An important issue in population ecology is to disentangle different density-dependent mechanisms that may limit or regulate animal populations. This goal is further complicated when studying long-lived species for which experimental approaches are not feasible, in whose cases density-dependence hypotheses are tested using long-term monitored populations. Here we respond to some criticisms and identify additional problems associated with these kinds of observational studies. Current caveats are related to the temporal and spatial scales covered by population monitoring data, which may question its suitability for density-dependence tests, and to statistical flaws such as the incorrect control for confounding variables, low statistical power, the distribution of demographic variables, the interpretation of spurious correlations, and the often used stepwise series of univariate analyses. Generalised linear mixed models are recommended over other more traditional approaches, since they help to solve the above statistical problems and, more importantly, allow to properly test several hypotheses simultaneously. Finally, several management actions aimed to recover endangered species, such as supplementary feeding, might be considered as field experiments for further testing density-dependence hypotheses in long-lived study models. We expect these opportunities, together with the most adequate statistical tools now available, will help to better our understanding of density-dependent effects in wild populations.

The density-dependent nature of demographic parameters is widely accepted as an important mechanism determining population sizes. The pathways underlying negative density-dependent feedbacks are, however, not fully understood and have become the focus of an intense debate about the relative importance of regulation and limitation for wild populations (White 2001, 2004, Berryman et al. 2002, Berryman 2004). Experimental studies have added insight by, for example, manipulating territory quality through food supplementation (Arcese and Smith 1988) or modifying breeding densities of small passerines (Both 1998). However, proximate density-dependent mechanisms may vary among species with different life histories and/or among populations under different ecological context. Thus, a variety of model organisms are necessary for a better understanding of population regulation processes.

In the case of large predators, not only logistic constraints but also management limitations legally imposed to endangered species preclude experimental approaches (Balbontín and Ferrer 2008), thus most research on the effects of density-dependence on demographic parameters has been performed through observational studies based on long-term monitoring of populations. Balbontín and Ferrer (2008) are timely warning against potential methodological problems linked to these kinds of studies. Indeed, the related literature is plagued with flaws, which are more numerous and not necessarily those pointed out by these authors. Here, we would like to go further on their comments to help to set the bases for a correct analysis of density-dependent effects in population monitoring studies. In our opinion, most serious caveats are related to 1) the spatial and temporal scales of population monitoring, and 2) the application of inadequate statistical analyses. We will illustrate these limitations by using as examples several of the publications cited and considered as ‘classics papers’ by Balbontín and Ferrer (2008), which used large raptor species as study models.

Population monitoring: spatial and temporal issues

The quality and suitability of data for testing density-dependence hypotheses is the first critical step to be considered when relying on observational information. One of the hypotheses that has attracted the most attention for explaining density-dependent effects on reproduction assumes the existence of habitat heterogeneity among breeding sites within a population: individuals are expected to first settle in the best-quality territories and further occupy lower-quality ones as breeding density increases.
This entails the need for identifying what a population is—a non-trivial task in studies of population ecology. Research often relies on artificial boundaries (e.g., administrative ones) comprising relatively small ‘populations’, which are at least spatially separated from others and/or show differential demographic parameters or trends, to detect within-population heterogeneity (Carrete et al. 2006a, Martínez et al. 2007, Sergio et al. 2007). In other cases, a grid-squared study area is arbitrarily selected to accurately monitor part of a population (Krüger and Lindström 2001). Conversely, the use of very large study areas may exceed the population scale, thus potentially confounding within-population heterogeneity in habitat quality and breeding performance with large-scale geographic and climatic effects on reproduction or particular metapopulation dynamics, such as source–sink phenomena. As an example, Balbontín et al. (2003) studied the age of Bonelli’s eagles Hieraaetus fasciatus breeding in the administrative region of Andalusia (southern Spain), proposing that the progressive increment in the proportion of immature breeders was an early warning signal of its endangerment. The study area comprised 298 breeding territories distributed across 87 597 km² (Fig. 1 in Balbontín et al. 2003), a surface showing extremely wide geographic and climatic ranges (altitude: 80–3482 m a.s.l. and annual rainfall: 200–1500 mm, Balbontín et al. 2003). The same data set was also employed to explore the relative importance of age and territory quality on the productivity of this ‘population’ (Penteriani et al. 2003). Given the well-known effect of climate on the breeding performance of this eagle (Ontiveros and Pleguezuelos 2003), and that testing for density-dependence was reduced to just 37 out of the 298 territories (i.e., those for which a 5-year demographic data set was available, Balbontín et al. 2003, Penteriani et al. 2003), large-scale geographic variations (not considering in the analytical procedures) might better explain variability in breeder’s ages and breeding performance than within-population habitat heterogeneity (see further arguments by Gil-Sánchez et al. 2005, and a statistically demonstrated example in Carrete et al. 2006a, p. 687).

The above example also illustrates additional monitoring flaws from a temporal perspective. The sampling unit for studying this kind of demographic processes, as correctly pointed out Balbontín and Ferrer (2008), must be the breeding event. By monitoring 298 territories during 18 years (Balbontín et al. 2003, Penteriani et al. 2003), these authors only observed— for the best sampled demographic variable — 11% of the breeding events (591 of ca 5300 breeding events, Penteriani et al. 2003). It is evident how such an incomplete monitoring may create temporal biases, also probably unevenly distributed in space, that seriously question its suitability for testing density-dependent fecundity. Summing-up, the above studies show how an inadequate spatial scale, together with an incomplete and non-stratified population monitoring design, may compromise the usefulness of a data set for testing density-dependent effects on demography when both space and time are not taken into account in analytical procedures.

The correct spatial delimitation of a population together with its accurate spatial and temporal monitoring is also crucial to obtain a key demographic parameter: the population trend. Balbontín and Ferrer (2008) stated that only stable or increasing populations are suitable for testing density-dependent fecundity, thus criticising Carrete et al. (2006a) for using two out of four populations with negative trends. This quotation seems to result from a misinterpretation of data, ‘Low density areas (LDA)’ referred to those showing relative low breeding densities independently of past population trends (Carrete et al. 2006a). In fact, the golden eagle Aquila chrysaetos in the LDA remained stable over the monitoring period, while the Bonelli’s eagle experienced first a decline, then stability, and lastly increased over the 19 monitored years (Carrete et al. 2006a p. 682). Moreover, when analyses are restricted to the increasing period in the last population, results remain unchanged (GLMM: age effect $F_{1,135} = 13.78, p = 0.0003$, ns year and territory effects fitted as random terms) compared to those based on the entire data set (Table 3 in Carrete et al. 2006a). Nonetheless, there is not a rationale for excluding declining populations from density-dependent tests of hypothesis, as claimed by Balbontín and Ferrer (2008). Under the habitat heterogeneity hypothesis, it is expected that during a process of population decline low-quality territories are the first to be deserted and only the best-quality sites will remain used (Rodenhouse et al. 1997), implying or not a redistribution of breeders among territories. This possibility has been evidenced even in large geographic areas exceeding the local population scale, by controlling for spatial autocorrelation (Carrete et al. 2007). From the point of view of the interference hypothesis (also called individual adjustment hypothesis), population decline is expected to be accompanied by a reduction in breeding density and a positive density-dependent feedback after crowding is relaxed (Sill et al. 2004).

**Statistical flaws detecting density-dependent patterns and underlying mechanisms**

A first, logical step in the study of density-dependent patterns is to simply examine the relationships between varying temporal and/or spatial densities and the resulting demographic parameters (e.g., breeding output, survival), where both negative (Arcese and Smith 1988, Rodenhouse et al. 2003, Carrete et al. 2006a) and positive relationships (e.g., due to Allee effects, Serrano et al. 2005) can be found. A more difficult task is to ascertain the fine-tuning mechanisms underlying these patterns. Balbontín and Ferrer (2008) highlight the need for examining the relationships between annual fecundity and its variance (measured as its coefficient of variation) to discriminate between alternative hypotheses. This claim has been held since the paper by Ferrer and Donázár (1996), where authors proposed that if conflicts among neighbours increase with breeding density, the mean annual fecundity should decrease without changing its variance, thus supporting the interference competition (or individual-adjustment) hypothesis. Contrarily, the habitat heterogeneity (or buffer) hypothesis would predict a negative relationship between mean fecundity and its variance, since individuals would be forced to occupy poorer habitats as population increases. This influential paper (ca 70 citations registered on ISI Web of Knowledge in March 2008) was however flawed by a statistical artefact not evidenced until recently: the spurious negative correlations.
that always arise between mean fecundity and measures of its variance because they are non-independent variables (Brett 2004, Beja and Palma 2008).

To partially solve the above problem, Ferrer et al. (2006) proposed the use of the skewness of the distribution of brood sizes as an adequate statistic that showed significant negative relationships with mean fecundity only in populations regulated according to the habitat heterogeneity hypothesis. However, Beja and Palma (2008) performed simulations based on a long-term monitored population of Bonelli’s eagles, all resulting in spurious correlations which were unable to distinguish the habitat heterogeneity from the interference hypotheses. The distribution of brood sizes used for these last simulations was left-skewed, as derived from the population monitored and as expected for long-lived species with low fecundity rates (Moreno et al. 2003). Recently, Ferrer et al. (2008) argued that such a fecundity distribution is typical of saturated populations under the habitat heterogeneity hypothesis, showing that the brood-size data used by Beja and Palma (2008) follow a quasi-normal distribution in territories occupied at low densities (assumed as good-quality ones) but a left-skewed one in later occupied territories (assumed as low-quality ones at high densities), thus questioning their results. We would like to point that these assumptions are, however, not fully justified. Although the existence of habitat heterogeneity is unquestionable in nature, the occupancy of low-quality territories may not follow a linear relationship with increasing densities (Johnson 2007); instead, it may be necessary to exceed a given population size (equal to the number of breeding pairs needed to fulfilling high-quality sites) that may not be reached in many of the studied populations even when they are increasing. Below this threshold, variance in territory quality may be not enough to regulate populations through habitat heterogeneity mechanisms. Besides, the left-skewed brood-size distribution showed by Ferrer et al. (2008) for new occupied territories may result from alternative hypotheses unrelated to habitat heterogeneity, such as the reduction of the age of first reproduction in increasing populations of long-lived species with delayed maturity (Newton 1979, Martinez et al. 2007) and the concomitant breeding performance depression associated to younger breeders (Forslund and Pärt 1995, Carrete et al. 2006a, Martínez et al. 2007). Nonetheless, apart of the discussion on brood-size distributions for simulations (Beja and Palma 2008, Ferrer et al. 2008), at least half of the contribution on the previously published coefficients of correlation between mean fecundity and its variance is attributed to spurious correlations ($r^2 = -0.52$–$-0.54$, Brett 2004, Ferrer et al. 2008), and thus conclusions from a variety of studies must be re-examined by testing these correlations against null models randomly generated (Ferrer et al. 2008). A major concern also arises when considering that the long-term research looking at (partially) spurious correlations supporting the habitat heterogeneity may have eroded efforts for testing alternative hypotheses.

A number of additional problems are detected in the literature when looking for mechanisms which may explain density-dependent regulation patterns. A common challenge is to disentangle the effects of territory and individual quality on reproduction. It is commonly assumed that poor-quality individuals (e.g. the younger in long-lived species) are forced to settle in the poorest territories (Balbontín and Ferrer 2008). This may be true in many cases (Serrano and Tella 2007), but should not be considered as a universal rule. Individuals often lack perfect knowledge of habitat quality (Schlaepfer et al. 2002, Johnson 2007), show non-ideal habitat selection (Arlt and Pärt 2007), or ecological traps may be present (Schlaepfer et al. 2002). Moreover, the fundamental (intrinsic) quality of territories may largely differ from their realised quality (i.e. that actually translated to the fitness of their occupants, Johnson 2007), which changes over space and time due to the interference with conspecific and heterospecific competitors (Carrete et al. 2005, 2006a, 2006b, Johnson 2007). This fact has been largely ignored by the static view of the habitat heterogeneity hypothesis. Individual quality can also change through time, something that can be only recognised by using long-term surveys of marked individuals (Serrano and Tella 2007), and density-dependent effects may change over the life of the individuals (Ratikainen et al. 2008). However, when marked individuals are not available for population studies, age of breeders has been used as a surrogate of quality, and a common way to disentangle age effect from that of the territory has been to compare the performance of adults and subadults breeding in the same territories through paired tests (Ferrer and Bisson 2003, Penteriani et al. 2003). This method suffers from both statistical and biological problems. First, such a simple comparison needs to control for potential confounding variables. Between years variability in reproduction has been considered as a variable to be controlled for in these paired-tests, adjusting breeding parameters for year effects by subtracting annual means from raw data. This may create a statistical violation, since no tests are provided to support a normal distribution of breeding parameters (Ferrer and Donázar 1996, Penteriani et al. 2003) which would justify the use of means, while actual distributions in long-lived species, including the studied eagles, are known to be left-skewed following a quasi-Poisson distribution (Moreno et al. 2003, Beja and Palma 2008). This statistical flaw may not greatly affect results and conclusions in the presence of strong year effects. However, two major problems are that in some cases between years variability was not tested before of controlling for (Ferrer and Bisson 2003, Penteriani et al. 2003), and that in some of the studied species no year effects have been detected (Ontiveros and Pleguezuelos 2003, Carrete et al. 2006a, Martínez et al. 2007). Therefore, blindly correcting for annual means lacks of sense and may produce unexpected artefacts in data used for analyses. Besides, the apparent between years variability in breeding parameters can be explained by annual changes in the proportion of subadult breeders, as shown by multivariate analyses (Carrete et al. 2006a). In the last case, subtracting annual means would have eliminated age effects rather than year effects on reproduction, a fatal endpoint taking into account that the resulting data is indeed used to test age effects (Ferrer and Bisson 2003, Penteriani et al. 2003). The same standardisation has been also blindly used for testing differences in quality among territories (Penteriani et al. 2004), even knowing that there were no year effects on reproduction in the studied population (Penteriani et al. 2002). Finally, it is worth noting that it is not biologically correct to homogeneously subtract mean productivity to
both good-quality and poor-quality territories when testing the habitat heterogeneity hypothesis, since their relative contribution to the mean productivity of the population is different (Rodenhouse et al. 1997) and the realised quality of territories can change over time (Johnson 2007).

There are more statistical concerns associated to the paired tests. The obvious one, as recognised by Balbontín and Ferrer (2008), is their low statistical power since sample sizes for paired-tests correspond to the number of breeding territories used by both adults and subadults and those are often not many in the study of long-lived species (e.g. n = 10 in Ferrer and Bisson 2003, 2003 n = 9 in Penteriani et al. 2003). These tests need power analyses to evaluate the robustness of their conclusions, a practice still not widespread in ecology (Martínez-Abrain 2007). Likewise, the inability of these approaches to simultaneously test alternative hypotheses and/or control for additional confounding variables, forcing to perform a stepwise series of univariate analyses using different data subsamples (Ferrer and Bisson 2003, Penteriani et al. 2003).

**Current methodological approaches and further directions studying density-dependence**

Recent studies have evidenced that a single density-dependent pattern may be explained by different processes (e.g. the direct effect of habitat heterogeneity, or the differential settlement of subadult breeders, due to variability in mortality rates unrelated to intrinsic habitat characteristics, on productivity, Carrete et al. 2006a), and that several density-dependence mechanisms may act together (Rodenhouse et al. 2003). An illustrative example of complexity in density-dependent population regulation is how the long-term breeding depression of a territorial vulture, the bearded vulture Gypaetus barbatus, is related to the combination of habitat heterogeneity, interference with territorial neighbours and non-breeding floaters, and changes in the mating system of the population (Carrete et al. 2006b, 2006c). Multivariate analyses are required to correctly disentangle the contribution of different mechanisms acting simultaneously.

Multivariate approaches have several advantages. First, they allow the use of individual breeding events as units for data analyses, thus increasing statistical power, while controlling for confounding variables. Among these approaches, generalised linear mixed models (GLMM) are the best alternative against more traditional methods such as ANOVA, ANCOVA, or multiple regression models. First, GLMMs allow fitting the most appropriate error and link functions to data, such as the Poisson distribution to some data (instead of average values per territory) and generalise results obtained from a sample of territories and years to the population as a whole (Littell et al. 1996). Moreover, in the case of individually-marked studies, individual effects may be accounted for by fitting individual identity as a random term. This may allow a fine-grain examination of the effects of aging on reproduction and their interactions with habitat quality and density while controlling for inter-individual variance in quality. Unfortunately, the use of general linear models is less familiar to ecologists than the more traditional analyses (Darlington and Smulders 2001), and researchers may still misunderstand the use of random terms. This seems to be the case of Balbontín and Ferrer (2008) when claiming a re-analysis of data by Carrete et al. (2006a) because they fitted territory as a random term in GLMMs to disentangle the effects of breeder age and territory on the fecundity of four populations of eagles. However, as argued above, the random effect is the best choice when sampling several events from the same territories in different years mostly if a sample instead of the whole population is monitored, thus allowing a further generalisation of results. Nonetheless, all territories were sampled in the criticised case of Carrete et al. (2006a), and after all results are expected to not vary qualitatively by fitting territory as a fixed instead of a random term. Accordingly, results remain the same had they used the fixed term option (GLMMs for each one of the four populations simultaneously testing the contribution of age and territory on brood size as fixed effects, while controlling for year as a random term; age: F-range: 7.75–18.4, all p < 0.004; territory: F-range: 0.75–1.30, p-range: 0.11–0.83), thus solving doubts posed by Balbontín and Ferrer (2008). Fortunately, GLMMs are increasingly being the method of choice in recent density-dependence studies (Tella et al. 2001, Rodenhouse et al. 2003, Carrete et al. 2006a, Soutullo et al. 2006, Martínez et al. 2007, Sergio et al. 2007) and we expect it will be the rule in the future.

Finally, we would like to remember readers that this short cautionary review has been focused on population studies of long-lived, usually top-predator species often threatened or near threatened, a fact for which manipulative experiments are not ethically nor legally allowed. There are however certain management actions by wildlife agencies aimed to their recovery or conservation that may be interpreted as field experiments and used for testing density-dependence hypotheses. This is the case of supplementary feeding stations that have altered the breeding and non-breeding densities of a threatened vulture, the bearded vulture (Carrete et al. 2006b). Other unexplored, promising possibilities include supplementary feeding programs in breeding territories that improve the quality of some of them (experimental) while leaving others unmanaged (controls) (González et al. 2006, Robertson et al. 2006, Robb et al. 2008). These approaches, together with adequate statistical tools, should help for a further understanding of density-dependence population regulation in a variety of life styles and ecological contexts.

Acknowledgements — We thank J. A. Donázar and D. Serrano for their comments on the manuscript. MC was supported by the Spanish Ministry of Education and Science through the postdoctoral program “Juan de la Cierva”. This study was partially supported by a mep project (cgl2006-10689/bos).
References


