Project no. 006463
EuMon
EU-wide **mon**itoring methods and systems of surveillance for species
and habitats of Community interest

Instrument: SSP

Thematic Priority: Biodiversity conservation

D18: Compilation and illustration of recommended methods for analysis of combined data from different monitoring schemes

Due date of deliverable: 31 October 2006
Actual submission date: 15 December 2006

Start date of project: 1.11.2004   Duration: 42 month

National Museum of Natural History (MNHN), Paris, France

Revision : final

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	PU
PP	Restricted to other programme participants (including the Commission	
RE	Restricted to a group specified by the consortium (including the Commission	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Compilation and illustration of recommended methods for analysis of combined data from different monitoring schemes

Deliverable 18 of EuMon's Work Package 2.2

By Pierre-Yves HENRY (MNHN) and Piotr NOWICKI (UJAG)
With contributions by Dirk S. SCHMELLER (UFZ), Klaus HENLE (UFZ) and Chris A. M. VAN SWAAY (VLINDERS)

Summary

Combining data from different schemes improves representativity, precision, and robustness of conclusions derived from monitoring data on biodiversity changes. Benefits, limits, and available methods for combining data from different schemes were compiled in D16. Here we illustrate recommended methods for data combination with reference to case studies. Raw data combination is recommended when data were collected with similar field methods and correspond to the same data type (i.e., data document the same biological process and have the same theoretical distribution). When estimates are to be combined, geometrical mean, and meta-analysis methods are recommended. Despite of its great potential to assess general patterns, the meta-analysis approach remains very poorly used for biodiversity trend assessment with species or habitat monitoring data. In all cases, whether data or estimates are combined, weights to account for differences in sampling design and sampling effort across schemes need to be included in computations of means and statistical analyses. Finally, different monitoring schemes may produce complementary information; e.g., combining estimates from count survey and capture-mark-recapture survey allows linking population trends to demographic processes.

Table of Contents

Compilation and illustration of methods for analysis of combined data from different monitoring schemes.

Deliverable **18**, WP2.2, EuMon

I Objectives

Combining data from different schemes improves representativity, precision, and robustness of conclusions driven from monitoring data on biodiversity changes. It also increases understanding of mechanisms of biodiversity change when schemes to be combined are complementary, i.e. they contribute data on different parts of the life cycles, or different population mechanisms. Benefits, limits, and available methods for combining information from different monitoring schemes were introduced in Chapter 7 of D2, and compiled and discussed in D16.

The purpose of the present deliverable is to illustrate recommended methods for data combination. Basically, there are two ways to combine schemes. Firstly, combination of data into a single dataset, which is possible when data are similar and can be analysed simultaneously in a single statistical model (i.e. the variable to be analysed follows the same theoretical distribution and quantifies the same biological process). Secondly, estimates from different data types can be combined. When data are too different to be combined into a single dataset but they document the same biological parameter, estimates of this parameter can be produced separately from each dataset. Then, estimates are combined to produce average estimates or to test for a given effect or trend across schemes.

Two other important issues have to be considered when combining data. Firstly, do we only need to compute averages (cf. Chapter II) or do we additionally need to implement statistical tests across schemes (cf. Chapter III)? Depending on the goals of data combination, the method to be applied differs. Secondly, do some data need to have a higher importance than others in the analysis? If yes, then data or estimates need to be weighted prior to computation of values averaged across datasets/estimates or prior to fitting statistical models for tests (Chapter IV). The latter is the case when differences in sampling design and sampling effort across schemes need to be adjusted for.

II Combining information across monitoring schemes to produce average estimates

II.1 Combining data from different schemes

The first step is to identify whether data can be combined or not. For combining data across schemes, the measurement unit needs to be the same (e.g. number of individuals per unit of sampling effort). For simultaneous analysis with the same statistical parametric model, data need to follow the same theoretical distribution. Even if data types are different, they can be jointly analysed to produce an estimate averaged across all schemes for the parameter of interest, e.g. probability of occurrence or population growth rate.

The second step is how to proceed to combine data. Given that data have the same type, combining data is trivial: they just need to be entered into a single dataset. Afterwards, estimates averaged across all datasets can be computed.

An example is provided in the case study 4 developed in D12 (Appendix 1). Julliard *et al.* (2004a) analysed simultaneously in a single statistical model the number of bird individuals counted at French Breeding Bird Survey plots and the number of bird individuals captured at French Constant Effort Ringing Sites. From the analysis, they obtained a national estimate of population growth rate averaged across all species, all sites, independent of the monitoring technique. Combining data is thus straightforward, but note that differences in representativity across datasets are a separate issue that should not be ignored (cf. part IV hereafter).

Compilation and illustration of methods for analysis of combined data from different monitoring schemes.

Deliverable **18**, WP2.2, EuMon

II.2 <u>Analysis of combined estimates</u>

The first step is to identify whether estimates can be combined or not. In cases where the original data are incompatible (different data types, different theoretical distributions), integration across schemes is possible by combining estimates of the same biological parameter obtained separately from each monitoring scheme. For instance, trends in population size can be estimated both from counts of individuals per unit of time or with presence-absence data (Strayer 1999; Pollock 2006). If the goal is to obtain an integrated measure of population growth rate, then both estimates derived from presence-absence data and count data can be combined.

a. <u>How to combine estimates?</u>

The second step is how to proceed to combine estimates. Geometrical mean, i.e. averaging on a log-scale and exponentiation of the average, is the recommended method to compute an average value with estimates from different schemes (Buckland *et al.* 2005). Advantages of the geometrical mean compared to the arithmetic mean are that, given a set of schemes with a constant number of estimates, given evenness of representation across schemes, and given that there is variation of the biological parameter per scheme, the average value of the measured biological parameter should show no trend. However, it should show a negative trend when the overall value actually decreases, or when the number of schemes with zero values actually increases (Fig. 1). Rewording these theoretical concepts in the context of seeking an average estimate of population growth rate, these properties would become: given a set of schemes documenting population size with a constant number of estimates of population growth rate per scheme, given evenness of the number of estimates per scheme, and given that population growth rates vary across schemes, the geometrical mean population growth rate should not show any trend. However, it should show a negative trend when there is an overall decrease in population size or when the number of schemes with zero values (i.e. extinctions of the species) is increasing.
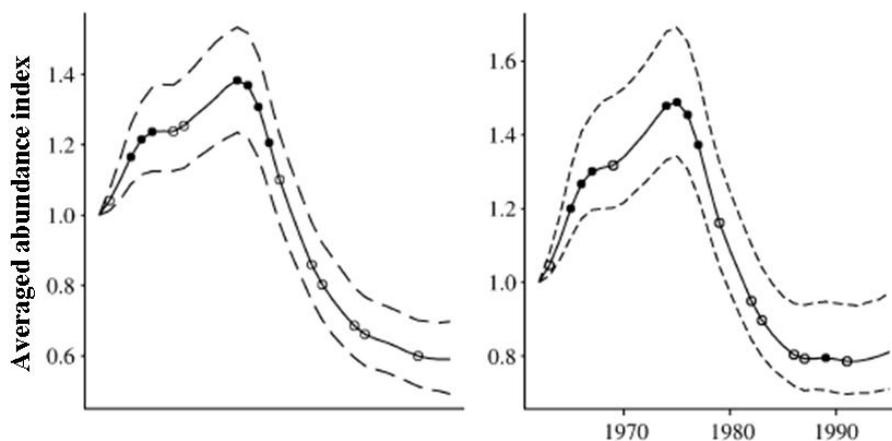

Fig. 1. Temporal variations in geometrical mean (left) and arithmetic mean (right) of abundance indices for 12 farmland bird species in the United Kingdom (data from the Chritsmas Bird Count; figures from Buckland *et al.* 2005). Dotted lines delimit 95% bootstrap confidence intervals obtained by resampling monitoring sites. Increases in the averaged abundance index are weaker, decreases are stronger, and estimates are more precise for geometrical mean than for arithmetic mean.

The Log-transformation included in the computation of the geometrical mean is incompatible with null values, but adding a small positive constant to zero values is a reasonable solution. The issue of how to average estimates in the context of producing biodiversity indices is fully

4

Compilation and illustration of methods for analysis of combined data from different monitoring schemes.

Deliverable **18**, WP2.2, EuMon

developed in Buckland *et al.* (2005). Formula to compute standard errors for geometrical means are provided in Appendix A of Gregory *et al.* (2005).

To document changes in abundance across all biomes and realms, Loh *et al.* (2005) combined time series of relative abundance indices into a set of datasets. These combined datasets are the basis for computing the Living Planet Index (LPI). Then, they computed population size trends for each species. When several time series were available per species, they averaged them so that only one estimate per species was used. Finally, they integrated average population growth rate across all species datasets to obtain a worldwide, mean rate of changes in abundance. Note that all heterogeneity existing across regions, realms, species, and length of time series (etc.) was ignored.

Another case study is the estimation of the average trend of breeding bird populations per major habitats in Europe (Gregory *et al.* 2005). In 2000, up to 18 EU countries had a national breeding bird survey running. All count numbers of individuals per species, but with a variety of field methods. Thus, data could not be combined into a single dataset from which trends could be estimated. The procedure that was chosen had three steps: (1) Each country, produced national estimates of population growth rate per year for each species, while missing data were interpolated with log-linear models; (2) these estimates were combined with independent estimates of national population sizes to produce yearly estimates of the European population size, allowing to compute the population growth rates at the European level for each species, with missing data being created by interpolation, and, finally, (3) species were attributed to broad habitat categories according to expert knowledge, and estimates of changes in population size were averaged across species by the geometrical mean to produce estimates of trend per habitat in Europe (Fig. 2). Nonetheless, although combination is rather straightforward, one should not omit to account for important differences among schemes that may bias the results. Here, national estimates per species were weighed by national population size to secure that trends in countries with a major part of the European population per species have a higher contribution to the European estimate than countries with minor populations. Thus, weighting raw information from separate schemes (i.e. either raw data or estimates) must not be neglected (cf. part IV).
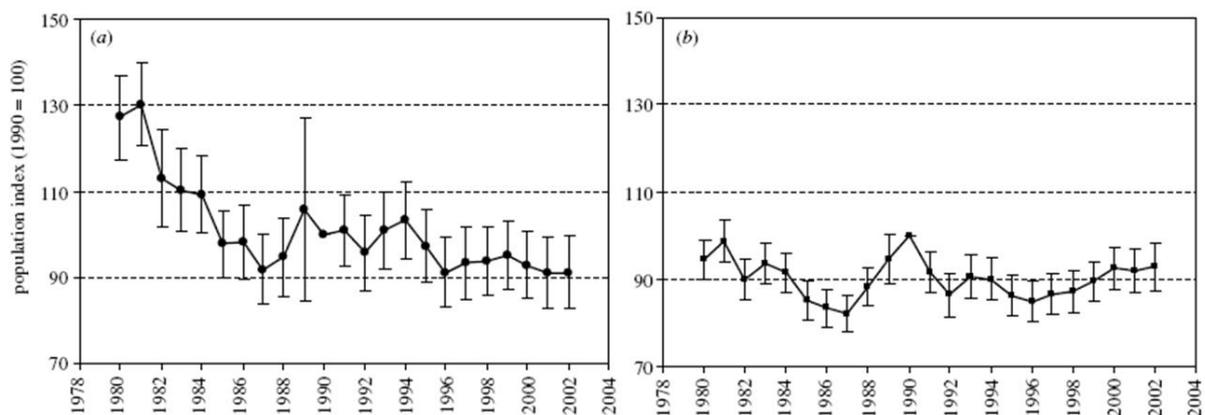

Fig. 2. Temporal variations in geometrical mean abundance indices for farmland (left, 23 species) and woodland, park, and garden bird species (right, 24 species) in Europe (± 1.96 s.e.; combined estimates from 18 national Breeding Bird Surveys; figures from Gregory *et al.* 2005). Farmland birds decreased more, and allover Europe, than woodland birds. The index was arbitrarily set to 100 for 1990.

III Testing effects with information combined across monitoring schemes

To quantify changes in biodiversity, one may be interested in obtaining a measure of the change only, without testing any hypothesis. For instance, the primary goal of the Living Planet Index (Loh *et al.* 2005) is measuring global trends in population size in the world. Testing for differences across groups through time is not central for the approach.

However, when linking changes in biodiversity components to putative causes of changes or across taxonomic groups, average estimates are not sufficient. An appropriate, robust way of testing is necessary. Two main approaches can be used: combining data and implementing statistical tests on the combined dataset or collecting estimates from independent studies and applying statistical tests to combined sets of estimates.

III.1 How to analyse combined data?

Analyzing a dataset combining data from different schemes is trivial. It resumes to applying a single statistical model to the combined dataset. Again, the weighing issue becomes central so that differences in sampling designs or field methods are adjusted for in the analysis (cf. part IV).

The case study 4 in Deliverable 12 (i.e. Julliard *et al.* 2004a) illustrates the analysis of combined data from rather different schemes and sampling techniques. Indeed, the number of individuals counted at French Breeding Bird Survey plots and captured at French Constant Effort Bird Ringing Sites were analyzed simultaneously with a single statistical model, thus producing a national estimate of population growth rate per species, averaged over sites and sampling techniques. Both data types could be combined since, for both methods, the unit of count was the number of individuals of a given species at a given site in a given year.

III.2 How to analyse combined estimates? Meta-analysis methods

Basically, the difference between analysing raw data or estimates is that raw data are supposed to be known without error, whereas estimates are known with error, and this error cannot be ignored. Thus, a proper analysis using estimates as dependent variable should simultaneously account for estimates and standard error of the estimates. This is illustrated in the case study 4 of D12 (i.e. Julliard *et al.* 2004a), where estimates of population growth rate for different species were combined into a single ANOVA to test which species traits could explain among-species variation in population size trend. Differences of precision among species were accounted for by including an appropriate weight in the statistical model for each species estimate (cf. part IV.2).

More generally, the idea behind meta-analysis is that results of independent studies are treated as input units for the analysis of a general pattern (Gurevitch *et al.* 2001). Such an approach allows combining information coming from various monitoring schemes regardless of the differences in their sampling designs, objects monitored, data characteristics, and to some extent even statistical methods applied. Nevertheless, pre-conditions are typically set on the information quantity (i.e. minimal spatial or temporal coverage) and quality (e.g. validity of data collection methods used) that is required for the study to be used in the meta-analysis. If the statistical methods applied within the studies allow the calculation of the effect size for each of them then the average effect size can be applied to infer the general pattern, otherwise the pattern can be assessed with non-parametric tests (see section 3.III of D16 for details).

Despite of its great potential the meta-analysis approach remains very poorly used for biodiversity trend assessment with species or habitat monitoring data. Literature examples

Compilation and illustration of methods for analysis of combined data from different monitoring schemes.

*Deliverable **18**, WP2.2, EuMon*

include studies of ecological patterns from the theoretical point of view, such as factors affecting population fluctuations (Marsh 2001), the effect of fire on individual survival (Pons *et al.* 2003), the effects of global warming on distribution range (Parmesan & Yohe 2003) or phenology (Menzel *et al.* 2006), or the impact of invasive species on communities (McCarthy *et al.* 2006). We have not found a single published study of meta-analysis dealing with classic monitoring objectives, i.e. quantifying general trends in distribution or population size, even though meta-analyses would be straightforward to apply: the effect size of interest would be that of time (i.e. years, taken as a continuous, linear variable). The deficit of literature examples of monitoring studies using meta-analyses is likely to be an indication of the low popularity of this approach among monitoring professionals, perhaps due to the relatively high level of statistical expertise required.

III.3 Complementarity across schemes: combining heterogeneous data to increase understanding of biological processes

Previous methods focussed on extracting the common information contained in separate datasets. But relevant information is also contained in differences across datasets (developed in D16). In the case study 7 in D12 (Julliard 2004), population growth rates for four bird species were estimated from count data. With a similar monitoring, but collecting capture-mark-recapture data, temporal variations in survival and recruitment were estimated per species. Then, variations in demographic traits were modelled with population growth rates as dependent variable. This analysis identified the dominant demographic process in driving changes in population size for each species. Thus, combining estimates from count survey and capture-mark-recapture survey allowed linking population trends to demographic processes.

IV Accounting for differences in sampling design across datasets: weighting data and estimates

Generally, combining data (i.e. datasets or estimates) is rather straight forward. Meta-analysis methods are standard and fully detailed in text books. An important issue remains to be addressed: adjusting data for major differences among schemes that could bias results of analyses involving separate schemes. This adjustment is implemented by weighting data in analyses so that each scheme contributes to the average estimate and final results accordingly to its representativity. Common weights to be used when combining monitoring data are compiled in Chapter 3.IV of D16. Here, we illustrate the use of weights of wide use when combining data across schemes or when correcting for non-representativity of data.

IV.1 Post-stratification to account for unequal sampling effort in different habitat types

When monitoring trends for a taxonomic group at a national scale, it is important that all habitats are accounted for. When designing the monitoring, this is planned by choosing monitoring sites so that they are representative of what occurs at a national scale (cf. discussions in D2, 16, and 17). However, if no sampling design was used (e.g. free choice of sites by observer), it is likely that habitats will be unequally sampled. To obtain a global trend estimate that can be considered as representative of the trend all over the country, weights need to be applied to the data so that each habitat is represented according to its surface area at the national scale (named post-stratification of the data). Unequal sampling effort among habitats can also be planned for optimizing field effort according to monitoring goals of the schemes. Such sampling designs involve stratification of the sampling effort across habitats.

Integration of schemes with stratification according to habitat is straightforward since the weight to be given to each habitat was defined when designing the monitoring.

For example, when producing national trends for butterfly population size in The Netherlands (Van Swaay *et al.* 2002), indices of population size per monitored site have to be post-stratified according to habitat availability at the national scale. Such a procedure was necessary, because butterflies and transects were not equally distributed over the country and habitats (Fig. 3).
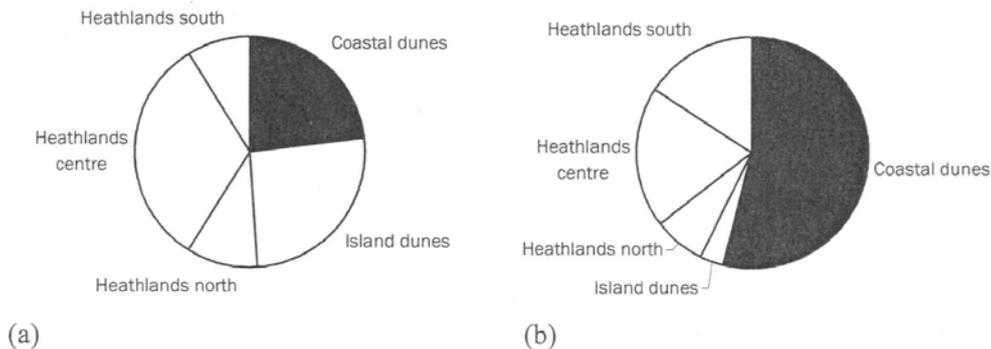

Fig. 3. Distribution (a) of the estimated habitat area of the butterfly *Hipparchia semele* over the five habitat-geographic strata where it occurs in the Netherlands, and (b) distribution of transects where the species was recorded (from Van Swaay *et al.* 2002). For raw data to be representative of the species at the national scale, both distributions should be the same. In practice, coastal dunes were over-sampled, whereas island dunes were under-sampled. Post-stratification for habitat and geographic strata adjust national estimates for these deficits/excesses of sampling.

The process had four steps: (1) the country was divided into five geographical regions, and five habitats types, defining 25 strata; (2) national distribution of butterfly species was extracted from an atlas to identify which strata had to be taken into account for which species, (3) weights were computed as the quotient of the number of transects per stratum as compared to the estimated area of the stratum and, finally (4) weights were applied to abundance indices estimated per site in the statistical model estimating temporal trends (Fig. 4). In the example in Fig. 3 this means that the oversampled mainland dunes are down weighted by 0.43, whereas for the undersampled dunes on the Wadden islands are upgraded with a factor 8.0.
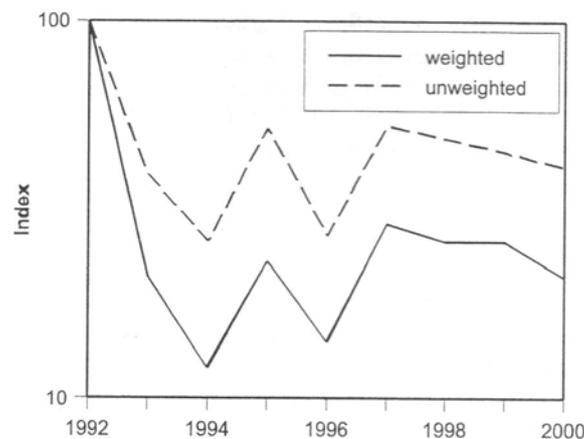

Fig. 4. Temporal trend in abundance index for the butterfly *Hipparchia semele*, in the Netherlands, with data post-stratified for geographic-habitat strata area (weighted) or raw data (unweighted; from Van Swaay *et al.* 2002). Post-stratification revealed that the decline at national scale was much stronger than suggested by raw data.

8

IV.2 Accounting for differences in precision of estimates across datasets

When testing for existence of a temporal trend with estimates from different schemes, precision of estimates per scheme need to be accounted for. A precisely estimated moderate decline should contribute more to the global estimate and test of the temporal trend than a very imprecise steep decline, which cannot be differentiated statistically from stability or even an increase. This is achieved by analysing combined estimates and including the inverse of the squared standard error as weight for each estimate per scheme in the statistical model.

This is illustrated in the case study 4 in D12 (Julliard *et al.* 2004a). Although estimates were not combined across schemes but across species, the principle is the same. Population growth rates of breeding birds were estimated from count and capture data. To test if variations in estimated growth rates among species could be explained by inter-specific differences, population growth rate estimates per species were entered in an ANOVA as dependent variable. Each estimate i was weighted by $1/se_i^2$ and explanatory variables were tested (Fig. 5). Conclusions of Julliard *et al.* (2004a) on determinants of population trend were thus adjusted for differences in precision and methods across species.
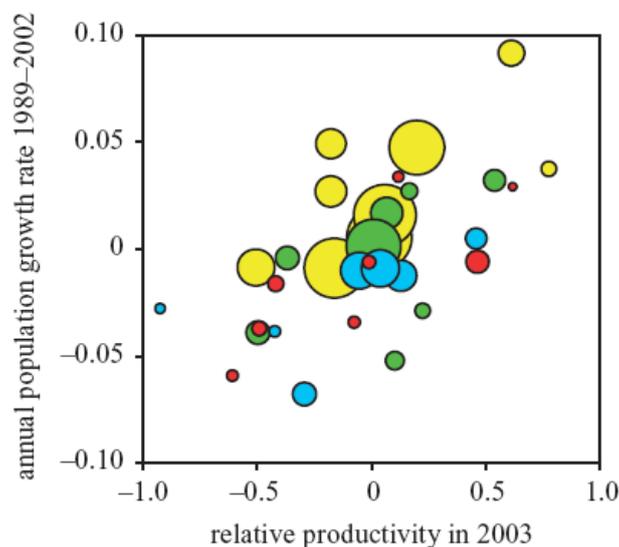

Fig. 5. Inter-specific relationship between medium-term population trend and relative productivity in 2003 for French birds. Relative productivity is derived from age-ratio at bird ringing sites. The higher the productivity in 2003 (exceptionally warm spring), the higher the medium-term population growth rate (from Julliard et al. 2004b). This analysis accounted for differences in precision across species of the estimate of relative productivity; circles are proportional to the standard error $(1/(se^2))$. Colors correspond to species traits: red, seed-eating; green, sedentary; blue, trans-Saharan migrants; yellow, partial or short-distance migrants.

Compilation and illustration of methods for analysis of combined data from different monitoring schemes.

Deliverable **18**, WP2.2, EuMon

V References

Buckland S.T., Magurran A.E., Green R.E. & Fewster R.M. (2005) Monitoring change in biodiversity through composite indices. Philosophical Transactions of the Royal Society B-Biological Sciences, 360, 243-254

Gregory R.D., van Strien A., Vorisek P., Gmelig Meyling A.W., Noble D.G., Foppen R.P.B. & Gibbons D.W. (2005) Developing indicators for European birds. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 360, 269-288

Gurevitch J., Curtis P.S. & Jones M.H. (2001) Meta-analysis in ecology. In: Advances in Ecological Research, Vol 32, pp. 199-247

Julliard R. (2004) Estimating the contribution of survival and recruitement to large scale population dynamics. Animal Biodiversity and Conservation, 27, 417-426

Julliard R., Jiguet F. & Couvet D. (2004a) Common birds facing global changes: what makes a species at risk? Global Change Biology, 10, 148-154

Julliard R., Jiguet F. & Couvet D. (2004b) Evidence for the impact of global warming on the long-term population dynamics of common birds. Proceedings of the Royal Society of London Series B-Biological Sciences, 271, S490-S492

Loh J., Green R.E., Ricketts T., Lamoreux J., Jenkins M., Kapos V. & Randers J. (2005) The Living Planet Index: using species population time series to track trends in biodiversity. Philosophical Transactions of the Royal Society B-Biological Sciences, 360, 289-295

Marsh D.M. (2001) Fluctuations in amphibian populations: a meta-analysis. Biological Conservation, 101, 327-335

McCarthy J.M., Hein C.L., Olden J.D. & Vander Zanden M.J. (2006) Coupling long-term studies with meta-analysis to investigate impacts of non-native crayfish on zoobenthic communities. Freshwater Biology, 51, 224-235

Menzel A., Sparks T.H., Estrella N., Koch E., Aasa A., Aha R., Alm-Kubler K., Bissolli P., Braslavska O., Briede A., Chmielewski F.M., Crepinsek Z., Curnel Y., Dahl A., Defila C., Donnelly A., Filella Y., Jatcza K., Mage F., Mestre A., Nordli O., Penuelas J., Pirinen P., Remisova V., Scheifinger H., Striz M., Susnik A., Van Vliet A.J.H., Wielgolaski F.E., Zach S. & Zust A. (2006) European phenological response to climate change matches the warming pattern. Global Change Biology, 12, 1969-1976

Parmesan C. & Yohe G. (2003) A globally coherent fingerprint of climate change impacts across natural systems. Nature, 421, 37-42

Pollock J.F. (2006) Detecting population declines over large areas with presence-absence, time-to-encounter, and count survey methods. Conservation Biology, 20, 882-892

Pons P., Henry P.-Y., Gargallo J., Prodon R. & Lebreton J.-D. (2003) Bird survival after fire in Mediterranean shrublands: combining capture-recapture data over several bird species. Population Ecology, 45, 187-196

Strayer D.L. (1999) Statistical power of presence-absence data to detect population declines. Conservation Biology, 13, 1034-1038

Van Swaay C.A.M., Plate C.L. & van strien A.J. (2002) Monitoring butterflies in the Netherlands: how to get unbiased indices. In: Experimental and Applied Entomology, pp. 21-27, Amsterdam, NL