

Ecography 32: 237–246, 2009

doi: 10.1111/j.1600-0587.2008.05354.x

© 2009 The Authors. Journal compilation © 2009 Ecography

Subject Editor: Miguel Araújo. Accepted 29 July 2008

## Using survival regression to study patterns of expansion of invasive species: will the common waxbill expand with global warming?

Luís Reino, Jordi Moya-Laraño and António Cláudio Heitor

L. Reino, Centro de Estudos Florestais, Dept Engenharia Florestal, Inst. Superior de Agronomia, Tapada da Ajuda, Univ. Técnica de Lisboa, 1349-017 Lisboa, Portugal, and ERENA – Ordenamento e Gestão de Recursos Naturais, Lda, Rua Robalo Gouveia, 1-1A, 1900-392 Lisboa, Portugal. – J. Moya-Laraño ([jordi@eeza.csic.es](mailto:jordi@eeza.csic.es)), Depto de Ecología Funcional y Evolutiva, Estación Experimental de Zonas Áridas, CSIC, General Segura, 1, Almería, ES-04001 Almería, Spain. – A. C. Heitor, CONFAGRI, Rua Maria Andrade, 13, 1199-013 Lisboa, Portugal.

Different approaches can be used to model the spread of invasive species. Here we demonstrate the use of survival regression, an approach that can be used to study a variety of events, not just death, to model the time to colonization. The advantage of survival regression to study colonisation of new areas is that information on those areas that have not been invaded by the end of a study can be included in the analysis, thus potentially increasing the accuracy of parameter estimation. We use proportional hazards regression (PHR; a type of survival regression) to model the spread of the common waxbill *Estrilda astrild* in Portugal. The species invaded Portugal in two peaks of invasion between 1964 and 1999. We built a PHR model with the information available up to the first invasion peak, then used this model to predict the pattern of invasion in the second peak. PHR had useful forecasting capabilities: areas that were actually colonised by 1999 had significantly higher hazards of colonization based on information from the first wave of invasion than areas that were not colonised. We then built a final model of expansion of the common waxbill that combined all available data up to 1999. Among climate variables, the most important predictor of colonization was temperature, followed by relative humidity. We used this model to estimate the invasion potential of the species under climate change scenarios, observing that an increase of 1°C in mean annual temperature increased the risk of a new invasion by 47%. Our analyses suggest that survival regression may be a useful tool for studying the geographical spread of invasive species. However, PHR was conceived as a descriptive technique rather than as a predictive tool, and thus further research is needed to empirically test the predictive capabilities of PHR.

Biological invasions and introductions have long fascinated biologists and ecologists around the world (Darwin 1859, Elton 1958, Williamson 1996). It is generally accepted that disturbed habitats and communities are more likely to be affected by biological invasions from introduced species than are pristine habitats (Simberloff 1995). In the last two decades, growing concern about this problem has stimulated research on the spread and naturalization of exotic species (Hengeveld 1989, Smallwood 1994, Williamson 1996). In particular, accurate predictions of future areas of invasion may be useful in prioritizing the management of areas under immediate threat.

Several statistical models have been employed to model biological invasions, mainly based on presence/absence data (Skellam 1951, Blackburn and Duncan 2001a, Silva et al. 2002, Muñoz and Real 2006). However, when data on the timing of invasion are available they can be used as an additional variable in models. This variable is useful because it allows forecasting future areas of invasion within a certain time frame. Since time is a continuous variable that can

usually be normalized by transformation, ordinary least squares (OLS) regression would seem appropriate to model the timing of invasion. However, OLS is not appropriate because only the spatial areas that have been invaded by the time censuses are finished can be included in the analysis. The inclusion of data from areas that have not been invaded by the target species at the time when censuses have been completed may be useful if these data add information that might increase accuracy in parameter estimation, thus improving forecasts of colonisation of new areas in the future. Survival regression, in contrast to OLS, allows the incorporation of censored data, points that have not been invaded by the time observations were discontinued.

Survival regression was originally implemented in medical research to study the factors affecting the timing of the death of patients (Cox 1972, Kalbfleisch and Prentice 1980, Cox and Oakes 1984), although this statistical approach can also be applied to study the timing of any type of event (Allison 1984, 1995). In fact, although this technique has not yet been widely used by ecologists, it is

suitable for studying a variety of ecological phenomena occurring in space and time (Muenchow 1986, Fox 1993, Moya-Laraño and Wise 2000). Of the several survival regression techniques available, the most extensively used is the Cox proportional hazards regression model (PHR; Cox 1972). The importance and broad applicability of this technique are attested to by the number of times Cox's original paper has been cited (>20 000 citations). Even though PHR can be used to study the occurrence of any kind of event (not just death), it is still not widely used in ecology.

In survival regression one can study either the distribution of survival times (as in accelerated failure time models; Fox 1993) or the hazard rate  $h(t)$ ,

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr\{t \leq T < t + \Delta t | T \geq t\}}{\Delta t}$$

or, in words, the instantaneous rate of failure at  $T = t$  conditional upon survival to time  $t$  (Kalbfleisch and Prentice 1980). As noted above, an advantage of using survival regression to model biological invasions is that it allows inclusion of data points (localities) that have not been invaded by the time a study ends. Hence, the analysis includes two types of data: censored (in which invasion has not occurred) and uncensored (in which the invasion has occurred). Therefore, the analysis will not be truncated to only those already-invaded spaces (as in OLS), but will include the whole geographic area of interest. Thus, survival regression models include a data structure of three vectors: a) time to events, b) censorship (usually taking values of 1 = occurring or 0 = censored) and c) covariates.

The cox proportional hazards regression model has the form  $h(t) = h_0(t)\exp(\beta X)$ , where  $h_0(t)$  is the baseline hazard function (which remains unspecified for covariate calculation),  $X$  is a vector of covariates, and  $\beta$  is a vector of regression coefficients (the  $b$ 's). The PHR model is semiparametric in that the distribution of the baseline hazard can take any form without affecting the estimation of the other parameters (i.e. the  $b$ 's). Thus, unlike other survival regression methods, PHR does not make any assumptions about the distribution of the baseline hazards (i.e. the distribution of hazards assuming that all covariates are zero; Kalbfleisch and Prentice 1980, Cox and Oakes 1984).

One of the most important features of PHR for the study of invasion patterns is that the estimated regression coefficients can be translated into the effects of that particular variable on the proportional change in the hazard of the event by the expression  $100 \times (e^b - 1)$ . For instance, if the coefficient for temperature (in Celsius) is  $b = 0.5$ , for each additional increase of one degree Celsius in temperature the hazard of invasion will increase by 65%. Any predictor included in the regression model can be translated into its relative effect on the hazard while maintaining all the other predictors in the model constant. In practical terms this means that the researcher (and the public land administrator) can know which variables, and to what extent, influence the risk of invasion by an introduced species in a given portion of land at a given moment.

## The common waxbill example

The common waxbill *Estrilda astrild* is one of the most successful introduced birds not only in the Iberian Peninsula, but in the whole Mediterranean Basin. The first documented records came from central coastal Portugal in the mid 1960s (Xavier 1968). From there, the species spread to the centre of Portugal, northwest Spain (Galicia), and southwest Spain (Extremadura and Andalucía) (Reino and Silva 1998, Silva et al. 2002). The species expanded quite slowly until the mid 1970s, but thereafter it spread rapidly to almost all coastal areas and progressively to inland areas in the south and the centre of Portugal. Reino and Silva (1998) reported that the expansion rate towards the north was twice that towards the south ( $13.02 \text{ km yr}^{-1}$  vs  $5.84 \text{ km yr}^{-1}$ ). Moreover, although most of the expansion occurred along a continuous front, some gaps in the distribution suggested the possibility of long-distance dispersal events (Silva et al. 2002). Previous studies suggested that both environmental and spatial-temporal variables were correlated with the pattern of expansion of this species in Portugal (Silva et al. 2002, Reino 2005).

The aims of this study are threefold. First, we aim to show the usefulness of survival regression in modelling the expansion of invasive species by using available data on the common waxbill to build a PHR model, and then test its predictive capability through a simple validation exercise. Second, we aim to build a model to forecast the future expansion of the common waxbill at the nation-wide level (Portugal) as a function of both spatial-temporal and environmental variables (climatic, topographical, and habitat correlates). Finally we discuss how global warming may affect the expansion of the common waxbill.

## Methods

### Data sources and sampling design

The expansion of the common waxbill was monitored between 1964 and 1999. Occurrence of the species was plotted on a UTM grid of  $943 \text{ } 10 \times 10\text{-km}$  cells that covers the entire range of continental Portugal (Fig. 1). The data that we use in this paper overlap substantially with those analysed in prior papers (Silva et al. 2002, Reino 2005). However, here data were carefully downsampled and additional records were included.

For modelling purposes we used all the presences available between 1964 and 1999, and for simplicity we considered that once the species settled in a cell it never disappeared from that cell, at least during the period considered here. Ultimately, we obtained a data set encompassing 126 occurrences and 817 absences that covered all continental Portugal.

### Environmental data

To determine the relative importance of different variables in explaining the expansion of the common waxbill in Portugal, we considered a set of environmental components assumed to be potentially relevant (Table 1). Taking

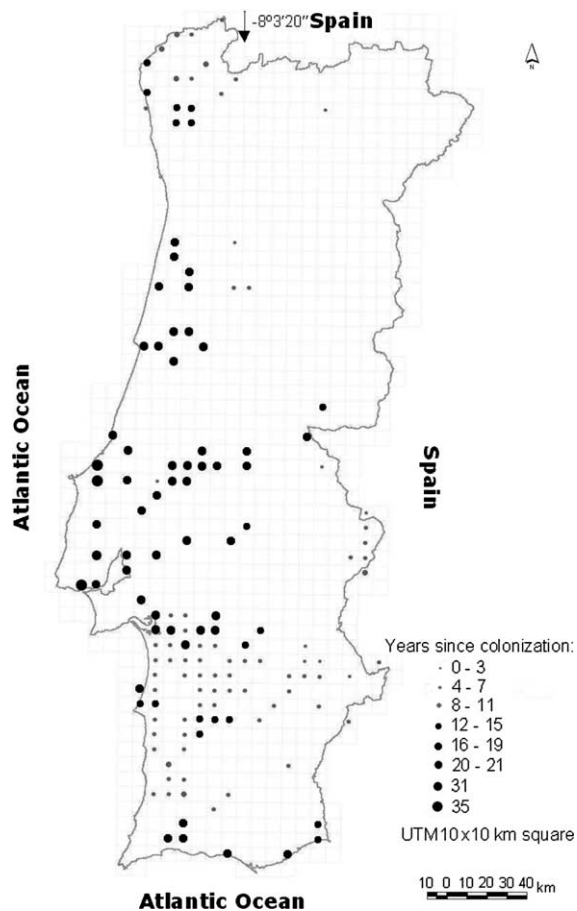


Figure 1. The expansion of the common waxbill *Estrilda astrild* in Portugal from 1964 to 1999 in a 10 × 10-km UTM grid. Grey and black dots reflect recent and older colonisations respectively (according to the bimodal distribution in Fig. 2). The arrow on the top shows the longitude which separates the area considered for our validation study (west of the arrow) from that not included (east of the arrow – see text for further details).

the selected grid system of Portugal, we scored each 10 × 10-km grid cell for its value for twelve environmental variables. Climatic variables were compiled from the available digital layers within the “Atlas do Ambiente” <<http://www.iambiente.pt/atlas/est/index.jsp>>. The data were then extracted and converted from the available vector maps (original scale 1:1 000 000) and resampled to a raster UTM grid of 10 × 10 km. We used zonal function in ArcView ver. 3.2 (ESRI 1999) to calculate the mean of each variable in each cell. The information on altitude for each cell was extracted from the US Geological Survey <<http://edc.usgs.gov/products/elevation/gtopo30/gtopo30.html>> with a horizontal grid spacing of 30 arc seconds (ca 1 km).

## Statistical analyses

### Testing and validating PHR: can PHR predict the future?

Invasion patterns could potentially occur either in a continuous, smooth manner or in bouts (peaks). If invasion occurred in different bouts, it would be important to know if the data collected in one bout could be used to predict

(forecast) patterns of invasion in the following bout. Thus, we first searched for bouts of colonisation in all colonisation records from the period 1964–1999. We tested for the simplest pattern of discontinuous invasion, i.e. bimodality in the distribution of times to invasion. Bimodality of distributions can be detected by analysing kurtosis when distributions are perfectly symmetrical (zero skewness) (Wyszomirski 1992). Thus, we first transformed the distribution of times to invasion by means of a Box-Cox transformation, which makes the distribution as symmetrical as possible. Significant negative kurtosis in a symmetric distribution means that the shoulders of the distribution are higher than they would be in a normal distribution, and a value beyond  $-1.2$  means significant bimodality (Hildebrand 1971, Wyszomirski 1992). After Box-Cox transformation we found significant negative kurtosis in the distribution of times to invasion (kurtosis =  $-0.98$ , SE = 0.4,  $p = 0.007$ ). A visual inspection of the symmetrised distribution clearly suggested bimodality in the distribution (Fig. 2), with two invasion peaks, one occurring 1979–1983 (Peak 1) and another between 1991 and 1995 (Peak 2).

We treated data from the two invasion peaks differently. To assess the performance of PHR in predicting invasions, we tested how a model built for the first peak would predict the pattern of invasion in the second peak. Even though the predictive capability of PHR is often low (O’Quigley and Flandre 1994, Allison 1995), especially when the proportion of censored observations is high (Allison 1995), we tested whether PHR could have at least some useful predictive capabilities. We first built a PHR regression analysis for the first invasion peak, for which we considered cells non-invaded by 1989 as censored. Since most of the cells colonised during the first peak were located in western Portugal, with only 7 out of 72 colonised cells having been invaded in the eastern part by 1989, we minimized the number of censored observations in our study by including only cells that were located west to  $-8^{\circ}3'20''$  longitude (Fig. 1). The PHR model built for western Portugal was used to predict the relative change in hazard rates (PHR) for western cells that had not been invaded by 1989. We considered the first year of observations (1964) to be year zero, and calculated the time of invasion of a cell as years elapsed since 1964. Since some of the environmental variables were highly correlated with each other, and because we are unaware of any statistical package that allows an iterative search for best-subset models for PHR, we reduced the spatial and environmental variables using principal component analysis (PCA) with Varimax axis rotation (Tabachnick and Fidell 2001, Quinn and Keough 2002), followed by preliminary regression analyses using principal components scores (PCs – Philippi 1993, Graham 2003). Prior to PCA all environmental variables were log-transformed to improve linearity with each other (Tabachnick and Fidell 2001, Quinn and Keough 2002). We obtained 6 PCs that together explained 92% of the variance (Table 2), which we named according to which environmental variables loaded highly on them: Latitude vs Radiation, Altitude, Frosts, Temperature, Humidity, and Longitude. These 6 PCs included most of the variation in spatial and environmental variables across western Portugal.

Table 1. Environmental variables considered for modelling the expansion of the common waxbill in Portugal. Descriptive statistics for both absences and presences are provided (mean  $\pm$  SD/max – min). Sources: <sup>1</sup> obtained from US Geological Survey (2002); <sup>2</sup> from Atlas do Ambiente.

Environmental variables (source-unit)	Code	Absences mean $\pm$ SD (max – min)	Presences mean $\pm$ SD (max – min)
Longitude	UtmE	–	–
Latitude	UtmN	–	–
Mean altitude <sup>1</sup> (m)	AltX	358.63 $\pm$ 255.99 (1499.5–8)	135.08 $\pm$ 127.70 (1116–4)
Maximum altitude <sup>1</sup> (m)	Altm	540.67 $\pm$ 347.07 (1918–20)	268.82 $\pm$ 219.27 (1911–6)
Annual average of water drainage <sup>2</sup> (mm yr <sup>-1</sup> )	Runoff	410.24 $\pm$ 361.50 (2400–25)	300.86 $\pm$ 336.71 (2200–25)
Evapotranspiration <sup>2</sup> (mm yr <sup>-1</sup> )	Evap	584.77 $\pm$ 115.58 (1000–400)	583.79 $\pm$ 142.26 (1000–400)
Total days with frost <sup>1</sup> (d yr <sup>-1</sup> )	Frosts	41.73 $\pm$ 18.41 (90–1)	31.83 $\pm$ 17.34 (90–1)
Annual average of humidity <sup>2</sup>	Humi	76.69 $\pm$ 4.88 (90–65)	79.79 $\pm$ 4.33 (90–70)
Number of hours of sunlight per year <sup>2</sup> (h yr <sup>-1</sup> )	Inso	2715.04 $\pm$ 265.21 (3200–1800)	2844.14 $\pm$ 254.64 (3200–2000)
Total annual rainfall <sup>2</sup> (mm)	Rain	973.31 $\pm$ 439.55 (2800–400)	856.55 $\pm$ 425.38 (2800–400)
Annual average radiation <sup>1</sup> (kcal cm <sup>-2</sup> )	Radi	153.04 $\pm$ 8.04 (170–140)	155.74 $\pm$ 9.96 (198–140)
Annual mean temperature <sup>1</sup> (°C)	Temp	15.05 $\pm$ 2.12 (19–7.5)	16.38 $\pm$ 1.32 (19–12.5)
January mean temperature <sup>1</sup> (°C)	TempJan	7.23 $\pm$ 2.33 (12.5 $\pm$ 0.2)	8.95 $\pm$ 1.30 (12.5–3.8)
July mean temperature <sup>1</sup> (°C)	TempJul	21.20 $\pm$ 2.37 (25–15.1)	21.77 $\pm$ 1.86 (25–17.5)

To test whether most of the potential explanatory variables were included in our regression analyses, we also included a complex combination of longitude (x) and latitude (y) as new variables  $x^2$ ,  $y^2$ ,  $xy$ ,  $xy^2$ , and  $x^2y$  in all regression analyses (Legendre 1993). If including these variables along with the PCs in the regression analyses resulted in any of them being significant, some unmeasured heterogeneity may be responsible for the pattern of invasion. To minimize multicollinearity, we subtracted the mean from x and y prior

to the calculation of the new variables (Neter et al. 1996). None of these newly created spatial variables were significant (results not shown), indicating that the PCs were sufficient to explain most of the non-random variation in the spatial-temporal pattern of invasion by waxbills. The significance of each predictor was corrected for multiple tests by applying the Holm's correction for dependent p-values (Legendre and Legendre 1998, Moya-Laraño and Wise 2007).

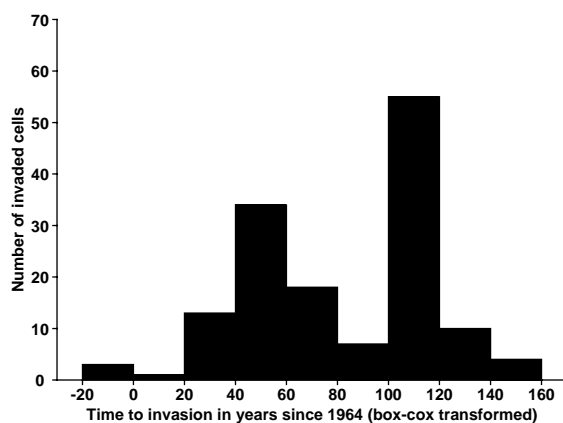


Figure 2. Bimodal distribution of invasion peaks for the common waxbill *Estrilda astrild* in Portugal. To visualize bimodality the values of the distribution have been optimally transformed according to a Box-Cox transformation following the expression  $[(t+1)^{1.5} - 1]/1.5$ , where t is time to colonisation measured in years elapsed since 1964 (see text for further details). The two peaks correspond to year ranges 1979–1983 and 1991–1995 as read from left to right.

Predicted hazard values were estimated by calculating how many times the hazard is predicted to increase or decrease relative to a cell with minimum values for all covariates in the sample. To accomplish this, we first calculated the relative contribution to the overall hazard for each predictor and cell as  $h_i' = (x_i - \min_X) \times (e^b - 1)$ , where  $x_i$  is the value of the predictor for cell i,  $\min_X$  is the minimum value of the predictor in the data set, and b is the PHR regression coefficient. We then added the contribution of each predictor (the  $h_i'$ ) to yield a predicted value for each cell. Note that since the b coefficients are exponentiated the overall effect, calculated by summing the effects of all the predictors, is multiplicative rather than additive. An alternative equation that makes the non-additive nature of the predictors in the model easier to follow would be (as an example for two predictors):  $h_{2i}' + h_{1i}' = \exp\{b_1 \times (x_{1i} - \min_{X1}) + b_2 \times (x_{2i} - \min_{X2})\} - 1$ . Since the inference of PHR is on the hazard rate and not on survival time, and because the relationship between the survivor function and the hazard function depends on the shape of the baseline hazard, it is simpler to make predictions on the relative hazard, as it is usually done in the medical sciences. However, to validate this procedure,

Table 2. Results of a principal component analysis of 14 spatial and climate variables for western Portugal (see text). Refer to Table 1 for a description of the environmental variables. The 6 PCs have been named according to which variables loaded more heavily on each. Loadings equal or above 0.7 (in absolute value) are shown in bold.

PC-number PC-name	1 Latitude radiation	2 Altitude	3 Frosts	4 Temperature	5 Humidity	6 Longitude
Variable						
utmE	-0.02	0.37	-0.23	0.08	0.18	<b>0.84</b>
utmN	<b>0.90</b>	0.12	-0.13	0.29	-0.17	0.05
Runoff	<b>0.72</b>	0.56	0.06	0.27	-0.04	0.11
Evap	<b>0.94</b>	0.14	-0.05	0.12	-0.03	0.07
Frost	0.19	0.08	<b>-0.90</b>	0.09	0.22	0.18
Humi	0.16	-0.04	0.18	0.02	<b>-0.95</b>	-0.13
Inso	-0.69	-0.35	0.11	-0.42	0.19	0.27
Rain	<b>0.78</b>	0.47	0.06	0.30	-0.10	0.14
Radi	<b>-0.86</b>	-0.14	0.26	-0.11	0.01	0.21
Temp	-0.43	-0.28	0.12	<b>-0.75</b>	0.08	-0.09
altm	0.17	<b>0.90</b>	-0.13	0.22	0.06	0.20
altx	0.32	<b>0.90</b>	0.01	0.19	0.00	0.13
TempJan	-0.31	-0.43	0.20	-0.68	-0.19	-0.21
TempJul	-0.50	-0.33	-0.32	-0.57	0.16	0.23
Eigenvalue	4.76	2.86	1.18	1.89	1.14	1.08
Variance	0.34	0.20	0.08	0.13	0.08	0.08

we also ran a log-normal parametric survival regression analysis (Fox 1993, Moya-Laraño and Wise 2000) including the same predictors as in PHR, and calculated the predicted median times to invasion for each cell following Allison (1995). The correlation between the predicted values (Peak 2) of both procedures was very high ( $r > 0.98$ ,  $p < 0.0001$ ). However, for the sake of simplicity we show only the results of PHR here. The significance of the overall PHR model is tested by a log-likelihood ratio test (G-test) obtained by partial likelihood (Cox 1972, Allison 1995), in which the likelihood of the model without the covariates is subtracted from the likelihood of the model with covariates and the result, which is previously multiplied by  $-2$ , contrasted with the chi-square distribution with as many degrees of freedom as covariates are included in the model.

A main assumption of the PHR is the proportionality of hazards, which means that the hazard changes proportionally in all individuals across time. Several methods have been described to test this assumption (Nagelkerke et al. 1984, Allison 1995). One of the easiest and most straightforward ways to test this assumption for each of the covariates is to run a PHR model that includes interactions between the time to the event and the covariate. If the interaction term is significant, this means that for that covariate the hazards are not proportional and the interaction term must be included in the analysis to correct for non-proportionality (Allison 1995). After a final model was found, we tested for the proportionality assumption using this later procedure. If there was a significant “covariate  $\times$  time” interaction, we included this interaction in the final model.

Table 3. Results of a principal component analysis of 14 spatial and climate variables for all continental Portugal. Refer to Table 1 for a description of the environmental variables. The 6 PCs have been named according to which variables loaded more heavily on each. Loadings equal or above 0.7 (in absolute value) are shown in bold.

PC-number PC-name	1 Rainfall	2 Longitude/altitude	3 Frosts	4 Humidity	5 Temperature	6 Latitude vs Radiation
Variable						
utmE	-0.35	<b>0.80</b>	0.30	-0.17	0.07	-0.09
utmN	0.32	0.39	0.21	0.07	0.26	0.73
Runoff	<b>0.88</b>	0.29	0.02	0.04	0.21	0.22
Evap	<b>0.86</b>	-0.13	0.07	0.07	0.09	0.36
Frost	0.02	0.27	<b>0.92</b>	-0.15	0.09	0.13
Humi	0.10	-0.18	-0.13	<b>0.97</b>	0.00	0.05
Inso	-0.56	-0.16	0.03	-0.10	-0.24	-0.67
Rain	<b>0.90</b>	0.20	0.01	0.08	0.21	0.26
Radi	-0.49	0.06	-0.10	0.01	-0.15	<b>-0.80</b>
Temp	-0.31	-0.36	-0.16	-0.04	<b>-0.75</b>	-0.25
altm	0.20	<b>0.89</b>	0.17	-0.13	0.22	0.10
altx	0.34	<b>0.86</b>	0.05	-0.08	0.21	0.18
TempJan	-0.17	-0.54	-0.20	0.10	<b>-0.71</b>	-0.18
TempJul	-0.62	-0.02	0.20	-0.05	-0.64	-0.24
Eigenvalue	3.77	3.02	1.15	1.05	1.82	2.10
Variance	0.27	0.22	0.08	0.07	0.13	0.15

Table 4. Regression results (PHR) predicting the time to colonisation of new 10 × 10 km cells by the common waxbill before 1989 in western Portugal. The independent variables are principal components extracted from 14 variables. Refer to Table 1 and 2 for a description of the variables and the PCs extracted from them. P-Holms are p-values obtained after a sequential Bonferroni correction (see text for details).

PC	Beta	SE	t-value	Exp-Beta	Wald	p	P-Holms(1)	effect (exp(b) – 1 × 100)
Latitude_vs_Radiation	–0.1	0.1	–0.8	0.9	0.6	0.4283		
Altitude	–1.2	0.2	–6.2	0.3	38.5	0.0000	0.0000	–70.8
Frosts	0.0	0.1	–0.1	1.0	0.0	0.9381		–0.9
Temperature	–1.4	0.3	–4.5	0.2	20.3	0.0000	0.0001	–75.7
Humidity	–0.4	0.2	–2.7	0.7	7.2	0.0073	0.0292	–33.3
Longitude	–0.3	0.2	–1.8	0.8	3.2	0.0744		–23.9

(1) P-values corrected for multiple tests. Only significant p-values are shown.

We validated the predictive capability of the PHR model in two ways. First, using only the data points invaded in Peak 2 we correlated the hazards predicted by the PHR model with the observed times of invasion. Due to the low forecasting capability of PHR, correlation between these two values was expected to be low; therefore, in a second step we validated the model by testing whether the predicted relative hazard rates were higher for the cells in Peak 2 that had been invaded by 1999 than for the cells that had not been invaded by 1999. Significant differences in the predicted direction imply that the model can at least approximate which cells on average will be more likely invaded within a period of 10 yr.

We used SAS 8.02 (Allison 1995) for the survival regression models and Statistica 7.1 (Statsoft 2006) for PCAs and the analysis of bimodality.

#### ***Predicting the future: a PHR model including invasions through 1999 and a predictive map of future invasions***

After validating the use of PHR we ran a regression model for all of Portugal that included all available records through 1999. Cells that had not been colonised by 1999 were included as censored data. In total, 943 cells were included, 126 of which had been invaded by 1999. The procedures used to build the model were similar to the preceding. First, we performed a PCA analysis after log-transforming all the predictor variables. Again, we obtained 6 PCs that together explained 92% of the variance (Table 3), and named them according to which environmental variables loaded highly on each: Rainfall, Longitude, Frosts, Humidity, Temperature, and Latitude vs Radiation. We then developed a preliminary PHR model using scores of the PCs. We then performed a PHR analysis using raw data for variables that loaded more heavily on the PCAs and were significant (after Holm’s correction) in the preliminary

PHR. As above, we also included a complex combination of longitude (x) and latitude (y) ( $x^2$ ,  $y^2$ ,  $xy$ ,  $xy^2$ ,  $x^2y$ ) in the regression analysis (Legendre 1993), but again none of these variables was significant (not shown). The significant predictors were then used to calculate predicted values ( $h_i$ ) as explained above. As before, if the assumption of proportionality of hazards was not met for some predictors we included the interaction “covariate × time” interaction in the model. Predicted values were then assigned to class intervals and, using Arcview 3.2 (ESRI 1999), we built a predictive map for the hazard of future invasions in Portugal.

## **Results**

### **Testing and validating PHR: can PHR predict the future?**

When modelling the pattern of invasion of common waxbills in western Portugal up through 1989, the PHR model using scores from the PCs was highly significant ( $p < 0.0001$ ). After Holm’s correction for multiple tests, the PHR model included three significant PCs: Altitude, Temperature, and Humidity (Table 4). The model including the raw values of the variables that loaded more heavily on each of the PCs differed slightly from the model including PCs, since only Altitude remained significant after Holm’s correction. The model predicted that an increase of 1 m in altitude would decrease the hazard of invasion by 1% (Table 5). However, the PHR assumption was violated because there was a highly significant “Altitude × time” interaction ( $p < 0.0001$ ), indicating that the hazard changes non-proportionately with Altitude. Thus, the final model included the interaction term. We used this model including the known invasion times for Peak 1, to predict the patterns of invasion for Peak

Table 5. Regression results (PHR) predicting the time to colonisation of new 10 × 10 km cells by the common waxbill through 1989 in Portugal. The independent variables are the variables that heavily loaded on the principal components that were significant in Table 4. P-Holms are p-values obtained after a sequential Bonferroni correction (see text for details).

Variable	Beta	SE	t-value	Exp-Beta	Wald	p	P-Holms(1)	effect (exp(b) – 1 × 100)
Humidity	0.0	0.0	0.8	1.0	0.7	0.4126		
Temperature	0.2	0.1	1.7	1.2	2.9	0.0878		
Altitude	0.0	0.0	–4.8	1.0	23.1	0.0000	0.0000	–1.0

(1) P-values corrected for multiple tests. Only significant p-values are shown.

Table 6. Regression results (PHR) predicting the time to colonisation of new 10 × 10 km cells by the common waxbill through 1999 in Portugal. The independent variables are the principal components extracted in Table 3. Refer to Table 1 and 3 for a description of the variables and the PCs extracted from them. P-Holms are p-values obtained after a sequential Bonferroni correction (see text for details).

PC	Beta	SE	t-value	Exp-Beta	Wald	p	P-Holms(1)	effect (exp(b) - 1 × 100)
Rainfall	0.0	0.1	-0.4	1.0	0.2	0.6904		
Longitud/altitude	-1.2	0.1	-8.6	0.3	73.8	0.0000	0.0000	-69.9
Frosts	-0.2	0.1	-2.7	0.8	7.4	0.0066	0.0133	-17.2
Humidity	0.6	0.1	6.0	1.8	36.5	0.0000	0.0000	84.5
Temperature	-0.9	0.2	-4.5	0.4	19.9	0.0000	0.0000	-60.4
Latitud_vs_Radiation	-0.4	0.1	-4.3	0.7	18.6	0.0000	0.0000	-33.9

(1) P-values corrected for multiple tests. Only significant p-values are shown

2. The correlation between the predicted and observed values for Peak 2 was not significant ( $r = -0.13$ ;  $n = 51$ ;  $p = 0.348$ ). However, using PHR the mean predicted values of cells that were invaded by 1999 were significantly higher (i.e. they were predicted to have a higher hazard of invasion; Mean  $\pm$  SE,  $h_i' = -4.94 \pm 5.07$ ) than those that were not ( $h_i' = -7.02 \pm 5.94$ ; Mann-Whitney U-test,  $Z = 3.1$ ,  $p = 0.002$ ), suggesting important predictive capabilities for PHR.

### Predicting the future: a PHR model including invasions until 1999 and a predictive map of future invasions

The PHR model using PCs was highly significant ( $G^2 = 169.6$ ;  $DF = 6$ ;  $p < 0.0001$ ) and all predictors but Rainfall were significant (Table 6). A model including only the environmental and spatial variables that loaded more heavily on the PCs was also highly significant ( $G^2 = 130.3$ ;  $DF = 6$ ;  $p < 0.0001$ , Table 7). After Holm's correction, Longitude, Humidity, and Temperature remained significant. However, due to violations in the assumption of proportionality of hazards, the final model included the interactions "Longitude × time" and "Humidity × time" (both  $p < 0.0001$ ) while excluding the raw value of Humidity ( $p = 0.118$ ). Thus, the effects of Longitude and Humidity on the hazard changed with time, indicating non-linear effects for these covariates. The final model was thus,  $h(t) = h(0)\exp(-0.042 \times \text{Longitude} + 0.384 \times \text{Temperature} + 0.001 \times \text{Longitude} \times \text{Time} + 0.004 \times \text{Humidity} \times \text{Time})$ . Temperature remained as the only predictor that did not violate the proportionality assumption (i.e. it had a linear effect on the hazard). In the final model, an increase of 1°C in

mean annual temperature increased the hazard of invasion by 47%. Figure 3 shows the predictive map depicting the areas from highest to lowest risk of invasion after 1999.

## Discussion

### A PHR model has some predictive power

Our simple validation exercise (i.e. artificially censoring data of known timing of invasion) demonstrated that PHR can be used to predict the patterns of future expansion of invasive species. Although PHR was not able to accurately predict the timing of invasion, using a single predictor (Altitude) and its interaction with time, PHR did predict that cells that were occupied by the common waxbill by 1999 had a higher hazard for invasion than cells that were not occupied by 1999. Although PHR (like most techniques of statistical modelling) is usually not able to predict the timing of events beyond the study time frame, our results suggested that PHR may have some value as a predictive tool even beyond the temporal scope of the study. However, these results may only apply to this particular study, and thus the generality of PHR as a potential tool to predict future events needs to be investigated with further research including both simulation and application to a wide range of real scenarios.

Even though the predictive ability of PHR is generally expected to be low (O'Quigley and Flandre 1994, Allison 1995), some authors argue that the predictive capability of survival models is most closely related to the magnitude of the standard errors of the coefficients (Allison 1995). All our models showed very small standard errors relative to the

Table 7. Regression results (PHR) predicting the time to colonisation of new 10 × 10 km cells by the common waxbill through 1999 in Portugal. The predictive map in Fig. 3 is based on this model. The independent variables are the variables that heavily loaded on the principal components that were significant in Table 6. Refer to Table 1 and 3 for a description of the variables and the PCs extracted from them. P-Holms are p-values obtained after a sequential Bonferroni correction (see text for details).

Variable	Beta	SE	t-value	Exp-Beta	Wald	p	P-Holms(1)	effect (exp(b)-1 × 100)
Longitude	0.0	0.0	-3.7	1.0	13.6	0.0002	0.0008	-1.0
Latitude	0.0	0.0	-0.2	1.0	0.0	0.8366		
Frosts	0.0	0.0	1.5	1.0	2.1	0.1442		
Humidity	0.1	0.0	4.5	1.1	20.5	0.0000	0.0000	10.2
Radiation	0.0	0.0	0.6	1.0	0.4	0.5482		
Temperature	0.4	0.1	4.4	1.4	19.6	0.0000	0.0000	43.3

(1) P-values corrected for multiple tests. Only significant p-values are shown.

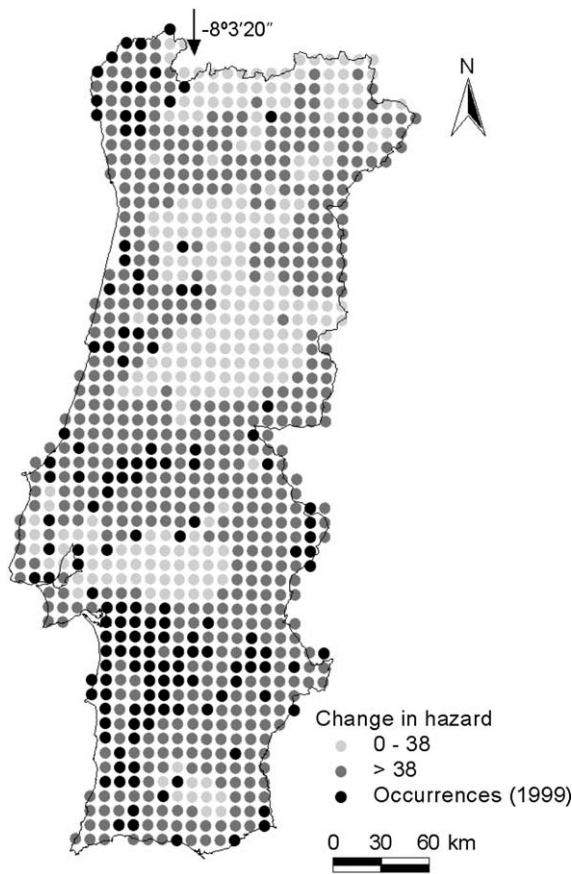


Figure 3. Predicted change in hazard (relative to minimum hazard) considering non-colonised cells after 1999 only. See the text to see how the values here displayed in two categories (above and below the median) were obtained. As an example, a value of 20 means that the cell has a relative hazard of invasion 20 times higher than an cell with minimum values for all covariates (Allison 1995). In practical terms this means that the darker areas in the map will be more likely invaded before the lighter ones.

magnitude of the coefficients, as demonstrated by the small p-values of the predictors (Table 4–7). As noted above, the predictive capability of survival regression models for biological invasions needs to be tested with further research. However, we must stress that the accommodation of censored data in PHR allows for much higher accuracy in parameter estimation, and as such PHR has potential to become an important tool for purely descriptive purposes (*sensu* Araújo and Guisan 2006). Also, following Araújo and Guisan (2006), who defined a predictive model as “one where confidence in the hypothesized relationships allows projections of observed patterns into independent situations”, we acknowledge that our model validation is bound to be optimistic, since the patterns of invasion in two different time bouts within the same waxbill invading population are not completely independent.

### A PHR model for the future expansion of the common waxbill: the role of global warming

The final model predicting the future expansion of the common waxbill in Portugal included the variables Long-

itude, Temperature, and the interactions between Longitude and Time and Humidity and Time. Temperature was the most important predictor and the only one that held to the proportionality assumption. An increase in 1°C in mean annual temperature increasing the hazard of invasion by 47%. The importance of temperature for this species was also highlighted in previous work from Portugal (Silva et al. 2002). Figure 3 shows the map of predicted future expansion for the common waxbill if temperature does not change. Since the relationship between hazard and temperature does not change with time, an increase in temperature of 1°C with global warming will mean that the partial hazard of invasion in the whole range of Portugal (controlling for the partial effect of the other covariates) will increase by 47%. Thus, according to our model, the common waxbill will most likely benefit from global warming (Climate Change 2001). This may affect the likelihood of expansion, mainly to inland areas over all of the country (Fig. 3) that previously were less suitable for occupation. Humidity was also strongly correlated with the expansion of this species. However, since there is an interaction between humidity and time, the effect of humidity is not as clear cut. Nevertheless, an increased scarcity of water with global warming, especially in the southern and inland regions of Portugal, could counteract the effect of increasing temperature on the range expansion of the common waxbill. Projections for southwest Europe suggest that climate will be dryer, affecting animal groups such as reptiles and amphibians (Araújo et al. 2006). These results highlight the importance of knowing the underlying mechanisms of dispersal for the common waxbill. In the past, dispersal has resulted in major shifts in the distribution of species as climate has changed (Watkinson and Gill 2002). Since this species has shown a considerable rate of expansion in Portugal and in certain regions of Spain as well (Reino and Silva 1998), climate change may play an important role in shaping the expansion dynamics of the common waxbill in Iberia. We are currently working in expanding our model with a larger data set and using data for different global climate change scenarios to build an array of different predictive maps. In addition to our prediction that the common waxbill may expand to warmer areas, other factors, such as the presence of suitable habitat, landscape fragmentation, and degradation, might reduce the ability to disperse to new areas. Native Mediterranean passerine species, such as the trumpeter finch *Bucanetes githagineus*, also have been hypothesized to increase their range with global warming (Carrillo et al. 2007).

The common waxbill is an eclectic estrildid of widespread distribution in mesic habitats throughout sub-Saharan Africa (Hall and Moreau 1970). The widespread expansion of this species in Iberia has been well documented (Reino and Silva 1998), being one of the most successful among the passerine species introduced in the Mediterranean area. It has been argued that there are some difficulties associated with the establishment of bird species from other continents as regular breeders in the Mediterranean area (Blondel and Aronson 1999). The success of the common waxbill in this region may be linked to different factors. First, its natural range in Africa is quite large, and successful introductions elsewhere have been linked to species with high tolerance and larger geographical ranges



(Brown 1984). Second, successful introductions have also been linked to species with a wider range of resource use (Gaston 1994); the common waxbill is considered a nearly ubiquitous species, which avoids only areas without surface water or dominated by bare ground (Bernard 1997). Ubiquitous species are probably more likely to be successful in newly colonised areas (see for example Blackburn and Duncan 2001b).

The data that supported our results are based on >30 yr of historical records. This study suggests that these kinds of data are valuable for investigating the reasons for introduction success or failure (Blackburn and Duncan 2001b). Nonetheless, reaching firm conclusions based on these records may sometimes prove difficult (Williamson 1996). We believe PHR will be an important tool for developing accurate assessments of mechanisms of invasions in future studies.

In conclusion, we have shown that if information on the timing of colonisation of new space by invasive species is available in addition to presence/absence data, survival regression in general, and PHR in particular, may be a useful tool to describe the patterns of colonisation. In addition, PHR showed significant predictive capabilities beyond the time frame of the study and revealed temperature to be the most likely important predictor of the future expansion of the common waxbill. These results allowed us to make the general prediction that in the face of continuing global warming, the common waxbill may be able to colonise new areas in Portugal, which may also apply to certain areas of Spain as well. However, some care is needed because other climatic (e.g. humidity) and landscape factors may have important, possibly counteractive, effects as well. Survival regression may prove to be an important descriptive (and perhaps predictive) tool in invasion studies that may help both researchers and public land administrators to understand why some introductions succeed while others fail (Blackburn and Duncan 2001b). In particular, if some inference about future invasions can be made, PHR may help the land administrator to identify locations of immediate invasion threat and initiate preventive actions early on.

*Acknowledgements* – We are indebted to D. H. Wise and J. T. Rotenberry for their helpful comments and revising the English of this paper. We are also grateful to Helena Simões and Pedro Segurado for GIS support and help in map production; Tiago Silva for working together since the studies on the introduction of common waxbill started in 1994. LR was funded by a Ph.D. grant (SFRH/BD/14085/2003) from the “Fundação para a Ciência e Tecnologia” of the “Ministério da Ciência, Tecnologia e do Ensino Superior”. JML holds a Ramón y Cajal research contract from the “Ministerio de Educación y Ciencia”.

## References

- Allison, P. D. 1984. Event history analysis: regression for longitudinal event data. – Beverly Hills, Sage Publ.
- Allison, P. D. 1995. Survival analysis using the SAS system. – Cary, SAS Inst.
- Araújo, M. B. and Guisan, A. 2006. Five (or so) challenges for species distribution modelling. – *J. Biogeogr.* 33: 1677–1688.
- Araújo, M. B. et al. 2006. Climate warming and the decline of amphibians and reptiles in Europe. – *J. Biogeogr.* 33: 1712–1728.
- Bernard, P. 1997. Common waxbill *Estrilda astrild*. – In: Harrison, J. A. et al. (eds), The atlas of southern African birds – Volume 2: passerines. Birdlife South Africa, Johannesburg, pp. 612–613.
- Blackburn, T. M. and Duncan, R. P. 2001a. Establishment patterns of exotic birds are constrained by non-random patterns in introduction. – *J. Biogeogr.* 7: 927–939.
- Blackburn, T. M. and Duncan, R. P. 2001b. Determinants of establishment success in introduced birds. – *Nature* 414: 195–197.
- Blondel, J. and Aronson, J. 1999. Biology and wildlife of the Mediterranean region. – Oxford Univ. Press.
- Brown, J. H. 1984. On the relationship between abundance and distribution of species. – *Am. Nat.* 124: 255–279.
- Carrillo, C. et al. 2007. Northward expansion of a desert bird: effects of global warming? – *Ibis* 149: 166–169.
- Climate Change 2001. Third Assessment Report of the Intergovernmental Panel on Climate Change. – IPCC (WG I and II), Cambridge Univ. Press.
- Cox, R. D. 1972. Regression models and life-tables. – *J. R. Stat. Soc.* 2: 187–202.
- Cox, D. R. and Oakes, D. 1984. Analysis of survival data. – Chapman and Hall.
- Darwin, C. 1859. On the origin of species by means of natural selection, 2nd ed. – John Murray.
- Elton, C. S. 1958. The ecology of invasions by animals and plants. – Methuen.
- ESRI 1999. ArcView GIS help (v3.2). – Environmental Systems Research Inst., New York.
- Fox, G. A. 1993. Failure-time analysis: emergence, flowering, survivorship, and other waiting times. – In: Scheiner, S. M. and Gurevitch, J. (eds), Design and analysis of ecological experiments. Chapman and Hall, pp. 253–289.
- Gaston, K. J. 1994. Rarity. – Chapman and Hall.
- Graham, M. H. 2003. Confronting multicollinearity in ecological multiple regression. – *Ecology* 84: 2809–2815.
- Hall, B. P. and Moreau, R. E. 1970. An atlas of speciation in African passerine birds. – British Museum (Natural History), London.
- Hengeveld, R. 1989. Dynamics of biological invasions. – Chapman and Hall.
- Hildebrand, D. K. 1971. Kurtosis measures bimodality? – *Am. Stat.* 25: 42–43.
- Kalbfleisch, J. D. and Prentice, R. L. 1980. The statistical analysis of failure time data. – Wiley.
- Legendre, P. 1993. Spatial autocorrelation: trouble or a new paradigm? – *Ecology* 74: 1659–1673.
- Legendre, P. and Legendre, L. 1998. Numerical ecology, 2nd English ed. – Elsevier.
- Moya-Laraño, J. and Wise, D. H. 2000. Survival regression analysis: a powerful tool for evaluating fighting and assessment. – *Anim. Behav.* 60: 307–313.
- Moya-Laraño, J. and Wise, D. H. 2007. Two simple strategies of analysis to increase the power of experiments with multiple response variables. – *Basic Appl. Ecol.* 8: 398–410.
- Muenchow, G. 1986. Ecological use of failure time analysis. – *Ecology* 67: 246–250.
- Muñoz, A.-R. and Real, R. 2006. Assessing the potential range expansion of the exotic monk parakeet in Spain. – *Divers. Distrib.* 12: 656–665.

- Nagelkerke, N. J. D. et al. 1984. A simple test for goodness of fit of Cox's proportional hazards model. – *Biometrics* 40: 483–486.
- Neter, J. et al. 1996. *Applied linear statistical models*. – McGraw-Hill.
- O'Quigley, J. and Flandre, P. 1994. Predictive capability of proportional hazards regression. – *Proc. Nat. Acad. Sci. USA* 91: 2310–2314.
- Philippi, T. E. 1993. Multiple regression: herbivory. – In: Scheiner, S. M. and Gurevitch, J. (eds), *Design and analysis of ecological experiments*. Chapman and Hall, pp. 183–210.
- Quinn, G. and Keough, M. 2002. *Experimental design and data analysis for biologists*. – Cambridge Univ. Press.
- Reino, L. 2005. Variation partitioning for range expansion of an introduced species: the common waxbill *Estrilda astrild* in Portugal. – *J. Ornithol.* 146: 377–382.
- Reino, L. M. and Silva, T. 1998. The distribution and expansion of the common waxbill (*Estrilda astrild*) in the Iberian Peninsula. – In: Spina, F. and Grattarola, A. (eds), *Proceedings of the 1st Meeting of the European Ornithologists' Union*. *Biol. Conserv. Fauna* 102: 163–167.
- Silva, T. et al. 2002. A model for range expansion of an introduced species: the common waxbill *Estrilda astrild* in Portugal. – *Divers. Distrib.* 8: 319–326.
- Simberloff, D. 1995. Introduced species. – *Encyclopedia of Environmental Biology*, Vol. 2, Academic Press.
- Skellam, J. G. 1951. Random dispersal in theoretical populations. – *Biometrika* 38: 196–218.
- Smallwood, K. S. 1994. Site invasibility by exotic birds and mammals. – *Biol. Conserv.* 69: 251–259.
- Statsoft 2006. *Statistica* (data analysis software system), ver. 7.1. – <[www.statsoft.com](http://www.statsoft.com)>.
- Tabachnick, B. G. and Fidell, L. S. 2001. *Using multivariate statistics*, 4th ed. – Harper Collins.
- US Geological Survey 2002. GTOPO30 30 arc seconds digital elevation model from US Geological Survey. – <<http://edc.usgs.gov/products/elevation/gtopo30/gtopo30.html>>.
- Xavier, A. 1968. Bicos de Lacre em Óbidos. – *Cyanopica* 1: 77–81.
- Watkinson, A. R. and Gill, J. A. 2002. Climate change and dispersal. – In: Bullock, J. M. et al. (eds), *Dispersal ecology*. Blackwell, pp. 410–428.
- Williamson, M. 1996. *Biological invasions*. – Chapman and Hall.
- Wyszomirski, T. 1992. Detecting and displaying size bimodality: kurtosis, skewness and bimodalizable distributions. – *J. Theor. Biol.* 158: 109–128.