and habitats (Baker 1965). Sahara mustard, which rapidly spread to cover a broad range of climatic and habitat conditions is a strong candidate for carrying such a plastic genotype. Sahara mustard reproduces primarily by self-fertilization, which is a strategy often associated with maintaining a general purpose genotype (Baker 1965), and limiting the capacity of new populations to rapidly adapt to local conditions (Barrett et al. 2008). Those ideas are supported by the general agreement between the climatic conditions over its native and invaded range, suggesting that the species’ strategies for coping with climatic variation are through well-preserved plasticity rather than novel adaptation. Therefore, its invasion may reflect niche conservatism (Wiens et al. 2010).

Nevertheless, just how much self-fertilization limits Sahara mustard’s genetic variability is an open question. Its native populations in Iraq were found to retain the potential for outcrossing (Al-Shehbaz 1977). Used as female parents, Sahara mustard was also able to artificially hybridize with other Brassicaceae species (Hinata et al. 1975). These opportunities for genetic recombination leave open the possibility that traits may evolve in invading Sahara mustard, with potential impacts on its distribution. Such genetic changes are particularly likely if repeated introductions are made from different parts of the native range, introducing new adaptive variation into the invasion (Ellstrand and Schierenbeck 2000, Dlugosch and Hays 2008, Dlugosch and Parker 2008), though in this case Sahara mustard appears to already have invaded most habitats that reflect the breadth of climates occupied in its native range.

Management implication for Sahara mustard and other invasive species

The climatic limits on the regional expansion of Sahara mustard and the species’ continuing spread at 5–50 km scales suggest that control efforts should focus on containing ongoing local expansion. Our SDM prediction highlights areas where climate provides ample room for local-scale expansion of the species. Prevention, early detection and eradication should focus on those high risk areas to slow down its local expansion. Nevertheless, regional scale monitoring is needed

to detect any new shifts of its range boundaries under the changing global climate.

The rapid spread of this species is fueled by effective long distance dispersal, which may be carried out by vehicle transportation. Early detection efforts should focus on suitable habitats with strong linkage, such as those accessible by roads.

Strength and novelty of our approach

Using widely available data from herbarium collections, the box-counting method successfully revealed the scale dependent spread of an invasive species, whose invasion history was not closely monitored. The method can be readily adopted to study the scale dependent expansion of many other species, for which the only data for interpreting range dynamics are from long term herbarium collections. Some studies have used the box-counting method to assess the spatial structure of species distribution across multiple spatial scales (Foxcroft et al. 2009, Veldtman et al. 2010, Akasaka et al. 2012), but all have focused only on the current distribution and neglected the temporal range dynamics. Our work presents the first case of applying the box-counting method to studying scale-dependent range expansion.

Our analysis effectively uncovered the scale dependency of an invasive spread. Quantifying such scale dependency and its contributing factors not only fosters understanding of species distribution but is also essential for effective control of any biological invasion. By highlighting critical scales where an invasion is contained or unconstrained due to specific factors, one can direct the management efforts to the appropriate range and scales where invasion is to continue. However, so far few studies have examined expansion rate of invasive species using multi-scale approaches (Lonsdale 1993, Pysek et al. 2008). Those that did use the approach limited their investigation to a very few discrete scales (e.g. local to regional to continental scale in Pysek et al. 2008) with a large range of intermediate scales missing in-between. Failing to assess a continuous range of scales can result in missing critical scales where major change in the expansion rate takes place. Due to these limitations, we have poor knowledge of the prevalence and general pattern of the scale dependency of invasive expansion. Less do we know the factors that lead to such scale dependency (Pysek and Hulme 2005). The methods demonstrated in our study provide a powerful tool for filling this research gap in order to better inform the management of many invasive species.

Caveats of our approach

Herbarium collections provide inexpensive and widely available long-term data for inferring species distributions. However, using those data can be challenging because they were not collected with the intention to discover multi-scale distribution patterns. We have incorporated in our methods means to reduce the influence by potential sampling biases. Nevertheless, several sources of error may still influence this type of analysis.

First, many of the herbarium collection records have imprecise or inaccurate geographic coordinates. We showed

that such errors at smaller scale could quantitatively impact estimates of expansion at larger scales, though they would be unlikely to alter patterns qualitatively.

Second, various spatial sampling biases could affect our analysis. Our method used the 'expansion' rate of native species with assumed stable range to correct for difference in spatial sampling efforts between decades. However, if the nature of spatial sampling bias is different between invasive and native species, our method will not be able to adequately correct the bias specific to the invader. For instance, if herbarium collectors paid more attention to a rapidly expanding invasive species more recently but meanwhile consistent attention to the natives, our method will overestimate the expansion rate for the invasive.

It is also possible that collectors might avoid collecting species in the same local area over time. This sampling bias does not necessarily affect the accuracy of our estimates because our method relies on measuring the change in total number of occupied cells in each time step rather than tracking the geographic movement of those cells. Making a first collection in one cell and the second in a different cell will not change the total number of occupied cells between the two time steps and thus will not change the estimated expansion rate at the scale represented by those cells. However, such method will affect our estimates if it forces multiple collections to concentrate in one cell at one time and to diverge to different cells at another time. Also, if collectors specifically avoid returning to the same places, those locations will not be recorded in the following decade. Therefore, there may be a smaller number of occupied cells than there should be, leading to an underestimate of the expansion rate. Our method would not be able to fully correct the bias, if such bias is different between the invasive and the native collections.

Our method assumes equal sampling efforts over space. It is evident that few records of Sahara mustard exist in Mexico and southern Central Valley of California despite the prediction of high likelihood of its presence by our SDMs (Fig. 2). Scarcity of records in Mexico may be a direct result of lower collection effort or the lack of digitally available data from Mexican herbaria. Scarcity of records in southern Central Valley of California might be due to limited habitat. Most of the valley has been developed for industrial scale farming. Sahara mustard might lack habitat there and herbarium collectors might avoid this region due to the lack of natural plant habitat. A search for all specimens of Asteraceae and Brassicaceae recorded in California Consortium of Herbaria show that the number of records from the three counties in the valley (Kern, Merced and San Joaquin County) was orders of magnitude lower than those in adjacent counties with more abundant natural plant habitat. Any large disparity in sampling efforts between regions will result in biased estimates of expansion rate across the regions. In our case, we know very little how the historical and current distribution of Sahara mustard in Mexico and Central Valley, California would have affected the estimated expansion rate across North America. Fortunately, the majority of its predicted range was relatively well sampled by herbarium collectors (Fig. 2).

Spatial sampling bias will also affect the prediction by MaxEnt SDMs, which assume equal sampling efforts over

the entire area of interest (Phillips et al. 2009). If herbarium records are spatially biased towards better surveyed area for the species (e.g. areas where an invasive species is a serious concern), less surveyed areas will have less influence on the SDMs, resulting in higher omission errors in those areas. Indeed, our SDMs using southwestern North America as the background fail to predict the presence of Sahara mustard in some of its recorded eastern and southern range boundaries. Fortunately, our models using North America as the background do include those areas as potential range for Sahara mustard (Supplementary material Appendix 4, Fig. A3). The latter models use a larger background with a wider range of values in explanatory variables. Values from the recorded range boundaries become less extreme in models using this enlarged background. An alternative solution to minimize such problem caused by spatial sampling bias is using spatial filtering to reduce the number of occurrence records in oversampled areas (Kramer-Schadt et al. 2013).

Finally, our analysis only indirectly infers the limiting factors that drive the scale dependency of invasive range expansion. Our methods look for correlation between certain ecological factors and observed scale dependency of range expansion. It provides useful insight for designing future studies that can directly test mechanisms that drive or limit range expansion through manipulative experiments conducted at various scales.

Conclusions

In this study we have demonstrated the validity of a novel methodology that can be broadly applied to uncover the scale dependency of invasive range expansion and its contributing ecological and evolutionary factors. Large online herbarium databases provide free and accessible data to which the box-counting method can be applied to quantify scale-dependent spread of many invasive plant species (after adjusting for variation in sampling efforts). Coupling this multi-scale analysis with MaxEnt SDMs, one can effectively determine on what scales the variables chosen in the models are limiting or contributing to the spread. Controlling efforts can then focus on the scales where the invasion is not constrained and areas where ecological factors are still favoring spread. Moreover, using SDMs based on variables in the invasive range to project its native distribution allows inference of the adaptation of an invasive species to novel environmental conditions. Such knowledge is essential for predicting the risk of further invasion powered by evolution.

Acknowledgements – John Donoghue II, Irena Simova and Jin Wu guided us to cross the mysterious terrain of ArcGIS. Naia Morueta-Holme circled many MaxEnt pitfalls that might have trapped us. Cal-IPC and Cameron Barrows shared important data of Sahara mustard recorded in the Californian deserts. Mara MacKinnon and Kimberly Byers spent months finding the latitude and longitude coordinates of hundreds of herbarium records. We thank Cameron Barrows, Aaryn Olsson, and Peter Chesson, who inspired critical improvement on the manuscript. We also thank all the herbarium collectors for documenting the history of Sahara mustard invasion. Funding was proved to YML by the US Marine Corps Air Stations, Yuma and Luke Air Force.

References

- Akasaka, M. et al. 2012. Invasive non-native species attributes and invasion extent: examining the importance of grain size. – *J. Veg. Sci.* 23: 33–40.
- Al-Shehbaz, I. A. 1977. Protogyny in the Cruciferae. – *Syst. Bot.* 2: 327–333.
- Arim, M. et al. 2006. Spread dynamics of invasive species. – *Proc. Natl Acad. Sci. USA* 103: 374–378.
- Baker, H. G. 1965. Characteristics and modes of origin of weeds. – In: Baker, H. G. and Stebbins, G. L. (eds), *The genetics of colonizing species*. Academic Press, pp. 147–169.
- Bangle, D. N. et al. 2008. Seed germination of the invasive plant *Brassica tournefortii* (Sahara mustard) in the Mojave Desert. – *West. N. Am. Nat.* 68: 334–342.
- Barrett, S. C. H. et al. 2008. Plant reproductive systems and evolution during biological invasion. – *Mol. Ecol.* 17: 373–383.
- Barrows, C. W. et al. 2009. Effects of an invasive plant on a desert sand dune landscape. – *Biol. Invasions* 11: 673–686.
- Baskin, C. C. and Baskin, J. M. 2006. The natural history of soil seed banks of arable land. – *Weed Sci.* 54: 549–557.
- Brown, J. H. et al. 1996. The geographic range: size, shape, boundaries, and internal structure. – *Annu. Rev. Ecol. Syst.* 27: 597–623.
- Catford, J. A. et al. 2009. Reducing redundancy in invasion ecology by integrating hypotheses into a single theoretical framework. – *Divers. Distrib.* 15: 22–40.
- Caughley, G. et al. 1987. Kangaroos and climate: an analysis of distribution. – *J. Anim. Ecol.* 56: 751–761.
- Crooks, J. A. and Soulé, M. E. 1999. Lag times in population explosions of invasive species: causes and implications. – In: Sandlund, O. T. et al. (eds), *Invasive species and biodiversity management*. Kluwer, pp. 103–125.
- Davis, M. B. and Shaw, R. G. 2001. Range shifts and adaptive responses to Quaternary climate change. – *Science* 292: 673–679.
- Delisle, F. et al. 2003. Reconstructing the spread of invasive plants: taking into account biases associated with herbarium specimens. – *J. Biogeogr.* 30: 1033–1042.
- Dlugosch, K. M. and Hays, C. G. 2008. Genotypes on the move: some things old and some things new shape the genetics of colonization during species invasions. – *Mol. Ecol.* 17: 4583–4585.
- Dlugosch, K. M. and Parker, I. M. 2008. Invading populations of an ornamental shrub show rapid life history evolution despite genetic bottlenecks. – *Ecol. Lett.* 11: 701–709.
- Elith, J. et al. 2010a. The art of modelling range-shifting species. – *Methods Ecol. Evol.* 1: 330–342.
- Elith, J. et al. 2010b. A statistical explanation of MaxEnt for ecologists. – *Divers. Distrib.* 17: 43–57.
- Ellstrand, N. C. and Schierenbeck, K. A. 2000. Hybridization as a stimulus for the evolution of invasiveness in plants? – *Proc. Natl Acad. Sci. USA* 97: 7043–7050.
- Estrada-Villegas, S. et al. 2011. Climate, habitat, and species interactions at different scales determine the structure of a Neotropical bat community. – *Ecology* 93: 1183–1193.
- Foxcroft, L. C. et al. 2009. Patterns of alien plant distribution at multiple spatial scales in a large national park: implications for ecology, management and monitoring. – *Divers. Distrib.* 15: 367–378.
- Gaston, K. J. 2009. Geographic range limits: achieving synthesis. – *Proc. R. Soc. B* 276: 1395–1406.
- Greig-Smith, P. 1979. Pattern in vegetation. – *J. Ecol.* 67: 755–779.
- Hijmans, R. J. et al. 2005. Very high resolution interpolated climate surfaces for global land areas. – *Int. J. Climatol.* 25: 1965–1978.
- Hinata, K. and Nishio, T. 1980. Self-incompatibility in Crucifers. – In: Tunoda, S. et al. (eds), *Brassica crops and wild allies*:

- biology and breeding. Japan Scientific Societies Press, pp. 223–234.
- Hinata, K. et al. 1975. Interspecific crossability in the tribe Brassicaceae with special reference to the self-incompatibility. – *Tohoku J. Agric. Res.* 25: 58.
- Holt, R. D. 2003. On the evolutionary ecology of species' ranges. – *Evol. Ecol. Res.* 5: 159–178.
- Holt, R. D. and Barfield, M. 2011. Theoretical perspectives on the statics and dynamics of species' borders in patchy environments. – *Am. Nat.* 178: S6–S25.
- Holt, R. D. et al. 2005. Theories of niche conservatism and evolution. – In: Sax, D. F. et al. (eds), *Species invasions: insights into ecology, evolution, and biogeography*. Sinauer Associates, pp. 259–290.
- Kinlan, B. P. and Hastings, A. 2005. Rates of population spread and geographic range expansion: what exotic species tell us. – In: Sax, D. F. et al. (eds), *Species invasions: insights into ecology, evolution, and biogeography*. Sinauer Associates, pp. 381–419.
- Kirkpatrick, M. and Barton, N. H. 1997. Evolution of a species' range. – *Am. Nat.* 150: 1–23.
- Kot, M. et al. 1996. Dispersal data and the spread of invading organisms. – *Ecology* 77: 2027–2042.
- Kramer-Schadt, S. et al. 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. – *Divers. Distrib.* 19: 1366–1379.
- Lonsdale, W. M. 1993. Rates of spread of an invading species – *Mimosa pign*a in northern Australia. – *J. Ecol.* 81: 513–521.
- MacArthur, R. 1972. Geographical ecology: patterns in the distribution of species. – Harper and Row.
- Marie, R. 1965. *Flore de L'Afrique du Nord*. – Paul Lechevalier, Paris.
- Mandelbrot, B. B. 1983. The fractal geometry of nature. – W. H. Freeman.
- Marushia, R. G. et al. 2012. Phenology, growth, and fecundity as determinants of distribution in closely related nonnative taxa. – *Invasive Plant Sci. Manage.* 5: 217–229.
- McGill, B. J. 2010. Matters of scale. – *Science* 328: 575–576.
- Miller, A. G. and Cope, T. A. 1996. *Flora of the Arabian Peninsula and Socotra*. – Edinburgh Univ. Press.
- Minnich, R. A. and Sanders, A. C. 2000. *Brassica tournefortii* Gouan. – In: Bossard, C. C. et al. (eds), *Invasive plants of California's wildlands*. Univ. of California Press, pp. 68–72.
- Morse, D. R. et al. 1985. Fractal dimension of vegetation and the distribution of arthropod body lengths. – *Nature* 314: 731–733.
- Parker, I. M. et al. 2003. An evolutionary approach to understanding the biology of invasions: local adaptation and general-purpose genotypes in the weed *Verbascum thapsus*. – *Conserv. Biol.* 17: 59–72.
- Pearson, R. G. and Dawson, T. P. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? – *Global Ecol. Biogeogr.* 12: 361–371.
- Phillips, S. J. et al. 2006. Maximum entropy modeling of species geographic distributions. – *Ecol. Model.* 190: 231–259.
- Phillips, S. J. et al. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. – *Ecol. Appl.* 19: 181–197.
- Phillips, B. L. et al. 2010. Evolutionarily accelerated invasions: the rate of dispersal evolves upwards during the range advance of cane toads. – *J. Evol. Biol.* 23: 2595–2601.
- Pysek, P. and Hulme, P. E. 2005. Spatio-temporal dynamics of plant invasions: linking pattern to process. – *Ecoscience* 12: 302–315.
- Pysek, P. et al. 2008. Comparing the rate of invasion by *Heracleum mantegazzianum* at continental, regional, and local scales. – *Divers. Distrib.* 14: 355–363.
- Ricklefs, R. E. 2005. Taxon cycles: insights from invasive species. – In: Sax, D. F. et al. (eds), *Species invasions: insights into ecology, evolution, and biogeography*. Sinauer Associates, pp. 165–199.
- Ritchie, M. 1998. Scale-dependent foraging and patch choice in fractal environments. – *Evol. Ecol.* 12: 309–330.
- Root, T. 1988. Environmental factors associated with avian distributional boundaries. – *J. Biogeogr.* 15: 489–505.
- Sakai, A. K. et al. 2001. The population biology of invasive species. – *Annu. Rev. Ecol. Syst.* 32: 305–332.
- Sexton, J. P. et al. 2009. Evolution and ecology of species range limits. – *Annu. Rev. Ecol. Syst.* 40: 415–436.
- Trader, M. R. et al. 2006. Seed production by the non-native *Brassica tournefortii* (Sahara mustard) along desert roadsides. – *Madrono* 53: 313–320.
- Townsend, C. C. and Guest, E. 1980. *Flora of Iraq*. – Ministry of Agriculture and Agrarian Reform, Baghdad.
- Veldtman, R. et al. 2010. Using scale–area curves to quantify the distribution, abundance and range expansion potential of an invasive species. – *Divers. Distrib.* 16: 159–169.
- Wangen, S. R. and Webster, C. R. 2006. Potential for multiple lag phases during biotic invasions: reconstructing an invasion of the exotic tree *Acer platanoides*. – *J. Appl. Ecol.* 43: 258–268.
- Wetthey, D. S. 2002. Biogeography, competition, and microclimate: the barnacle *Chthamalus fragilis* in New England. – *Integr. Comp. Biol.* 42: 872–880.
- Whittaker, R. J. et al. 2001. Scale and species richness: towards a general, hierarchical theory of species diversity. – *J. Biogeogr.* 28: 453–470.
- Wiens, J. A. 1989. Spatial scaling in ecology. – *Funct. Ecol.* 3: 385–397.
- Wiens, J. J. et al. 2010. Niche conservatism as an emerging principle in ecology and conservation biology. – *Ecol. Lett.* 13: 1310–1324.
- Zohary, M. et al. 1980. *Conspectus Florae Orientalis: an annotated catalogue of the flora of the Middle East*. – The Israel Academy of Sciences and Humanities, Jerusalem.

Supplementary material (Appendix ECOG-00722 at <www.ecography.org/readers/appendix>). Appendix 1–5.