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

Instrument: Specific Targeted Research Project

Thematic Priority: Biodiversity conservation

D20: Recommendations for the coherence, scientific quality, and time and cost-effectiveness of habitat monitoring schemes

Due date of deliverable: 31st October 2006)
Actual submission date: 30th January 2007

Start date of project: 1.11.2004 Duration: 42 months

Organisation name of lead contractor for this deliverable
University of Debrecen, Department of Evolutionary Zoology and Human Biology

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 Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Recommendations for coherence, scientific quality, and time and cost-effectiveness of habitat monitoring schemes

Deliverable 20 of EuMon's Work Package 3

By Szabolcs LENGYEL (UNDEBR)

with contributions by Eszter DÉRI (UNDEBR), Pierre-Yves HENRY (MNHN), Eva PAPASTERGIADOU (UNPAT), Andrej KOBLER (SFI), Lado KUTNAR (SFI), Klaus HENLE (UFZ), Andrej SELIŠKAR (ZRC SAZU), Valerija BABIJ (ZRC SAZU)

Table of contents

0.1. Policy summary ... 3
0.2 Background and general comments ... 4
1. Introduction (objectives and definitions) ... 5
2. Coherence ... 5
2.1. Coherence between goals and data used in monitoring ... 6
2.2. Coherence with EU directives and 2010 expectations ... 6
3. Criteria for scientific quality (habitat monitoring) ... 7
3.1. Criteria derived from entries to the DaEuMon database ... 7
3.1.1. Reason for launching ... 7
3.1.2. Spatial representativity of data from monitoring ... 8
3.1.3. Biodiversity representativity of habitat monitoring schemes ... 9
3.1.4. Ability to detect trends: statistical power ... 10
3.1.4.1. Measurement precision ... 10
3.1.4.2. Background variation in habitat range and area ... 11
3.1.4.3. Effect size ... 12
3.1.5. Reliability judged by coordinator vs. reliability estimated from database entries ... 12
3.1.6. Habitat quality evaluation ... 13
3.1.7. Scientific/biological knowledge requirements for collection of monitoring data ... 14
3.1.8. Use of state-of-the-art field and statistical methodologies ... 14
3.2. Other optional criteria for scientific quality ... 16
4. Criteria for time and cost-effectiveness ... 16
4.1. Potential indicators ... 17
4.1.1. Coverage of habitat monitoring ... 17
4.1.2. Ecological and taxonomical extent of monitoring ... 17
4.1.3. Scientific value ... 17
4.2. Effort indicators ... 18
4.2.1. Indicators for time requirement of monitoring schemes ... 18
4.2.2. Financial resources indicators ... 19
4.2.2.1. Personnel costs ... 19
4.2.2.2. Material/equipment costs ... 19
4.2.2.3. Total costs ... 19
4.3. Indicators for time and cost-effectiveness ... 20
5. Synthesis: recommended logic for evaluations ... 20
6. Case studies ... 23
6.1. Comprehensive case study: all habitat monitoring schemes ... 23
6.2. A preliminary evaluation of monitoring schemes for Natura2000 site designation/in Natura2000 sites from Greece, Spain, and the UK ... 27
7. References ... 30

0.1. Policy summary

The general aim of this deliverable is to help in identifying suitable approaches for the establishment of new habitat monitoring schemes, for the improvement of existing schemes, and for moving towards higher coherence among schemes in Europe. The direct objectives are to provide a detailed list of criteria for judging and quantifying the scientific quality of habitat monitoring schemes, to outline ways for establishing coherence within and among schemes, and to estimate the time- and cost-effectiveness of schemes. All criteria proposed can be qualitatively characterised or quantified from data entered in the DaEuMon database available at http://eumon.ckff.si/monitoring.

Coherence (chapter 2) can be defined in several ways. The DaEuMon database can be used to characterise schemes using criteria related to:

- the internal consistency of individual monitoring schemes,
- compatibility among monitoring schemes, and
- coverage of Natura 2000 habitat types for which countries have medium to very high responsibility by the monitoring schemes.

Scientific quality of monitoring schemes (chapter 3) is characterised by the following measures:

- spatial and statistical representativity of data collected,
- biodiversity representativity
- statistical power: ability of monitoring schemes to detect trends,
- measurement precision of monitoring schemes: reliability of data from monitoring,
- the extent to which habitat quality can be monitored,
- scientific knowledge requirements for data collection, and
- use of state-of-the-art field and statistical methods.

Time and cost-effectiveness (chapter 4) is proposed to be measured as the ratio of (i) the level of information obtained by the scheme and (ii) the effort necessary to conduct the scheme. The level of information encompasses the quantity and quality of information provided by the scheme and can be measured by:

- the areal coverage of habitat monitoring,
- ecological and taxonomical extent, and
- scientific quality of monitoring schemes.

Effort required to run the schemes can be of two kinds:

- Time requirements, measured by manpower, including both professionals and volunteers and
- financial costs, which are composed of
  - personnel costs, which can be estimated by manpower plus salaries (not in DaEuMon), and
  - costs of materials and equipment.

We provide these guidelines to filter out schemes that can be recommended as examples of “best practice” schemes given the trade-offs described or schemes that are particularly suitable for integration into broader (geographically or taxonomically) monitoring schemes. This should allow coordinators of monitoring schemes to identify scopes for improvements of their schemes given their specific constraints. A full evaluation should involve the simultaneous evaluation of scientific quality, coherence, and time and cost-effectiveness for monitoring schemes. Because monitoring schemes differ largely in geographic scope, extent of habitat types or quality assessed, and time and cost requirements, we argue that full

evaluations should be carried out on smaller sets of similar schemes. Guidelines for a synthesis approach using composite measures of quality for such smaller sets are given in chapter 5 of this document.
In chapter 6 we illustrate the evaluations in two case studies from the EuMon database. In one of the case studies, we assess all habitat monitoring schemes in the database and in the second one, we evaluate schemes from Greece, Spain, and the UK.

0.2 Background and general comments

This document aims to develop criteria for evaluating the coherence, scientific quality, and time and cost-effectiveness of habitat monitoring schemes. This document has been developed in parallel with Deliverable 17 (Recommendations for coherence, scientific quality, and time and cost-effectiveness of species monitoring schemes), because many of the criteria used for evaluations are similar for monitoring schemes regardless of whether they focus on species or on habitats. Therefore, the structure and wording of the two documents is similar at some places. However, the documents differ in their topics as they concentrate either on species (D17) or habitat monitoring (D20). Finally, it obviously follows that each document contains criteria applicable only to either species or habitat monitoring.

The DaEuMon database can provide data to quantify some of the criteria given below, whereas some criteria cannot be evaluated because the database will not have information on the topic. However, it is important to list the latter criteria in addition to the ones that can be tested directly from the database in order to provide a complete set of criteria both for WP5 and for future external reference. When data from the EuMon database (DaEuMon) can be used for the evaluation, specific references for this possibility are given or will become obvious from the text (e.g. when referring to database question numbers).

Some of the ideas presented may seem at first as overly optimistic and simplistic to be used in a general evaluation of monitoring schemes. However, we considered it important to include all ideas because these recommendations will provide a foundation for future work in WP5, which will focus more on testing and selecting the best approaches for the general evaluation of monitoring schemes.

1. INTRODUCTION (OBJECTIVES AND DEFINITIONS)

Habitat monitoring is often difficult and challenging in several aspects. Besides the problem of what to measure, coordinators are often faced with problems related to which sampling design to choose, how to collect data, whether remote sensing or ground-based serves the purposes better, how to overcome constraints (man-power, resources), and how to interpret and report results. Therefore, it is essential that the EuMon project, which collects and analyses information on monitoring schemes from European countries develop criteria as to the coherence, scientific quality, and time and cost-effectiveness of habitat monitoring schemes.

The general aim of this deliverable, therefore, is to help identifying suitable approaches for the establishment of newly initiated schemes, for the improvement of existing schemes and to contribute to moving towards higher coherence among monitoring schemes in Europe. The direct objectives are to outline ways for establishing coherence within and among monitoring schemes, to provide a detailed list of criteria for judging and quantifying the scientific quality, and time- and cost-effectiveness of habitat monitoring schemes. The outputs from this document may help in the development of the basis for integration of monitoring schemes (input to D19), and to provide practical input to WP5 in the development of indicators and tools.

Before we start, it is necessary to define the concepts central to the correct interpretation of this document. Therefore, each of the following chapters starts with defining what is meant by coherence, scientific quality, and time or cost-effectiveness of monitoring schemes. Most of the conclusions are made on extractions from the DaEuMon database. The conclusions thus depend to a certain extent on the quantity and quality of the information present in the database. At the start of the preparation of this document (October 25, 2006) 140 habitat monitoring schemes were available in the database, while in this final version submitted in January 2007 were 146 schemes taken into consideration. However, it can be expected that the conclusions of this document remain valid, even if re-analysed on a larger sample size of habitat monitoring.

2. COHERENCE

Coherence within and among monitoring schemes can be defined in several ways. Within schemes, coherence can exist between the stated goals and methods used in the monitoring scheme (‘aspect 1’). This aspect will especially be important for the suggestion of optimal approaches, which will need to have high internal coherence (i.e., whether the recommendations about quality, time constraints etc. fit each other). Coherence among schemes may indicate the level of compatibility among the schemes (‘aspect 2’). The main meaning of coherence here is that monitoring schemes are convergent enough to allow the drawing of common inferences from them. For instance, schemes that have a similar scope, similar goals, and similar methods may be more coherent with each other than schemes differing in these aspects. Coherence among the schemes is directly related to integration, i.e., more coherent schemes may also be better suited for integration. The examination of coherence from this perspective is the focus of Deliverable 19. Therefore, we will not go into much detail in this aspect here.
Finally, coherence of the monitoring schemes may be interesting from the perspectives of the Natura 2000 (‘aspect 3’). Questions that are important in this aspect concern the coverage of Natura 2000 sites by the monitoring activities, as well as the coverage of Natura 2000 habitats

by the monitoring schemes. This aspect of coherence is also related to tasks of Work Package 4 (defining national responsibilities in monitoring), which studies in detail the coverage of species and habitats of Community interest by the existing monitoring schemes. Here, we only examine whether the monitoring schemes provide sufficient information to assess the complementarity of Natura 2000 sites. This information will be placed in a broader context of information gaps for the establishment of reserves site networks in deliverable D14.

2.1. Coherence between goals and data used in monitoring

In this part, we review coherence between the stated goals (objectives) and the main data collected in a monitoring scheme (‘aspect 1’). We use information entered for questions H1 (what is monitored: distribution or composition) and H2 (main data collected: species presence/absence, species abundance). We investigate all potential combination of goals/data types and evaluate the appropriateness of data types as a function of the goals stated (**Table 1**). “Appropriateness” is scored in three ways: appropriate (high coherence between goal and data type), poorly appropriate (low coherence), and inappropriate (negligible coherence; criteria extracted from Appendix 2 of Deliverable 2).

Table 1. Coherence between goals and data types collected in habitat monitoring schemes.

H1: goals	H2: data type	Derived information	Evaluation
distribution	species presence/absence	spatial representation	appropriate
	species abundance	representation in space and in numbers	poorly appropriate (too detailed)*
composition	species presence/absence	spatial representation	appropriate
	species abundance	representation in space and in numbers	poorly appropriate (too detailed)*

*The majority of collected/included habitat monitoring schemes does not provide or specify information about species abundance (H2), which is limited and not detailed enough. Therefore, further evaluation in this view could not be appropriate. However, information about structure of habitat should never be ignored. When other sources or data are included, it is obvious, that determining quantitative relationships between species abundances is essential to estimate structure or composition of habitat/community.

2.2. Coherence with EU directives and 2010 expectations

To determine the coherence of monitoring schemes with the expectations derived from the Natura 2000 (“aspect 3”) and with the goals of the 2010 target (jointly referred to as objects of Community interest), it is essential to quantify the status and trends in composition, quality, and distribution. For the following evaluation of coherence, only those schemes can be used that provide an opportunity for such quantification (which is similar to Natura 2000 reporting obligations for the member states).

The objective of evaluating coherence is to quantify the percentage of schemes in DaEuMon that are clearly relevant for the 2010 target. The coherence of schemes with EU targets can be scored using H22 (List of habitats monitored) in relation to the list of all habitats listed in Annex 1 of the Habitats Directive, additionally taking into account the coverage of the priority habitats. The question is how many of the available habitat types are present within each sub-group of habitat types. For example, there are 72 habitat types in group 9 (forests). If 51 of them are covered by monitoring schemes, then coherence towards EU targets is 70%.

With regard to point 2.1. (Coherence between goals and data types), a further separation is necessary for schemes that monitor habitat distribution and composition. This process needs to involve the following steps:

1. Compilation of lists of habitats under the Habitats Directive for each country.
2. Identification of habitats that are actually monitored. This needs to be based on a routine to be developed using data entered in DaEuMon.
3. Separation of actually monitored habitats based on coherence between the goals and data types used in monitoring (point 2.1).
4. Calculation of proportion (%) of habitats of Community interest that are actually monitored and then separate calculation of proportion of habitats whose monitoring is considered appropriate by the criteria proposed in 2.1 (Coherence between goals and data types) and 3 (Criteria for scientific quality).

One main problem with the steps above is that the results are largely dependent on how exhaustive DaEuMon data are (Step 2). The evaluation of coherence will be reliable if all main monitoring schemes are entered for a country, whereas it will be less reliable if many monitoring schemes operating in a country are missing from DaEuMon. Thus, for countries well represented in DaEuMon (as of October, 2006: e.g. Spain, Greece, UK, Poland, France, Hungary, and Germany with more than 7 habitat monitoring schemes), the above approach will provide reliable results and a higher potential for compatibility and integration. In any case, this approach provides the basis for a kind of gap analysis: "Which habitats of Community interest are NOT monitored adequately?" Some preliminary examples of such gap analyses can be found in chapter 6 (Case study).

Another way to characterise coherence would be to consider if a scheme monitors many or only a few habitats of Community interest. For instance, if one scheme monitors 10 Directive habitat types, whereas other schemes for the same country monitor only a few (or even only a subset of the 10 habitat types), the latter schemes can be considered as less coherent than the first scheme. In such cases, a suggestion of improvement would be to monitor more habitats than actually are monitored or to better align existing schemes to increase cost-efficiency and coherence with national monitoring goals.

3. Criteria for scientific quality (habitat monitoring)

3.1. Criteria derived from entries to the DaEuMon database

3.1.1. Reason for launching

Chances are that if a monitoring scheme is launched because of scientific interests, it is probably better designed in a scientific sense and its results are analysed better (or may be better analysable). Therefore, if the reason for launching (question 5.) is "scientific interests", there is a good indication that the monitoring scheme has a higher-than-average scientific value. Because these claims are often debated, the assumption needs to be tested with answers given to other questions in the database. Questions that may provide a surrogate measure of scientific quality are outlined below in sections 3.1.2. through 3.1.8. A 'signal' for scientific quality may be present if there is a difference in some measure of scientific quality among schemes launched for different reasons, or if variables associated with higher scientific quality are more frequently associated with the reason of "launched from scientific interest". Therefore, the approach is interesting also if reversed, as it provides an opportunity to ask whether studies not launched for scientific reason differ in sampling design, representativity, or other relevant criteria.

3.1.2. Spatial representativity of data from monitoring

Representativity is relevant in judging scientific value for several reasons. A spatially representative monitoring is either conducted in the entire target habitat/region/country or uses an appropriate sampling design. The sampling design should ensure that the results of habitat monitoring are representative at the national or regional scale. The scale of representativity always depends on the goal of the monitoring scheme. Thus, one of the main questions of interest here is question 6 (geographical scope). From the perspective of representativity in terms of coherence with the EU directives and the 2010 target, schemes that are international or national in scope need to be considered as representative, in contrast to regional or local schemes. If there is a protected area across two countries, it should not be considered international, but is actually regional across borders and as such not representative.

Possibilities for estimating spatial representativity of monitoring schemes from DaEuMon are limited. One possible measure of representativity is the relative area monitored per habitat per country. From this perspective, habitat monitoring schemes in the DaEuMon database can be of two types. One type is when all habitats within a region or country are monitored simultaneously (‘holistic approach’) and the other type is when a certain habitat type (or group of similar habitats) is monitored within a country or region (‘specific approach’). In the holistic approach, the number given in H10 (actual area monitored) should be close to the area of the country (question 7.; country areas are necessary to obtain from other sources). If this criterion is fulfilled, then the scheme should be scored as representative at the national level. If, however, the actually monitored area is restricted to a part of the country, representativity will depend on the actual representation (e.g. proportion, spatial pattern etc.) of the target habitats within the region studied. Scoring representativity in such cases, as well as in cases of the specific approach of habitat monitoring, thus requires information not available in DaEuMon.

In cases of the specific approach, however, additional information in DaEumon can be used. For example, if the sampling design is stratified or is based on randomisation, there is a good chance that the scheme has high spatial representativity. The main and additional criteria for representativity are summarised in **Table 2**. The scoring can be based on a simple nominal scale (as in **Table 2**) or on the relative position of a scheme compared to the best one according to this criterion (e.g. number of actually monitored sites per maximum number possible).

Table 2. Spatial representativity scores for answers to questions in the DaEuMon database.

Question in database	**If answer is:**	**Representativity is scored as:**
6. Geographical scope	international	representative
	national	representative
	regional	not representative (except of endemic/restricted habitats) *
	local	not representative (except of endemic/restricted habitats) *
H4: Documentation of spatial variation by	remote sensing	generally representative
	field mapping	generally not representative (rarely implemented at national scale)**
H5: Sampling design	stratified	representative
	not stratified	may or may not be representative
H7: Choice of sites to be monitored	exhaustive	representative
	systematic	representative
	random	representative
	based on expert knowledge	not representative
	other	not representative

* The EuMon database does not contain information on whether a scheme monitors endemic or restricted habitats.
**Majority of collected/included habitat monitoring schemes does not provide or specify information about documentation of spatial variation by field mapping (H4). Therefore with further evaluation we could not meet the required representativity.
Some small-scale habitats (e.g. endemic/restricted habitats) could only be mapped in field because remote sensing is unable to provide sufficient and accurate information about their spatial variation.

It is important to draw attention to two potential problems. In H10, some coordinators may have given the actual area monitored, e.g. sum of plot areas and not the area at which results can be extrapolated (although explicit in the comment to this question). Another problem is when a habitat is endemic or restricted, in which case even a small percentage value of national area can secure high representativity for the monitoring. However, this problem is relatively small; if the restricted/endemic status is known (as is the case with many such habitats), it is easy to go back to DaEuMon and assess representativity.

3.1.3. Biodiversity representativity of habitat monitoring schemes

An important aspect of the representativity is biodiversity representativity. The biodiversity representativity can be scored using H22 (Lists of habitats monitored) in relation to the list of all habitats listed in Annex 1 of the Habitats Directive (HD). The coverage of the priority habitats, if present in the focal country or region could additionally be taken into account. The main question in this aspect is how many of the available habitat types are monitored by the schemes within each sub-group of HD habitat types. For example, there are 72 habitat types in group 9 (forests) in the HD. If 51 of these habitats are covered by the monitoring schemes, the biodiversity representativity will be 70%. In **Table 3** we propose a scoring rule.

Table 3. Biodiversity representativity scores for answers to question H22 in the DaEuMon database.

Question	If answer given suggests that	Representativity is scored as
H22: Lists of habitats monitored	less than 33% of the habitat types within HD group* are monitored	Not representative
	33% or more of HD habitat types within group* AND less than 100% of priority habitats within group* are monitored	Not representative
	33 % or more of HD habitat types within group* AND 100% of priority habitats within group* are monitored	Representative

* Habitat Directive groups are the high level groupings of habitat types in Annex 1, e.g., group 9 (forests).

3.1.4. Ability to detect trends: statistical power

A central criterion for the evaluation of a monitoring scheme is whether it can statistically detect a change of certain size in the range and area of habitats. The probability of detecting a trend or change of certain size is statistical power. Statistical power concerning habitat monitoring is a function of measurement precision, annual background variation in the variable of interest, and effect size (the strength of the trend or certain size of the change one wishes to detect as significant).

3.1.4.1. Measurement precision

Measurement precision is the most important component of statistical power. Measurement precision (or "error", in analogy with experimental design theory) is a composite measure of several features of the monitoring system, but is mostly influenced by sample size (number of measurements repeated in space and time). Measurement precision can be estimated quite well from data in DaEuMon. The following variables are available for this purpose:

- Data type:
  - H2: nature of data collected: presence/absence data vs. counts,
- Temporal aspect:
  - number of years monitored (frequency of monitoring, to be related to among-year variation): given by difference between H17 (ending year) and H16 (starting year)
  - H13: among-year frequency of sampling (= to be related to among-year variation)
  - H14: number of visits per site (= to be related to within-year variation)
- Spatial aspect:
  - H11: number of sampling sites monitored per year (or precision of annual status estimates; = sampling effort + spatial variation)
  - H12: number of samples per site (= replicates per visit)

Sample size, i.e., number of measurements repeated in space and time, is negatively related to measurement error; the more measurements, the smaller the errors and the higher the precision of the measurement. The data given in DaEuMon are not directly suitable for a quantitative evaluation of measurement precision across many different monitoring schemes.

The data in DaEuMon, however, can be used in a qualitative evaluation of measurement precision as follows. An index for the ability to measure sampling and measurement error can be devised by setting an arbitrary value (here, the minimum number of sampling sites was taken to be 15, and the minimum frequency of monitoring was taken to be at least once in four years). The following four cases possible are:

- poor precision: H11 < 15 OR H13 > 4
- medium precision: H11 ≈ 15 AND (H12 = 1 OR H14 = 1) AND H13 ≈ 4
- good precision but measurement error not considered: H11 > 15 AND ((H12 = 1 OR H14 = 1) AND 1 < H13 < 4)
- good precision and measurement error accounted for: H11 > 15 AND ((H12 > 1 OR H14 > 1) AND H13 = 1)

As a shortcoming of the method proposed, the results largely depend on the quality of information given in DaEuMon. Based on the database entries, it may be that precision cannot be reliably estimated due either to the lack of detailed information or inapplicability of the declared sampling design.

Special consideration is warranted for habitat monitoring schemes that use remote sensing as their main type of data acquisition and interpretation. This is because several of the above-proposed correlates of measurement error are not readily available in schemes based on remote sensing (e.g. “number of sampling sites” cannot be interpreted for space imagery).

One of the most common means of expressing remotely sensed habitat map accuracy or change-map accuracy is the preparation of an error matrix, sometimes also called confusion matrix or contingency table (Lillesand and Kiefer 1994). Error matrices compare the relationship between known reference data (e.g., ground truthing) and the corresponding results of remotely sensed habitat identification. Several characteristics of habitat identification can be expressed by an error matrix. For example, the overall accuracy is computed by dividing the total number of correctly identified samples (the sum of elements along the major diagonal) by the total number of reference samples. An often used accuracy measure is the kappa statistic (Congalton and Green 1999), which indicates the extent to which the percentage correct values of the error matrix are due to “true” agreement vs. “chance” agreement.

3.1.4.2. Background variation in habitat range and area

Habitat status and trends, by their nature, may change slowly through time especially over large landscapes. The spatial patterns of large landscapes strongly influence population dynamics and community structure (Johnson et al. 1992). The spatial patterns can be quantified using a variety of landscape structural metrics (e.g., O'Neill et al. 1988, Turner 1990, Turner and Gardner 1991, Baker and Cai 1992, McGarigal and Marks 1995), based on a conceptual model known as the patch-corridor-matrix model (Forman 1995).

Estimates for the year-to-year variation (background temporal variance) in the variable of interest (habitat range and area) cannot be derived from DaEuMon. We propose to extract them (at least orders of magnitude) from published data on successional/anthropogenic conversion rates changes of habitat type (H22).

Data on habitat changes should be available for habitat monitoring schemes using remote sensing, and when data are available from the same region/country from at least two points in time. For example, using remote sensing, the temporal variation in habitat areas and trends can be evaluated using land cover change analysis, at least for broad groups of habitat types, e.g., as broad-leaved forests in the CORINE Land Cover project (Bossard et al. 2000). Land cover change analyses can be done using various approaches, including time series analysis, change vector analysis, time series correlation, temporal image differencing, and temporal image rationing (e.g., Lillesand and Kiefer 1994, Eastman 2003). Typically, to validate the results of a land cover change map one needs to determine what the true land cover was for the time periods being compared, or what was the true nature of change between the time periods. This can be done using some reference data, either ground truthing data (e.g., field data gathered at a sample of permanent plots) or high resolution remotely sensed imagery (e.g., multitemporal aerial photographs).

It is also relevant which criterion of habitat change is used for the reporting versus the precision that is achievable regarding the habitat remaining (e.g. there is considerable literature on remote sensing estimates of changes in the cover or distribution of tropical rainforests and possibly on some key habitats also in Europe).

3.1.4.3. Effect size

Effect size or the strength of the trend or extent of change one wishes to detect as significant is often an arbitrary value obtained from informed guesses. For monitoring schemes, the effect size is likely to be influenced by what is considered as an alarming change in the range and area of the habitats of interest. We propose to use threshold values from the Notes & Guidelines of Assessment of Monitoring and reporting under Article 17 of the Habitats Directive by the Scientific Working Group of the Habitats Committee at DG Environment of the European Commission:

- trend in range of the habitat(s): 1% annual decrease OR 5 % annual decrease of coverage in favourable reference range,
- trend in habitat area: 1% annual decrease OR 10% annual decrease in the favourable reference area.

If the three parameters (measurement precision or “error”, background variation in habitat range and area, and effect size) can be quantified, then the statistical power to detect a trend can be computed. The precision of the power estimate will depend largely on measurement precision as this is the component most strongly related to the features of the monitoring schemes. Such an estimated statistical power can then be evaluated against a set of criteria developed to judge efficiency over a range of monitoring objects in a range of countries, regions etc.

3.1.5. Reliability judged by coordinator vs. reliability estimated from database entries

One important measure of the quality of a monitoring scheme is how well it functions according to its designers/implementers and how well it functions in reality according to independent criteria. The match between statistical power assumed by the coordinator and the requirements set by the Natura2000 goals needs to be an important factor in making recommendations as to a “best practice” monitoring scheme. The “reliability” of the scheme as judged by the coordinator is given in question H24 (“Minimal annual change you think you can statistically detect”). Three scenarios can be deduced from the answer to this question:

- If the data entry is missing, it likely means that the statistical power to detect trends is not considered important by the coordinator, e.g. he/she has no idea of what it might be or it is very difficult to estimate.
- If an entry is given and it is lower than the thresholds for conservation status evaluation, then statistical power is supposedly high (sufficient).
- If an entry is given and it is higher than the thresholds for conservation status evaluation, then statistical power is supposedly low (insufficient).

Schemes with an experimental design can be considered as monitoring schemes designed to detect specific temporal changes, due to given causes of change. In such cases, the trend is no more the slope for a continuous temporal change, but an effect size associated with a specific treatment. If we view it in this way, the issue of experimental design may be discussed here, in the part “ability to detect a trend”. Question H6 (“Use of experimental design”) is of relevance here. If the answer is “yes”, the scheme is highly appropriate for monitoring purposes as it contains an element of an experiment that can be appropriate to discern causes and effects. Thus, a scheme containing experiment(s) probably have a way to quantify their statistical power. If the answer is “no”, statistical power is less likely to be considered, although the scheme can still be good or appropriate for monitoring.

3.1.6. Habitat quality evaluation

Several questions in DaEuMon offer an opportunity to evaluate habitat monitoring schemes on the basis of the quantity and quality of information collected on the properties of the habitats (**Table 4.**).

Table 4. Scores of scientific value for answers to questions related to habitat quality.

Question in database	**If answer is:**	**Scientific value is scored as:**
H1: Composition monitored	yes	higher
	no	lower
H3: Environmental parameters collected	yes	higher
	no	lower
H25: Causes of change can be inferred	yes*	higher
	no	lower
H27: Habitat quality criteria monitored	yes**	higher
	no	lower

* additional criterion is that H26 (the causes of changes) is not an empty list and H6 (the scheme uses an experimental design) = “yes”
** additional criterion is that H28 (quality criteria) is not an empty list

Even though the monitoring of the quality of the habitat(s) provides scientific value to a monitoring scheme, it is difficult to judge whether the type of habitat quality criterion is appropriate. This is mainly because the relevant measure of the quality of the habitat will vary greatly among habitats (e.g. fragmentation may be important for forests but less for some marine systems). Therefore, using the same type of habitat quality measure (fragmentation, structural changes, species composition, physical-chemical environment, indicator-keystone-umbrella-typical species) for all monitoring schemes is not adequate. However, it can be relevant for smaller sets of monitoring schemes that are carried out on similar ecological systems or in similar environmental settings.

Another issue is the use of typical species (indicator species in the questionnaire – H29) for characterization of habitat quality.

The basic problem is that currently a widely accepted, recognized, and independent list of typical species linked to habitat types does not exist. Therefore, the judgment on whether scientifically sound species are monitored and on whether an adequate number of such species is monitored requires expert knowledge. To make the evaluation work, such subjective judgments should be avoided.

It should be mentioned here that the EU guidelines foresee the typical species as such indicators and that work is still in progress attempting to identify typical species for different habitats and the appropriateness of the approach.

3.1.7. Scientific/biological knowledge requirements for collection of monitoring data

If we assume that the higher the scientific/biological knowledge requirements in a monitoring scheme is the higher is the scientific quality of the scheme, then proportion of professionals participating in the monitoring scheme can indicate scientific quality. The proportion of professionals can be calculated from numbers given in answer to questions H30 and H31 as: H30/(H30+H31). Furthermore, monitoring schemes that require training or expert knowledge (H32) may have a higher scientific quality than those that do not require such measures. However, before using these criteria to evaluate scientific quality, the above assumptions need to be tested by using data from DaEuMon. Scientific output from the monitoring schemes is difficult to obtain from DaEuMon, although answers given to question 8 (References) can give a hint of whether scientific publications are prepared as an output. Therefore, as an example on testing the relationship between involvement of professionals and scientific quality, the proportion of professionals can be used as an independent variable and existence of scientific publications (e.g. a binary variable from answers to question 8 References) can be used as response variable in a logistic regression. A possible bias exists if many coordinators did not provide references on their scheme even though they have some. Alternatively, some coordinators may also upload many references not having high scientific value (e.g. project descriptions, grey literature etc.). One option to estimate this bias is to relate the level of statistical analysis (H19) to occurrence of scientific publications in question 8 (References), assuming that results from more sophisticated analyses are published more often.

It is necessary to emphasize again that all of the above needs a rigorous testing before they can be proposed for the general evaluation of scientific quality. Especially, the relationships between either the number of professionals or the presence (and possibly quality) of references and high scientific quality needs to be tested and judged robust to use in estimating scientific quality.

If the assumptions do not hold, the scientific knowledge criterion is not recommended for direct use in the evaluation. In this case, the information given in questions H30-H32 can be used to build cost-efficiency indicators, assuming that whoever collects monitoring data (H30-31: professional/volunteers, H32: trained/not trained), their value to quantify state and trends in distribution and population size is the same.

3.1.8. Use of state-of-the-art field and statistical methodologies

The scientific quality of a monitoring scheme is likely to be higher if the data are collected and/or analysed by more up-to-date, more sophisticated methods. For example, combining remote sensing with field mapping can be viewed as modern in contrast to using only field mapping (question H4), since the two are complementary to each other (see below). Similarly, data analysis is more sophisticated than lack of data analysis and if it involves General Linear Models or other advanced statistics beyond simple graphics and descriptive statistics (H19). A special case is when data are analysed by persons/institutions different from those collecting the data (which is often the case with large-scale national or international projects). Data quality and analysis in the latter cases usually are both sophisticated.

It is important to note here that the general relationship between modern methodology and scientific quality is in the direction modern methods $\rightarrow$ (usually) high scientific quality. However, high scientific quality may also be achieved by relatively simple data collection and analysis; therefore, one cannot declare poor scientific quality for a monitoring scheme that does not use the above modern methods of data collection and analysis. This difference in the relationship between modern methods and scientific quality needs to be considered in the evaluation of monitoring schemes or analysis of data from DaEuMon.

Remote sensing methods are complementary to field methods. By repeatedly providing overview of spatial aspects of habitats the remote sensing methods are elucidating spatial processes. On the other side, the field methods provide better insight into qualitative aspects of habitats, as well as a better thematic resolution. For example, one can easily identify the type of vegetation (habitat type) in the field, whereas remote sensing methods are limited to more general habitat groupings, e.g., deciduous forests.

Satellite or aerial imagery is therefore useful for showing thematically generalized, coarse-scale changes in vegetation, patterns of development, and major disturbances. However, there is no national data set that is adequate to show changes in habitats at fine spatial (less than 30 meters) and fine temporal (annual) scales. One caveat, no matter what system is used, is that an intermediate number of habitat or land cover categories are probably best for the purposes of countrywide monitoring. Some countries with diverse ecological landscapes may have more than 50 habitat types, whereas other countries with more uniform ecological landscapes may not have much more than a dozen. Some widely used examples of habitat / land cover nomenclatures are CORINE Land Cover (Bossard et al. 2000), Anderson’s land use and land cover classification system (Anderson et al. 1976), and FAO land cover classification system (http://www.africover.org/LCCS.htm). Identifying trends in habitat status requires repeated observations. Using Landsat Thematic Mapper imagery (http://www.landsat.org) has been shown to be a cost effective way for countries to track medium-scale habitat change (Project CORINE Land Cover - http://dataservice.eea.europa.eu/dataservice/corine). To purchase additional Landsat imagery or use aerial photography each country must bear additional costs.

The collection of appropriate information, either by automated classification or by visual interpretation of remotely sensed imagery, can be costly, time-consuming, and technically difficult. The advantage of visual interpretation is the ability of human interpreter to glean thematically more detailed information from the imagery, by making effective use of contextual and ancillary information. On the other hand, the main advantage of automated image classification methods is the speed of processing. Ground truthing is often needed to interpret or classify the remotely sensed imagery. Therefore, decisions about what imagery to collect and how to collect it need to be made carefully with due consideration given to what is feasible, practical, and effective, and appropriate for the scale of analysis.

There are at least two ways to incorporate methodological complexity in a general evaluation of monitoring schemes. Firstly, the questions related to methodological complexity can be considered separately and each of the more “modern” or “sophisticated” entry value can get a higher weight in the evaluation. Secondly, an index of methodological complexity can be developed, which could summarise information from several questions. Questions that are relevant for this aspect are H4 (remote sensing/field mapping), H19 (type of data analysis), and possibly H32 (training/expert knowledge required).

3.2. Other optional criteria for scientific quality

The criteria listed in this part cannot be derived directly from the DaEuMon database, but we list them here for the sake of completeness. The criteria may be used on a subset of monitoring schemes with details of suitable depth or in future attempts to evaluate monitoring schemes once data become available.

Are indicator groups appropriate for the goal of monitoring habitat quality? Which groups of species may be suitable as indicators for different habitat types? [E.g. diurnal butterflies are inappropriate for forest monitoring] These questions, important for habitat monitoring, are the subject of other research endeavours besides the EuMon project (e.g. EC DG ENV.B2 "Habitats" Scientific Working Group, EU Biodiversity Working Group on Indicators, Monitoring, Reporting, Data management and Information sharing, EC CHM - European Community Biodiversity Clearing House Mechanism:
http://biodiversity-chm.eea.europa.eu/information/indicator,
SEBI2010 - Streamlining European 2010 Biodiversity Indicators:
http://biodiversity-chm.eea.europa.eu/information/indicator/F1090245995).

- Other refinements of the sampling design:
    - in time: Are various aspects of habitats considered? For example, is the phenology of indicator species monitored considered? Is frequency of sampling adjusted to the temporal (e.g. aspectual or seasonal) changes of the habitat or to the phenology of the indicator species monitored?
    - in space: Are size and heterogeneity of the habitat monitored considered? Is the sampled area representative of the size and heterogeneity of the total habitat?
- Are habitat properties monitored?
    - Are compositional variables (typology, syntaxon) of the habitat monitored? (α-diversity)
    - Is the spatial structure of syntaxa considered? (β-diversity)
    - Is the structural diversity of habitat (e.g. canopy layers) measured/estimated?
- What is the external validity of the monitoring scheme? Is the output from the scheme applicable to other systems (e.g. collecting data that can be fed in EU-wide systems such as EUNIS or CORINE)? This issue is treated in detail in Deliverable 19 (Integration of schemes), which attempts to identify, which schemes could be combined to set up international monitoring schemes that can incorporate existing local, regional, or national schemes.

4. Criteria for time and cost-effectiveness

Efficiency or effectiveness is usually determined as the achievement relative to some investment allocated to do the work. For monitoring schemes, an analogy can be to evaluate Potential of a scheme per Effort. Potential is the quantity and quality of processed information obtained in the scheme, whereas Effort can be any measure of investment, e.g. time requirements expressed in manpower (person months) or money. The Potential per Effort (or PpE) can then be used as a measure of time or cost-effectiveness of monitoring schemes. The present document defines both Potential and Effort indicators but will only mention the possibility of the division to obtain PpE because the exact estimation or calculation of PpE will largely depend on what the user is asking. Different answers may be reached, for example, if one divides different Potential indicators with different Effort indicators. Therefore, users can conduct assessments using different PpE ratios suitable to the questions they address. The aim of this part of the document thus is to provide guidelines for readers to evaluate for themselves how time/cost-effective their approaches are and how they could optimise the compromise between Potential and Effort.

4.1. Potential indicators

Potential variables are numerators in the ratio of "achievement/investment" proposed to describe the time and cost-effectiveness of monitoring schemes. There are several measures of Potential available from DaEuMon. We propose that the Potential of a monitoring scheme depends on the area involved, the ecological and taxonomical spectrum of monitoring, and the scientific quality of the monitoring. Simply put, the main questions are how large the monitored area is, how wide is the range of habitats monitored and how well is monitoring conducted. Most of the variables important as Potential variables are thus also relevant for evaluating the scientific potential of monitoring schemes. Therefore, the following overview will refer back to chapter 3 regarding scientific potential and its components (representativity, statistical power, measurement precision etc.).

4.1.1. Coverage of habitat monitoring

In an ideal case, the proportion of habitat monitored per total area of the target habitat ("coverage" or "spatial representativity") is an important measure of the potential of the monitoring scheme. DaEuMon contains information on the total area to which the results of monitoring can be applied (extrapolated, question H10). Thus, although the actual area monitored is not known, the coordinators, ideally, have already provided the area to which their results can be extrapolated. Then, only the distribution range of the habitat types monitored is required to estimate the proportion of the total area covered by the habitats that is actually monitored (areal coverage of monitoring).

4.1.2. Ecological and taxonomical extent of monitoring

Answers to several questions offer an opportunity to quantify the ecological and taxonomical coverage of monitoring schemes. Besides the areal extent of monitoring, this basic information is the primary source of estimating the Potential produced by monitoring schemes. For example, the Potential is higher of a monitoring scheme if it monitors not one but all the major habitats within a region. Within habitat monitoring schemes involving indicator/keystone/umbrella etc. species, a scheme involving several indicator species can be considered to contain more Potential than a scheme involving only one indicator species. Therefore, the number of habitats (or number of indicator species) monitored is a basic information of direct relevance to Potential of a monitoring scheme. The questions that can be used for such purposes are as follow.

- Number of habitats or indicator species involved in monitoring ("ecological" coverage of habitat monitoring schemes):
    - H20: all habitats or not all habitats;
    - H22: number of habitat types monitored;
    - H29: number of indicator species monitored.

4.1.3. Scientific value

The scientific value (i.e., how monitoring is conducted) is a composite measure of several attributes describing scientific quality reviewed in chapter 3 (especially section 3.1.). Of the variables listed there, representativity of data (questions 6., H4, H5, and H7), statistical power

(H2, H11, H13, H16, and H17) and measurement precision (H12, H14, and H24) appear primarily relevant to estimating the Potential or "extent" of a monitoring scheme.

In addition to these variables, the availability of information on potential drivers or background variables (H25, H26) and information on habitat quality (H27, H28) can be important, as this information enables one to address causes of changes observed during monitoring. The availability of such information greatly adds to the scientific Potential of monitoring, whereas this information is not necessarily and directly related to scientific quality (*sensu* chapter 3). A monitoring system not involving background variables may be independently high, medium, or low in scientific quality.

It has to be noted here that two measures of scientific quality (3.1.7. Scientific/biological knowledge requirements for collection of monitoring data and 3.1.8. Use of state-of-the-art field and statistical methodologies) will need to be considered separately. The reason behind this is that the involvement of many professionals and/or many state-of-the-art methods may result in higher costs for a monitoring scheme, but also can disproportionately increase the Potential or "extent" of the scheme. Therefore, the involvement of professionals and modern methodology needs to be studied separately for each monitoring scheme or within a group of similar schemes.

4.2. Effort indicators

Effort variables are denominators in the ratio of "achievement/investment" proposed to describe the time and cost-effectiveness of monitoring schemes. There are two important measures of Effort available from DaEuMon. One is time (estimated by manpower in persondays), and the other is money.

4.2.1. Indicators for time requirement of monitoring schemes

Time or manpower variables can be deduced from the database in two ways. Firstly, the answer to question H33 (manpower [in personday] needed per year to run the scheme) gives an estimate of the yearly effort in time. To account for the fact that some monitoring schemes are not run every year (question H13: frequency of monitoring), manpower is averaged per year (i.e., divided by H13). The total time requirement to run a monitoring scheme is then given by H33 / H13.

Another group of questions allows the quantification of the time requirement for fieldwork per year (in persondays). The number of sampling sites (H11), the number of sampling occasions per year (H14), and the time requirement necessary for one sampling occasion (H15) can be used to calculate the time requirement for fieldwork in a year. The total time requirement for fieldwork per year is therefore given as H11 * H14 * H15 / H13.

The two indicators (total time required to run scheme and fieldwork involved) give an opportunity for several interesting comparisons. Firstly, if H33 = H11 * H14 * H15 / H13, this means that the monitoring scheme consists entirely of fieldwork and H33 is sufficient for further use. If H33 < H11 * H14 * H15 / H13, i.e., fieldwork is more than the total time required per year, at least one measure of manpower given by the coordinators is wrong and the estimates will not be reliable. Finally, if H33 > H11 * H14 * H15 / H13, we can calculate the manpower necessary for laboratory work (e.g. sorting samples, identifying species, data entry and analysis) as H33 - H11 * H14 * H15 / H13.

4.2.2. Financial resources indicators

Two kinds of basic information on the financial resources of monitoring schemes can be retrieved from the DaEuMon database. Personnel costs and material/equipment costs given by the coordinators can be used to characterise the extent of monitoring schemes in a financial sense.

4.2.2.1. Personnel costs

Personnel costs are not directly present in the database but may be estimated from time variables. The basic idea is to compute the equivalent of human resources in money. The total time requirement for running the monitoring scheme (question H33) multiplied by a daily salary (Y) indirectly estimates personnel costs. Thus, personnel costs can be estimated as H33 / H13 * Y. DaEuMon does not cover information about salaries. Therefore, such information has to be extracted from national statistical sources. Alternatively several categories of "arbitrary salary" may be used for preliminary assessment made directly from DaEuMon. If other costs are negligible, schemes could be compared in terms of effort by directly using manpower involved instead of converting manpower to monetary values.

It should be kept in mind that personnel costs strongly depend on country, by status of the participants, and employers. Forgoing such distinctions may result in biases such as Eastern European schemes being cost-effective merely due to lower average salaries compared to those in Western Europe. Similarly, not taking into account that volunteers or non-trained personnel may earn less than professionals involved in the schemes may overestimate personnel costs. Thus, comparisons of schemes from countries with similar salaries and that involve similar ratios of volunteers and professionals are the most informative comparisons. One also could standardize the evaluation in terms of the share of monitoring costs in the overall spending of a country for biodiversity conservation or of its GDP (gross domestic production) which would allow a meaningful comparison of countries with very different salaries and economies.

The ratio of professionals and volunteers can be used to compute how much money is saved in the monitoring schemes thanks to the involvement of volunteers. To make such an evaluation, we may use two salaries, Y1 for trained and professional participants and Y2 for untrained and volunteer participants (and include Y1 and Y2 in formulae hereafter). The total amount equivalent to total human resources per year is then given by H33 / H13 = Y1 * H30 + Y2 * H31.

The percent of the total manpower that is saved thanks to volunteer involvement is given as H33 / H13 * H31/(H30+H31). If we want to calculate the amount of money saved thanks to volunteer involvement per year, it can be expressed as H33 / H13 * Y1 * (H30/[H30+H31]).

4.2.2.2. Material/equipment costs

The annual costs of materials and/or equipment used in the monitoring scheme is directly given in answers to question H34 ("How much do you spend on material and equipment per year [in €]?").

4.2.2.3. Total costs

Personnel costs and material/equipment costs can be readily incorporated into "Total costs of monitoring per year" and can be given per year as H33 / H13 * Y + H34.

4.3. Indicators for time and cost-effectiveness

After the quantification of indicators for Potential and Effort, composite indicators for time and cost-effectiveness can be devised. The general idea is to establish a ratio of “achievement/investment” or Potential per Effort (PpE). Considering that three types of Potential indicators and two types of Effort indicators are deduced from the DaEuMon database, six combinations can be envisioned, illustrated in **Table 4**.

Table 4. Examples of potential pairing of Potential and Effort indicators.

	Effort indicators	
Potential indicators	*Time requirement*	*Financial resources*
Coverage of habitat monitoring	Area monitored per year	Area monitored per unit manpower/money
Ecological and taxonomical extent of monitoring	Number of habitats / taxa monitored per year	Number of habitats /taxa monitored per unit manpower/money
Scientific value	High or low scientific value per year	High or low scientific value per unit manpower/money

The simplified composite measures shown above are only recommendations. The users (coordinators) may want to define their own indicators rather than strictly following the guidelines proposed here to make them most relevant for their evaluation goals.

Several issues arise with a simplified composite measure. The most important is that a simplified composite measure cannot be meaningfully compared across many monitoring schemes that differ greatly in geographic scope, taxonomical extent, time and cost requirements etc. Therefore, it is important to stress that such composite measures should be calculated only for sets of schemes that reach a certain level of similarity in their stated goals and methods unless the measures are standardized e.g. relative to the total amount of money available for monitoring in a country/program or the economic potential of different countries. Time and cost-effectiveness should be the last step in the evaluation once coherence and scientific quality of the schemes are evaluated.

5. SYNTHESIS: RECOMMENDED LOGIC FOR EVALUATIONS

In this section, we suggest a synthesis-like approach to the evaluation of monitoring schemes. The evaluations suggested in chapters 3&4 are not the only ones possible and users of DaEuMon may want to make different evaluations (even among very different monitoring schemes, e.g. distribution of lichens and population trends of birds). The aim of this section is to provide some guidelines on how users can do such an evaluation for themselves. We do not believe in the usefulness of evaluations of schemes that attempt to develop overall ranks for all schemes. Rather, we provide guidelines that can be used to filter out schemes that can be recommended as examples of “best practice” schemes given the specific trade-offs described above or other trade-offs a use is faced with.

For the evaluation, we suggest using a filter-like approach. As an example, one should imagine a table containing monitoring schemes as rows and different features of the schemes as columns. Then for each relevant entry for a given scheme, qualifications on scientific quality, time/cost efficiency, and coherence as suggested above can be added. For example, for representativity (section 3.1.2.), one should add “representative” in a separate column for schemes showing "international" or "national" entries in question 6 (Geographical scope) and "not representative" for schemes answering "regional" or "local". Moving on to e.g. H5 (Sampling design), one then can mark schemes with entries "stratified" as "representative" and those with entries "not stratified" as "not representative". Repeating these steps for each of the relevant columns (database fields), one should be able to use filtering to search for schemes that are marked as "representative" for each relevant question. In the example above, filtering for "representative" would yield the schemes that are international or national in scope and that use a stratified sampling design.

This approach would enable the users/coordinators to focus on the questions they feel relevant and/or to use their own ranking system, if they wish. For EuMon, such an approach would provide an opportunity to calculate average values (for quantifiable criteria) or ideal values (for qualitative criteria) and to quantify the proportion of schemes reaching an above-average rank (for quantitative criteria) or a satisfactory rank (for qualitative ones). Furthermore, there is an opportunity to reveal trade-offs (or synergies) between criteria in case we find negative (or positive) correlations. Finally, the relationships found can be repeatedly tested for all the relevant habitat types.

For the quantifiable indicators (scientific quality, Potential and Effort indicators), numeric values can be calculated that may characterise the feature one is interested in. To calculate these indicators, one could follow the formulae given at the respective feature (chapters 3 and 4) or could develop one's own formulae.

A synthesised approach for the evaluation of monitoring schemes needs to be hierarchical starting with the questions most relevant to the end-users and studying questions of increasingly less importance and larger detail in subsequent steps. Therefore, it is recommended that the following logic be applied on a smaller set of similar monitoring schemes.

1. The coherence between stated goals and data types should be established. This step is the first filter, as appropriate schemes are selected for further assessment, and poorly appropriate ones receive a lower score or are treated separately.
2. The coherence between EU goals (Natura 2000, 2010 target) and monitoring should be established. Here the ability of the schemes to quantify status and trend in the range and area of habitats of Community interest needs to be evaluated.
3. Build a composite indicator for scientific quality based on the following primary indicators: representativity, statistical power, and precision. Secondary indicators (habitat quality, requirement of scientific knowledge, and state-of-the-art methodology) should also be considered, but it is probably useful to treat primary and secondary indicators separately.
4. Calculate or estimate Potential indicators.
5. Calculate or estimate Effort indicators.
6. Build appropriate composite PpE measure by contrasting Potential with Effort.

A full evaluation therefore will involve evaluations of coherence, scientific quality, and time and cost-effectiveness for monitoring schemes.

However, monitoring schemes differ greatly in coherence, scientific quality, and Potential and Effort indicators and a joint evaluation based on detailed analysis is likely to lead to high variation in the indicator values and hard-to-explain relationships among schemes. Therefore, for any concrete evaluation the above-listed logic needs to be carried out for every scheme that has similar goals and focuses on similar types of habitats. If such sets of schemes are identified, schemes could be compared according to their achievement of their goal and their costs. This way many of the characteristics evaluated along the above logic would be the same (or at least similar) for the schemes under study, which would reduce the list of relevant indicators to a set of 3-4 indicators per scheme. An evaluation of the similar schemes based on 3 or 4 indicators should then be straightforward.

6. Applying the recommendations

6.1. Case study 1: All habitat monitoring schemes

Here we apply the given recommendations to the metadata on habitat monitoring collected in DaEuMon to test them in practice and to examine the possibility of DaEuMon for evaluations. In total, currently 146 such schemes (*database date 10.1.2007*) have been described in the database.

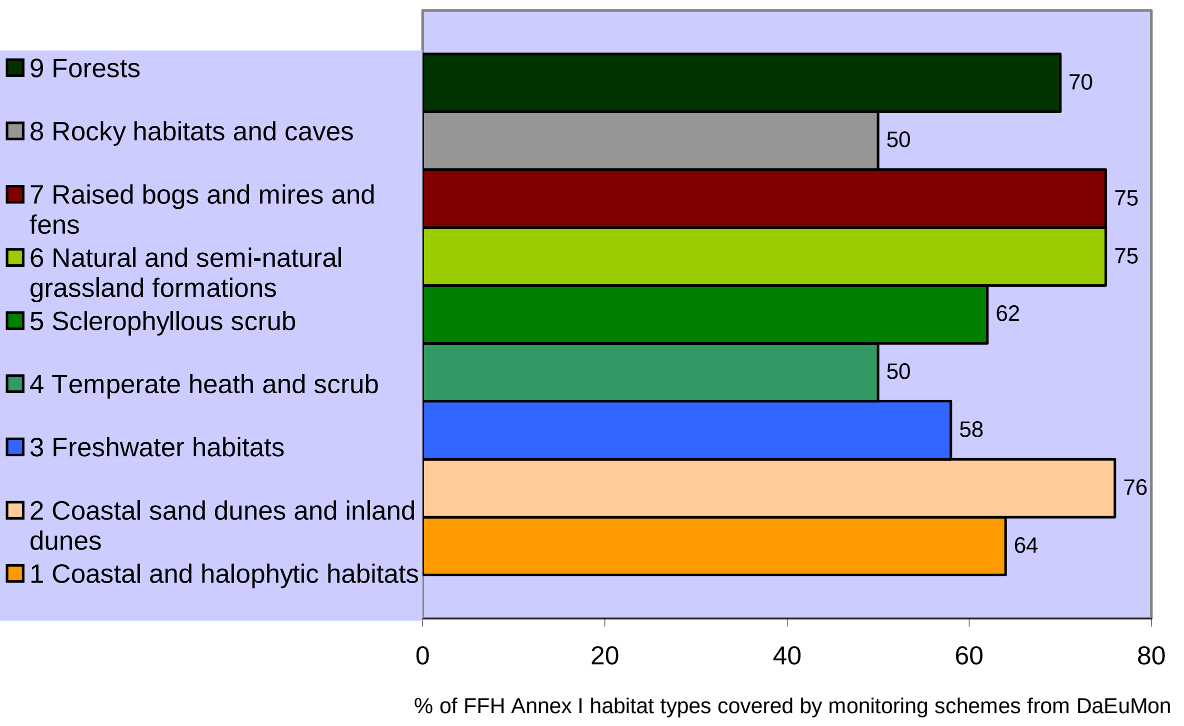
Coherence

Coherence between goals and data used in monitoring was calculated based on **Table 1** (questions H1 and H2 are relevant here). There was no answer to H2 in 10% of the schemes, so there is no chance to test the coherence of those schemes. 71% (103) of the schemes are appropriate because they monitor presence/absence data.

Coherence with EU Directives and 2010 expectations:

The coherence of schemes from DaEuMon with EU targets is scored using H22 (List of habitats monitored) in relation to the list of all habitats listed in Annex 1 of the Habitats Directive. There are 28 habitat types in group 1 (coastal and halophytic habitats). In DaEuMon 18 of them are covered by monitoring schemes, therefore coherence towards EU targets is 64% (**Figure 1**). Equally calculated the coherence for group 2 (coastal sand dunes and inland dunes) is 76%, group 3 (freshwater habitats) 58%, group 4 (temperate heath and scrub) 50%, group 5 (sclerophyllous scrub) 62%, group 6 (natural and semi-natural grassland formations) 75%, group 7 (raised bogs and mires and fens) 75%, group 8 (rocky habitats and caves) 50%, group 9 (forests) 70% (**Figure 1**).

Fig. 1: The coherence of habitat schemes from DaEuMon in relation to the habitat groups of the FFH Annex I



Criteria proposed for scientific quality

If a monitoring scheme is **launched** because of scientific interests, it may be better designed in a scientific sense and results may be easier to analyse (score given: 2 points). With respect to scientific quality, the next launching reason is management or restoration (we gave 1 point) and obligation (national law, EU law) has the lowest scientific value (0 points) in our understanding. Only 13% (19 out of the 146) of the habitat monitoring schemes were launched because of scientific interests, more than half of the schemes (54%) were launched because of different obligations, and 45 schemes (31%) were conducted because of management/restoration reasons.

The spatial **representativity rate** of schemes was calculated based on **Table 2** (questions 6, H4, H5, H7). In addition, H10 (actual area monitored) was also used in comparison to country area. Country areas were obtained from other sources (wikipedia) for this evaluation. Only 8 (6%) schemes were evaluated as representative on the national or international level, 64 (44%) schemes were not representative, because the monitored area was much smaller than the country area. Finally, half (73) of the schemes had no answer for H10. Therefore, H10 is a good example of the gaps in the database.
As for representativity tested according to **Table 2**, none of the schemes was representative in terms of all four criteria. This was mostly because of the many empty rows in the database; usually H4 was without answer. Interestingly, 35% (51) of the schemes were not representative in terms of any of the four criteria. Fourteen percent (20) of the schemes were representative in terms of three criteria, 15% (22) of the schemes were representative in terms of two criteria, and 32% (47) in terms of one criterion. There were 5 schemes, which we were unable to test, because one or more answers were missing from four questions.

Measurement precision can be estimated based on six questions: nature of data collected (H2), number of sampling sites (H11), number of samples per site (H12), annual frequency of sampling (H13), number of visits per site (H14), and number of years monitored until 2006 (H17). Precision was tested based on the instructions in section 3.1.4.1. The answers of H11 and H14 (number of sites and visits) are 0 in many of the schemes. In these cases, it was not known whether the value is a real zero or a N/A type of answer (e.g. for schemes based entirely on remote sensing, it is possible that H14 is 0). Unfortunately, most of the studies (122, which make up 84%) fell in the poor precision category. Of these 122, only 10 have more than 15 sampling sites, but in those schemes annual frequency of monitoring is less than once in four years (between 5 and 10 years). Only 6 schemes (4%) had a high precision score (with measurement error quantifiable), 7 schemes (5%) had “good precision but measurement error not considered”, and 10 (7%) schemes were scored as of medium precision.

Reliability judged by coordinator is given as the minimal annual change that can be statistically detected (H24). For 12% (18) of schemes the statistical power was supposedly low (insufficient), for 20% (29) of schemes the change was 10% or less, so the statistical power was high (sufficient), and for 68% (98) the data entry was missing (so the precision was difficult to estimate or it was not considered important by the coordinator).

Statistical power based on the experimental design used (H6). Schemes with an experimental design can be considered as monitoring schemes designed to detect specific temporal changes, due to given causes of change. So if the answer was ‘yes’ to H6, than we evaluated the scheme as appropriate, and evaluated it as inappropriate when the answer was ‘no’. Roughly half or 52% (76) of the schemes were appropriate based on the above-mentioned criterion and 48% (69) were inappropriate.

Habitat quality evaluation can be made based on questions H1, H3, H25, and H27-29. Imperfection of the questionnaire/database can be seen here again, because there were some schemes, where the answer for H27 ('habitat quality monitored') was no, but a list was given in H28 ('list of quality criteria monitored'). In such cases, there could be a distinction as "high" and "higher" for those that answered "yes" and provided answers to 'quality criteria' monitored. Based on the six questions, 33 (23%) of the schemes received as score 'higher' for all questions. No scheme received 'lower' for all questions. Most schemes received more 'higher' scores and fewer 'lower' scores. However, for six questions, many combinations are possible, which should be subject of further studies.

The proportion of professionals can be calculated from numbers given in answer to questions H30 and H31 as: H30/(H30+H31). Furthermore, monitoring schemes that require training or expert knowledge (H32) may have a higher scientific quality than those that do not require such measures. There were only 2 schemes, which require only volunteers (No. 368 and 648) and do not need any expert knowledge, and 3 schemes, in which only professionals are involved but expert knowledge is not needed. The latter appears strange, but also may be caused by some misunderstanding or mistake in data entry. It may also reflect that generally volunteers seem to be less involved in habitat monitoring. 28% (41) of the schemes include only professionals and most of them require also training, what may indicate the higher scientific quality of these schemes. Seventeen (12%) of the schemes needed professionals and volunteers as well and all of them required expert knowledge. Of the 17 schemes, in 8 more volunteers and in 7 more professionals were required (almost 50%), whereas in 2 schemes the required numbers were equal. Answers were not available for most of the schemes (85, which make up 59%). Only 28 (19%) schemes did not require any expert knowledge (or, maybe, the answers are just missing).

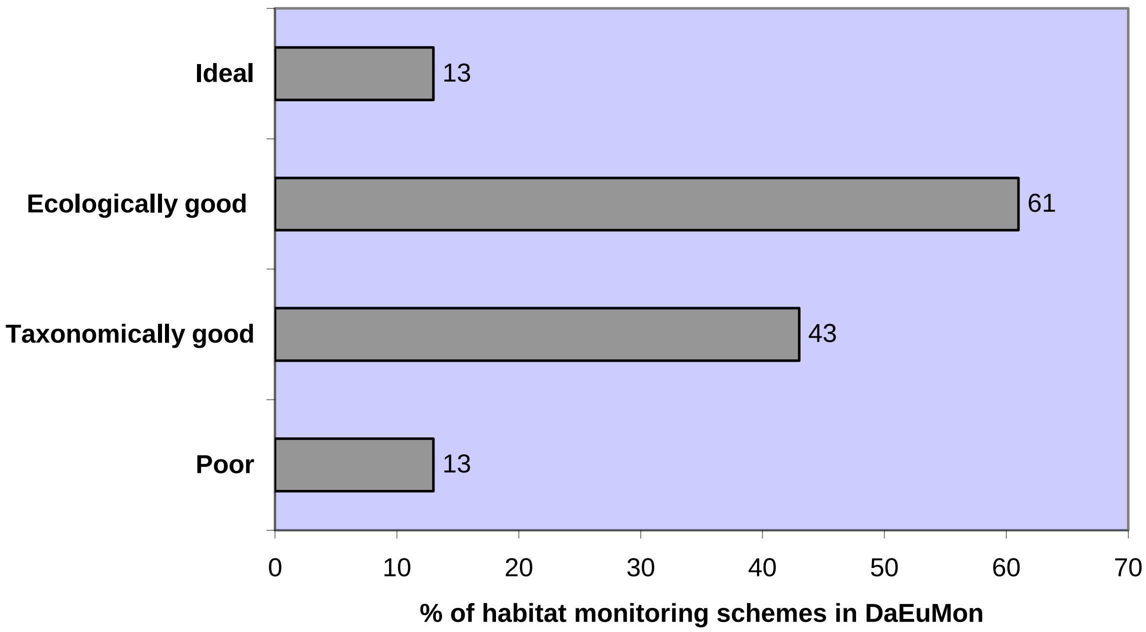
The scientific quality of scheme can be higher if field and statistical methods are more up-to-date (relevant questions here are H4, H19, and H32). Only 21 (14%) of the schemes use remote sensing methods. But there are plenty of schemes (57%) without any answer for H4. In 51% (74), it is not known what statistical analyses happen in the scheme (answers: 'other' and 'data are analysed by someone else'). Only 17% (24) use advanced statistics and the rest (32%) has poor statistical (and thus possibly poor scientific) value.

Criteria for time and cost-effectiveness

The important measure of the monitoring potential is the **proportion of habitat monitored** per total area of the target habitat. Since it requires the estimate of the total area covered by the habitat (which is not present in the database and we know no other easily accessible compilation available), it was not possible to calculate this measure for this case study.

When considering **the taxonomical and ecological extent of monitoring**, H20, H22 and H29 can be relevant questions. We used H20, H22 and H29 with a coding as follows: poor: H20=0 and H22=0 and H29=0/1, taxonomically good: H20=0/1 and H22=0/1 and H29>1, ecologically good: H20=0/1 and H22=1 and H29=0/1, ideal: H20=1 and H22>1 and H29>1. There were 21 (13%) schemes, which are estimated as ideal, 70 (43%) as taxonomically good, and 100 (61%) as ecologically good schemes (**Figure 2**). Finally, 21 (13%) schemes were scored as poor.

Figure 2: Taxonomical and ecological extent of monitoring
(No. of schemes = 164)



There are several points available to measure the **effort indicators** of schemes (such as time requirement and financial resources indicators). Total time requirement to run the monitoring scheme (calculated as H33/H13) ranges approximately from 0 to 5000 (!) (manpower in person days per year) and total time requirement for fieldwork per year (H11*H14*H15/H13) from 0 to 1 698 000 (!). There were no answers to H33 in 91 (63%) cases, so it is hard to estimate the time required for running the schemes. After comparing these two in the remaining 54 schemes, 30 (21%) of the schemes could be viewed as ‘complex’, because the total time needed for the entire monitoring is more than the time needed for fieldwork, so these schemes include other work than fieldwork as well. 23 (16%) of the schemes were estimated wrongly by the coordinators because the total time needed is less then the time needed for fieldwork. Only one scheme (No. 815) consists entirely of fieldwork.

Personnel costs as well as material/equipment costs are highly variable among monitoring schemes. If we take for example 40 € for an arbitrary daily salary, the schemes personnel costs (H33/H13*salary) range approximately between 0 and 200 000 €. Total costs ([H33/H14*salary]+H34) ranged between 0 and 525 520 €.

We have not made a synthesis of the measured data, so there is no chance to pick the ‘best’ scheme from the 146 monitoring schemes. However, there are several important lessons that can be learnt from this case study and testing. The most important is that the database in this format is of limited use for testing habitat monitoring schemes and for choosing best practices. There are several data missing from the database, and all the schemes should be checked before testing whether the missing data are really missing. Finally, a species monitoring scheme (372) from Hungary occurred in the habitat database table.

6.2. Case study 2: A preliminary evaluation of monitoring schemes for Natura2000 site designation in Natura 2000 sites from Greece, Spain, and the UK

Here we comment on problems and conclusions on the methods that can be used to evaluate the coherence, the scientific quality, and time and cost-effectiveness of monitoring of natural habitat types of Community interest, whose conservation requires the designation of special areas of conservation.

As regards habitat types of Community interest the establishment of monitoring schemes followed the implementation of Directive 92/43/EC and the designation of the Natura 2000 in Europe. In the first phase of the Natura 2000 12 European countries have participated and the lists of proposed sites of Community interest was completed at the end of 1995. So, the starting year of monitoring schemes is between 1994 and 2005 for Spain, 1994 and 2004 for Greece, and 1991 and 1998 for UK. In Cyprus up to the present only one project has been financed by the EU for the completion of the proposed sites for the Natura 2000. Almost all the monitoring projects examined are reported to have lasted 3 years.
For Spain, we found 30 main monitoring projects for habitat types and species of Community interest, which mainly stem from the database of LIFE projects (EU) and from personal contacts. In Greece, to our knowledge, there were no national-wide monitoring schemes for habitat types. We found about 30 projects mainly focusing on conservation and management of the proposed Natura 2000 sites (SCI) of Greece that included some actions on monitoring of habitats. Finally, for the UK we found 20 monitoring schemes well organized and funded mainly from EU and national sources. For Cyprus the only available schemes (3) are concerning the recognition and the listing of the habitat types in order to be included in the revised Natura 2000.

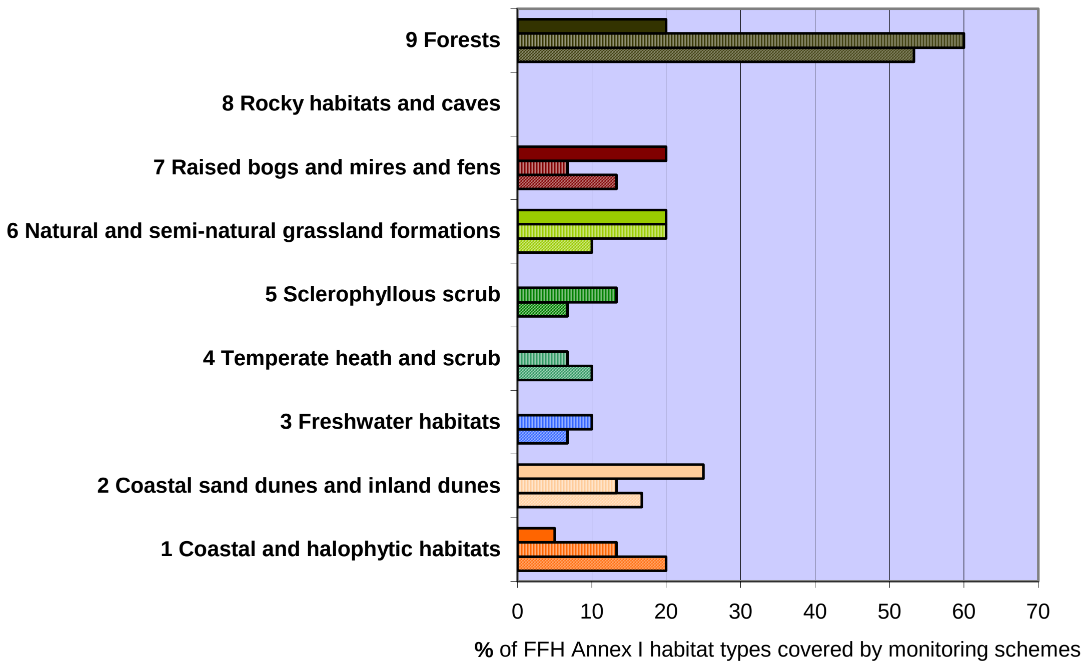
An example for coherence between habitat and species monitoring schemes is that schemes monitor not only the habitat type but also a target species. This may reflects in reduction of time and costs.

The total number of natural habitat types of Community interest whose conservation requires the designation of special areas of conservation is 218 (Directive 92/43/EEC; Treaty of accession 2003). Greece belongs to the Mediterranean region and of the 87 natural habitat types of Community interest, the 43 schemes entered in DaEuMon monitored 49%. Spain belongs not only to the Mediterranean region but also to the Atlantic and Alpine region. Of 99 natural habitat types of Community interest, the 42 schemes entered in DaEuMon monitored 42%. In the UK, 76 natural habitat types of Community interest that belong to the Atlantic region are registered. The 21 schemes entered in DaEuMon monitored 28%. The coherence of schemes from UK, Greece and Spain in relation to the list of all habitats listed in Annex 1 of the Habitats Directive is shown in **Figure 3**.

The majority of the habitat monitoring schemes examined was launched mainly under European Directives and management/restoration actions. Only one scheme in Spain was launched for scientific interests and concerns the recovery of the habitat of amphibians and the European pond turtle (*Emys orbicularis*).

The majority of schemes address the entire habitat. Twenty-nine (97%) of a total of 30 monitoring schemes in Spain, 26 (87%) of 30 projects in Greece and 18 (90%) of 20 projects that were mainly funded by the EU monitor the entire habitat concerned.

Figure 3: The coherence of habitat schemes from UK (■), Greece (III) and Spain (///) in relation to the habitat groups of the FFH Annex I



Of the total number of monitoring schemes for habitat types and species of Community interest, sampling design was stratified for 57% of the schemes and an experimental design was followed in 73% of the schemes in Spain. In Greece, the comparable numbers were only 3% and 16%. UK projects also frequently pursued a stratified sampling (75%) and used an experimental design (75%).

Unfortunately, data concerning the reliability of measurements have been lacking for most schemes. The same was observed also for the repeatability of measurements. Most measurements are taken only once a year, thus, repeatability was probably low, and as a consequence, the extent of error cannot be quantified or estimated. There was no information in the monitoring schemes examined on the percentage of change detectable in the monitored habitat.

References to the main causes that resulted in habitat degradation, however, have usually been given, such as extensive land use, pollution, fragmentation, climatic change, and invasive species. Almost all the projects attempted to uncover the main threats to the conservation of habitats.

The question on the number of professionals and volunteers participating in the monitoring schemes has rarely been answered. Furthermore, there were no references about the manpower or the total time that is necessary to run the monitoring of habitat types. These deficiencies may later cause problems when using the criteria that are based on this information. Only the total budget of the monitoring program was given.

The presence of professionals and volunteers in the monitoring of habitats and species of Community interest is very relevant and needs further discussion. Project managers should try to cooperate with professionals with the relevant expertise in order to evaluate the condition of each ecosystem and to design and implement in management and restoration actions afterwards. The EU could put an emphasis on enforcing the idea of involving professionals in areas and schemes concerning species and habitats of Community interest.

New technological developments, such as automation of analytical methods, new software for data analysis, spatial aspects (GIS), non-parametric methods, multivariate analysis, and modelling, are available for monitoring purposes. Habitat mapping is an expensive undertaking and using remote sensing to augment field survey is the most cost-effective means of achieving outputs for scientific and management purposes (Green & al. 2000). Four types of cost are encountered when undertaking remote sensing: (1) set-up costs, (2) field survey costs, (3) image acquisition costs, and (4) the time spent on analysis of field data and processing imagery. The largest of these are set-up costs, such as the acquisition of hardware and software, which may comprise 40–72% of the total cost of the project depending on specific objectives (Green & al. 2000). Thus, remote sensing is applied in a high percent of the schemes (50%) only in the UK, whereas another 10% used field mapping in the UK. In Greece, 7% of the schemes apply new technologies (remote sensing), other schemes are without answer to that question. In Spain 3% of the schemes use field-mapping and other schemes are without answer to that question.

The majority of the monitoring data were not analysed, probably due to lack of data or due to the fact that a number of projects have not been finished yet. Percentages range from 20 to 33%. Only 3% of the schemes analyse data by advanced statistics, and a large number of projects did not report usage of statistical analysis at all.

The main habitat classification scheme used was CORINE. In Spain, it was used in about 73%, in Greece in about 63%, and in UK in about 56% of the habitat monitoring schemes. These high percentages may be advantageous for the possible integration of the monitoring schemes, which requires that data are in similar format to make inferences common to all schemes involved. This fact, thus, should facilitate the development of a EU-wide common framework for habitat monitoring. The question that remains is whether each country interprets CORINE habitats in the same way and whether CORINE is detailed enough. Although these are issues outside the scope of the EuMon project, this needs be taken into consideration in the integration part of the project (WP5).

There was no information available about the exact total area of the habitat that was monitored. However, it may be reasonable to assume that in each monitoring program the proportion of the habitat that is monitored cover the total area of the habitat. Each monitoring scheme, however, mentioned the number of habitats monitored and the number of indicator species. In Spain, 42 habitat types, in Greece 42, and in the UK 21 habitat types were monitored.

7. REFERENCES

Anderson, J. R., Hardy, E. E., Roach, J. T. & Witmer, R. E. 1976. A Land Use And Land Cover Classification System For Use With Remote Sensor Data. US Geological Survey Professional Paper 964, United States Department of the Interior.

Baker, W. L. & Cai, Y. 1992. The r.le programs for multiscale analysis of landscape structure using the GRASS geographical information system. Landscape Ecology 7: 291-302.

Bossard, K., Feranec, J. & Otahel, J. 2000. CORINE Land Cover Technical guide – Addendum 2000. European Environment Agency.

Congalton, R. G. & Green, K. 1999. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. CRP Press.

Eastman, J. R. 2003. Idrisi Kilimanjaro Guide to GIS and Image Processing. Clark Labs, Clark University.

Forman, R. T. T. 1995. Land Mosaics: The Ecology of Landscapes and Regions. Cambridge University Press.

Green, E. P., Mumby, P.J., Edwards, A. J., Clark, C.D. 2000. Remote Sensing Handbook for Tropical Coastal Management (extracts). Unesco.

Johnson, A. R., Milne, B. T., Wiens, J. A. & Crist, T. O. 1992. Animal movements and population dynamics in heterogeneous landscapes. Landscape Ecology 7:63-75.

Lillesand, T. M. & Kiefer, R. W. 1994. Remote sensing and image interpretation. John Wiley & Sons.

McGarigal, K. & Marks, B. J. 1995. FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. Gen. Tech. Report PNW-GTR-351, USDA Forest Service, Pacific Northwest Research Station, Portland.

O'Neill, R.V., Krummel, J. R., Gardner, R. H., Sugihara, G., Jackson, B., DeAngelis, D. L., Milne, B. T., Turner, M.G., Zygmunt, B., Christensen, S.W., Dale, V.H., and Graham, R.L. 1988. Indices of landscape pattern. Landscape Ecology 1:153-162.

Turner, M. G. 1990. Spatial and temporal analysis of landscape patterns. Landscape Ecology 4:21-30.

Turner, M. G., & Gardner, R. H. (editors). 1991. Quantitative Methods in Landscape Ecology. Springer-Verlag, New York.