EXTERNAL SCIENTIFIC REPORT

Review of statistical methods and data requirements to support post market environmental monitoring of agro ecosystems[1]

Peter A. Henrys[1], Jill Thompson[1], Greet Smets[2], Patrick Rudelshiem[2], Rosie Hails[1], Stephen Freeman[1], Debora C.M. Glandorf[3], Ron Smith[1]

Centre for Ecology and Hydrology[1], Perseus[2], Rijksinstituut voor Volksgezondheid en Milieu[3]

SUMMARY

The EFSA "Science strategy 2012-2016" identified Post Market Monitoring (PMEM) as an area requiring further methodological development in order to support the evaluation of regulated products. Post market monitoring aims to identify changes and trends in ecological indicators as a consequence of an unexpected change in the environment. This could be, for example, a result of the release of an authorised product which showed no cause for concern in an environmental risk assessment. The nature of PMEM and the unexpected influence of environmental factors, imply that no new monitoring is put in place in order to track an effect. Therefore, to analyse the influence of any individual factor, one must rely on Existing Surveillance Networks (ESNs) and the data they collect. It has been suggested that these networks have the potential to evaluate the impact of regulated products, identifying unanticipated adverse effects on the environment. It is possible, with a focus on appropriate indicators and with specific analyses, to explore the capabilities and the limitations of ESNs and their potential power to detect change. For this information to provide useful guidance and recommendations for PMEM, both the statistical methodologies and the existing data sources need to be identified and reviewed.

To fully review the statistical methods and data requirements needed to support PMEM of agro ecosystems, we:

- Performed a literature review of appropriate statistical methodology;
- Collated an inventory of statistical approaches;
- Created an inventory of environmental data sources;

[1] Question No EFSA-Q-2012-00597.
Any enquiries related to this output should be addressed to amu@efsa.europa.eu

Suggested citation: Centre for Ecology and Hydrology, Perseus, Rijksinstituut voor Volksgezondheid en Milieu; Review of statistical methods and data requirements to support post market environmental monitoring of agro ecosystems. Supporting Publications 2014:EN-582. [151 pp.].

Available online: www.efsa.europa.eu/publications

- Assessed the power that available data sets have to detect any potential effect; and
- Provided conclusions and recommendations based on our findings.

A literature review of statistical methodologies used in the field of agro ecosystems was conducted and results saved, though this often provided search criteria that were too wide and generic to establish sufficient detail on appropriate statistical methodologies. Therefore an assessment and review of methods was made using expert knowledge and this was externally reviewed by a group of international statisticians. This review, together with all references to the relevant statistical literature, formed the basis for the statistical methods considered in work that followed and the basis of the statistical inventory.

These selected methods were then each assessed for their use within the field of agro ecosystem science and for the comparative power to detect changes in a specific indicator over time. This was conducted in two stages. Firstly, a literature review to establish the sources and applications of each statistical method and then secondly a simulation study to look at the power of each method to detect change under differing scenarios. The literature review found that the most referenced statistical method in the field was generalised linear models, whereas more niche and bespoke methods were rarely cited in the relevant literature. The power analysis showed that under differing scenarios, different methods perform better, for example non-parametric methods perform better when data is sparse. Guidance is provided in this report as to when to use certain methods.

The inventory of environmental data sources was completed using an online web based tool and a separate excel spreadsheet of this inventory has been provided to EFSA. The inventory contains detailed information covering indicators across a whole range of categories from human health to soil function. The detail collected in the inventory on differing aspects of ESNs enable us to define and search for networks that meet certain criteria. Idealised criteria were defined that demanded a network to have:

- European-wide data
- Multisite, even distribution of sites
- Observations at least annually
- Standard protocol
- Protocol well documented
- Trained surveyors
- Validated data
- Analysis method well documented
- Access to raw data

Querying the inventory showed that there were a number of ESNs that met this criteria and hence would provide sound data for use in PMEM. Over 500 networks have been entered into the inventory covering 27 different countries across the EU.

A detailed power analysis showed that there are existing ESNs that, under certain scenarios, demonstrate high statistical power to detect a “treatment” effect over time. Generic equations are presented that allow for simple and quick estimation of power given a particular scenario and the details of a specific indicator as measured by an ESN. These equations, coupled with the details from the particular ESN as contained in the inventory, enable quick assessment of the suitability of any ESN in PMEM.

In light of the evidence presented, recommendations for the future of PMEM were made. These included: highlighting the importance of surveyor training / quality control of field surveys; the need to link results from statistical analyses to more experimental / causality based work; the importance of keeping the ESN inventory up to date; and clear guidance on schemes that are useful and have

sufficient power for PMEM. Many of the recommendations presented relate to the need for increased standardization of data collection, availability of documentation, collaboration across networks, supporting established networks, developing new networks and keeping track of all networks.

Ultimately, we conclude that there exists data collected by environmental surveillance networks that have great potential to increase our investigations into PMEM to assess the unforeseen consequences of some change in the agro ecosystem on the environment and, coupled with additional investigation and experiments, to potentially enhance our ability to avoid such consequences.

KEY WORDS

Statistical Methods, Trend detection, Environmental Surveillance Network (ESN), Environmental Surveillance Programme (the ESP), Post-Market Environmental Monitoring (PMEM), Protection goal, Influencing factor, Inventory, General Surveillance, Genetically Modified Organisms (GMO), Post Market Monitoring,

TABLE OF CONTENTS

BACKGROUND AS PROVIDED BY EFSA

EFSA "Science strategy 2012-2016" has identified Post Market Monitoring (PMEM) as an area requiring further methodological development in order to support the evaluation of regulated products. PMEM aims at identifying possible unanticipated adverse effects on the environment which were not anticipated in the environmental risk assessment prior to the authorisation of products for use within the agro ecosystem. Environmental monitoring networks have been established at European, national and local level in order to measure biodiversity and other key indicators related to environmental protection goals and ecosystem services, therefore it is recommended to make use of existing monitoring data for PMEM. Most importantly a holistic and integrative approach for PMEM should be developed which considers PMEM within a framework of general environmental protection monitoring within the EU.

Environmental monitoring is included in a number of EU directives and regulations:

• Member State reports to the European Environment Agency (EEA) under e.g. the Council Directive 92/43/EEC on the conservation of natural habitats and of wild fauna and flora (Habitats Directive) and the Council Directive 2009/147/EC on the conservation of wild birds (Birds Directive) which are combined under the Natura 2000 framework; Council Directive 2000/60/EC establishing a framework for Community action in the field of water policy (Water Directive). These data, and other environmental datasets, are collated using the EEA ReportNet framework and made publicly accessible via the Biodiversity information system for Europe (BISE) platform.

• Bee health surveillance data collected under Commission Regulation No 87/2011 designating the EU reference laboratory for bee health, laying down additional responsibilities and tasks for that laboratory.

• Data collected to support the assessment of pesticide sustainability under Regulation 1185/2009/EC concerning statistics on pesticides and Directive 2009/128/EC establishing a framework for Community action to achieve the sustainable use of pesticides.

• Data on plant pathogens and pests collected under the Directive 2000/29/EC on protective measures against the introduction into the Community of organisms harmful to plants or plant products and against their spread within the Community

Agricultural systems display considerable dynamics in time and space, change in the environment is natural and variation due to natural effects may be substantial. For monitoring networks it is important to define the size of biological change that is biologically relevant prior to commencing measurement activities to ensure the survey design has the power to detect such changes (Statistical Significance and Biological Relevance). Plus for ecological and environmental monitoring determining whether an observed change exceeds existing variability can be difficult especially as multiple stressors are present in agro ecosystems. For this reason sophisticated analysis methods would be required to support PMEM. Comparative methodologies such as Before – After Control- Impact (BACI) design, multi-criteria and Observational and Simulated Evidence approaches have been recommended to support PMEM. However the success of any statistical approach is dependent on the availability and quality of the data used for the analysis. The temporal and spatial frequencies of the monitoring observations are a critical component of data quality assessments.

TERMS OF REFERENCE AS PROVIDED BY EFSA

This contract seeks to investigate whether data obtained from existing monitoring networks and programmes can effectively contribute to PMEM of new and existing agricultural products authorised for use within Europe.

The specific objectives of the contract resulting from the present procurement procedure are:

• Review of published statistical methods used in the analysis of ecological and environmental datasets to

i. determine whether observed change exceeds existing variability

ii. investigate spatial correlation with environmental stressors

• Inventory of statistical approaches in ecological and environmental monitoring and identification of data requirements for the items in the inventory

• Inventory of European, National and Regional environmental monitoring networks assessed according to a criteria defined by the analysis methodologies data and survey design requirements and data quality standards

• Recommendations of the most appropriate analysis methodologies for PMEM based on available environmental monitoring data in Europe (EU 27 + Norway and Iceland).

This contract/grant was awarded by EFSA to:

Contractor/Beneficiary: Centre for Ecology and Hydrology

Contract/grant title: Review of statistical methods and data requirements to support post market environmental monitoring of agro ecosystems

Contract/grant number: OC/EFSA/SAS/2012/02

INTRODUCTION

Europe has well developed Environmental Surveillance Networks (ESNs), many of which have extensive coverage and long term data sets. These networks are used to report on the status of the environment. It has been suggested that these networks also have the potential to evaluate the impact of regulated products, identifying unanticipated adverse effects on the environment. One example of this is the potential of ESNs to contribute to the General Surveillance (GS) of genetically modified (GM) organisms: particularly GM plants released into agro ecosystems. The aim of GS is to identify any adverse effects that are unanticipated after a rigorous environmental risk assessment, so there is no specific hypothesis linking the intervention (e.g. GM plant) to the measurement endpoint (e.g. abundance of a species of cultural or conservation significance). Consequently it is logical to select measurement endpoints that are good indicators of agro ecosystem health (or harm), and such endpoints will therefore have broad applicability in detecting adverse impacts caused by a range of interventions, including changes in agricultural practice, agro chemicals, the uptake of other novel crops etc. In conclusion, the outputs from ESNs could have broad and significant policy-making influence with respect to evaluate the post market environment impact of regulated products.

However, it is important to select networks, and measurement endpoints within networks, that will genuinely inform the observer, against the background of environmental variability and other external pressures and drivers. It is now well-established in the scientific peer-reviewed literature that certain changes in agriculture have adversely affected biodiversity, natural resources and some ecosystem services in lowland agricultural habitats to varying degrees across Europe over the last 50 years. The evidence supporting this has come in part from existing ESNs, supplemented with more detailed experiments and observations to link cause and effect. In the absence of more detailed studies, it can be more difficult to determine the impact of individual drivers as many operate synchronously, and for this reason, ESNs have not been used routinely to investigate relationships between cause and effect. Despite these challenges, it is possible, with a focus on appropriate indicators, and with specific analyses (e.g. Before – After, Control – Impact or BACI design, multi-criteria, paired comparisons and other analyses) to explore the capabilities and the limitations of ESNs and their power to detect change.

Two important elements of such a study should be a) the size of change that would be considered biologically relevant; and b) a consideration of the power of the proposed analysis to detect change. The first is a judgment that may be based on a degree of expert opinion, but which may be informed by knowledge of the proximity of a resource or species to defined limits or thresholds. Once the magnitude (or range) of change that it is necessary to detect has been determined, then a power analysis can be conducted for each measurement endpoint, network and analysis method. This will allow the probability of detecting that observed change to be determined. In many, if not most, cases, the analysis of the influence of a new agricultural intervention upon a specific measurement endpoint will conclude that there is no statistically significant effect. This could mean one of two things: that there is no influence of that driver on that endpoint; or that the driver does influence the endpoint but the analysis has failed to detect this. In the absence of a power analysis it is difficult to distinguish between these two possible conclusions.

To fully review the statistical methods and data requirements needed to support PMEM of agro ecosystems, we: performed a literature review of statistical methodology; collated an inventory of statistical approaches; created an inventory of environmental data sources; assessed the power that available data sets have to detect any potential effect; and provided conclusions and recommendations based on our findings.

1. Literature Review of Statistical Methods with potential for PMEM

1.1. Introduction and Objectives

The first objective was to assemble a database of literature incorporating statistical methods utilized in ecosystem monitoring of organisms and the environment, which could be used to inform policy makers, regulators and scientists, on the post market impact of EFSA regulated products and other substances. These products include, for example; genetically modified organisms, herbicides, pesticides, fertilizers and feed additives. The aim was that the statistical methods identified were subsequently investigated to assess their statistical power and usefulness for conducting post-market monitoring of product impacts on agro ecosystems. This follows in Section 2.

It has been proposed that Environmental Surveillance Networks (ESNs) could be utilized to assess the impact of products released into agro ecosystems. ESNs fit into two broad categories; the more systematic networks, where locations of assessment sites follow a predetermined geographical strategy and are more likely to be regularly assessed by trained assessors; and non-systematic networks, where sites are selected by individual Citizen Scientists, enthusiasts and volunteer groups depending upon their interests and where they live and are less likely to be regularly assessed (see Section 3 for a more detailed discussion of the networks). There are relatively few systematic networks, such as the UK Countryside Survey (Carey et al. 2008), as these are expensive to set up and maintain but they are also more likely to include abiotic measurements (e.g. weather and physical and chemical characteristics). The non-systematic networks, such as those covered by Citizen Science activities (Roy et al. 2012), may be widely distributed but the results may be biased by geographical region, habitat quality or species abundance as the sites are sometimes selected by individual preferences (Freeman et al 2007). The impact of these different types of networks and their relative sensitivity and utility for assessing the impact of products and policies on organisms and the environment is an important consideration when assessing the power of statistical techniques for PMEM.

To find potentially useful literature describing statistical methods, we conducted a systematic literature search in Web of Knowledge using a number of search terms suggested by people active in environmental monitoring. It became clear that (a) the search terms suggested for the project were returning an unrealistically large number of references that could not be fully assessed, and (b) sample assessments of the resultant publications found showed that the systematic literature review approach was identifying very few papers with any useful information on statistical analyses. We have described the systematic literature search in Section 1.2 of this report and provided the results of the searches in appendices and as an EndNote library for reference. In light of complications resulting from a systematic literature review, we then re-focused our efforts and identified the range of statistical techniques that would be the most appropriate to use for the types of analyses that would be useful for PMEM (see Section 1.4). These techniques are generally from the standard modern statistical toolbox used across a wide range of applied disciplines. We have supplied supporting references and a selection of example publications in which these techniques were applied. The section of this report that involved the statistical review was circulated to a number of international statisticians who provided an independent review of the document as well as additional material that has been incorporated into our report. We believe this study does give a robust assessment of current statistical practice, albeit at a summary level, which, alongside the literature reviews conducted, allowed the project to achieve its objectives of providing advice on statistical methods that could be used.

1.2. Materials and Methods for the systematic literature search for statistical techniques

1.2.1. Scoping the review using expert knowledge

We sent email requests to 45 experts in 10 EU countries who were interested in post-market monitoring or are involved in Citizen Science studies and other ESN activities. Requests were sent to people involved in post market monitoring and who had attended EU working meetings of Member States linked to GMO cultivation, or had published agro environment assessments. Other experts contacted are involved in the establishment of ESNs and the analysis of Citizen Science networks and data collected by them. We asked these experts for contributions on three aspects we believed useful for this study:

- To suggest key words that might be useful for a systematic literature review into statistical methods useful for post-market monitoring of the environment.
- Suggest publications that would contribute to the project, which would help identify key words for a systematic literature review and provide examples of statistical methods that would be useful for PMEM.
- Identify other experts that might contribute key words for the literature review.

The suggestions for literature search terms made by experts are demonstrated below (Section 1.2.2.3) in the details of the systematic literature reviews.

1.2.2. Systematic literature reviews

A systematic literature review to find appropriate statistical techniques for PMEM was conducted according to the methods incorporating the Cochrane protocol established in Higgins and Green (online version 2011 www.cochrane-handbook.org; EFSA 2010). A systematic literature review involves the following aspects:

a) Defining the research question
b) Establishing the information sources
c) Identification of appropriate search terms
d) Criteria for inclusion or exclusion or search results
e) Management of the references and database of results
f) Documenting and reporting the search

In addition, a systematic review should be accessible, repeatable and reproducible. O'Conner et al. (2012) advised that depending upon the volume of evidence a systematic literature review may take on average 1 year, but they also note that Khangura et al., (2012) suggested that 6 months to 2 years research is required.

1.2.2.1. Defining the research question for the systematic literature review of statistical methods

The main objective was to identify literature that describes the statistical methods used to assess data collected through Environmental Surveillance Networks (ESNs) and which might be useful to investigate the impact of products such as genetically modified organisms and agro chemicals on organisms and ecosystems. The research question aims to conform to the requirements of a systematic review see (http://www.environmentalevidence.org/Authors.htm). The definition of terms in our research question are in Table 1.

Our specific research question for the systematic literature review was: What statistical methods and types of data are used by environmental surveillance networks to look for evidence of change in organisms and ecosystems?

Table 1: Definitions of terms in the literature search objectives and research question:

Term in objective	Definition of objective term

Statistical methods	Any analysis that provides a statistical test of a change in a data series
Data	Qualitative or quantitative measures of population, growth, survival, abundance, density
Monitoring	Repeated measurements taken over a space and time and that record data which might be statistically analysed
Change	Increase, decrease, decline, reduction, expansion, variation, difference, trend
Organism	any living plant or animal
Ecosystems	environmental type including terrestrial, freshwater and salt water e.g. forest, river stream, marine

1.2.2.2. Establishing the information sources

We assessed the availability of literature for this study by performing preliminary searches to test search terms. O'Conner et al. (2012) noted that many systematic reviews miss out the gray literature and efforts to include these should be incorporated. Much of the gray literature and output from Citizen Science activities and ESNs are provided in local newsletters and annual reports that are only available to members. The gray literature is often rapidly out of print or not downloadable from the web site (if there is one to support the network). As Citizen Science develops into a totally web based activity then we anticipate that more of the gray literature will become publically available. Some websites encountered such as SORA the "Searchable Ornithological Research Archive" based at the University of New Mexico in the USA http://sora.unm.edu/ are now providing "gray" and old ornithological literature with pdf's of papers online. These facilities, however, are not yet widely available for other networks or organisms. We also investigated http://www.opengrey.eu/ a system for information on grey literature in Europe. This supplied titles and location information such as libraries that could be visited to view documents, publishers, places where documents could be purchased or ordered from the British library, but no useful publications describing statistical techniques were found freely available. Many of the titles of articles in the "gray literature" could be manually included in our reference database, but given their intended audience these publications would be unlikely to have detailed statistical information. Google also finds many papers, but the search results are not repeatable as the results change as Google learns the operator's preferences. Millions of results are obtained for every search attempt with Google, so we did not use Google for our systematic search.

We determined that Web of Knowledge (subsequently referred to as WoK; http://wok.mimas.ac.uk/), which incorporates a full range of environmental, agricultural, and mathematical literature and produces repeatable search results, would be the most appropriate reference database for a systematic literature search. WoK restricts searches to those using the English language.

1.2.2.3. Identification of appropriate search terms.

The search terms supplied by the experts we contacted included very broad terms, such as butterflies, reptiles, birds, habitats, biological indicators, and many names of varieties of genetically modified organisms. The ESN web sites and literature demonstrated that a very wide range of search terms would be required to effectively cover the range of organisms and habitats studied by the networks and any changes to these that might be statistically analysable. The search terms needed would effectively cover the whole range of ecology and environmental studies. We also determined that we could not use specific statistical technique or method names as search terms, as this would only have given us the statistics that we expected to find and statistical techniques are rarely named in the WoK searchable parts of the publications including title, abstracts or key words.

Our initial attempts at a systematic search based upon key words from experts and the literature they referenced (see Section 1.2.1) suggested three search targets that could be used as topic strings in WoK. These would identify literature that included results from Environmental Surveillance Networks, those that may indicate a change in species abundance, population distribution, growth or reproductive activity, and which may be useful in post market monitoring of agro ecosystems. We divided the search topics into three categories: 1) Ecosystem type (e.g. agriculture, agri-environment, grassland, forest etc.); 2) organism type (e.g. birds, butterflies, reptiles etc.); and 3) data type that might indicate statistical analysis had been conducted, or that a change had been measured during environmental surveillance (e.g. abundance, monitoring, response, mortality, increase, decrease trend etc). We also planned to include literature searches on topics such as water quality, soils and plant health but these were not attempted before we decided to refocus our efforts away from a WoK systematic search. We included habitat types to target the searches towards environmental research and to reduce the number of topics such as taxonomy and genetics that were judged not useful for this study. Comprehensive information on research terms tested for the systematic literature reviews are described further below.

To accommodate a range of formats of search terms in the WoK we tested combinations of formats for individual terms including “wild cards”. Examples for agri-environment and saltmarsh in different combinations with wild cards and the use of quotation marks are given in Table 2. The numbers of references extracted for agri-environmental ranged from 1-1134 and for saltmarsh 601 to 9744 (Table 2). These results depend upon how WoK recorded these. For organisms such as birds we added wild cards*, so that bird would include songbird and birds. Terms were also truncated to accommodate English spelling so that butterfly was covered by butterf* to encompass butterfly and butterflies. The word and wild card combinations used for the initial searches are given in Appendix 1.

Table 2: Example results after testing wild card symbol * combinations for key word searches of the literature in Web of Knowledge

Term and wild card combination	Description	Number of references
“Agri*-*environment”	Wild card for connection between hyphen and words and term with quotation marks	1134
Agri*-*environment	May have a space before or after hyphen or both	1134
Agri * environment	with wild card for space or hyphen	1024
Agri – environment	Agri hyphen environment with spaces between words and hyphen	1024
“agri-environment” returns	no spaces around hyphen and in quotes	783
Agri–environment	no spaces around hyphen	768
Agri*environment	Wild card with no spaces between words and wild card	9
“agri * environment”	With wild card between agri and environment and term within quotes	1 record = agrisystem environment
Salt * marsh	Space either side of wildcard for hyphen	9744 (also results in marsh on its own)
Salt*-*marsh	Space on either side of hyphen	7700
Salt*marsh	No space on either side of wildcard for hyphen	602
Term and wild card combination	**Description**	**Number of references**
Saltmarsh	One word	601

1.2.2.4. Criteria for selection using search filters and exclusion search terms.

The agreed search period was 2002 to 2012. In addition to the search terms described above, we refined WoK searches by "Research Domains" such as Science and Technology and specific "Research Areas" such as biology or plant sciences or forestry or mathematics. Further details are described below, but in general we found that:

- refining to Research Domain "Science and Technology" reduced the number of medical and social orientated references obtained.
- refining to Research Area within the Research Domain "Science and Technology" in WoK resulted in different filter list of Research Areas to select from that depended upon the search terms used so that *bird* may produce a different set of research area filters to select than fish* which makes these searches inconsistent. Most searches reported here were conducted without "Research Area filters" unless otherwise indicated in Appendix 3.
- references that concerned DNA, taxonomy or genetics were excluded using a NOT term in the search as these would be unlikely to contribute to ESN data.

1.2.2.5. Management of the references and database of results

References were primarily managed using EndNote 5X software which has sufficient capacity for holding the large reference libraries. The libraries are stored on secure CEH networks with automatic backup together with their associated database of publications and copies of the reference libraries have been delivered as a key deliverable alongside this report.

The searches using the criteria described above from WoK (in English) were saved in separate EndNote libraries (exported 500 references at a time). EndNote was then used to search for and automatically upload documents (searches for pdfs of publications can only be done for 250 references at a time). Documents not found automatically by the EndNote search tool that looked as if they might be especially useful were looked for individually and if available manually uploaded to the EndNote library. All available WoK information including number of citations was stored. Key words, numbers of citations and URL's for the location of the publication location were also included in the EndNote library if provided through WoK.

- an EndNote library has been created with separate groups within it, to cover different search selections and stages in the reference selection process.
- custom groups in EndNote are used to demonstrate different searches.
- automatic removal of duplicates that sometimes occurs when WoK uses different databases was achieved using EndNote tools.
- books and book chapters produced by WoK search results were not individually looked for.
- some references that from title and abstracts we believed were unlikely to describe statistical methods. (e.g. notes on individual occurrences of organisms, distribution maps without analysis etc) were rejected.
- some references where we were confident from title and abstract that were not about an ecological or environmental topic were also rejected.

For non systematic searches, where we obtained references from web sites or annual reports and newsletters, the references were either input into EndNote manually, or searched for in WoK and the reference information exported to EndNote. For manually entered papers some of the reference information may not be available.

1.2.2.6. Documenting and reporting the literature searches

This report describes the literature searches that produced the final database of literature references obtained by the project that have been provided as an EndNote library. The EndNote library is divided into separate groups for the different searches that are summarized in Table 3, with further details about the individual searches that produced the libraries in Appendices 1-3.

1.3. Information on the literature searches conducted to prepare the reference library for the statistical methods

We conducted a systematic search in WoK in an effort to produce a library of references that should contain useful publications with information on statistical methods, and that utilized the type of information collected by ESNs. These initial WoK searches used key word search terms suggested by experts and revealed in ESN literature (see details below and Appendix 3). We followed this initial attempt with a series of systematic searches to refine the reference list results to manageable proportions and to remove irrelevant publications. We also conducted non systematic searches for documents displayed on ESN web sites (see details below) and compiled references suggested by project members, colleagues and experts.

The literature searches conducted are here divided into five types of which three were considered systematic searches using WoK, and two that were considered non-systematic. These five literature search types are:

- systematic searches in WoK using key words suggested by project members and other external experts and selected from ESN literature (see 1.3.1)
- systematic but restricted searches in WoK using organism type and very few keywords that might indicate statistical methods are described and a change was detected (see 1.3.2).
- systematic searches in WoK using the name of the network as the search term (see 1.3.3)
- non systematic searches for literature that were advertised on, or could be obtained from the network web sites or by seeking publications referred to on the specific ESN web site (1.3.4)
- non systematic searches for publications that were suggested by project members, colleagues and experts that had been contacted (1.3.5)

1.3.1. Systematic searches of WoK using key word search terms suggested by experts and from initial checks of ESN publications

We tested a range of systematic searches using search terms obtained from experts and ESN publications and describe five examples here:

- initial searches using suggestions from experts and from ESN publications (Appendix 2) produced over 5.5 million references (Search terms in Appendix 2). We restricted the search to the Research Domain "Science and Technology" and divided the search into three topics to ensure combinations of organism types, AND ecosystem types AND terms for data collection and analysis of change. Although this result undoubtedly included some duplicates, there were too many references to export (500 at a time) from WoK into EndNote to find the duplicates.
- using the same searches but filtering the Research Domain, Science and Technology, and one Research Area i.e. Environmental Sciences and Ecology, gave 1,098,303 references. Using this Research Area filter may also have missed statistical computational and mathematical references. As an example of tests the effect of filters with the term environm* monitoring

produced 60,100 refs, filtering by Science and Technology 59,864, filtering by both Science and Technology and then by Environmental Sciences and Ecology gave 46,144. Adding another 24 research areas filters in biology and mathematics did not reduce the number of records below 46,144 references.

- using the same search terms as in a) and selecting European countries returned 317,699 references, but it was not clear how the country was being searched. As much literature that involves European research sites is not prepared only by European authors we decide that this type of search was not appropriate (See example Appendix 2).

- further searches using greater numbers of Research Area filters (Appendix 2) also returned huge numbers of references that were not useful. As noted earlier the type of filters that were allowed in WoK change according to the search terms.

- many other searches were tested to exclude a variety of literature that appeared not relevant but still produced 2 to 3 million references. As an example we attempted to exclude terminology that covered people, medicine and veterinary science with a "NOT" exclusion in the WoK search Topics.

NOT Topic=((clinic* or medic* or veterinar* or dna* or immunology or "stem cell" or cellular or genetic or genes or human or teenager or baby or babies or people or "human*behaviour" or "human*behaviour" or pig* or cow* or sheep or lamb* or calf or calves or chicken* or enzyme* or hen*))

As a result of these investigations that demonstrated that a systematic literature search would not be useful we conducted very limited searches with few key words to sample the literature and then adopted a more directed approach by searching sample ESNs themselves on the internet.

1.3.2. Systematic WoK literature searches conducted using restricted search terms including organism types and limited statistical terms

We conducted further searches by organism type and very restricted key words that suggested a change in the population or environment and including the terms 'statistic' and 'meta-analyses' as they might specifically indicate that a statistical test was used (see examples below). These systematic searches gave few references that could be related to ESNs and we know from our own experience that useful references are missing. The following are examples of restricted searches:

- Topic=((bird*)) AND Topic=((ecology or environment* or ecosystem*)) AND Topic=((increas* or decreas* or variation* or difference* or trend* or reduc* or expan* or decline*)) AND Topic=(statistic*)
  This gave 5,464 references with 64 added if 'meta-analysis' was included in addition to 'statistic'.

- Topic=((butterfly or butterflies)) AND Topic=((ecology or environment* or ecosystem*)) AND Topic=((increas* or decreas* or variation* or difference* or trend* or reducti* or expan* or decline*)) AND Topic=(statistic*)

- Topic=((insect* or *arthropod*)) AND Topic=((ecology or environment* or ecosystem* or population*)) AND Topic=((increas* or decreas* or variation* or difference* or trend* or reducti* or expan* or decline*)) AND Topic=((statistic* or "meta*-*analysis" or "meta*analysis"))

NOT Topic=(gen* or dna or taxonomy or illness* or disease* or human or people or econom* or "food loss" or cancer)

This example for 'insect or arthropod' with 'statistic' and 'meta-analysis' included and also a "NOT" filter to exclude genetics, DNA and disease etc., produced 845 references.

Libraries of references produced by these restricted searches were saved in EndNote and are presented with this report (see summary Table 3, Appendix 3).

We believe that the results of searches outlined in 1.3.1 and 1.3.2 demonstrated that a systematic review using key word search terms was not a useful strategy to access statistical methods in the ecological and environmental literature. The results from systematic searches were either unmanageable in the number of references suggested, or produced many references that were not relevant to this study. Further discussion outside of the project team corroborated our conclusion. In scoping statistical experts for input into the systematic literature review, it became apparent that most experts consulted believed that any analysis that might provide information on the impact of EFSA regulated products would sit within a Generalised Linear Model (GLM) framework, and that the literature on specific organisms or habitats would not easily reveal any new techniques, or reveal the relatively minor, although potentially interesting, differences in approaches that statisticians were not already familiar with. Furthermore, it was noted that the ecological and environmental literature would likely have only a fleeting reference to the statistical techniques employed in their data analysis, because GLM methods are the standard tools. In light of this, we added a specific section to this report on the statistical techniques that professional statisticians agreed were most appropriate and should be used to assess changes in organisms and the environment using the type of data that is collected in ESNs (see below Section 1.4).

To compliment Section 1.4 on appropriate statistical methods, it was agreed that we should change the literature search strategy to focusing on the web sites of the networks themselves to look for information on statistical technique. This ensures that all methods and approaches are captured even if the ecological publications do not specifically mention statistical methods in the key words or abstract and if the statistical publications do not specifically use ecological data in any example or application. A comprehensive list of ESNs across Europe, and the information they contain is in Section 3 of this report. We sought literature from networks in two ways. Firstly by a systematic search in WoK using the network name (1.3.3) and secondly by visiting network web sites to see what literature they advertised (1.3.4).

1.3.3. Availability of ESN literature that includes statistical methods obtained from systematic searches in WoK using the network name as a search term

Using the network list that was in preparation for the environmental data source inventory, we used WoK to search for references to literature by network name both written in English and the language of the country. We searched for 66 different ESN network names and found 1943 publications in WoK. Several networks publish information in journals that are not listed in WoK as they were not in English or may have been considered gray literature.

We discovered that many of the publications, even in peer reviewed journals, did not include the network name in the title, key words or abstract, so that a systematic search in WoK did not identify publications from ESNs. Some documents did include the network name in the acknowledgments, but this is not searchable in WoK. A list of web sites searched in WoK is given in Appendix 3. Networks for this were chosen to represent a range of network types from EU wide networks that report under various EU directives, represent a range of EU countries if they appeared from their name to be more than a local nature organization, and also a sample of local smaller scale organizations.

1.3.4. Non systematic searches to assess the availability of ESN literature, which may include statistical methods, listed on ESN web sites

We visited 31 network web sites in 6 countries that we believed covered a national scale to investigate the availability of document references. We also searched web sites with information such as that collected for the European Environment Agency and the habitats, water framework and birds directives etc. Information on these networks was also collected for the environmental data source inventory where more information about different network activities and a more comprehensive description of the ESNs is described in Section 3. The search used ESNs that were in the process of being compiled for the data inventory and EuMon (http://eumon.ckff.si/index1.php), Centre for Ecology & Hydrology, and publications, such as the review of Citizen Science Review by Roy et al., (2012). Our initial observations from looking at the network web sites for statistical techniques used were:

- the publications available reflect the nature of the organization and may only include photographs, distribution maps, newsletters, annual reports, species help guides, notes on taxonomy and observations without any apparent statistical analyses.
- many organizations that maintain ESNs are meant for local and national organizations so are written mainly in the local language and are therefore not available in WoK.
- publications that only have titles and not abstracts or key words are difficult to assess for statistical methods and insufficient time was available to search for all listed references.
- information was often in publications that were not publically available.
- the publication list may not cover our search period from 2002 to 2012.

Where potentially useful publication lists were found we did the following:

1) Looked for publication titles written in English
2) Assessed from the title (or abstract if available) whether the publication was likely to contain statistical methods. We excluded records of individual sightings of organisms, taxonomic information, species guides, magazines and newsletters and general descriptions of parks and gardens etc.
3) Downloaded and saved any available document that achieved 1) and 2).
4) If the publication was downloaded in 3) or looked potentially of use for this project the information was entered into an EndNote Library and the published document searched for through WoK, Google and journal sites etc.

A summary of the references that we found to be publically available after viewing some 31 individual ESN web sites from 6 countries is given in Table 3 and Appendix 3.

1.3.5. References obtained through non systematic searches following suggestions from experts.

In the EndNote library produced for the project we included an EndNote library group of publications recommended by experts while scoping the literature review. In addition to ESN specific literature we included other potentially useful ecological and environmental literature including references to post-market monitoring, genetically modified organisms and agro chemicals (see summary Table 3).

We also included in the library here the references used in this document and those specific to Section 1.4 which describes expert opinion on the most appropriate statistical techniques to use for assessing the impact of EFSA regulated products on organisms and the environment.

1.4. Statistical considerations for post-market analysis of ecological surveillance networks

1.4.1. Summary

The aim of this first task was to identify statistical methods that have been or could be used to detect an environmental change from a surveillance network. We argued that most statisticians asked to analyse these types of data would use an effectively standard modern statistical toolbox based on extensions of the generalized linear model (GLM) concept, and whether the model is fitted in a classical or Bayesian framework is not of primary importance. We have given a very brief overview of these model developments from GLM to GAMM, GLMM, etc. and noted, as well, other types of statistical analyses that may be appropriate. However we have seen no specific subsets of statistical terms that would appear, necessarily, in any scientific papers reporting these types of analyses, indicating that any literature review would have to be very specifically targeted to deliver an assessment of current practice. By identifying relevant methods and a small sample of literature identified by experts, this report section and an associated literature review provides a basis for further work in the project.

1.4.2. Introduction

The aim here was to identify relevant statistical techniques that might reasonably be expected to indicate or identify a change within data from a typical ESN, and that such a change might be related to an unexpected post-market effect of a product being introduced into the environment. The term product is used here in a very general sense to indicate something new in the ecosystem. In terms of the ecology there are 2 reasonable scenarios:

1) the product goes into use everywhere in the target ecosystem at effectively the same time (either at one location or at several simultaneously), leading to widespread change within a relatively short time scale;

2) the product goes into use at different times at different places, leading to a more gradual change in the ecosystem.

The second scenario is more difficult to detect and would require more sophisticated spatial modelling. From the statistical perspective there are 2 possibilities:

1) a change is detected through the Normal overview of monitoring data from ESNs over and above background variation and initially there is no specific hypothesis for that change;

2) following the introduction of a product to the environment, changes in ESN monitoring data occur over and above background variation.

There are two additional caveats in the statistical analyses for these situations. First, if the hypothesis (or model) that is being explored was set up after a change in the data series had been observed, then it is more likely that a statistically significant difference will be detected unless you adjust your analysis appropriately. The second caveat is that the data series will have been subjected to multiple testing, particularly in the first possibility above, and that multiple testing will require adjustment of the test statistics, for example by controlling the False Discovery Rate (FDR) (Benjamini, 2010; Efron, 2010).

What is very clear is that there is no classical statistical designed situation (which would occur with pre-market testing), there are no readily available cause-effect relationships (because that would be covered by planned post-market assessment), and a meta-analysis or equivalent may be required to find sufficient statistical power from small effects in multiple, possibly independent, data series.

1.4.3. Standard modern statistical procedures

1.4.3.1. Model structures

For the initial analysis of a dataset collected over time, almost all statisticians would use some variant of the generalisations of the linear model. The term "linear" does not mean fitting straight line regressions, but that the model is a linear function of the parameters. The generalised linear model (GLM) (McCullagh and Nelder, 1989) brought together a series of loosely related developments into a now standard framework where there is a systematic component or model equation, a link function to transform scales between the systematic component and the random component, and a flexible set of error structures (within the exponential family) to specify the form of the random component. For a standard linear regression, the model would be a set of linear relationships between the response and the predictors, the link would just be the identity function (so there is no scale change) and the error would be the Normal distribution function. The error structure choice allows continuous, discrete, and binary to be fitted within the same framework, so the distinction between the analysis of continuous measurements and classified or categorical data was removed. In particular this allowed counts and data on proportions to be included within a standard modelling paradigm. Typically, however, count data are over-dispersed in practice compared to the Poisson distribution, so the distributional assumptions are modified to include over-dispersion parameters and the idea of quasi-likelihood to fit the model only by specifying the mean-variance relationship is introduced (Wedderburn, 1974). Recent further extensions include generalized non-linear models (Turner and Firth, 2012) and Hierarchical GLMs (Lee and Nelder, 2001).

This framework was later extended to Generalised Linear Mixed Model (GLMM) where the assumption of no correlation between observations was superseded by more complex variance-covariance structures with both fixed and random effects allowing explicit estimation of the different components of the covariance matrix (Breslow and Clayton, 1993; Venables and Ripley, 2002; Pinheiro and Bates, 2000). This flexibility allows for many different sets of assumptions including autocorrelation between successive observations over time, estimation of variances within nested models to separate population and individual effects (such as in genetics, many environmental surveys and multilevel modelling in the social sciences), and extensions into spatial-temporal models. If the estimates of the random effects are not themselves important then Generalized Estimating Equations (GEE) (Liang and Zeger, 1986) can fit the averaged population parameters while having an unknown or mis-specified correlation structure, and this is one of a class of semiparametric regression techniques available which allow relaxation of some of the GLMM model assumptions. The systematic component of a GLM can be made more flexible by combining parametric and nonparametric regression within Generalised Additive Models (GAM) (Hastie and Tibshirani, 1990; Green and Silverman, 1994; Wood, 2006; Ruppert et al., 2003) where the nonparametric element introduces smoothing functions derived from the data such as spline curves, local linear and polynomial regression, nearest neighbour and kernel smoothing. These extend in the same way as GLMs to Generalized Additive Mixed Models (GAMM) (Wood, 2006; Lin and Zhang, 1999) with the more complex variance structure assumptions.

1.4.3.2. Model fitting

The ideas behind the model structures chosen are the same whatever process is applied to fit the model to the measurements. The initial impetus for these developments was the Nelder and Wedderburn (1972) paper where an iteratively reweighted least squares method was proposed to computationally allow maximum likelihood estimation of the model parameters in GLMs, a natural extension of the simpler least squares methods commonly used for linear regression (see Edwards (1972) for a theoretical discussion of likelihood). Alternative approaches to fitting the statistical model include more general maximum likelihood procedures, generalised estimating equations, partial least squares (Vinzi et al., 2010), pseudo-likelihoods (Besag, 1975), restricted maximum likelihood (Patterson and Thompson, 1971), partial likelihood (Cox, 1975), penalised likelihood (e.g. Wood, 2000), general Bayesian approaches (Congdon, 2006) or the more recent developments of integrated nested Laplace approximations (INLA) (Rue et al., 2009), with many, particularly Bayesian methods, benefitting from the vast increase in computing power since 1970. However the technicalities and choice of the model fitting process are not primary considerations in the selection of a conceptual model to detect a change, and there are currently a wide range of models which can be fitted by a variety of methods.

Modern methods have brought an increased reliance on extended assessment of the adequacy of the model, with auxiliary techniques such as bootstrap to provide robust sample estimates with confidence intervals (Davison and Hinkley, 1997) or generalised cross-validation (Stone, 1974; Wood, 2006) to explore the choice of smoothing parameters are often required, so model development is often an iterative process.

Other developments in statistics that do affect the fitted models include challenging the independence of the responses by introducing the assumption of autocorrelated errors in time (and various forms of time series structures specified according to the autoregressive integrated moving average (ARIMA) framework) (see Chandler and Scott (2011) for many examples) and in space (with classical geostatistics and point process models now leading to more flexible space-time structures (Wackernegel, 2003; Diggle et al., 2010; Cressie and Wikle, 2011). These approaches and many further topics in detecting change in environmental and ecological data are covered in recent texts on model fitting for environmental applications, such as Chandler and Scott (2011) and the series of books from Highland Statistics: Zuur et al. (2007), Zuur et al. (2009), and Zuur et al. (2012). There is also literature related to software packages developed for specific audiences, e.g. TRIM (van Strien et al., 2004) used for Bird Census data is an application of GLM, and the particular software is not necessary for the application of the statistical method.

1.4.3.3. Generic approach

Therefore in modern statistics there is now a general structure with both fixed and random effects, some containing systematic components, with correlated error structures for time dependence, spatial dependence and possibly space-time interaction. How you fit the model is a technical issue that most applied statisticians would resolve by a pragmatic choice of whatever classical or Bayesian method is available and delivers the result. There is a lot of debate about how we choose the model, e.g. model selection methods (e.g. using Akaike Information Criteria (AIC) and related methods), structural equations, etc., but, at a gross level of simplification, these are variants of similar approaches intended to provide the best model for the application while dealing with the combinatorial explosion of having huge numbers of potential variables and retaining explanatory interactions. These may, however, be optimising the wrong criteria unless care is taken to make sure that change detection is the primary purpose of the analyses.

1.4.4. Additional Methods

Alongside the general framework of the GLM exists a class of methods that do not assume a distributional form of the data itself, although the procedures often assume a distribution on a derived statistic. These methods are classed as non-parametric and can form a favourable alternative to parametric methods when sample size is low or distributional assumptions are clearly violated. There are many simple ways to obtain unbiased parameter estimates non-parametrically, such as minimising the sum of squared residuals or using standard minimum variance unbiased estimators. However, drawing inference on the estimated parameters non-parametrically (i.e. distribution free) requires a different approach - the most powerful of which is the use of resampling methods. The idea behind these methods is to resample the data under the null hypothesis in order to build up a range of plausible test statistics to which one can compare the observed derived test statistic and hence draw inference. Examples of resampling methods are the bootstrap (Efron and Tibshirani, 1994), the jack knife (Efron, 1982), cross validation (Efron, 1983) and permutation tests (Fisher, 1954; Pitman, 1937), all of which use similar principles in order to test the hypotheses under scrutiny. In many more extensive studies, both classes of methods may be used at different stages of the same analysis.

Other non-parametric tests of association, differences and coherence often involve deriving a test statistic based on ranks or differences, which is then shown algebraically to follow some mathematical distribution. The Wilcoxon signed rank (Wilcoxon, 1945), for example, can be used to compare two related samples by examining a statistic defined using the sum of the signed ranks of differences between the paired samples, which is shown to follow Normal distribution under the null hypothesis of no difference between the samples. Similar non-parametric methods have been developed to investigate whether the mean between two samples is different (Friedman, 1937; Kruskal and Wallis, 1952) and how a sample conforms to some hypothesised distribution or expectation (Cochran, 1952; Kolmogorov, 1933; Smirnov, 1939).

Some more niche areas of statistics where one may find before/after effect analyses conducted include survival analysis (Hosmer, Lemeshow and May (2011) provide an excellent overview), extreme value theory (De Haan and Ferreira, 2006), spectral analysis (Priestley, 1981) and multivariate techniques. Within multivariate techniques there are numerous methods for representing data, but in order to look for a specific effect the methods of Canonical Correspondence Analysis (CCA) (ter Braak, 1986; Manly, 1992), Redundancy Analysis (RDA) (van den Wollenberg, 1977) and Multivariate regression trees (De'Ath, 2002) serve as the most useful tools.

Hastie, Tibshirani and Friedman (2009) provide some very interesting ventures into the Data Analysis, Computer Science and Engineering worlds, and categorize statistical methods into supervised or unsupervised learning methods. In the former the goal is prediction of a measured outcome relative to input measures (effectively the modern statistical procedures above), while the unsupervised learning relates to association and pattern, as in the niche areas. This important contribution to the overall prediction problem also incorporates a lot of data-intensive methods. Other interesting techniques include control charts (such as CUSUMs) which are a standard tool if the testing of the measurements are repeated every day (e.g. Mei, 2010) and various epidemiological methods to detect outbreaks of diseases (e.g. Zhu and Wang, 2012).

1.4.5. A targeted approach to the literature review

Most statisticians would know what they would do to analyse data from ESNs, the majority of analyses would sit within the generic modelling frameworks described above, and the literature will not easily reveal the relatively minor but potentially interesting differences in approaches. Indeed it might well have only fleeting reference to the statistical techniques which are just the standard tools used every day. Therefore, an untargeted literature review based on statistical terms in ecological or environmental papers is unlikely to deliver the library of techniques this project aims to identify.

Within the statistical literature, the references will not necessarily come from environmental papers as techniques for non-linear trend detection, break points in a series, possible heterogeneity of variance, time series drifting outside their previous confidence intervals, etc. are more generic. A scattergun approach here will not yield useful results but there is guidance available from papers in the ecological and environmental literature published by those who have been tracking trends in ESNs.

Here we have indentified key analyses approaches to form the basis of further statistical assessment and will be used in combination with the literature review described. A number of experts have contributed information during the consultation on this note. There are 4 examples of papers using some of the discussed statistical techniques highlighted at the end of the references.

1.4.6. Examples of papers applying some of the statistical analysis techniques

1.4.6.1. Example 1: Application of GLM

Kleijn, D., Berendse, F., Smit, R., and Gilissen, N. (2001) Agri-environment schemes do not effectively protect biodiversity in Dutch agricultural landscapes. Nature, 413: 723-725.

Abstract

We surveyed plants, birds, hover flies and bees on 78 paired fields that either had agri-environment schemes in the form of management agreements or were managed conventionally. Management agreements were not effective in protecting the species richness of the investigated species groups: no positive effects on plant and bird species diversity were found. The four most common wader species were observed even less frequently on fields with management agreements. By contrast, hover flies and bees showed modest increases in species richness on fields with management agreements. Our results indicate that there is a pressing need for a scientifically sound evaluation of agri-environment schemes.

The data on individual species and species groups were analysed using GLM with a logistic link function and assuming a Binomial error distribution, followed by a likelihood ratio test (or G-test). The models included the factors area, pair and management agreement, where both area and pair were considered as replications. As effects of management agreements on the insect groups differed markedly per sampling period, analysis of these species groups was performed on individual sample periods. Furthermore, the factor sward height at the time of sampling was included in the model. All proportional data were arcsin transformed before analysis.

1.4.6.2. Example 2: Application of GLMM

Rundlöf, M. and Smith, H.G. (2006) The effect of organic farming on butterfly diversity depends on landscape context. Journal of Applied Ecology 43: 1121–1127.

Abstract

We used generalized linear mixed models (SAS macro Glimmix) with Poisson error distribution and log-link function to analyse the effects of farm practice and landscape type on butterfly species richness. The butterfly abundance (individuals per 50 m of transect) was log- transformed [ln(x+0•1)] to achieve residual Normal distribution, and analysed using general mixed models with Normal error distribution. Data were analysed at the segment level to account for the slightly unequal sampling effort at different farms (all results were qualitatively the same if analysed at transect level). The fixed factors in the models were year, landscape type, farm practice and landscape type × farm practice; the random factors were farm pair and farm identity; the repeated factor was visit (nested within year). We selected the covariance structure for the repeated factor based on AIC (Akaike Information Criterion), which in all cases resulted in a first-order autoregressive structure being used. We used the Satterthwaite method (Littell et al., 1996) to approximate denominator degrees of freedom. Pearson correlation was used to assess the association between proportion of organic arable land and productivity of the arable land (yield of spring barley; kg ha−1). All statistical analyses were performed in SAS 9•1 for Windows.

Although we can conclude that both farming practice and landscape heterogeneity significantly affects butterfly species richness and abundance, the most interesting result is the effect of the interaction between the two. A similar relationship has been proposed for arable weeds (Roschewitz et al., 2005).

1.4.6.3. Example 3: Application of GAM

Fewster, R. M., Buckland, S. T., Siriwardena, G. M., Baillie, S. R. and Wilson, J. D. (2000) Analysis of population trends for farmland birds using generalized additive models. Ecology, 81: 1970-1984.

Abstract

Knowledge of the direction, magnitude, and timing of changes in bird population abundance is essential to enable species of priority conservation concern to be identified, and reasons for the population changes to be understood. We give a brief review of previous techniques for the analysis of large-scale survey data and present a new approach based on generalized additive models (GAMs). GAMs are used

to model trend as a smooth, nonlinear function of time, and they provide a framework for testing the statistical significance of changes in abundance. In addition, the second derivatives of the modelled trend curve may be used to identify key years in which the direction of the population trajectory was seen to change significantly. The inclusion of covariates into models for population abundance is also discussed and illustrated, and tests for the significance of covariate terms are given. We apply the methods to data from the Common Birds Census of the British Trust for Ornithology for 13 species of farmland birds. Seven of the species are shown to have experienced statistically significant declines since the mid-1960s. Two species exhibited a significant increase. The population trajectories of all but three species turned downward in the 1970s, although in most cases the 1980s brought either some recovery or a decrease in the rate of decline. The majority of populations have remained relatively stable in the 1990s. The results are comparable with those from other analysis techniques, although the new approach is shown to have advantages in generality and precision. We suggest extensions of the methods and make recommendations for the design of future surveys.

1.4.6.4. Example 4: Application of Redundancy Analysis

Walker, K. J., Critchley, C. N. R., Sherwood, A. J., Large, R., Nuttall, P., Hulmes, S., Rose, R. and Mountford, J. O. (2007) The conservation of arable plants on cereal field margins: an assessment of new agri-environment scheme options in England, UK. Biological Conservation, 136: 260-270.

Abstract

Agri-environment (AE) schemes aim to arrest declines in arable biodiversity through cereal field margin management options. We evaluated the effectiveness of uncropped cultivated margins (UCM), spring fallow (SF) and cropped conservation headlands with (CH) or with- out fertiliser inputs (CH(NF)) in sustaining plant species diversity and rare species, in Eng- land, UK. Sampling was stratified at 1 m, 3 m and 5 m from the edge of the margin and in eight regions to assess environmental influences on species composition. Species diversity, including rare species, was highest on UCM, followed by SF and CH(NF) margins. Diversity was generally lower on cropped margins due to competition from the crop. Fertilised CH margins were the least diverse option and were similar to cereal crop controls. Species diversity was greatest at the edge of all except UCM margins and there was a strong latitudinal decline in overall diversity and rare species. AE management accounted for more variation in species composition than habitat context, physical/climatic variables, soil properties or region.

Environmental influences on vegetation composition were carried out on log-transformed species frequencies using Canoco V.4.02 (ter Braak and Smilauer, 1998). Variation partitioning (Økland and Eilertsen, 1994; Økland, 2003) was performed using redundancy analysis (RDA) and partial redundancy analysis (PRDA) with variance partitioned between five subsets of environmental variables (option, region, physical/climatic, habitat context, soil properties) and expressed as the % of total variation explained (TVE). For each subset, an RDA was carried out to select variables that contributed significantly to the model ($p < 0.05$). Forward selection (using 999 Monte Carlo permutations) was used to test each variable in turn and the significance of the first axis and the overall RDA.

1.5. Summary of literature searches for references to publications with statistical information

The aim in this section of the report was to describe the process we used, using a systematic literature review procedure and to give information about the published literature available, which would enable us to comprehensively survey statistical methods and assess data collected by Environmental Surveillance Networks (ESNs). This information would then be used to support analyses of the impact of products on the environment post-market distribution. As described above we tested systematic literature surveys using the WoK literature database and found this method unsuited to the task and concluded that a systematic literature review was not appropriate for the requirements of this project. In this section of the report we have provided information on the searches carried out and also provided reference information in an EndNote library that includes several different literature search methods to give an overview of the published information publically available that we believed may have provided the information required on statistical techniques (see summary Table 3 and refer to Appendix 3 for full details). This EndNote library was used during the project contributed to both the statistical and data inventories. Section 1.4 provides independently reviewed information on the techniques that professional statisticians describe the methods that they believe are appropriate to analyse the type of data collected by ESNs and describe the Generalised Linear Model (GLM) framework and other techniques. We believe that the complimentary information collected in the reference libraries and provided by the information on the GLM framework provides useful information that will assist those that plan to analyse data for PMEM to detect ecological and environment changes that have been caused by the release of EFSA regulated products.

Table 3: Summary of the EndNote library and groups within it submitted with this report.

EndNote Library group	EndNote Library group content	Number of references
SSO **Systematic search by organism type**	References produced under Section 1.3.2 very restricted search for organism type, AND terms that reflect change AND statistics and met-analysis (see Appendix 4). Searches made independently for each organism type including searches that might capture a different range of references, which reflect different recording schemes i.e.: amphibian, birds, bat, “bee+insects+moth”, butterflies, reptiles, “plant species” have been combined into one library.	7124
SSN **Systematic Search by network**	References produced under systematic WoK searches (Section 1.3.3). This includes network names for 66 networks using local language where available and also the English translation. Searches were either conducted individually or with multiple network names in one search. Some are listed separately and others are encompassed in separate groups.	3031
NSW **Non systematic web search on individual networks**	References produced by non-systematic searches by viewing 31 individual web sites (Section 1.3.4). With multiple links within. Some networks are in separate EndNote groups and others are combined into one group.	220
EXL **Expert literature suggested**	Reference library produced from references recommended by experts. Also contains references used in this document and those in the statistical note.	289
EndNote Library group	**EndNote Library group content**	**Number of references**
STN **Statistics note references**	References used in the statistics note	75

2. Inventory of Statistical Approaches

2.1. Introduction

Using the database of literature described in Section 1 and the information on appropriate statistical methods that could be used to analyse data for PMEM, we set out to compare different methods and prepare an inventory of where and how these methods are being used. An assessment of the statistical methods found in the literature reported in Section 1 is crucial if these methods are to be recommended for application by ESNs across Europe. The popularity and use of these statistical methods may differ among data types and networks and here we have highlighted these differences and also the potential performance of different methods for use in PMEM.

Here we compared the selected statistical methods in two ways. The first approach was a literature review of articles published within the sphere of agri-ecosystem science that cite the methods reviewed in Section 1. The second was a simulation study that enabled a fair comparison among the methods to highlight their relative performance for use in Post-Market evaluation.

2.2. Literature Review

We first compiled an inventory of the statistical approaches used in the field of agri-ecosystems, using a selection of the literature identified in Section 1 and then found related publications that directly cited these methods and approaches. Section 1.4 also includes some key references and highlights the main statistical approaches that are available for PMEM. Each method is accompanied by the source reference. The statistical methods described in Section 1.4 form the basis of the inventory developed here where we compare the methods identified by statisticians, and these are detailed below.

2.2.1. Define Source Articles.

We first compiled the references that were to be used in a cited reference search. These references were taken from Section 1.4, which describes appropriate statistical methods and are listed in Table 4. There are potentially other references that may be used for the same statistical method, but those we have selected were considered the key and the most frequently cited reference for the particular method as determined during the research for appropriate statistical methods in Section 1.4.

Table 4: List of statistical methods considered in the comparison and inventory, together with its key reference as stated in Section 1.4.

Statistical method	Key reference for the method
Bootstrap	Efron, B., and Tibshirani, R. J. (1994) An Introduction to the Bootstrap (Chapman & Hall/CRC Monographs on Statistics & Applied Probability).
GAM (Generalised Additive Model)	Hastie, T. J. and Tibshirani, R. J. (1990) *Generalized Additive Models*. Chapman & Hall/CRC.
GAMM (Generalised Additive Mixed Model)	Lin, X., and Zhang, D. (1999) Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society B*, 61: 381-400.
GEE (Generalised Estimating Equations)	Liang, K.Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika* 73: 13-22.
GLM (Generalised Linear Model)	McCullagh, P. and Nelder, J. (1989) *Generalized Linear Models, Second Edition*. Boca Raton: Chapman and Hall/CRC.
GLNM (Generalised Nonliner Model)	Turner, H. and Firth, D. (2012) *Generalized nonlinear models in R: An overview of the gnm package.* Cran-R Project.
GLMM (Generalised Linear Mixed Model)	Breslow, N.E. and Clayton, D.G. (1993) Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association* 88: 9–25. Pinheiro, J. and Bates, D.M. (2000) *Mixed-Effects Models in S and S-PLUS*. Springer.
GLMM – REML (REstricted Maximum Likelihood)	Nelder, J. and Wedderburn, R. W. M. (1972) Generalized Linear Models, *J. R. Statist. Soc.* A, 135: 370-384.
Hierarchical GLMs	Lee, Y. and Nelder, J.A. (2001) Hierarchical generalized linear models: a synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika* 88: 987-1006.
INLA (Integrated Nested Laplace Approximations)	Rue, H., Martino, S., and Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society B* 71: 319-392.
Kruskal – Wallis Test	Kruskal and Wallis (1952) Use of ranks in one-criterion variance analysis, *Journal of the American Statistical Association* 47: 583–621
Redundancy Analysis	Van Den Wollenberg, A. L. (1977). Redundancy analysis: an alternative for canonical correlation analysis. *Psychometrika* 42: 207-219.
Space-time point process	Diggle, P.J., Kaimi, I. and Abellana, R. (2010) Partial-Likelihood Analysis of Spatio-Temporal Point-Process Data. *Biometrics* 66: 347-354.
Spectral Analysis	Priestley, M. B. (1981) Spectral analysis and time series.
TRIM	Van Strien, A., et al. "A loglinear Poisson regression method to analyse bird monitoring data." *Bird* 1995 (2004): 33-39.
Wilcoxon Signed Ranks	Wilcoxon, F. (1945) Individual comparisons by ranking methods. *Biometrics Bulletin* 1: 80–83.

Having obtained the references with which to conduct the cited reference search, we then followed the steps as documented in Higgins and Green (2009) and O'Conner et al. (2012) to define the search criteria needed and to ensure that all results are repeatable.

2.2.2. Define search criteria for obtaining information on publications that used each statistical method:

2.2.2.1. Define research question

The aim of this inventory of publications that used a statistical method was to find the uses and examples where each of the methods was being used within the relevant scientific area. This was to inform scenarios where a method may be applicable for a range of data types. Therefore the main question that we asked for this literature search was:

"What are the different applications of the selected statistical methods that are relevant to agro ecosystem science?"

Where:

Term in objective	Definition of objective term
"selected methods"	All statistical methods documented in Section 1.4 and listed in Table 4
"applications"	All published instances of data analysis citing the particular statistical method in question.
"agro ecosystem"	All managed environments and associated natural ecosystems.

2.2.2.2. Establish information Sources

As with the literature searches carried out in Section 1 to establish appropriate statistical methods, we used WoK as the source of reference material. WoK has a full database of references incorporating literature from all fields important in answering the research question, produces repeatable results and also allows cited literature searches. This enabled us to search for all references citing particular articles, which was what was required in this case, and then refine and search within these results. As with Section 1, other reference databases such as Google Scholar and the Opengrey website were considered, but WoK was deemed the most comprehensive and repeatable source to use. These searches were conducted in addition to the searches carried out and described in Section 1.3.

2.2.2.3. Indentify Search terms

Having conducted the cited reference search, we refined our results to include only those articles that may have direct relevance to agro ecosystems, as defined in the search question. The search key words were broad categories that attempted to exclude irrelevant articles but not so restrictive as to exclude applications that could be of interest. The categories chosen were a subset of those from the systematic literature review carried out in Section 1 and included a selection from each of the two topics "search terms for habitat or ecosystem" and "terms that might indicate statistical analysis carried out". The search string used to refine the cited reference search results was:

{ecosystem* OR agri* OR ecology} AND {monitoring OR survey*}

As with many of the searches conducted when reviewing statistical methodologies in Section 1, this string included wild cards to allow for possible miss-spellings and plurals for some of the key words.

2.2.2.4. Search Filters

The only filter set on the results was to ensure that the date of publication was between January 1 2002 and December 31 2012. This was the same 10 year period as used previously in Section 1.2. This search time period, without extra refinement to research area or research domain categories, ensured that the search was repeatable and consistent.

2.2.2.5. Management of references

The references are stored in a single EndNote database with different groups defining the different statistical methods searched. There is one group of references for each statistical method searched for so that it is easy to distinguish among the examples for each separate method. All of the references found were exported directly from WoK into EndNote. This library "Stats_Method_Application.enl" forms a key deliverable for this work and is submitted together with this report.

2.2.3. Search Results

The literature searches yielded a large number of citations for some methods and a fairly low number for others. These results broadly follow the ideas presented in Section 1.4 on appropriate statistical methods, that suggested that some methods have a more bespoke application and others are a more general framework that are often used. Table 5 shows the number of publications that cite the key statistical method referenced in Table 4. From this Table 5, it is clear that the standard GLM model is the most commonly used method relating to the field of agro ecosystem science. The non-parametric bootstrap approach to estimating significance was also found relatively frequently. More bespoke (methods for very specific types of data) and modern methods (those published in the last 5 years) such as generalised nonlinear models (GLNM), Integrated Nested Laplace Approximations, space-time point processes and redundancy analysis are less frequently encountered. This reduced frequency may in part be due to: their relatively recent entry into the statistics literature (eg. INLA in 2009); the lack of readily available software (such as hierarchical GLMs); or the potential for incorrect or missing references (possibly the case with redundancy analysis as scientists may have cited software such as CANOCO rather than the method itself). Nevertheless, the database of literature we have constructed provided a useful comparison of the frequency of use of the different statistical methods and the type of data they are used for within the literature published from 2002 - 2012. This search facility in the EndNote library provided should enable a quick cross reference among statistical methods. For example, one could search each group of references in the EndNote library for "bird trend" to see examples of the statistical methods used and retrieve publications that have applied each of the methods.

Table 5: Table showing the frequency of publications citing each of the statistical methods listed in Table 4 and hence looked for during the literature search.

Method	Number of publications citing references about statistical method
GLM	679
Bootstrap	544
GLMM	484
Generalised Estimating Equations	391
GAM	271
GLMM - REML	197
Kruskal – Wallis Test	163
Wilcoxon Signed Ranks	140
Spectral Analysis	99
Method	**Number of publications citing references about statistical method**
TRIM	19
GAMM	17
Redundancy Analysis	11
Hierarchical GLMs	8
Integrated Nested Laplace Approximations	6
GLNM	3
Space-time point process	2

2.2.4. Data types used for each statistical method

The total number of publications (more than 3000) found in the literature search that cited the statistical methods was too many to fully assess and extract information required for this study. Therefore, a random sample of articles using each statistical method was assessed to extract information about the data types. We believe this provided enough information to look at the broad data types that each method has been used for. In defining the categories we wished to collect the meta-data for, we ensured that all information required was available in the selected manuscripts themselves. These categories included: the type of response variable; the temporal resolution; and the spatial resolution. These three important aspects can be represented by: the number of survey sites; the number of years of the study; the temporal frequency of data; whether the data was count, presence/absence or continuous; and how many variables were analysed. Table 6 shows the details of the publications sampled to assess the data type that was analysed using each different statistical method. These categories were also used to provide guidelines on the information that was needed from the ESNs, for these further results see the data source inventory described in Section 3.

Table 6: Table showing the types of data described in a random sample of publications that cite each of the statistical methods considered. Number of sites (No. of sites) refers to the spatial resolution of the data and whether it was a single study site or if data was collected from multiple spatial locations; No. of years refers to the length of time any monitoring had taken place; No. per year is how many observations were taken within a single year for the response variable that was modelled; Data Type is either "Count", "Presence/Absence" or "Continuous" that best describes the response variable being modelled; and No. of variables is how many response variables were modelled in the publication e.g. if multiple species were modelled.

Method	Reference	No. of Sites	No. of Years	No. per Year	Data Type	No. of Variables
GLM	Poyry et al. 2005	33	1	1	Count	11
	Bani et al. 2009	1	15	1	Count	51
	Trenkel et al. 2004	2	30	2	Count	23
GLMM	Augustin et al. 2009	1475	30	1	Continuous	?
	Krishnamurthy et al. 2011	294	3	2	Continuous	?
	Kery and Matthies 2004	25	2	1	Poisson	?
Generalised Estimating Equations	Amacher et al. 2008	12	3	1	Continuous	4
Method	**Reference**	**No. of Sites**	**No. of Years**	**No. per Year**	**Data Type**	**No. of Variables**
	Friedmann et al. 2003	1	3	1	Continuous	745
	Brook et al.	1	14	1	Count	
GAM	Benton et al. 2002	1	27	1	Count	12
	Potts et al. 2010	1	38	1	Count	214
	Virkkala et al. 2005	2810	1	1	Presence/Absence	10

Method	Reference	No. of Sites	No. of Years	No. per Year	Data Type	No. of Variables
GAMM	Devictor et al. 2007	180	15	1	Count	100
	Augustin et al. 2011	METHOD PAPER				
	Hazen et al. 2009	17	1	16	Continuous	7
Integrated Nested Laplace Approximations	Lopez-abente et al.	24	9	1	Count	1
	Shaddick et al. 2012	934	1	1	Continuous	1
	Prague et al. 2012	METHOD PAPER				
Space-time point process	Aarts et al. 2012	100-10000	1	1	Count	3000
	Gabriel et al. 2012	20000	1	1	Presence/Absence	1
TRIM	Voriesek et al. 2008	18	20	1	Count	48
	Arheimer and Svensson 2008	1	20	1	Count	1
	Kasahara and Koyoma 2010	1	14	1	Count	13
Bootstrap	Smart et al. 2003	9596	2	1	Count	?
	Grizzetti et al. 2008	182	7	1	Continuous	?
	Brown et al. 2005	2	26	1	Count	?
Wilcoxon Signed Ranks	Buckley et al. 2012	1	1	54	Count	13
	Houle et al. 2009	1	50	1	Continuous	3
	Johansson et al. 2002	40	1	1	Count	4
Kruskal – Wallis Test	Su et al. 2011	41	8	1	Continuous	13
	Bailey et al. 2012	1	5	1	Continuous	51
	Orekhova and Rasina 2012	2	6	1	Count	2
Spectral Analysis	Uri 1996	1	47	1	Continuous	1
	Ferguson et al. 2012	6	30	1	Count	10
Redundancy Analysis	Pellegrino et al. 2011	10	1	4	Continuous	
	Griffith et al. 2001	86	2	1	Continuous	45
	Fabien et al. 2007	243	1	1	Count	257

The inventory in Section 3 collated this data from the networks themselves and completes a similar table (with additional information).

2.3. Simulation study to compare methods for PMEM

The literature search presented above provided a comparison of each of the selected methods in terms of their current usage within environmental monitoring. However, the reasons for use of a particular method or the performance of each method described within the article could differ considerably for many unknown or undisclosed reasons that would cause “noise” in these data. We therefore attempted to factor out this noise to provide a fair comparison among the methods by conducting a simulation study on some of the most popular methods identified in Section 1.4 concerning appropriate statistical methods. This simulation study also allowed an assessment of the relative power of each method to detect a hypothetical change. In the simulated data sets, we held all factors constant that may differ among the applications of the different methods, in order to obtain comparable estimates of power.

The simulation study described below tested the null hypothesis that there was no change in a given indicator over time, against the alternative hypothesis that there was a linear trend over time. This hypothesis, that there was no trend over time, is tested because here we were interested in comparing the ability of each method to identify a specific trajectory and we were not comparing the effect of a specific treatment. Relative effects between treatments and controls are considered later in Section 4.

2.3.1. Chosen data sets

We used simulated data sets having known characteristics to test the different statistical methods. The range of different data types such as count and continuous data, with different periods of study (see Table 7) would ensure that they were comparable across statistical methods. The data sets were simulated to represent a range of characteristics that are likely to be found in ESN monitoring data and the change in data trend that might occur as a result of environmental change. One can then investigate “real” data sets from ESNs using the simulated data as a proxy to look into the affects of analysing monitoring data using different statistical approaches.

We simulated each of the three main data types in ecological data – count data, presence/absence data and continuous data. Simulating these different types of data is important as each type requires a different error structure that must be applied for the models used in the statistical analysis. This ensures that the correct statistical assumptions are met and that the associated standard errors estimated are not miss-specified, which could result in making an incorrect inference about the hypothesis under consideration and the results of the analysis. We also simulated different datasets to reflect different spatial scales, temporal scales and levels of variability, to ensure that our ‘simulated data sets’ are sufficiently varied to provide a comprehensive representation of real ES network data and that real data sets can be mapped onto the simulated data.

The full parameter set used to prepare the test data for the simulations is given below in Table 7, where the important high level parameters represent information that can be obtained directly from ESNs. In the simulations two start values, which represent the count (Poisson case) or likelihood of occurrence (Binomial case) of a particular species that are used to start generating the dataset. Two values were used because, for both the Binomial and Poisson case, variance is linked to the mean and it has been shown in multiple previous studies (for example Hails et al. 2012) that statistical power is highly dependent on the prevalence of a species in a data set. The % change may seem small but as this is a change per year and hence a continuous change over time, the change over the course of the time series can be large as it has accumulated.

Table 7: Full list of parameters defining the data sets created in the simulations study. The number of sites defines how many distinct spatial units are to be simulated; the start value provides the expected count, value for continuous data, and occurrence probability for presence/absence data of the count; the

error defines the standard deviation used in the Gaussian simulations for the continuous data; the length of monitoring and samples per year define the temporal resolution of the simulated time series; change % is the change in the expected value per year that might result from environmental change; and the number of variables represents how many time series were simulated for each data type, in each iteration of the simulation.

Data Type	No. of Sites	Start Value	Error	Length of Monitoring	Samples per Year	Change %	No. of Variables to Simulate
Count	10	100	NA	5	12	0.1	10
Count	50	100	NA	10	1	0.1	10
Count	1	100	NA	5	52	0.1	10
Count	10	100	NA	25	1	0.1	10
Count	1	100	NA	25	12	0.1	10
Count	1	100	NA	10	1	0.1	10
Count	10	10	NA	5	12	1	10
Count	50	10	NA	10	1	1	10
Count	1	10	NA	5	52	1	10
Count	10	10	NA	25	1	1	10
Count	1	10	NA	25	12	1	10
Count	1	10	NA	10	1	1	10
Continuous	10	100	25	5	12	0.1	10
Continuous	50	100	25	10	1	0.1	10
Continuous	1	100	25	5	52	0.1	10
Continuous	10	100	25	25	1	0.1	10
Continuous	1	100	25	25	12	0.1	10
Continuous	1	100	25	10	1	0.1	10
Continuous	10	100	5	5	12	1	10
Continuous	50	100	5	10	1	1	10
Continuous	1	100	5	5	52	1	10
Continuous	10	100	5	25	1	1	10
Continuous	1	100	5	25	12	1	10
Continuous	1	100	5	10	1	1	10
Presence/Absence	10	0.5	NA	5	12	0.1	10
Presence/Absence	50	0.5	NA	10	1	0.1	10
Presence/Absence	1	0.5	NA	5	52	0.1	10
Presence/Absence	10	0.5	NA	25	1	0.1	10
Presence/Absence	1	0.5	NA	25	12	0.1	10
Presence/Absence	1	0.5	NA	10	1	0.1	10
Presence/Absence	10	0.9	NA	5	12	1	10
Presence/Absence	50	0.9	NA	10	1	1	10
Presence/Absence	1	0.9	NA	5	52	1	10
Presence/Absence	10	0.9	NA	25	1	1	10
Presence/Absence	1	0.9	NA	25	12	1	10
Presence/Absence	1	0.9	NA	10	1	1	10

2.3.2. Simulation algorithm

The simulation study was carried out in the R statistical computing environment as this had the widest choice of available methods with which to run the simulations. If R did not have a published package on any particular method, then the method was deemed too rare and bespoke to use in this comparison and hence was excluded.

The methods that were selected for this simulation study represent a range of approaches that could easily be adopted by applied ecologists and agro environmentalists and that may provide positive benefits in terms of statistical power resulting from a slight change in approach. A full list of methods that were examined is presented in Table 8. Perhaps the only key method that was omitted for these simulations was the spectral approach, which had a reasonably high (99) frequency of hits in the literature search. This method was omitted as there is currently no available function to look for signs of significant change using spectral approaches. Methods do exist to estimate periodograms and users can use this to fit bespoke Fourier functions to their data, but to draw any inference, most users would then have to resort to using bootstrap based approaches which are included in our assessments. The lack of any "off the shelf" approach means this method is very unlikely to be used in ESN data analysis, so spectral analysis was omitted from the simulation runs.

Table 8: List of all methods used to analyse the data generated in the simulation study together with the R function and package used

Method	Function	R Package
Generalised Linear Model (known family, e.g. Poisson)	glm	mgcv
Generalised Linear Model (Gamma distribution assumed)	glm	mgcv
Generalised Additive Model (known family, e.g. Poisson)	gam	mgcv
Generalised Additive Model (Gamma distribution assumed)	gam	mgcv
Generalised Linear Mixed Model (Maximum Likelihood fit) | site = random	lmer	lme4
Generalised Linear Mixed Model (REML fit) | site = random	lmer	lme4
Generalised Linear Mixed Model (MCMC fit) | site = random	MCMCglmm	MCMCglmm
Generalised Linear Mixed Model (PQL fit) | site = random	glmmPQL	MASS
Generalised Linear Mixed Model (PQL fit, AR (1) process incl) | site = random	glmmPQL	MASS
Generalised Linear Mixed Model (REML fit, AR(1) process incl) | site = random	lmer	lme4
Generalised Estimating Equations (known family) | site = random	gee	gee
Generalised Estimating Equations (Gamma distribution assumed) | site = random	gee	gee
Bootstrap based resampling	lm + own code	Base
Wilcox Test	Wilcox.test	Base
Kruskal Wallis test	Kruskal.test	Base
Method	**Function**	**R Package**
Redundancy Analysis	rda	vegan
CUSUMS	sctest	strucchange

The simulated data are distributed (in the case of count data, though binary and continuous data follow in a similar way) according to the following specification. The count *C* at site *i* at time *t* follows a Poisson distribution with mean $\mu_{i,t}$ as shown below.

$$C_{i,t} = Pois(\mu_{i,t})$$

The expected mean count at a given site in a given year is a function of the induced change over time α, site level variation γ and observation level random error δ, which in the case of the Poisson and Binomial models represents the associated over dispersion. The mean value at a given site *i* at a given time *t* is given by:

$$\log(\mu_{i,t}) = \alpha \cdot \log(\mu_{i,t-1}) + \gamma_i + \delta_{i,t}$$

The data are simulated by evaluating the mean trend on the link scale before generating a realisation of the process. So the data are generated from the start population $\mu_{i,1}$ according to the value specified in Table 7 and draws according to $\gamma \sim N(0,\eta)$ and $\delta \sim N(0,\theta)$, where η is a random draw from the uniform distribution between 0.1 and $\log(\mu_1)/20$, where $\mu_1$is the expected initial mean abundance, and θ is a random draw from the uniform distribution between 0 and 0.1. From this start population a time series of the specified length and resolution was simulated by firstly generating a mean trend.

This mean trend, therefore, generates the required change over time, which is dependent on the previous value back in time. From this mean trend, each data series is a realisation of the mean values across sites and across time. This was then repeated for 10 variables in each test case.

The resulting dataset was then analysed using each specific statistical method and the significance of the change/attribution was stored. This was repeated 1000 times to obtain the percentage of times we observed a significant effect, to determine the statistical power of each method.

2.3.3. Results and Comparisons

The full results of the simulation study are presented in Appendix 4 and the main conclusions are presented here. The most obvious result from the simulation study was that generally the non-parametric tests (Bootstrap resampling, Wilcox and Kruskal-Wallis) performed poorly compared to the parametric approaches, except when the sample size was small. If the sample size is reasonable, then the parametric approaches offer far greater power to detect change than non-parametric resampling or rank based tests. This is perhaps unsurprising. It is clear that the definition of "small" and "reasonable" depends on many factors, but the key aspect is whether there are enough observations to produce reliable estimates of the parameters in the model-based parametric approaches. This is generally considered by statisticians to be at least 8 observations per parameter and that was backed up in the simulations described in this report. Our simulations show that when n=10 (no. of sites=1, no. of years=10, no. per year=1), the non-parametric approach is no poorer than the parametric approach and does not suffer from the same convergence problems that can cause the parametric models to fail.

Including random effects and autoregressive terms in a model can help tease out a signal in the data that could be masked by site to site differences or heavy serial dependence in the observations. If there were signs of site to site differences or temporal autocorrelation, then these terms should be included in the model. If, however, there are no signs of temporal autocorrelation then inclusion of an autoregressive term could lead to a reduction in power. It is, therefore, necessary to first test for this effect on each data set before deciding which modelling approach to adopt.

Generalised estimating equations (gee) performed well when there was a sufficient sample size (gee was unable to run for the n=10 option as defined above). This approach generally, as is the case here, performs better than standard likelihood based model fits due to the smaller, robust standard error estimates that it produces. Generalised estimating equations fit a marginal rather than conditional model and therefore any covariate inference is made at the population level rather than at the subject level. This therefore has implications on inference and can be quite a philosophical argument as to which level one wishes to draw conclusions. Nevertheless, it is an important consideration when deciding on the statistical modelling approach.

As always, one needs to take care when interpreting output from fitted GAMS and significant results could simply indicate over-fitting as opposed to a real trend. This can be particularly true with large sample sizes.

If data exists on multiple species / variables (we tested here 10 hypothetical species), a multivariate analysis can be more powerful if common trends in populations among the species are expected. Rather than conducting multiple analyses per variable, the analyst should consider whether a multivariate approach could be useful for looking at common signals. If there was an effect across multiple variables, as one may expect/hypothesise post authorisation of a regulated product if many species were affected in a similar way, then a multivariate redundancy analysis could prove to be the most powerful.

Figure 1 below provides clear guidelines as to which method to adopt in certain circumstances. In some cases this is more clear cut than others, but ultimately in following the decision tree the users should know what terms are needed in the model and hence which analysis route to take. The guidance in Figure 1 is a direct result from which methods and approaches performed best in what circumstance in the simulation study.

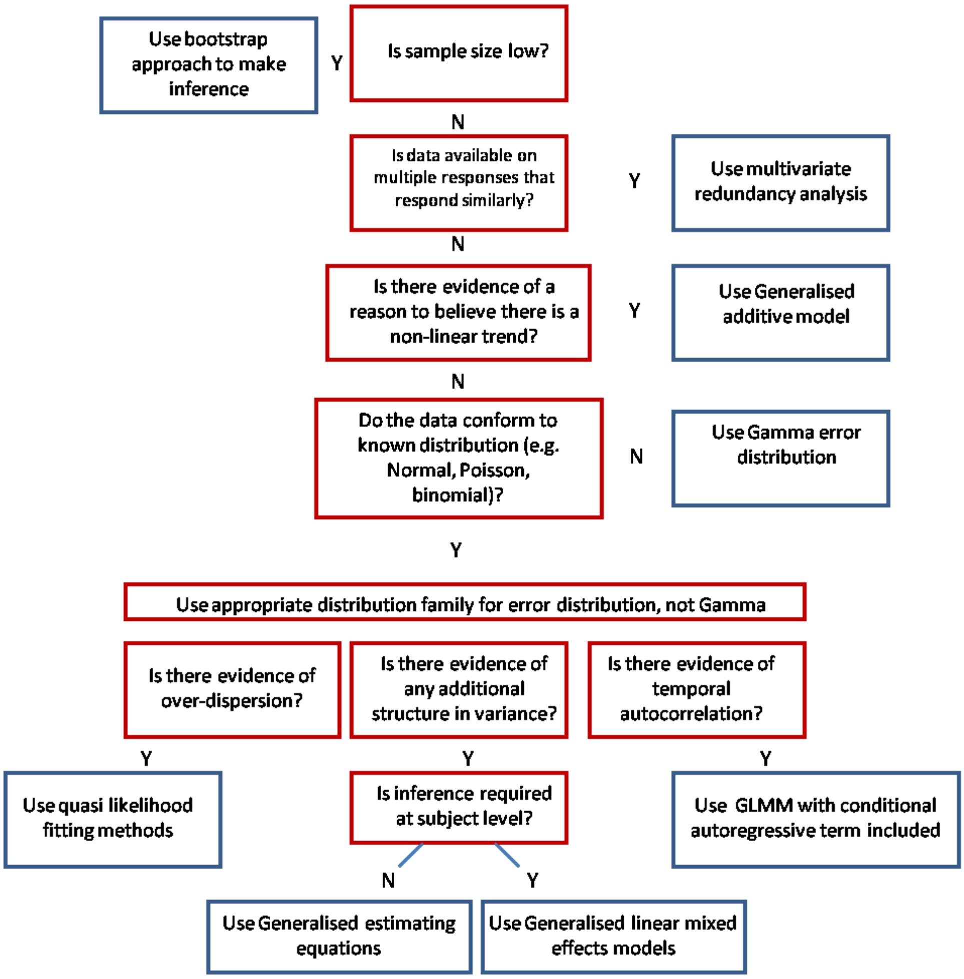


Figure 1: Decision tree to provide guidelines as to appropriate statistical methodologies in different scenarios based on methods tested in simulations study described in Section 2.3.

2.4. Summary of statistical analytical methods using simulated data.

The work presented in Section 2.2 has demonstrated the uses and applications of the methods discovered during the search and review in Section 1. It is clear that some methods are used more frequently for data analysis and described in published papers than others, and this could be for many reasons. What is also clear is that some methods performed better in some situations than others and there was not one method that was optimal over all data types. It is therefore important that when analysing each data set, care is given to the methods employed. This simulation study could act as a useful tool and will provide a checking system for data analysts to ensure that, given their particular data, scientists are using the most optimal method for statistical analysis using ESN data.

In the simulation study presented here the data generated represented fairly regular conditions. Although overdispersion about the mean was included on the link scale when generating the data sets, the resulting series remain reasonably well behaved. It can often be the case, however, that data sets with counts of species, particularly rare species, are dominated by zeros. Generally speaking, standard GLM models do not cope well with such data with many zeros and often miss-specify the standard errors associated with parameter estimates, which in turn can lead to poor inference. Poor inference in this case could lead to rejection of the null hypothesis when in fact it is true or vice versa. There are however, some readily available extensions to the GLM framework that allow for excessive numbers of zeros in count data and do not suffer from the same separation issues that can arise. Though not included in the simulation study here as a miss-specification may result in a model incorrectly appearing to be more powerful, models such as Zero Inflated Poisson (ZIP) or Zero Inflated Negative Binomial (ZINB) provide a solution to these issues. If, upon checking through a data set to be analysed, there are signs of excess of zeros, one should consider using ZIP or ZINB based models that can now be implemented in a variety of frameworks including GAMS in the VGAM R package.

It is also generally important to consider the effect that over/under dispersion could have on any inference made. This can easily be included when fitting a GLM type model by use of quasi distributions or, in the case of count data, use of the negative Binomial distribution. Given the simplicity of including this in the model, over/under dispersion should always be tested for in data prior to fitting the final model one wished to make inference on.

3. Inventory of Environmental Data Sources

3.1. Introduction

The third aspect of this project was the compilation of an inventory of existing environmental data providers, also called environmental surveillance networks (ESNs) that potentially support post-market environmental monitoring (PMEM) of agro ecosystems. ESNs are private or public organisations operated by professionals and/or volunteers that are monitoring a range of environmental and biophysical parameters like the presence of particular species of plants and animals, biodiversity, plant health, air quality, etc. For PMEM organisations that register agronomic practices, land use/management practices, weather conditions, etc. are also useful as they can provide insights on factors that may influence the sensitivity of any post market response.

The authors defined an ideal ESN that likely may serve PMEM as one that:

- collects data on a relevant subject over a broad geographical area of interest, with a standardized collection method, using a dense and even distribution of collection points, which are visited regularly at the time of interest.
- relies on professionals or at least people trained to collect information in a standardized way.
- documents data collection and data analysis methods.
- makes raw data publically available or allows access by request. Interpretative reports issued by the ESN were given extra value as they provided an expert-based analysis and assessment by people who were directly involved in setting up the data collection and were most knowledgeable about the data and the region in which it is collected.

The term ‘network’ was interpreted broadly: it covered organisations grouping people working in distinct geographical areas on specific study subjects as well as several people/entities collaborating in the study of one particular topic. ESNs at European, national and regional levels included:

- Governmental networks that were official initiatives focused on particular policy areas;
- Academic networks that provided platforms for scientific communications on particular projects or research;

- Trade networks that addressed specific professional issues; and
- Environmental networks that monitored, recorded, researched, educated and/or promoted diverse components of nature.

As part of this project, ESNs have been inventoried and for each identified ESN, information concerning the organisation, monitoring subject, specificity, methodology and reporting has been summarized for this report. Based on the collected information, ESNs were then assessed for suitability as an element of PMEM according to the criteria defined by EuroStat[2]. The outcome, especially the information on the methodology used to collect and analyse data, fed into the large-scale simulation studies and reviews carried out in Sections 2 and 4.

Figure 2 provides a schematic representation of the activities in this element of the work. In this report we present the methodology that has been developed to trace networks, collect information and determine PMEM suitability. Whereas the full inventory is submitted as a separate document to EFSA, summary information is included in this report.

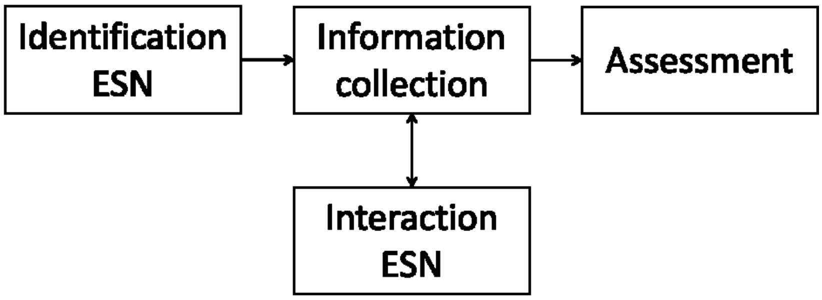


Figure 2: Diagram presenting the relation among the activities carried out for ESN data assessment

3.2. An Inventory of Environmental Surveillance Networks in Europe

3.2.1. ESN Identification

Three distinct approaches for identifying ESN were applied:

1) Given the role of competent authorities related to EU Directive 2001/18/EC[3] as risk managers in PMEM, it was essential that their views were integrated. In cooperation with EFSA an inquiry form was developed. EFSA distributed this form to its 'focal point network' and the EFSA GMO Network (February 2013, reminder May 2013). The Biotechnology Unit of the European Commission, Health and Consumers Directorate-General, discussed the project at the Regulatory committee of Directive 2001/18/EC (14 May 2013) and invited competent authorities to provide references of ESNs.

2) Different initiatives have already developed ESN lists for specific purposes (e.g. EuropaBio, LTER-Europe[4], ALARM Field Site Network[5], JRC-FATE[6], Biodiversity Information System

[2] http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/ess%20quality%20definition.pdf
[3] Directive 2001/18/EC of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC. OJ L106, 17.4.2001, p.1-38.
[4] http://www.lter-europe.net/
[5] http://www.alarmproject.net/alarm/fsn_start.php
[6] http://fate.jrc.ec.europa.eu/rational/home

for Europe (BISE)[7], Water Information System for Europe (WISE)[8], EuMon Database on monitoring schemes in Europe[9], European Biodiversity Observation Network[10]). These lists provided at a minimum the name of the network, sometimes supplemented with the web-address or contact point. In some cases only part of the information was up to date and publically available and it was necessary to establish a dialogue or make an agreement with the owners to access the information.

3) Additional networks were identified by the project partners and through exploring links among ESNs (e.g. via links and information on websites and publically available documents).

This first identification of and ESN typically includes the name of the ESN in a variety of languages, the website (if available) and contact information.

3.2.2. Definition of information elements

To structure the information a distinction was made between the organisation (the ESN) that contributes to environmental surveillance and the actual environmental surveillance programme (the ESP). Some ESNs performed or contributed to several ESPs each with specific objectives, methods and reporting. The inventory therefore needed to enable reflecting this complexity.

The information collected in the inventory had to be adequate to support the other components of the project (e.g. information on data type so that the work in Section 2 can be used to suggest which of the statistical methods highlighted by Section 1 might be the most appropriate) and to allow future use of the inventory for specific PMEM efforts. The inventory therefore combines organisational aspects of the ESN performing the monitoring or surveillance; the monitoring subject; specificity (geographic location); methodology of data collection and data analysis; and information on reporting.

The monitoring subjects were classified either as protection goal or as influencing factor. Protection goals include those aspects of the environment that are valued by society and need to be protected from harm. Examples included bats, butterflies, river water quality, health of woodlands, etc. Factors that may cause an effect on an object being protected were classified as influencing factors. The way farmland is managed, for example, may influence the presence and abundance of certain arthropods. Data on influencing factors are essential in interpreting variations in observed values relative to protection goals.

Information on data type, i.e. whether observations are recorded as presence/absence, or as counts of individual plants or animals, or as continuous values, will determine the most appropriate statistical analysis method as assessed in Section 2.

In order to allow systematic searches and analysis, it was important that the parameters were fixed, well defined and adequate for the intended use. An initial proposal was drafted and discussed among the project partners and subsequently with the project's promoter. The final lists are presented in Table 9 (information on the ESN) and Table 10 (information on the ESP).

[7] http://biodiversity.europa.eu/
[8] http://water.europa.eu/
[9] http://eumon.ckff.si/aims.php
[10] http://www.eubon.eu/

Table 9: List of information elements to describe an ESN

Item [(a)]	**Specification** [(b)]	**Values** [(c)]	**Meaning**[(d)]
Internal code	Internal code that allows easy referencing *e.g.* when establishing relationships between networks	ESN13-xxxx	Unique code automatically attributed
Official name of the ESN	The full legal name as registered (local language)	"Name"	
Other names	Abbreviation (local language); English translation and abbreviation, if available	Text	
Type of organisation	Identification of the type of organisation	- Governmental	Official networks, organized by governments, *e.g.* supporting particular policy areas, to fulfil national and/or European requirements (*e.g.* NATURA 2000)
		- Academic	Networks of scientific research (*e.g.* national academy of science, European funded research projects).
		- Environmental	Organisations involved in education on, promotion and observation of nature, often with a strong conservation orientation, often based on volunteers
		- Trade	Special interest groups addressing issues of trade (*e.g.* bee keepers, farmers, crop protection producers).
Contact person	Name and function of a contact person	Text	
Coordinates	Contact details	Text	
Website	Website address	Text	
Funding	Known sources of funding	- Company - Membership - Grants - Sponsoring - Unknown [(e)]	
Part of other network(s)	Relationship with other networks	Names and/or internal code(s) other ESNs	
Grouping of other network(s)	Relationship with other networks	Names and/or internal code(s) other ESNs	

(a): field denominator

(b): short description if required
(c): the type of value (text, numbers, further subdivisions), fixed options are indicated with "-"
(d): additional clarification of values
(e): 'Unknown' is given when the required information is not available from the ESN's website or reports

Table 10: List of information elements to describe an ESP

Item [(a)]	Specification [(b)]	Values [(c)]	Meaning[(d)]
Programme			
Internal code	Internal code that allows easy referencing e.g. when establishing relationships between programmes and with organisations	ESP13-xxxx	Unique code automatically attributed
Official name of the ESP	The full legal name as registered (local language)	Name	
Other names	Abbreviation (local language); English translation and abbreviation, if available	Text	
Organisation	Link to the organisation that performs the programme	Internal code ESN, Name	
Contact person	Name and function of a contact person	Text	
Coordinates	Contact details	Text	
Website	Website address	Text	
Funding	Known sources of funding	- Company - Membership - Grants - Sponsoring - Unknown (e)	
Member of other programme(s)	Relationship with other programmes	Names and/or internal code(s) other ESPs	
Grouping of other programme(s)	Relationship with other programmes	Names and/or internal code(s) other ESPs	
Monitoring Subject			

Item [a]	Specification [b]	Values [c]	Meaning[d]
Description	Brief description of the objectives and activities of the programme.	Text	
Purpose	The reason for conducting the programme may be a legal requirement	- Legal requirement - Other	
Protection goal(s)	Indication of which protection goal(s) are addressed by the programme (more than one option can be indicated)	- Biodiversity	Further specified as birds, mammals, fish, reptiles, arthropods (among which insects), plants and fungi
		- Human health	*e.g.* occupational diseases
		- Animal health	*e.g.* performance of livestock
		- Plant health	*e.g.* prevalence of certain pests
		- Soil function	*e.g.* organic content, pH
		- Water quality	*e.g.* pollution with nitrates, O_2 level
		- Air quality	*e.g.* pollution with NO_x
		- Sustainable agriculture	*e.g.* data on IPM, abundance of pollinators
		- Others	
Influencing factor(s)	Indication of which influencing factors are addressed by the programme (*i.e.* those factors that may contribute to a change observed in one of the protection goal related endpoints) (more than one option can be indicated)	- Agronomic practice	*e.g.* fertilizer use, ploughing vs. no-tilling
		- Plant protection	*e.g.* pesticide use in a specific crop
		- GMO cultivation	*e.g.* acreage, areas of cultivation
		- Land use/ management practices	
		- Other environmental conditions	*e.g.* weather records, climate change
		- Other human influences	*e.g.* urbanization, industry
Methodology			
Data collection	Method of sampling or data collection	- Standard protocol	According to defined protocols
		- At random	
		- Unknown	
Protocol description	Short description of essential parameters	Text	*e.g.* transect method, 1km long, 200m sections, etc.
Data collectors	The people doing the field work	- Trained, professionals	
		- Trained, volunteers - Untrained volunteers	

Item [a]	Specification [b]	Values [c]	Meaning[d]
		- Unknown	
Geography	Where does the programme collect information?	- European-wide	All EU Member States
		- Member State(s)	Selection of Member States
		- Regional	Federal states, provinces, geographic regions within a country
		- Local	Municipality(ies)
		- Unknown	
Geography: countries	Specifying countries	List of countries to choose from	
Geography: specific	Name of country, region, community	Text	
Geography: standardized format	The programme may by complying with international standards when collecting and publishing data.	- INSPIRE - GBIF - Other - Unknown	
Area type	The programme may be restricted to certain types of areas	- Protected - Agricultural - Other - Unknown	
Sampling: spatial distribution		- Single site	
		- Multi-site, even distribution	
		- Multi-site, uneven distribution	
		- Multi-site, rare habitats	
		- Unknown	
Sampling: choice of sites		- Free	
		- Grid-based	
		- Stratified random sampling	
		- Other	
		- Unknown	
Sample size	Sample size in analysis		number of samples, or *e.g.* quadrant size
Parameters	Identification of number of parameters or subjects,	- Single species/ parameters	

Item [(a)]		Specification [(b)]	Values [(c)]	Meaning[(d)]
		and which, are monitored	- Multiple species/ parameters	
			- Unknown	
	Data	Type	- Binary	Presence/absence
			- Count	*e.g.* species abundance
			- Continuous	*e.g.* soil moisture content
			- Unknown	
		Range - typical values for the data collected		*e.g.* average species count (12 starlings) or plausible data range (40%-75% soil moisture content)
	Temporal Info	Temporal resolution		*e.g.* weekly, monthly, annually
		Time period covered by latest report		*e.g.* 1960-2010
		Repeat Data (data collected from same sites over time)		Yes, same sites repeated (while new ones may be added), or No, different sites over time
	Network Amalgamation	Single scheme or amalgamation of multiple schemes?		Yes, several (sub-)programmes (*e.g.* inventorying habitats comprises monitoring of flora, soils, water quality etc. or No, only one scheme
	Analysis	Validation - Whether data are checked first before analysing	- Validation of data	
			- Unknown	
		Operator - Whether the analysis is done in-house or somewhere else	- In-house	
			- Unknown	
		Method	- Univariate	Analysis of one metric at a time
			- Multivariate	Analysis of multiple metrics simultaneously
			- Unknown	
		Frequency	- Bi-Annual	Twice a year *e.g.* in spring and in autumn
			- Annual	
			- Multi-year	
			- Unknown	
Reporting				
	Language	Language used in reporting		
	Reporting	Availability of reports	- Website - Scientific Journals	

Item [a]		Specification [b]	Values [c]	Meaning[d]
			- Books, etc. - No data available	
		Links to reports, if available		
	Frequency	When regular reports are issued		
	Historical reference	Data available since when?		*e.g.* to allow identification of long-term trends
	Future perspective	Longevity of the ESN - Estimation whether data will be available in the future		*e.g.* research projects may have an end date
		Future plans - whether sampling method, area, etc. will be changing		

(a): field denominator
(b): short description if required
(c): the type of value (text, numbers, further subdivisions), fixed options are indicated with "-"
(d): additional clarification of values
(e): 'Unknown' is given when the required information is not available from the ESN's website or reports

3.2.3. Organisation of information

A web-based password-protected tool was developed for recording information retrieved from the websites and literature, allowing multi-user input and access to the stored information. The structure of the tool reflected the difference between ESNs and ESPs, each having separate entry screens, while allowing the establishment of links between them. ESN and ESP entries automatically obtained independent unique codes. Finally, the editing history was fully captured so that changes in the inventory could be traced back.

One of the benefits of the web-based tool was a more rigorous standardization of the parameters that were into inventory. This is also a pre-requisite for enabling future searches and is crucial for quality control.

Export options from the web-based tool included both individual sheets per ESN or ESP as well as overviews and statistics. The parameters of the information exported was determined in relation to the needs of Section 4 and the statistical analysis and modelling performed. As agreed with the project's promoter, the inventory was provided in a spreadsheet format together with the final report of the project.

3.2.4. Collection of information

Following the initial identification (name, website and contact), the ESN's website, if available, was consulted to retrieve the missing information elements. Additional searches were required to respond to two challenges that were frequently encountered:

- The initial information was incomplete or incorrect;
- The initial information was inconsistent (e.g. the same ESN or ESP were referred to under different names and in different languages).

Once the ESN or ESP was accessed, the publically available information was analysed and an effort was made to extract the information elements. Rather than inviting ESNs to enter this information on their network and the observations they make, it was decided that Perseus would perform this initial categorization of data and data entry. The benefits of this approach included:

- Ensuring proper structuring of the information that was already available;
- Limiting the time required from ESNs (and thereby enhancing the possibility for obtaining their cooperation and contribution);
- Avoiding the bias that the inventory is limited to those ESNs that reply on an inquiry.

In addition to information on the webpages, manuals and reports were reviewed. Web addresses to (the latest) reports were captured and reports were stored electronically for later reference. Whenever websites were encountered that were only available in a local language not understandable to one of the project partners, an electronic translation tool was used to navigate the site.

As agreed with the project partners the ESNs operating EU-wide were given first priority, followed by ESNs that have national coverage.

Subsequently, the ESNs were directly invited to verify the information that had been captured, to correct data and to provide additional information. This invitation included a project support letter from EFSA, a short explanation of the different information elements in the inventory and the specific entries for the ESN and its related ESP(s). Any feedback was verified (if possible) and subsequently the records were updated in the web-based tool. With data inclusion continuing until the end of the project, ESNs/ESPs that were included after 16 December 2013 were not invited to react anymore.

An exception was made for information captured from the EuMon Database. As the information in this database is entered directly by the ESNs and ESPs, it was not deemed necessary to seek additional confirmation again.

3.3. Quality assessment of data and analysis

To allow a structured assessment, the project partners developed classification criteria to position the ESPs in relation to an ideal situation (monitoring data collected according to a fixed monitoring methodology for all performers, repeated at appropriate intervals at the same location by the same performer, in order to be able to observe changes, validated data and adequate, powerful statistical analysis, availability of data). The quality criteria were determined taking into account the Eurostat quality framework[11].

The project team recognized that ESNs/ESPs might perform highly relevant research without fulfilling all the criteria that would make them suitable to support PMEM. In consequence, the assessment was not expected to provide a value judgement on "good" or "bad" organizations or methods, rather to formulate a number of indications/recommendations that users (e.g. EFSA, authorities) can take into account when considering to use information collected by an ESN/ESP. These recommendations were to be standardized and directly linked with the information in the inventory.

3.4. Results

3.4.1. An Inventory of Environmental Surveillance Networks in Europe

3.4.1.1. ESN Identification

Existing lists established by other initiatives were expected to present a major source for identifying ESNs/ESPs. Table 11 provides the results of reviewing these initiatives in relation to the purpose of this investigation into potentially useful networks for PMEM.

Table 11: Input based on existing ESN lists

Initiative	Description	Conclusion for project
LTER-Europe	- National networks of the majority of EU Member States (41 organizations including some non-EU ESNs).	- No individual ESPs identified. - Partner organisations were included in the inventory as ESNs.
JRC-FATE	- Monitoring chemicals such as fertilizers, pesticides and herbicides, and impacts of these pollutants in terrestrial and aquatic ecosystems.	- No individual ESPs identified. - Data are presented as interactive maps.
BISE and WISE	- Portal to national biodiversity reporting activities and datasets.	- No individual ESPs identified.
EIONET	- Portal to national biodiversity reporting activities and datasets (440 partner organisations).	- 4 of the Topic Centres included in the inventory as ESPs. - Partner organisations were not

[11] http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/ess%20quality%20definition.pdf

Initiative	Description	Conclusion for project
		included.
6th Framework programme ALARM - focal field site network (FSN)	- Local research projects, each of them limited in surface, but as a whole evenly geographic distribution. The effect of land use is studied in pairs of 4x4km² squares, one dominated by natural or semi-natural communities (but including some agricultural land) and the other with intensive agricultural usage.	- Due to the limited coverage, the local research projects were not included in this analysis.
6th Framework programme EuMon	- EU-wide gathering monitoring methods and systems of surveillance for species and habitats of Community interest. - In 28 countries monitoring organisations were identified: for 25 of the then 27 Member States, plus Norway, Croatia and Macedonia. - The major part (97) was located in the United Kingdom followed by Poland (45) of 327. At present 643 monitoring schemes are listed.	- Partly these ESNs overlap with other existing lists. Where possible inventory data were compared and updated with data obtained by EuMon via questionnaires. - New ESNs were added following a search on the world wide web, as often no web address was indicated in te EuMon database. This proved particularly difficult when the network name was not given in the local language.
European Biodiversity Observation Network (EUBON)	- Lists about 30 partners, of which 25 EU partners, each with a short presentation and web address	- Included in the inventory.
EuropaBio collection of ESNs	- Applicant's initiative for addressing existing environmental networks as part of PMEM general surveillance. - This list contains networks from some West- and South-European countries and from a few Member States in the eastern part of the EU.	- Agreement established for having access to information. - Data were reviewed, updated to the latest information (e.g. reports, contact persons, web addresses, etc.) and included in inventory.

Individual partner organizations that are members of the "Long term ecological research" (LTER) network were included as separate ESNs in the inventory. Further research on the specific projects in each country were not prioritized as the majority of LTER projects in Europe are site-specific or involve a small number of sites.

JRC-FATE has several activities each year concerning pollutants in water. However, no data on concrete projects were displayed on the website. In consequence, no ESNs were identified based on this source.

BISE and WISE are valuable accession points to datasets, but did not allow for identification of underlying individual programmes. To fit in our approach these portals themselves were included in the inventory as ESPs linked to the European Commission, DG Environment as ESN.

EIONET is the portal for the European Environmental Agency (EEA). Four of the Topic Centres were included as ESPs (on air quality, water quality, biodiversity and land use/land cover). The partner organisations were not included in the inventory. The Habitats Directive[12] asks Member States to monitor habitat types and species considered to be of Community interest. Complying with Article 17 of the Directive Member States prepare reports every 6 years. EEA and ETC/BD perform a Quality Assessment and Quality Control on the data in these reports before accepting and publishing.

The research projects of ALARM were not given priority because of their limited spatial coverage.

The EuMon list contains local and Member State-wide environmental research programmes, monitoring species and habitats of Community interest. The organisations performing this research presented themselves by filling out a questionnaire. However, often no website or email address was given, hindering the search. Especially when the ESN name was provided in English instead of the local language by which it is known, it was difficult -and in many cases impossible- to trace it on the world wide web. In exchange with the EuMon coordinator an attempt was made to solve this problem. As the coordinator was not entitled to pass on data protected personal information, he contacted the networks on our behalf. Despite a reminder, none of them replied. From this list all ESNs and ESPs were included, although only providing basic information. Further updating was done for those ESNs operating at country level. A link with the data on the EuMon website was included in the inventory.

EUBON is an EU-FP7 project working on integration, harmonization and standardization of biodiversity information from on-the-ground to remote sensing data. The partners are members of networks of biodiversity data-holders, monitoring organisations, and scientific institutions. The EU partners were included in the inventory.

The EuropaBio list includes about 140 ESNs that are already screened and retained on the basis of relevance (covering protection goals and/or influencing factors) and availability of information.

In the design of this project, the input from authorities was judged to be extremely relevant as it reflects the view of the risk managers on surveillance. Through the dual approach by EFSA and EC's Biotechnology Unit information was received for Belgium, Denmark, Estonia, Finland, France, Germany, Greece, Ireland, Poland, Slovakia, Spain and Sweden. All of the inputs were included in the inventory.

Finally, based on their own research the project partners identified additional ESNs.

As expected, there was some overlap in information obtained from the various sources. Figure 3 illustrates that authorities indicated a limited number of ESN/ESPs (26) that were identified in the existing lists established by other initiatives. The major part of the ESN/ESPs indicated by authorities had not yet been inventoried by any of the existing initiatives. Some ESN/ESPs identified by project partners revealed to be included in the existing initiatives. The lack of overlap between the input from authorities and those identified by project partners can be attributed to the fact that the project deliberately searched for complementary ESN/ESPs. Finally, it must be highlighted that even among the lists available from the existing initiatives much overlap occurred. Removing double indications from more than 1000 ESN/ESPs obtained by compiling all the initiatives resulted in a reduction of over 25%. Unexpectedly, this elimination was highly resource demanding, as the same ESN/ESP had been referred to in differently ways in different lists and this was only revealed through a more detailed and time-consuming investigation.

[12] Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. OJ L206, 22.7.1992, p.7–50.

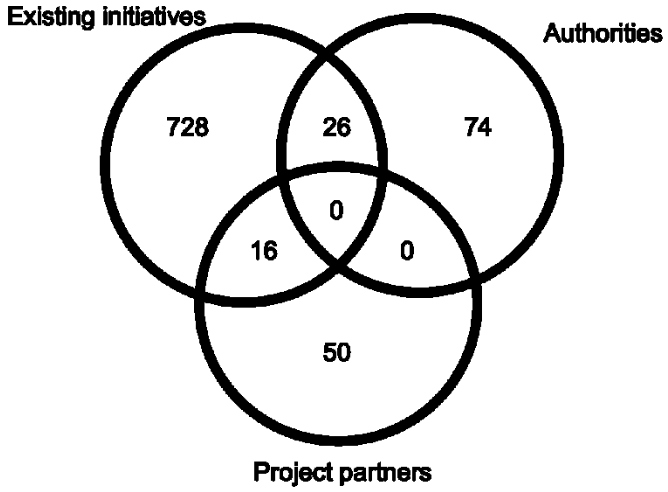


Figure 3: Diagram presenting the number of ESN/ESPs obtained via the different sources

A summary of the ESNs and ESPs that were entered per country and per topic is presented in Table 12 and Table 13.

Table 12: Number of ESPs per country for the specified protection goals (note one ESP may cover several protection goals and/or influencing factors and may therefore be counted more than once)

	Biodiversity	Biodiversity: Mammals	Biodiversity: Birds	Biodiversity: Insects	Biodiversity: Other Arthropods	Biodiversity: Amphibians	Biodiversity: Reptiles	Biodiversity: Fish	Biodiversity: Plants	Biodiversity: Fungi	Human health	Animal health	Plant health	Soil function	Water quality	Air quality	Sustainable agriculture
EU	3	3	9	4	1	2	2	2	2	1	1	3	2	4	3	1	4
Austria	1		4	1													
Belgium	20	6	21	12	5	5	3	2	4	2				1			2
Bulgaria	3	1	2			1	1										
Croatia		1				1											
Cyprus	1	3															
Czech Republic	3	1	4	2	1	1	1	1	1		1	2	2	1	2	1	
Denmark	3	1															
Estonia	4	3	4	2	1	1	1		3					1	2	1	
Finland	1	4	7	5	1				1					1	1		
France	18	21	51	16	3	8	6	7	9	1	1	2	2	10	2		4
Germany	14	9	14	9	2	6	4		14	1	3	5	4	5	8	2	1
Greece	31		2				1	1			3	3	1		5	1	
Hungary	11	8	9	9	1	2	3	2	4	1	2	2		1	2		2
Ireland	6	3		1					1								
Italy	2		2	3	1				2								
Latvia																	
Lithuania		14	13	2	1			2									
Luxembourg			1														
Malta	1																1

	Biodiversity	Biodiversity: Mammals	Biodiversity: Birds	Biodiversity: Insects	Biodiversity: Other Arthropods	Biodiversity: Amphibians	Biodiversity: Reptiles	Biodiversity: Fish	Biodiversity: Plants	Biodiversity: Fungi	Human health	Animal health	Plant health	Soil function	Water quality	Air quality	Sustainable agriculture
Netherlands	8	5	6	14	2	1	2	2	5	2	1	2					3
Poland	24	25	29	15	1	3	3	3	21				1	1	1	1	1
Portugal			2								1	1			2		
Romania	3		3	1							2	2	1	2	2	1	3
Slovakia	1		7	4	1	1	1	1	2			4	3	3	5	2	1
Slovenia	1	2	5	4				1									
Spain	47	1	19	7		2	2		5		1	2	3	4	2		6
Sweden	6		4	1					2	1	1		2	5	7	3	2
United Kingdom	74	3	2	7		1	1		3	1	4	5	3	6	9	5	6

Table 13: Number of ESPs per country for the specified influencing factors (note one ESP may cover several protection goals and/or influencing factors and may therefore be counted more than once)

	Agronomic practice	Plant protection	GMO cultivation	Land use/ management	Other environmental conditions	Other human influences
EU	7	3	1	2	6	3
Austria						
Belgium					1	
Bulgaria						
Croatia						

	Agronomic practice	Plant protection	GMO cultivation	Land use/ management	Other environmental conditions	Other human influences
Cyprus						
Czech Republic	2		1		3	
Denmark					1	
Estonia	2			3		
Finland				1		
France	11	1	1	4	5	2
Germany	8	4	2	3	5	2
Greece	1	1				
Hungary	2	1		1	1	
Ireland						
Italy						
Latvia						
Lithuania						
Luxembourg						
Malta	1					
Netherlands	3		1		10	1
Poland	1			1	1	
Portugal	1		1	1	2	
Romania	5	3	2		2	
Slovakia	6	1	1	2	2	1
Slovenia						
Spain	11	5	4	2	4	
Sweden	7	4	1	3	1	1

	Agronomic practice	Plant protection	GMO cultivation	Land use/ management	Other environmental conditions	Other human influences
United Kingdom	7	3	4	4	12	6

3.4.1.2. Data extraction

For each of the ESNs the information as presented in Table 9 and Table 10 was compiled. Priority was given to EU and countrywide ESNs. Furthermore, ESNs/ESPs suggested by authorities and ESNs/ESPs on existing lists that already contained more of the information elements (e.g. EuropaBio list) were first entered in the inventory.

Several different factors interfered with the information collection:

- Many ESNs were involved in more than one ESP. This was already anticipated by splitting the inventory of ESNs and the one for ESPs. Nevertheless, the relationships between ESNs and ESPs could be very complex and time-consuming to unravel.

- Some ESNs were grouped in Europe-wide umbrella organisation. Data were therefore summarized and presented for the central European organisation. While the format of the inventory anticipated hierarchical relationships, it was not clear how the exact contribution of each level could be properly described. E.g. if the Europe-wide organisation reports on amalgamated results based on observations by local ESNs, the role of the local ESNs may be difficult to discern.

- Alternatively, ESNs worked in collaboration with other organisations, each performing specific tasks, making the interrelationships very complicated. Often monitoring data were stored at yet another website (gateway or portal) or raw data was not available through the website. This further complicated searches for more detailed information on methodology and reporting as this was scattered over different organizations and websites.

- Data centres (gateway or portal) operated at national or European level, gathering information from many ESPs and even other data centres. Although they made validated (raw) data available in some form and under certain conditions, no information was available on the data collection methods.

- Essentially all identified ESNs needed to be searched for and reviewed one by one, whether they were already listed or newly identified for this project. Often only a website address was available and potential ESPs needed to be searched for individually. Even so in case the list contained extra information, this also had to be checked and the information transformed to meet our data format.

The following examples illustrate the complexity and difficulty to assign a specific ESP to an ESN and to collect information.

Example 1: Scattered information

The Dutch 'Stichting Veld Onderzoek Flora en Fauna' (VOFF)[13] co-ordinates 10 ESNs and collaborates with several other organisations. The ESNs have their ESP data analysed by the 'Centraal Bureau voor de Statistiek' (CBS[14]). The method of analysis is described on CBS' website. Results and reports can be consulted at the website of the 'Compendium voor de Leefomgeving' (Compendium for the Environment) that also reports on the environment using data from other organisations. Raw data

[13] www.voff.nl/
[14] http://www.cbs.nl/

of the VOFF member programmes are stored in the Nationale Databank Flora en Fauna and are available upon payment via yet another website, the 'Natuurloket'[15].

Example 2: Complexity in collaboration
The French Ministry of Agriculture, Food and Forests (MAAF) has initialised an Agriculture Action Plan in collaboration with the Ministry of Ecology, Sustainable Development and Energy and the Natural History Museum. The intention is to promote biodiversity in rural areas, while reducing the negative impacts that certain agricultural practices may have. Monitoring agricultural biodiversity is one of the five pillars of the Plan, aiming to observe the evolution of biodiversity in relation to agricultural practices. The Natural History Museum was to define the protocols. The training of the local coordinators was the responsibility of the Natural History Museum, the University of Rennes 1 and the Permanent Assembly of the Chambers of Agriculture. Also LADYSS, a CNRS laboratory is involved. General information on the initiative is given on the Ministry's website. Further information, more detailed data and results are to be found on the website of Vigie-Nature, a science programme of the Natural History Museum. Although the respective websites inform about the collaborations, no links are made available. Hence, the interested reader has to search for protocols and results without clear guidance.

Example 3: Data portals
The data portals at the EU level collate information on the environment from a diversity of sources. Some initiatives may be subportals of a broader entity, *e.g.* WISE as part of EIONET that on its turn is a partnership between the European Environment Agency and its member and cooperating countries. Others list on their data page, next to their own collated results, other sources of environmental information, e.g. EuMon is on the BISE data list.

Transferring data from previous ESN lists to this project's inventory pointed out the necessity of maintenance and regular updating of the database. Indeed, the availability of public information may change over time. ESNs may modify their websites according to new needs, resulting in new web links. Contact persons and contact details may change. ESPs initiated by one ministry may move to another ministry after elections and reorganisation of responsibilities. Reports may no longer be available or replaced by more recent issues.

On the closure of the inventory (28th of February 2014), the basic information set has been collected for over 500 ESNs and almost 1000 ESPs. The following general conclusions could be made based on the priority list:

- <u>Objectives</u> - The aim and purpose of the organisation could usually easily be retrieved from the homepage or the "about us" page. The same was true for the items "protection goals" and "influencing factors" and the geographical coverage. The purpose of the monitoring effort may result from legal obligations (Habitats, Birds and Water Framework Directives, legislation on pesticides or harmful organisms, etc.[16]) and/or as a basis for conservation actions.

[15] http://www.natuurloket.nl/
[16] Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora (Habitats Directive). OJ L206, 22.7.1992, p.7–50.
Council Directive 2000/29/EC of 8 May 2000 on protective measures against the introduction into the Community of organisms harmful to plants or plant products and against their spread within the Community. OJ L169, 10.7.2000, p.1-112.
Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy (Water Framework Directive). OJ L327, 22.12.2000, p.1-72.
Directive 2009/128/EC of the European Parliament and of the Council of 21 October 2009 establishing a framework for Community action to achieve the sustainable use of pesticides. OJ L309, 24.11.2009, p.71-86.

- Monitoring methodology - For ESNs working with volunteers, information about monitoring methods that were used to collect samples or other data could be found in manuals or guidelines, usually in the local language. Method descriptions were also found in reports, network magazines and publications as well as literature. However, these were usually summaries and details were rarely available to allow determining appropriate statistical analysis even if the actual data were available. Unfortunately, for many ESNs, no information on methodology was discovered through the website searches.

- Contributors - ESPs received data from volunteers or professionals.

  - Some monitoring programmes recorded random sightings. They received data from any person, trained or untrained.

  - Most often fauna and flora data collecting involved volunteers. These received training using manuals, species description charts or books; and by attending field meetings. Sometimes an instruction video was accessible on the website. They could also have been supported via a local ESP coordinator.

  - ESNs working on more complex subjects *e.g.* habitats, soil composition or water quality, seemed to be exclusively the work of professionals. For these themes typically a more advanced knowledge was needed and it involved special equipment or laboratory analysis post sample collection.

- Geographical coverage: Almost half of the ESPs are either locally or regionally performed. The majority of these were retrieved from the EuMon database. They often monitor in protected areas only. One fifth of all ESPs include agricultural land.

- Spatial resolution - Most networks operated across multiple sites. Most ESPs on the EuMon list and all of the ESPs from the ALARM consortium studied only one or two sites each. When working with volunteers, attempts were made to evenly distribute the observation sites by allocating “grids” to persons, although this was not always feasible. As a result the density of data points coincided with human population density. Other ESNs left the choice to the individuals, but an accurate description concerning the location of the sampling point was always required. Professional organisations adapted the design of the observation sites to the statistical needs of the study.

- Temporal resolution - It was not always clear how frequently samples or data were collected. Some ESPs were very strict in setting dates for observations to be made: e.g. overwintering waterfowl was counted 4 times a year at particular weekends and this was repeated every year. Other ESPs seemed to collect information at random. On the other hand, inventories of habitat types, like a forest inventory, ran in cycles of 5-10 years, thereby monitoring different parts of the territory each year.

- Reporting - Data were most often presented in summary reports as tables, graphs or maps. Few ESNs provided details on statistical analysis, and even then only references to statistical

Directive 2009/147/EC of the European Parliament and of the Council of 30 November 2009 on the conservation of wild birds (Birds Directive). OJ L20, 26.1.2010, p.7–25.

Regulation (EC) No 1185/2009 of the European Parliament and of the Council of 25 November 2009 concerning statistics on pesticides. OJ L324, 10.12.2009, p.1–22.

methods were offered. Usually only a description of the results was presented together with trends over time. Organisations that served as a portal to databases did not provide reports.

- Access to raw data – Access to raw data was never possible directly from the website. Some organisations, predominantly in the UK (e.g. ECN Data Centre[17]), but also in the Netherlands (Natuurloket[18]), offered the possibility of requesting specific raw data sets. Permission was then sought with the data owners and this could include establishing a licence agreement. Apart from a few cases this option was not provided for ESNs in other countries. Furthermore, access to information/data could be restricted to members only, requiring a subscription and/or a financial compensation. In other cases the users were allowed to compile their own overview or summary.

- Time to report - This varied considerably among ESNs. Annual reports were issued in the year following the year of observation, but often there was a time lag of several years. Reports could cover one or more years. Books, usually an atlas describing the distribution of certain species or taxonomic groups, were infrequently published, whereas scientific publications were issued irregularly and only focussed on certain aspects of the data or study.

- Except for some research projects with a specific end date, ESNs intended to continue their work.

Case study 1: Access to raw data

Raw data for a particular region were requested from one of the identified ESNs. The request was accompanied, as required, by a short explanation of the purpose of the request (i.e. assess suitability for PMEM within the framework of this study).

As explained on the ESN's website, such requests need authorisation from each of the sponsors responsible for the sites for which data are requested. This process took 6 weeks. In response a license agreement was presented with the following use restrictions:

- *The data can only be used by the requestor (a personal, non-exclusive, non-transferable, single site licence).*
- *The data can only be used during a 1-year period.*
- *Only processed data, such as summaries, may be published and the data centre must be acknowledged.*

The ESN requested additional information on the purpose of the study and they indicated that an earlier initiative considering applicability for PMEM concluded that their data collection had not the right sort of coverage of e.g. GM crop areas. After signing the agreement and additionally assuring that no persons other than the licensee would use the raw data, the data were received 4 weeks later in an electronic format.

*During the exchange, it was stressed that the purpose of use and the type of user are essential elements in granting access. This particular ESN offers data free-of-charge for non-commercial purposes. A request from an authorization holders (*e.g. *for marketing pesticides or GM crops), that is a commercial organisations, would likely be considered part of a commercial activity and consequently access to raw data for PMEM would be refused.*

[17] http://data.ecn.ac.uk/access.asp
[18] http://www.natuurloket.nl/natuurloket

The case study shows that gathering data for the purpose of a regular PMEM report requires time and effort even with a 'receptive' ESN. A search for data from other ESNs/ESPs, with the challenges of language, tracing methods, etc. would only lead to obtaining summary data (graphs, maps).

Case study 2: Pilot surveillance project on honeybee colony losses

The project's promoter pointed out a European Commission initiative to protect honey bee health. Copies of guidelines and visit forms were provided.

Searching EUR-Lex resulted in identification of Commission Regulation No 87/2011[19] *that designated the ANSES laboratory at Sophia-Antipolis as EU reference laboratory for bee health. Commission implementing decision of 21 December 2011 and of 4 July 2012 provide for a financial contribution to support voluntary surveillance studies*[20].

By searching the world wide web guidelines were obtained and one report by the UK National Bee Unit explaining the purpose of the pilot project. No central website where one would expect general information, guidelines and reports (interim reports due by 1 March 2013,) or visit forms were found. Only after contacting ANSES we were informed that a central website was under construction[21]. *A password protected part is available at the time our report is presented.*

This example shows that although interesting and valuable monitoring and surveillance work may be performed in the EU, it is not readily identified. To be useful for PMEM, these initiatives should be easily and publicly accessible and it should be clear which parties can have access to information.

3.4.2. Quality assessment of data and analysis

The quality criteria are listed in Table 14 together with a proposal for an ideal situation, the corresponding items from Table 10 that can be used in this assessment as well as a recommendation in case the actual situation deviates from this ideal. There are two main reasons for observing a deviation:

- The information has not been discovered within the framework of this project. The systematic review of ESNs/ESPs revealed many topics for which no answer could be entered either because it was not available or because it was not found during the survey. In this case, it may suffice to inquire further with the ESN/ESP and to update the database.
- The ESN/ESP approach deviates from the proposed ideal e.g. as it was not targeted for PMEM. The approach that is followed may be adequate for the initial objective. While many ESPs provide very valuable information, they are likely to be unsuitable for PMEM or may need to modify certain aspects in order to also suit PMEM objectives and quality criteria. In

[19] Commission Regulation (EU) No 87/2011 of 2 February 2011 designating the EU reference laboratory for bee health, laying down additional responsibilities and tasks for that laboratory and amending Annex VII to Regulation (EC) No 882/2004 of the European Parliament and of the Council. OJ L29, 3.2.2011, p.1-4.

[20] Commission implementing decision of 21 December 2011 concerning the adoption of a financing decision to support voluntary surveillance studies on honeybee colony losses (notified under document C(2011) 9597) (2011/881/EU). OJ L343, 23.12.2011, p.119-120.

Commission implementing decision of 4 July 2012 concerning a financial contribution by the Union to certain Member States to support voluntary surveillance studies on honeybee colony losses (notified under document C(2012) 4396) (2012/362/EU). OJ L176, 6.7.2012, p.65-69.Commission Regulation No 87/2011 of 2 February 2011 designating the EU reference laboratory for bee health, laying down additional responsibilities and tasks for that laboratory and amending Annex VII to Regulation (EC) No 882/2004 of the European Parliament and of the Council. OJ L29, 3.2.2011, p.1-4.

[21] https://honeybees.anses.fr/
https://sites.anses.fr/en/minisite/abeilles/laboratory

this case a decision will be required on possibly using the information although the ESN/ESP may not completely fit the criteria.

Relevance is the primary criterion. The subject of research and monitoring should fit in one or more of the protection goals and/or influencing factors as listed in Table 10. Ideally both protection goals and influencing factors are studied at the same time and place. Similarly the geographical scope should fit with the intended area or at least with an area that is comparable with it.

To ensure representativeness, several parameters such as the number of survey sites (spatial resolution); the number of years data are collected; the temporal frequency of data (temporal resolution); must be adequately chosen for the intended purpose. ESPs that only study very specific niche habitats are judged to be less suitable for PMEM purposes. Furthermore, if relationships between factors are expected to be established, then associated variables may need to be collected at each site (for example weather and flowering data concomitant with butterfly monitoring). It must be stressed that there are no general indications, as much of the details will have to be determined in function of the monitoring subject. Timing and frequency will largely be influenced by seasonal events: e.g. some animals and plants may only be present/visible during a specific period of the year. Similarly sample dimensions must be determined in function of the monitoring subject: the smallest sample size reported in this survey 76mm^3 is suitable for airborne pollen observations.

Accuracy and reliability depend to a large extent on the knowledge and skills of the data collectors and analysts. Professionals but also trained volunteers are expected to have the necessary skills. Self-assessment of data collectors would be a valuable criterion to assess the reliability of the ESP data as well, but this aspect is never mentioned on EPSs' websites. Furthermore, a defined protocol that is strictly followed by every performer improves the data quality.

When volunteers are free to choose the monitoring site, they often tend to opt for convenience by selecting sites close to their homes or particularly natural or attractive areas. Hence, the call for additional volunteers to fill the gaps is often seen on ESPs' websites.

Availability of raw data is considered very important as they provide detail on location and time. In reports data are already summarized and have the advantage that some expert interpretation in included. Reports usually become available in the year after the observation year. ESPs that report much later or infrequently will have limited value if PMEM is expected to provide an early warning system. Reports covering many years are able to demonstrate trends over time. Reports can be downloaded for free from the website, sometimes costs are involved. Raw data can be obtained on request for free, or is offered as a paid service.

ESPs that collaborate, or are part of umbrella organisations, usually have harmonized data collecting protocols and the same method of analysis (or have the data analysed centrally). This increases comparability and coherence.

Table 14: Matching quality criteria with list of items collected for ESNs/ESPs and recommendations

Quality criteria	Proposed "ideal"	Information for ESN/ESP	Recommendations
Relevance (Does it meet the needs?)	- Monitoring subject fits with the required protection goals or influencing factors	- Monitoring subject: Protection goals - Monitoring subject: Influencing factors	- Seek confirmation on fit with protection goals or influencing factors - Explore if ESN/ESP can adapt to fit better. - Explore if other ESN/ESP is better placed for the specific monitoring subject
	- Geography fits with the required geographical scope and landscape compartment.	- Methodology: Geography - Methodology: Geography (specific) - Methodology: Area type	- Seek confirmation on fit with geographical scope and landscape compartment. - Explore if ESN/ESP can adapt to fit better. - Explore if other ESN/ESP is better placed for the specific geographical scope and landscape compartment.
Representativeness	- A method to select sampling sites (*e.g.* grid-based, stratified random sampling) appropriate for monitoring objectives	- Methodology: Sampling - Choice of sites	- Seek additional information on sampling sites selection - Consider the impact of less representative data - Explore if ESN/ESP can adapt the sampling sites selection method
	- A dense and even distribution of collection points over the region of interest fitted to the specifics of the monitoring subject	- Methodology: Sampling – Spatial distribution	- Seek additional information on density and distribution of sampling sites - Consider the impact of less representative data - Explore if ESN/ESP can adapt density and/or distribution of sampling sites
	- Sample size fixed and justified in relation	- Methodology: Sampling – Sample size	- Seek additional information on

Quality criteria	Proposed "ideal"	Information for ESN/ESP	Recommendations
	to the intended observations and sensitivity of the analysis.		sampling size determination and justification - Consider the impact of less representative data - Explore if ESN/ESP can adapt the sampling size determination method
	- Regular observations at the time relevant for the monitoring subject	- Methodology: Temporal info – Temporal resolution	- Seek additional information on regularity and distribution of observations - Consider the impact of less representative data - Explore if ESN/ESP can adapt frequency and/or timing of observations
Accuracy & reliability (the closeness of computations or estimates to the exact or true values)	- People professionally collecting the information, or at least trained to collect information in a standardized way	- Methodology: Data collectors	- Seek additional information on data collectors and their training - Consider the impact of non/less accurate conclusions - Explore if ESN/ESP can rely on other data collectors
	- A well-defined and appropriate data collection protocol is available	- Methodology: Data collection - Methodology: Protocol description	- Seek additional information on data collection protocol - Consider the impact of non/less accurate conclusions - Explore if ESN/ESP can adapt the data collection protocol
	- A validation step is included to correct data	- Methodology: Analysis - Validation	- Seek additional information on data validation - Consider the impact of non/less accurate conclusions - Explore if ESN/ESP can implement a

Quality criteria	Proposed "ideal"	Information for ESN/ESP	Recommendations
			data validation step
	- Data analysis is done according to a validated method adequate for the data set and performed by people skilled in the method. The method is clearly identified and is verifiable.	- Methodology: Analysis – Method - Methodology: Analysis – Performance	- Seek additional information on analytical method and how it is performed - Consider the impact of non/less accurate conclusions - Explore if ESN/ESP can implement a suitable analytical method
Timeliness & punctuality	- Regular reporting at frequency adequate for PMEM use. - Time lag between data collection and availability of information (data/report) is limited to allow effective response in case of undesired effect.	- Reporting: Frequency - Temporal info: time period	- Seek additional information on reporting frequency and time lag. - Consider the impact of time lag on conclusions - Explore if ESN/ESP can modify reporting frequency
Accessibility & clarity	- Raw data are available publically or upon request (possibly with limitations) and can be used for meta-analysis.	- Reporting: Availability of data	- Seek additional information on availability of raw data and associated conditions of use - Consider if information in other form of report is adequate - Explore if ESN/ESP can modify reporting frequency
	- Reports are easily accessible that provide contextual information, results and conclusions	- Reporting: Availability of reports	- Seek additional information on availability of reports - Consider if incomplete reports can be used for PMEM - Explore if ESN/ESP can modify reporting process
Comparability & coherence	- Temporal comparability - historical information allows comparison over an adequate period. Repeat data (same/ similar sites sampled every time) allow	- Methodology: Temporal info – Time period - Methodology: Temporal info – Repeat data	- Seek additional information on historical information and repeat data - Explore if ESN/ESP can prolong observations

Quality criteria	Proposed "ideal"	Information for ESN/ESP	Recommendations
	comparative analysis.	- Reporting: Historical reference - Future perspective: Longevity of the ESN	- Explore if comparable data can be obtained from a different ESN/ESP
	- Geographical comparability - The methodology used for data collection and analysis are compatible and allows integration with other ESNs/ESPs addressing the same subject	- Programme: Part of other network(s) - Programme: Grouping other networks	- Seek additional information on compatibility of methodology and data - Explore if ESN/ESP can become compatible, if required for specific subject
	- Domain comparability - The methodology used for data collection and analysis are compatible and allows integration with other ESNs/ESPs addressing different subjects	This subject was not covered	

In the optimal situation programmes would monitor certain protection goals and influencing factors at the same location at the same time. Although sometimes the monitoring protocols ask for this background information, this combination was not apparent from the resulting data. Only in the summary reports the effect of some influencing factors was occasionally mentioned.

In order to find ESNs/ESPs that are next close to the ideal that could serve PMEM, a selection was made using the items that correspond to these criteria:

- European-wide (a broad geographical area),
- Multisite, even distribution (with a dense and even distribution of collection points),
- Temporal resolution of at least 1 year (visited regularly, *e.g*. once each year).
- Standard protocol (the method of data collection was standardized),
- Protocol description available (documentation of the data collection),
- Trained professionals and/or volunteers (the people collecting the information were professional, or at least trained to collect information)
- Validation of data is essential.
- Analysis method: univariate or multivariate (documentation of data analysis methods was required).
- Availability of data (access to raw data either by request or publically available was an ideal; interpretative reports give extra value).

Although many programmes aim to have their data collection points evenly distributed, this was rarely the case (a call for more volunteers to fill the ‘gaps’ was often seen). Therefore, also ‘multisite, uneven distribution’ was taken up in the selection. Also, ‘validation of data’ and ‘univariate or multivariate analysis’ are only marked in the database, if this was mentioned somewhere on the website or in a report of the monitoring programme. Also, the continuation of the programme in the future needs to be assured.

Resulting programmes:

- Pan-European Common Bird Monitoring Scheme, PECBMS (ESP13-0110)
- Butterfly Monitoring Scheme, BMS (ESP13-0112)
- European Topic Centre on Water, Eionet Water (ESP13-0137)
- Constant Effort Sites Ringing (ESP13-0250)
- The Biodiversity Information System for Europe (BISE) (ESP13-0267)
- EEA, Air quality (ESP13-0227)

Leaving out the requirement on analysis method resulted in more EEA (ESN13-0179) programmes (programmes for which the analysis method was unclear or unknown). All of them are umbrella organizations meaning that they guide and support monitoring and that collection and analysis methods are not always described in detail. The EEA programmes are an example of this. On the other hand the EEA programmes mention the possibility to obtain raw data. Furthermore, multinational programmes all report in English. Probably due to the many parties involved the time to report is rather long: information usually does not become available immediately the year after the observations.

Case study 3: Close to ideal ESNs

The Pan-European Common Bird Monitoring Scheme and the Butterfly Monitoring Scheme have been selected as examples of ESNs that appear to fulfil many of the criteria needed for an "ideal" ESP

The Pan-European Common Bird Monitoring Scheme is organized by the European Bird Census Council in cooperation with BirdLife Europe. Likewise, Butterfly Conservation Europe coordinates the Butterfly Monitoring Scheme. Both are European-wide umbrella ESPs that support their affiliated organisations in data collection and analysis. They aim to harmonize their monitoring programmes with regard to the methodology and statistical processing. The line or point transect method is used. Each year preferably the same individual travels a fixed route and observations are made on sections or at fixed points for a fixed period of time. Observation points are some distance apart to avoid double counts. The TRends and Indices for Monitoring data (TRIM) software package is used to determine species trends over time. This analytical method, based on generalized linear models, incorporates procedures to cope with missing values, as may often occur in large scale monitoring schemes conducted by volunteers (Pannekoek and Van Strien, 2005). Statistics are computed by professionals. On the European Bird Census Council website trends are presented in graphs and figures for Europe as whole. Links to country reports are also available.

As the PMEM area of interest may be smaller than the EU, a geographic coverage at the level of Member States is equally important.

Selecting the inventory records for:

- Member state
- Standard protocol
- Protocol description available
- Trained professionals and/or volunteers
- Multi-site
- Observations at least annually
- Validation of data
- Univariate or multivariate analysis
- Availability of data

Resulted in bird data collecting programmes (ESP13-0087, -0098, -0157, -0194, -0219, -0247, -0248, -0251, -0256) and the UK Butterfly Monitoring Scheme (ESP13-0126).

Leaving out the temporal resolution requirement for the observations and analysis requirement, will include programmes that monitor e.g. soils (monitoring every 5-10 years, type of analysis not specified), but also other wildlife monitoring schemes for which it was not clear how the analysis was done (choice was: univariate, multivariate or unknown).

Examples are (randomly chosen):

- Inventaire, Gestion et Conservation des Sols (ESP13-0003)
- Riksskogstaxeringen (ESP13-0148)
- United Kingdom Cereal Pathogen Virulence Survey (ESP13-0253)

The major drawback is that information relevant to the quality criteria is either not available or hard to find. Good programmes may not be picked up making selections on the available set of information. An example is the Dutch VOFF (Stichting Veld Onderzoek Flora en Fauna)(ESN13-0051) and its member networks. They have their data analysed by the Centraal Bureau voor de Statistiek (CBS), which suggests sound analysis, but more details on the method is only available on CBS' website (they use TRIM,). The same is true for reports on results (indices on population development) that are to be discovered at yet another website: Compendium voor de Leefomgeving[22]. Furthermore, raw data of these programmes are available upon payment (Nationale Databank Flora en Fauna, ESN13-0268).

Another type of networks that is not found in this way are the data portals that only host data without describing the collection methods and without any analysis of the data. Examples are (randomly chosen):

- Artportalen, Sweden (ESP13-0002);
- National Biodiversity Network, UK (ESP13-0128);
- Enviroportal, Slovakia (ESP13-0192);
- Nationale Databank Flora en Fauna, the Netherlands (ESN13-0268).

Only few ESPs approach this ideal situation. Although many ESPs provide very valuable information, they are likely to be unsuitable for PMEM or may need to modify certain aspects in order to also suit PMEM objectives and quality criteria.

ESPs that collect information on both protection goals and influencing factors at the same sampling sites are sometimes found. Most programmes that use fixed routes for monitoring animals also record weather conditions and environmental changes, if observed in succeeding years. Examples are the Spanish Amphibians and Reptiles Monitoring programme (ESP13-0332) and Plan d'Action Agriculture (ESP13-0096). The 'Observatoire de la biodiversité en milieu agricole' (ESP13-265) is an example of a programme that explicitly aims to link biodiversity to agricultural practices using indicator species. However, the programme is still too young to draw conclusions.

However difficult, one way to circumvent the issue of few ESPs working at the same time on protection goals and influencing factors is to search for complementary networks. Therefore, the value of an individual ESP is also determined by the availability of a complementing ESP.

3.5. Conclusions

Given the broad diversity of PMEM subjects, identifying a large number of initiatives (ESNs and ESPs) that may provide relevant information was not unexpected. Incorporating input from authorities and risk managers, with lists established by previous initiatives and networks identified by the project partners, a comprehensive inventory was established. The effort that was required for this compilation underlines the need to make an inventory publically available that provides such information in a transparent and reliable way. Further unravelling the relationships, searching for more detailed information on methodology and reports, confirming the validity of the entered information and keeping the inventory up-to-date is beyond the scope of the project mandate. Yet, for any future effort to be meaningful these four elements would need to be taken up. Without maintenance, this inventory, like other initiatives, will quickly become obsolete.

The inventory supported the identification of some ESNs/ESPs that seem relatively well suited for PMEM. Information on methodology and data sets was provided as input to simulations of power analysis. Recommendations were formulated for dealing with ESNs/ESPs that may not completely correspond to the "ideal" profile for PMEM. In most cases, this will be addressed by providing details on information that is missing based in the analysis. In other cases, it may require adaptations if the data are required for PMEM.

[22] http://www.compendiumvoordeleefomgeving.nl/dossiers/nl0090-natuurgraadmeters.html?i=2-76

The recommendations are pertinent within the context of PMEM. The ESNs/ESPs may irrespectively provide valuable information within the framework for which they were established. Still, it is noticed that more and more initiatives aggregate in umbrella organizations operating at an international if not EU level. As such it can be expected that along with integration and harmonization also quality control and methodological rigour will be more widely shared.

The description of an ESN/ESP "ideal" to support PMEM is largely determined by the monitoring subject. Rather than looking at specific parameters, the assessment considered to what extent the choices were inspired by a systematic approach and were justified. Almost none of the ESNs/ESPs provided a justification on why a certain approach was taken.

Raw data are only exceptionally available, and even then in some cases under strict conditions of use. At this moment, this would prevent meta-analysis or other forms of additional evaluation. On the other hand, summary reports, communications, distribution maps and other publications are issued. While they represent a contextual interpretation, they may not allow subsequent comparison with other reports (different methodology, different parameters etc.). Further integration of information via umbrella organizations may also address this aspect and lead to more uniform reporting.

Finally, it has been noted that initiatives may be looking at different environmental compartments within the same geography. Most ESNs/ESPs observe one aspect (e.g. a specific protection goal). Yet in order to obtain an indication on a possible influencing factor concomitant data would be required. Usually these are not included by the same ESN/ESP and may not even be collected in the same environmental compartment. If it is the intention to link changes in protection goals to modifications in influencing factors, the overall strategy of the ESNs/ESPs will need to be revisited.

4. Recommendations for PMEM

4.1. Introduction

The first three objectives of this work focussed on collating and analysing information on statistical methods useful for PMEM and compiling a detailed inventory of existing networks that may provide useful data for PMEM. Having reviewed the statistical methods, it is now necessary to look at the data requirements needed to support PMEM of agro ecosystems and hence to investigate statistical and associated meta-data properties that define suitable networks. Further to this, it is important that, drawing on all work carried out in this project, recommendations are made and considerations given to how PMEM may use data collected from ESNs, what caveats exist when using these data and networks and the potential for future development of ESNs to enhance the suitability for PMEM.

Defining the appropriate meta-data to describe ESNs suitable for PMEM (see additional information collected in the inventory in Section 3) is perhaps more subjective than a strictly statistical perspective, but it is important to develop some criteria demonstrated by existing networks that would be regarded as important properties of a network, if it was to be used for PMEM. Properties such as the availability of data and access to methodological reports of data collection and data treatment are critical when trying to interpret the results of statistical analyses. This information is essential for assessing the relative importance of environmental drivers, and mechanisms and to distinguish them from spurious correlations that may be obtained when analysing data across multiple ESNs, if they are to be successfully used in post-market monitoring.

The statistical properties that determine the suitability of an environmental monitoring network and the data they collect for PMEM can be defined as the power that the network has to detect a potential change in the ecosystem. The change in the agro ecosystem could be a result of changes in crop species or variety, pesticide or herbicide use or different management practice. In this context one would be interested on the effect this change has on the environmental indicator monitored by the network. We, therefore, in this section of the report described our investigations into the power that networks may have to detect an ecosystem effect, by conducting a large scale simulation study, which covered a wide range of scenarios that describe and characterise various potential networks, the data they collect and environmental indicators.

4.2. Simulation study to assess potential power of ESNs in PMEM

We conducted a large scale Monte Carlo simulation study to investigate the statistical power that a network may have to detect a specific effect due to change in an agro ecosystem. Data characteristics were selected according to the different types and temporal distribution of data collected by representative networks. The simulation study tested the null hypothesis that there is no difference in trends over time of two levels of a factor against the alternative that there is a difference. The two levels of a factor are referred to subsequently as treatment and control and could be the uptake of some management option or the uptake of a particular product for example. Further details are provided below.

4.2.1. Methods and Algorithm

We considered environmental data gathered annually at a large number of sites, according to a standardised protocol. According to the inventory of identified networks, data collected annually is by far the most common data type and hence the focus of the simulation (Figure 4). Other regularly spaced time intervals do, however, fit into the same modelling framework, so the methods for annual data frequency will apply to other temporal scales and we can proceed without loss of generality whilst focussing on the most common occurrence. Initially, we considered count data, again identified as the clear dominant data type collected by the networks according to the inventory (Figure 5), which is typically the number of a certain species recorded by a volunteer observer on each visit to a specified site.

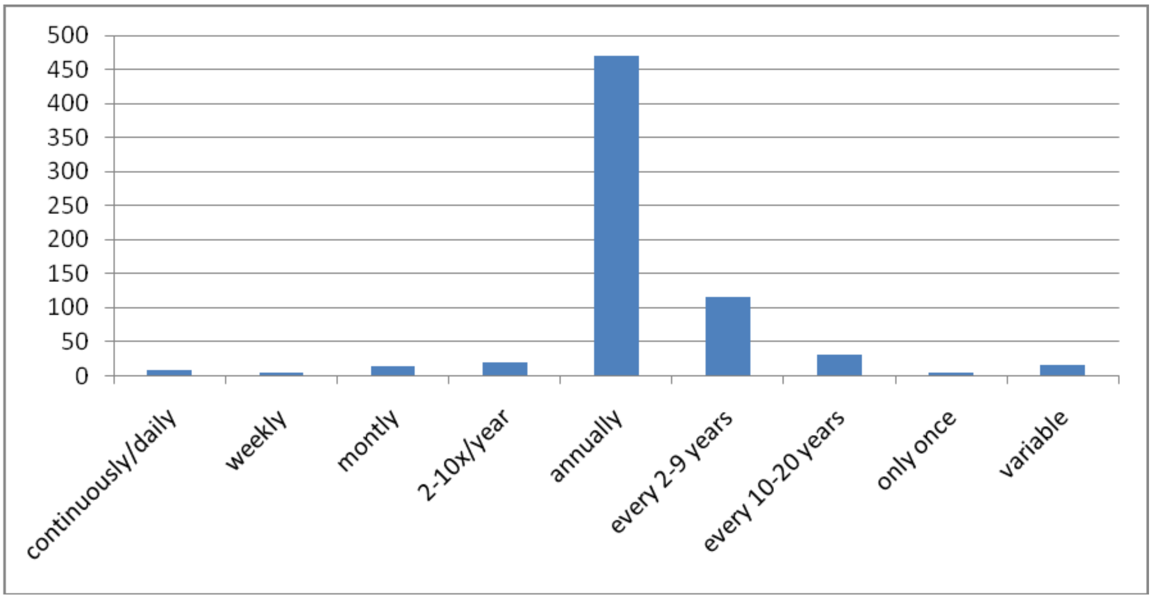


Figure 4: Barplot showing the frequency of occurrences of data collection found in the network inventory . Data collected annually is clearly the most common.

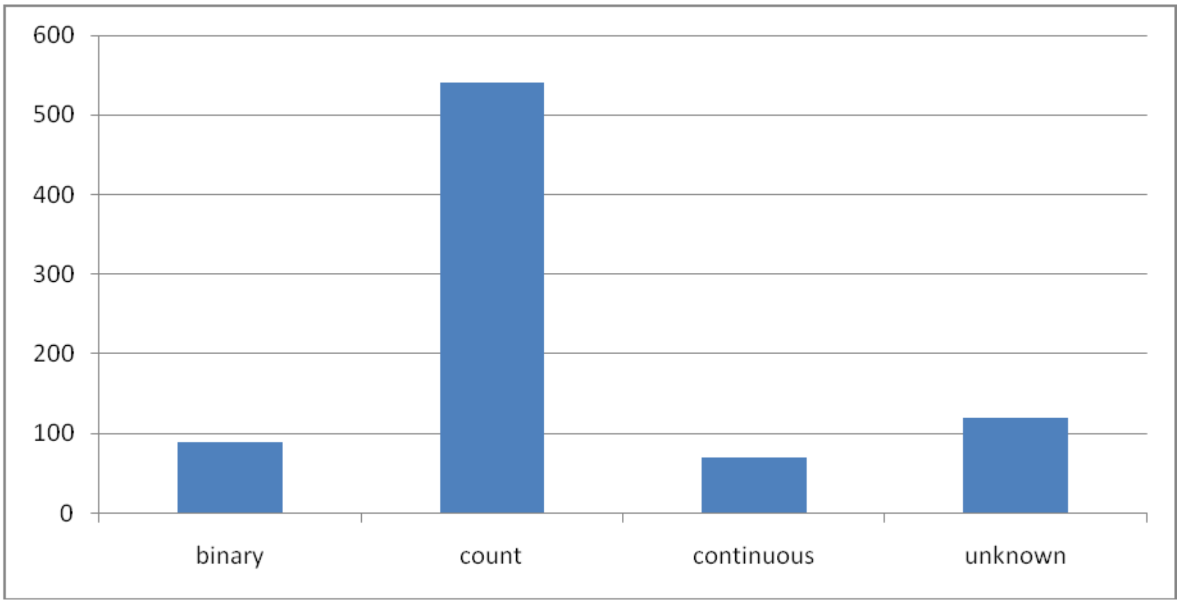


Figure 5: Barplot showing the relative frequency of occurrences of different data types according to the network inventory. Count data was clearly the most common type of data collected by the surveillance networks.

We denoted the count taken at site i in year j by C_{ij}. For data comprising of non-negative integers a Poisson model is intuitive. Freeman and Newson (2008) developed a model for such data, given in equation 1, in which the rate of growth (the ratio of two consecutive years' expected counts at the same site i) is a linear function of an annual 'effect', R, and a spatio-temporally varying covariate, P.

1) $$\log\left(\frac{\mu_{i,t}}{\mu_{i,t-1}}\right) = R_t + \alpha P_{i,t-1}$$

Thus R_t is the rate of growth at a site where $P_{i,t-1} = 0$, and α quantifies the effect upon growth of the covariate, so that if $P_{i,t}$ is a simple binary variable (0/1), growth where $P_{i,t-1} = 1$ is reduced (if α<0) to a fraction e^{α} of that at a site where $P_{i,t-1} = 0$.

We explored this model in an earlier report (Hails et al. 2012), where the predictor of growth, *P*, was a simple binary variable indicating whether the site in question was receiving a management regime or 'treatment' of some description. The coefficient of this variable (α in the notation as used in Equation 1, which we shall adopt throughout) is thus a measure of the change in numbers at a treated site relative to an untreated 'control'. A test of the null hypothesis α=0 is thus a test for the effect of the treatment. As the model of Freeman and Newson is a simple Generalized Linear Model, this is easily accomplished using an appropriate statistical software package. Here we use the glm function within the R statistical environment. In using this model framework in the simulation study we were consistent with the findings in this report of the statistical methodologies, which revealed GLMs to be the method most appropriate (see Section 1.4) and the most commonly used in the relevant literature.

Within the Hails et al. 2012 report we also explored a method of quickly estimating the power of this model in a variety of circumstances. This was achieved by simulating a large number of data sets assuming a particular survey structure and set of relevant parameters of interest. The process is straightforward, but computationally intensive even to estimate the power in a single hypothetical situation. We thus here further developed the calculation of a single 'generic' equation relating power to the set of variables used to simulate the data. This procedure again required considerable initial computational outlay, but once calculated the power could be estimated almost instantaneously for any set of realistic environmental and data conditions by using the estimated equation to make predictions. We here explored this option further to provide a useful tool that can establish an estimate of power that a network may have to detect the specific effect under investigation, during post market monitoring. The aim was to develop a model whereby we could predict power from a function of the key properties, initially as follows:

2) $$g(\theta) = a_0 + \sum_{i=1}^{k} a_i X_i$$

where $a_0 - a_k$ are unknown parameters corresponding to the baseline power and coefficients associated with the factors influencing power ($X_1 - X_k$), θ (0<θ<1) is the power (expressed as the probability of rejecting the hypothesis that α=0) and *g* is an appropriate link function constraining θ to lie in the valid range (0,1).

Power can of course depend upon a large number of factors, which when grouped together form a "scenario". Those factors considered here, based on previously collected information, were the number of sites in the survey, and the proportion of these to receive some treatment; the number of years over which they were visited and the proportion of visits that were missed; the annual rate of increase (assumed constant) at a control site, the average number of individuals (e.g. animals) counted at a site at the onset of the survey ('year 1') and the variability in this initial number between the sites; and of course the strength of the treatment effect α itself. For the simulations themselves to be useful in the context of PMEM and to provide information on suitable networks, it is important that the scenarios simulated are as representative as possible of the range of data that are available, as demonstrated by the network inventory. The scenarios should also cover factors concerning the actual influencing factor of agro ecosystem change (e.g. product release). In summary, a total of nine predictor variables, $X_1 - X_9$ as in Equation 2, are considered that are defined in Table 15. These predictor variables were chosen to represent different types of data and the different factors considered to influence statistical power. Where possible, the ranges used for each of the factors involved were cross referenced with the information from the network inventory, so that all realistic scenarios were covered in the parameter space of the simulations.

Table 15: Definition and range of nine predictor variables used to simulate data sets which would then be analysed.

Variable	Variable Name	Definition	Range considered
X_1	*Slope.*	The average annual rate of change ("growth") at an untreated site, $Slope_t$ (= *Slope*, assumed constant here)	-0.1 to 0.1
X_2	*N_sites.*	The number of sites visited	5 to 500
X_3	*Treated*	The proportion of sites treated (e.g. the proportion of sites that have adopted a particular product / management regime)	0.01 to 0.5
X_4	*Abundance_Mean*	The average log(abundance) at a site in year one of the survey	0 to 2
X_5	*Abundance_Var*	The variance of the log(abundance) measures at each site in year one of the survey; a measure of inter-site variability	0.1 to 5
X_6	*Missed*	The proportion of survey visits missed	0 to 0.5
X_7	*Duration*	The duration of the survey	2 to 10 Years
X_8	*Scale*	Scale parameter: A measure of the excess residual deviance, or overdispersion in the data (USED WHEN GENERATING COUNT DATA)	0 to 5
X_9	α *(alpha).*	The magnitude of the difference between the two treatments. This is change on the same scale as slope and therefore represents change over 1 year.	-0.1 to 0.1

Each scenario was therefore defined by a set of values for X_1 to X_9. We produced a large number of datasets for each scenario by generating random Poisson counts matched to the model of Freeman and Newson (2008) with the inclusion of site level differences and observation level error as shown in equation 3. The model was fitted to each scenario's 'set of datasets', and the proportion of significant values for α (indicating the identification of a treatment effect) from the replicates was recorded. This gave an estimate of power for each specific scenario.

3) $$\log\left(\frac{\mu_{i,t}}{\mu_{i,t-1}}\right) = R_t + \alpha P_{i,t-1} + \gamma_i + \delta_{i,t}$$

The simulation algorithm is as follows:

1. Generate expected counts over time by conditioning on realised values defined by the predictor set $X_1 - X_9$ and by using the population growth model defined by Equation 3.

2. Fit the model defined in Equation 1) to a realization of this expected time series using a GLM and test the hypothesis that α=0.

3. Repeat 1-2 using the same predictor set $X_1 - X_9$, generating 1000 distinct data sets and store the number of times a significant value of α is returned.

4. Repeat 1-3 using new draws from the stochastic distributions of the predictor set $X_1 - X_9$ to generate 200 values corresponding to how many of the 1000 simulated data sets returned a significant value of α.

With the process repeated 200 times, 200 estimates S_i of the number of significant results based upon 200 scenarios defined by the values X_{1i}-X_{9i}, i=1,2,… 200 were generated. Power θ was then estimated by assuming the S_i are binomially distributed, $S_i \sim \text{Binom}(1000,\theta)$ with θ related to X_1-X_9 via equation (2). Logistic regression was used to derive the equation predicting this power from the known values of the various parameters from which the data were generated.

The advantage of the method was that, once the parameters in (1) were estimated, they could be used quickly to derive an approximation of the power for any chosen combination of predictor variables X_1-X_9 and the relationship between power and any of these covariates was easily explored, without the need to repeat the computer-intensive simulations every time.

Freeman and Newson (2008) considered only the Poisson case. It was possible to extend the theory for Binomial models (applicable, for example, to binary data such as the simple presence/absence of records for a species at a site) or to continuous data assumed to follow a Normal distribution, as might be appropriate for measurements of many environmental variables. (Note that as the continuous data are generally positive, we have assumed that the log of the variable is normally distributed and the positively-valued data themselves are log-normally distributed). The model was easily extended to accommodate binary and continuous data, owing to its GLM structure, and the process was repeated for these data types.

4.2.1.1. Defining treatments and controls in PMEM

In this Section we have used the terms treatment and control to define the different regimes we were interested in and the different factors to test for an effect between. However, as we were concerned with post-market monitoring and evaluation, we did not mean control in the strict experimental design definition of the term. A true control would not exist for most ESNs as there is just country-wide surveillance. If one wants to determine what a true 'control' is, one needs to know where exactly the uptake of the agricultural product or practice takes place. Obviously, at the time of monitoring this would not necessarily be known. The term control here therefore refers to all the areas and observations that are essentially not in the treatment category. For PMEM, there will be information available on where the treatment areas were and what observations sat in this category. The analysis then compares these observations within the treatment group to the observations out of this group. It is this group that we have referred to as the control group. If the sample design is biased or the uptake of the treatment group is biased, this definition of control can have impact on interpretation. This is discussed further in Section 4.4.

There is a further scenario one could anticipate where there is no available information on where a particular product had been applied. In this case one would examine the trend in an ESN indicator before and after the introduction to see if they were equivalent. If the before and after time periods, were assumed to be the same, then this scenario is covered by the existing model defined in equation 3 where essentially we are testing for a difference in regression slopes between two “treatments”. For the scenario where only an introduction date was known, the treatment categories would be before and after this date. If the time periods were very different, then methods described and analysed in Section 2 can test for the hypothesised change in trend at the time of introduction.

4.2.2. Results

There were clearly unlimited ways in which estimates of power from these equations could be presented; indeed their value arose from the speed with which such a large range of data variables could be explored. A few examples, which encompassed the range of potential inputs according to the network inventory, are presented here. All graphs showed each individual factor plotted against statistical power to detect the “treatment” effect. It is typically accepted that power of approximately 0.7-0.8 is considered adequate for ecological data.

Figures 6-8 quantify the improvement in power of the Poisson model (applicable to count data) as the numbers of sites surveyed or the duration of the study is increased, and also shows that high power to detect change of reasonable magnitude is achievable from a study of realistic dimensions. Under the Binomial model (for binary data, Figures 9-10), power is much reduced and substantial numbers of sites were required to achieve more than modest power. This result was predictable as the information contained within the data was reduced: each annual observation at each site being merely a binary variable corresponding to presence or absence, and clearly considerable change in the population is required before it was manifest in such data. This implies that if an ESN has the ability to capture information on abundance, cover or value rather than simply whether or not something was present, then, for the same number of sites monitored over the same time period, far greater power exists to detect the effect of a specific treatment.

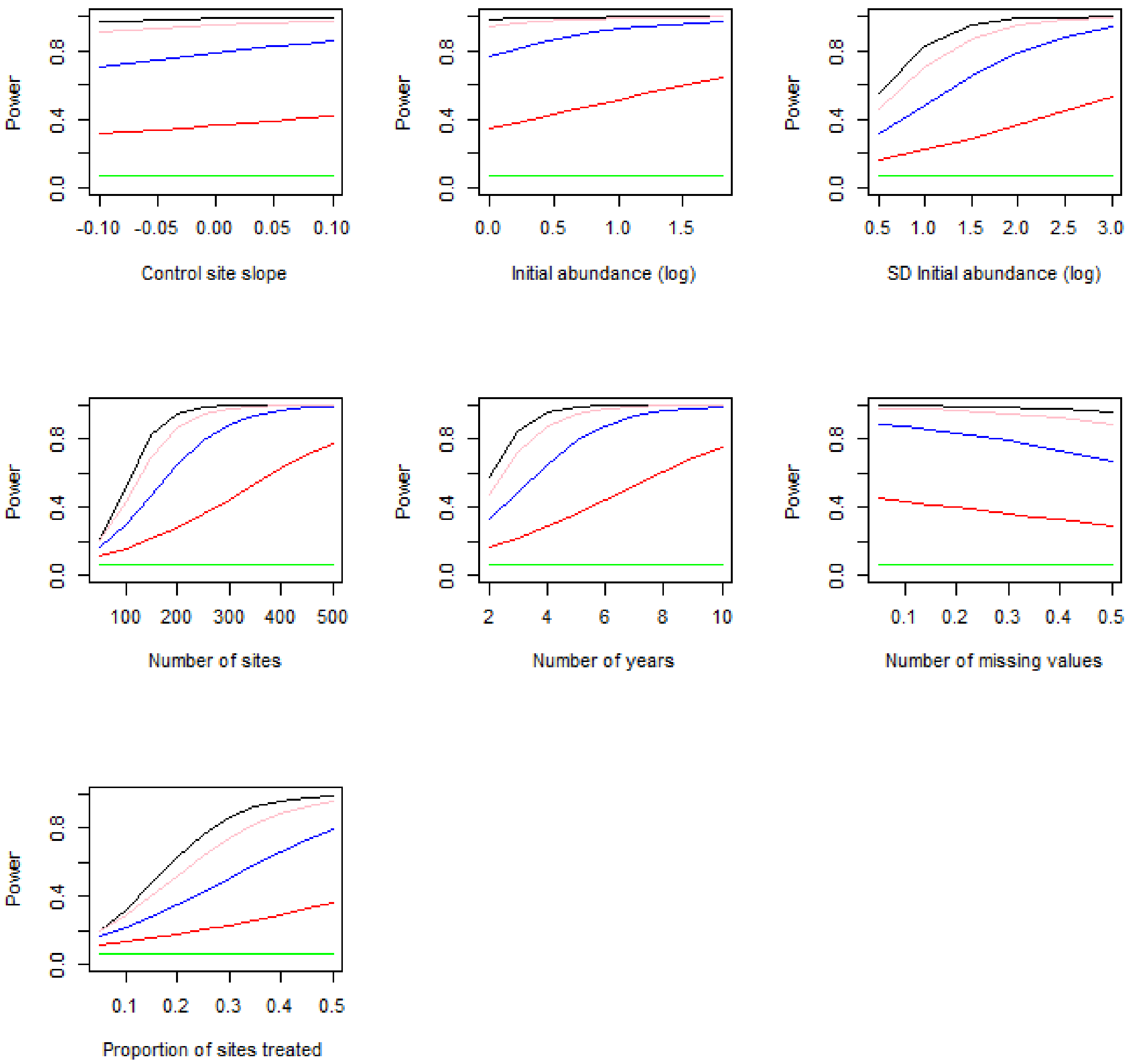
Note that for the Poisson/Binomial distributions, the variances were entirely determined by the model specification; this was not the case under the assumption of a Normal distribution and the variance is an additional free parameter to be estimated in the model, and was adopted as an extra predictor in the generic equation. The relationship between this error and the average value of the observations was critical in determining power. When the error is large with respect to the initial mean abundance, for example, power was low (Figure 12), although with an enlarged study with a greater number of sites, acceptable power was recovered (Figure 13).

Figure 6: Power of Poisson model to detect specified negative (a) or positive (b) treatment effects (α, in the notation of Freeman and Newson, 2008) versus assorted population- or scheme-specific parameters. All parameters are held constant, other than α and those varying along the x-axis in each diagram, these constant values being as shown below. Curves in (a) represent α = -0.1 (black), -0.075 (pink), -0.05 (blue), -0.025 (red) and 0 (no effect; green). Curves in (b) represent α = 0.1 (black), 0.075 (pink), 0.05 (blue), 0.025 (red) and 0 (no effect; green)

Other parameter values, where fixed:

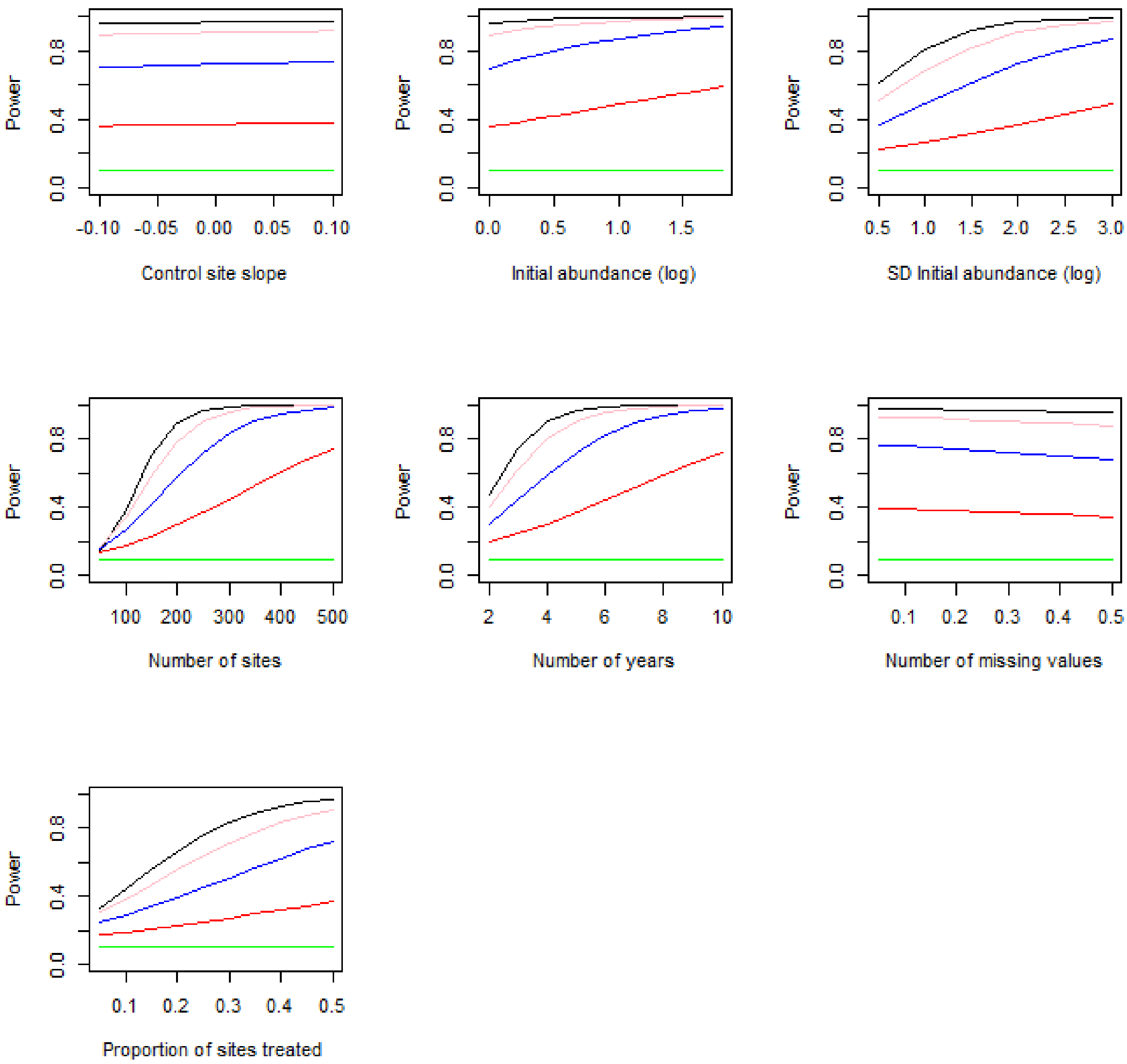
Control slope	0.0
Log Initial abundance (mean)	0.1
Log Initial abundance (SD)	2
No. sites	250
No. years	5
Proportion of visits missed	0.3
Proportion of sites treated	0.5

Figure 6 (a)



The graphs above show that, for this particular scenario of count data, the number of sites and the number of years of monitoring affect the power to detect an effect the most. Changes in these parameters greatly affected the power to detect change. The slope of the control sites had very little effect on power, which showed that even if there was a strong background relationship over time, an additional effect could still be adequately captured. Reasonable power (often considered to be 70-80%) is achieved in many cases, though rarely when α is of a small magnitude (e.g. red line representing α =-0.025). As the proportion of sites treated rose to the optimal value of 0.5 (which represents equal data in each of the treatment, control categories), the power significantly increased. For large effects (α=-0.1, black line) a very small proportion of sites treated (approximately 0.22) provided sufficient statistical power.

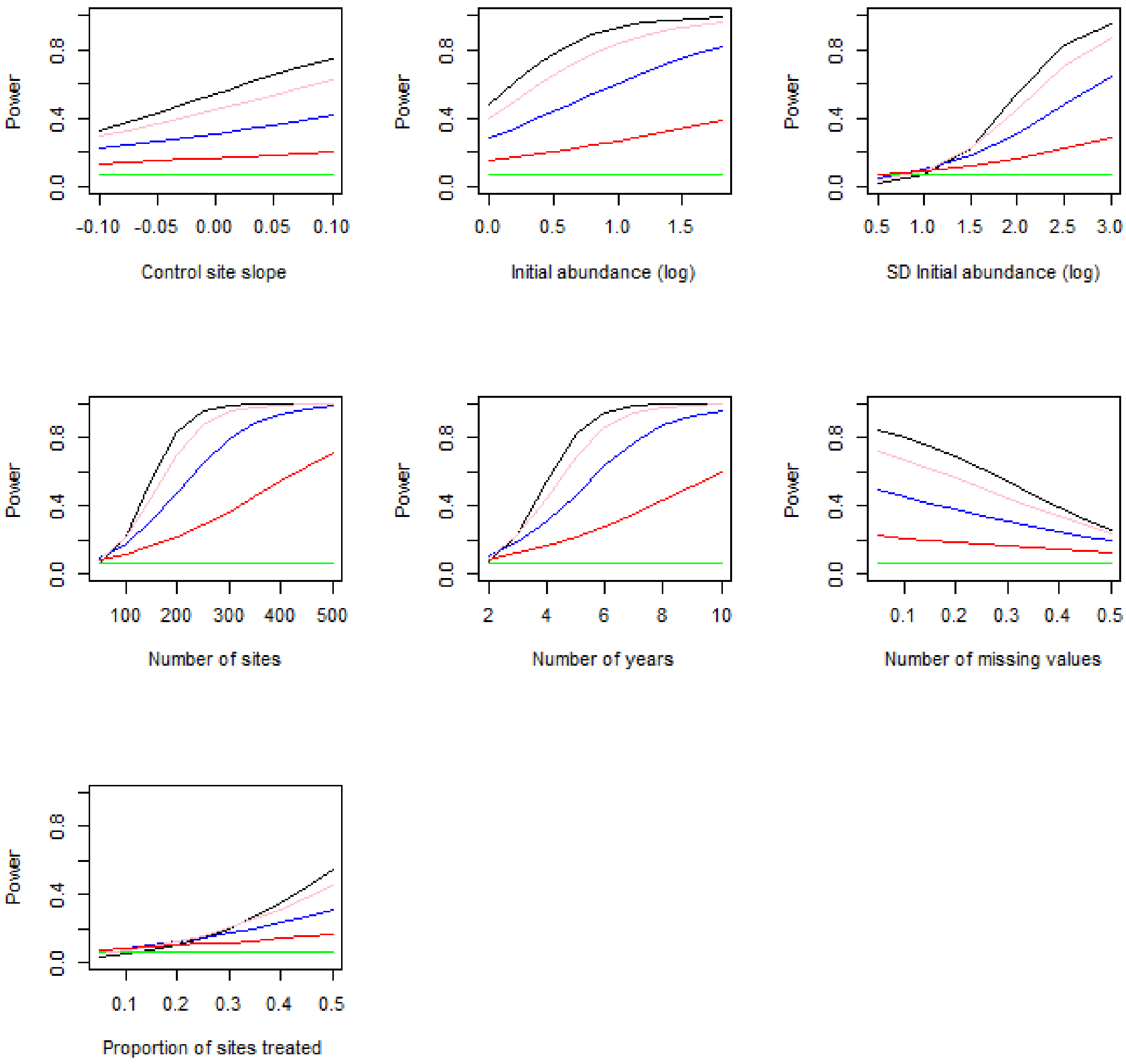
Figure 6 (b)



When the treatment effects were positive, i.e. when α>0, the results were very similar to when the treatment was detrimental. This mirroring was not unexpected, but as a separate model was built for the two cases, there was no reason that the results should be equivalent. Here we again see that the number of sites and the number of years of monitoring are the greatest influence on power.

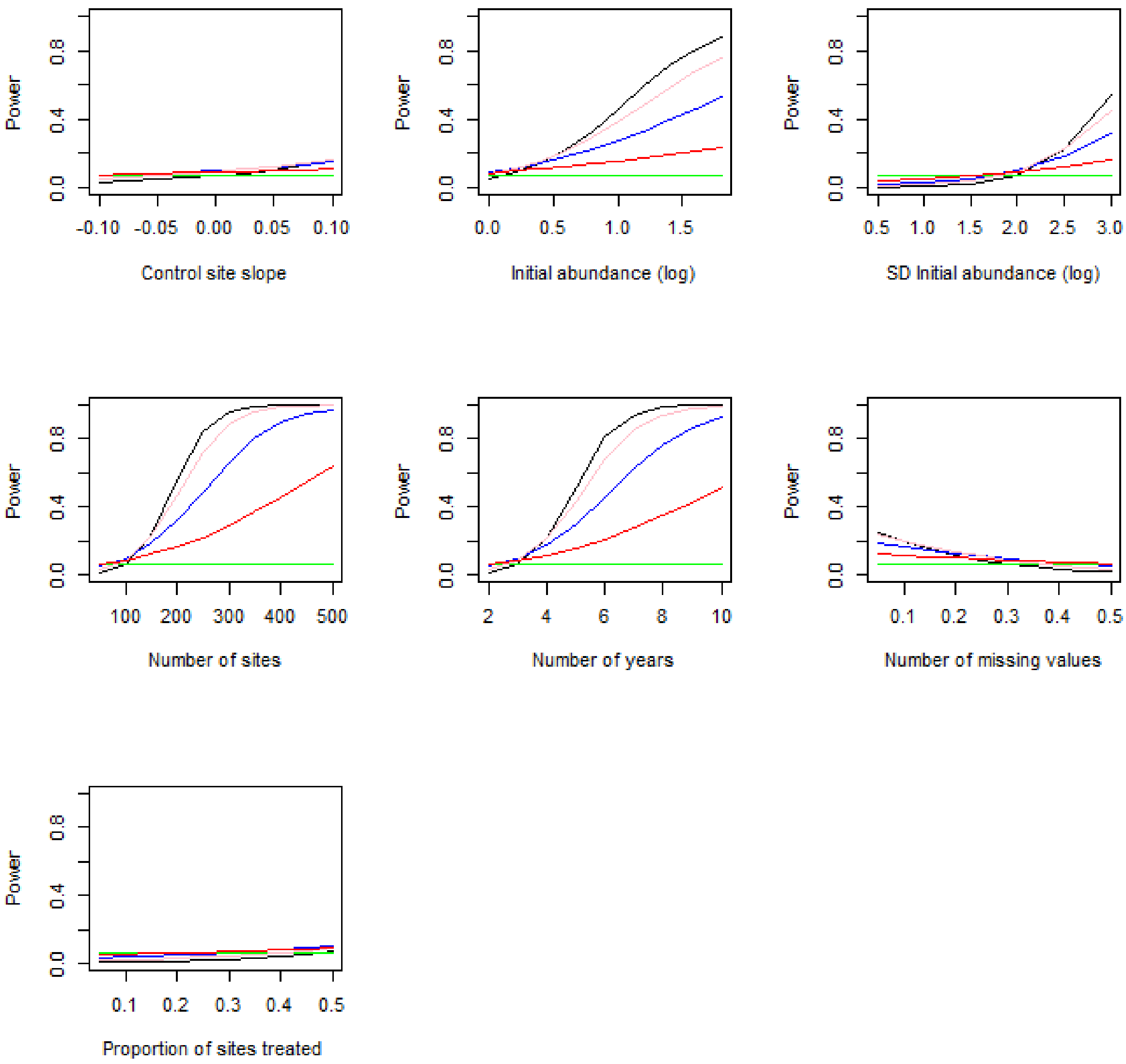
Figure 7: Power curves with fewer sites/years (where appropriate) compared to Figure 6. Numbers of sites and years are respectively 150 and 4 (a) and 100 and 3 (b). All other parameters as Figure 6; negative values of α only are shown from -0.1 (black) to zero (green), also as Figure 6.

Figure 7 (a)



When the number of sites and number of years was dramatically reduced (for the values that are held constant), we could see the greater influence of the variation in initial abundance and the effect of missing sites. When we had sufficient data, i.e. in the previous Figure 6, the variation in initial abundance had a lesser effect perhaps as the series had time to converge somewhat or reach some form of equilibrium. Likewise, as data observations were scarcer, if we had a high number of missing values, this decreased the power considerably. Both of these factors suggested that an ESN with few sites should take extreme care to ensure that as few observations are missed as possible.

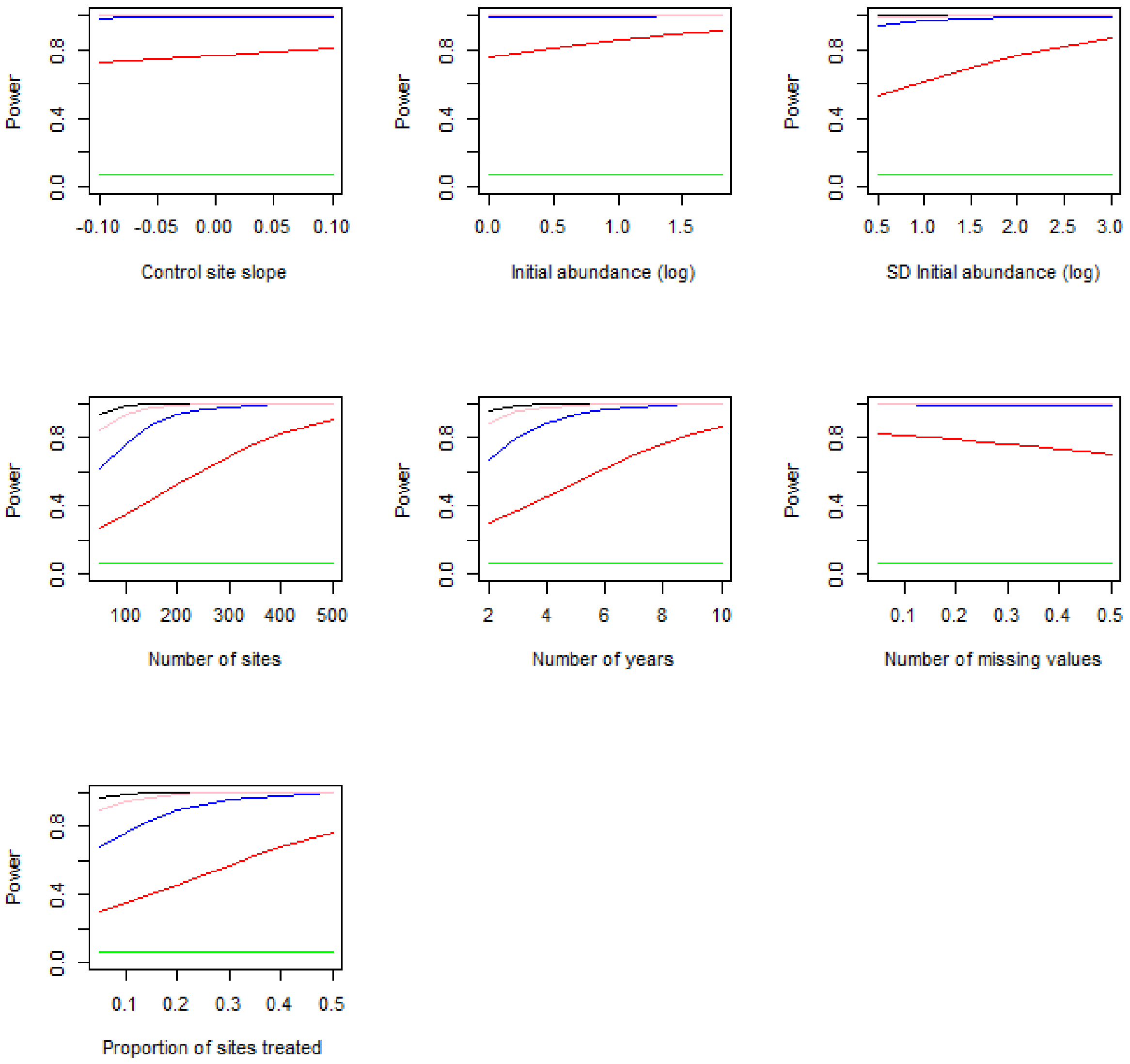
Figure 7 (b)



As the number of sites and years of monitoring were reduced even further, we could see that the relationships between each of the parameters and power remained the same, but that overall power was greatly reduced. Here, even for large values of α (e.g. α =-0.1, black line), sufficient power could only be obtained by significantly increasing the number of sites or the length of time series. No amount of care taken to avoid missing values or to optimise the number of observations in the treated category could generate sufficient power.

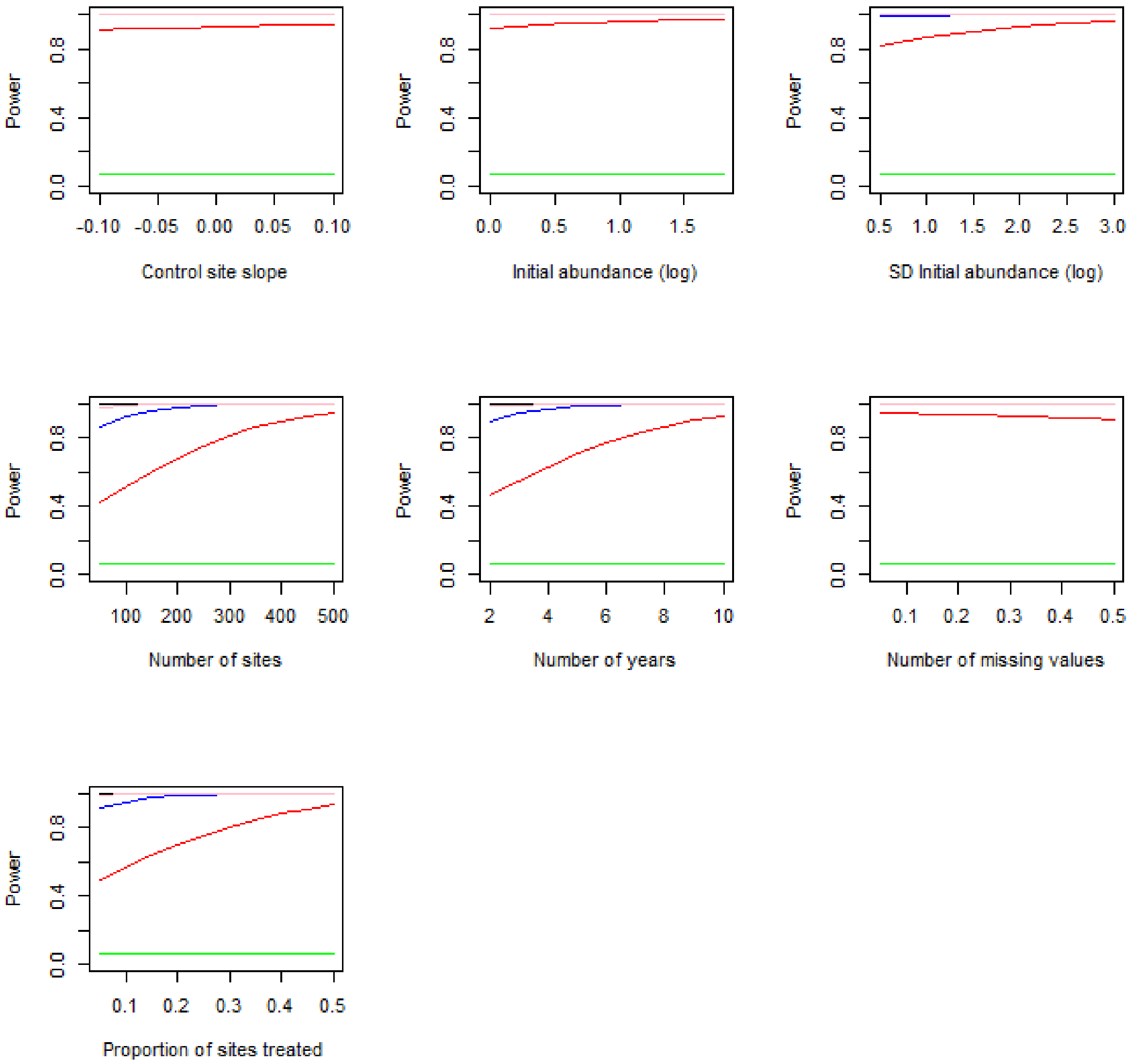
Figure 8: Power curves with more sites/years (where appropriate) compared to Figure 6. Numbers of sites and years are respectively 350 and 8 (a) and 450 and 10 (b). All other parameters as Figure 6; negative values of α only are shown from -0.1 (black) to zero (green), also as Figure 6.

Figure 8 (a)



When the number of sites and the length of monitoring were high, the vast majority of cases, even for low levels of α, resulted in reasonable statistical power. Only a low proportion of treated sites, when the effect of the treatment was low (red line), resulted in insufficient power.

Figure 8 (b)



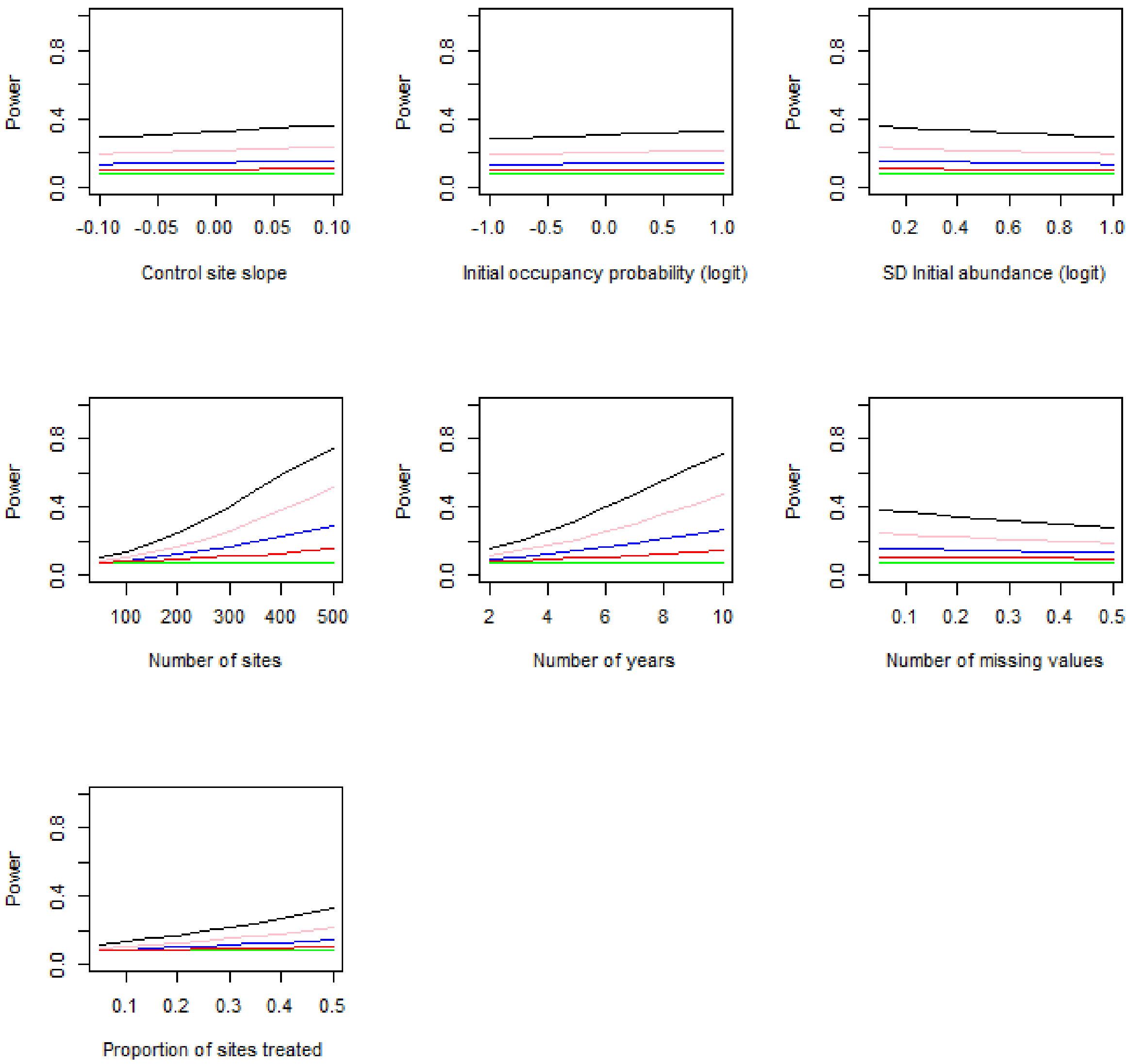
When the number of sites and length of monitoring increased even further, reasonable statistical power was achieved under all scenarios.

Figure 9: Power of Binomial model to detect specified negative (a) or positive (b) treatment effects (α, in the notation of Freeman and Newson, 2008) versus assorted population- or scheme-specific parameters. All parameters were held constant, other than α and those varying along the x-axis in each diagram, these constant values being as shown below. Curves in (a) represent α = -0.1 (black), -0.075 (pink), -0.05 (blue), -0.025 (red) and 0 (no effect; green). Curves in (b) represent α = 0.1 (black), 0.075 (pink), 0.05 (blue), 0.025 (red) and 0 (no effect; green)

Other parameter values, where fixed:

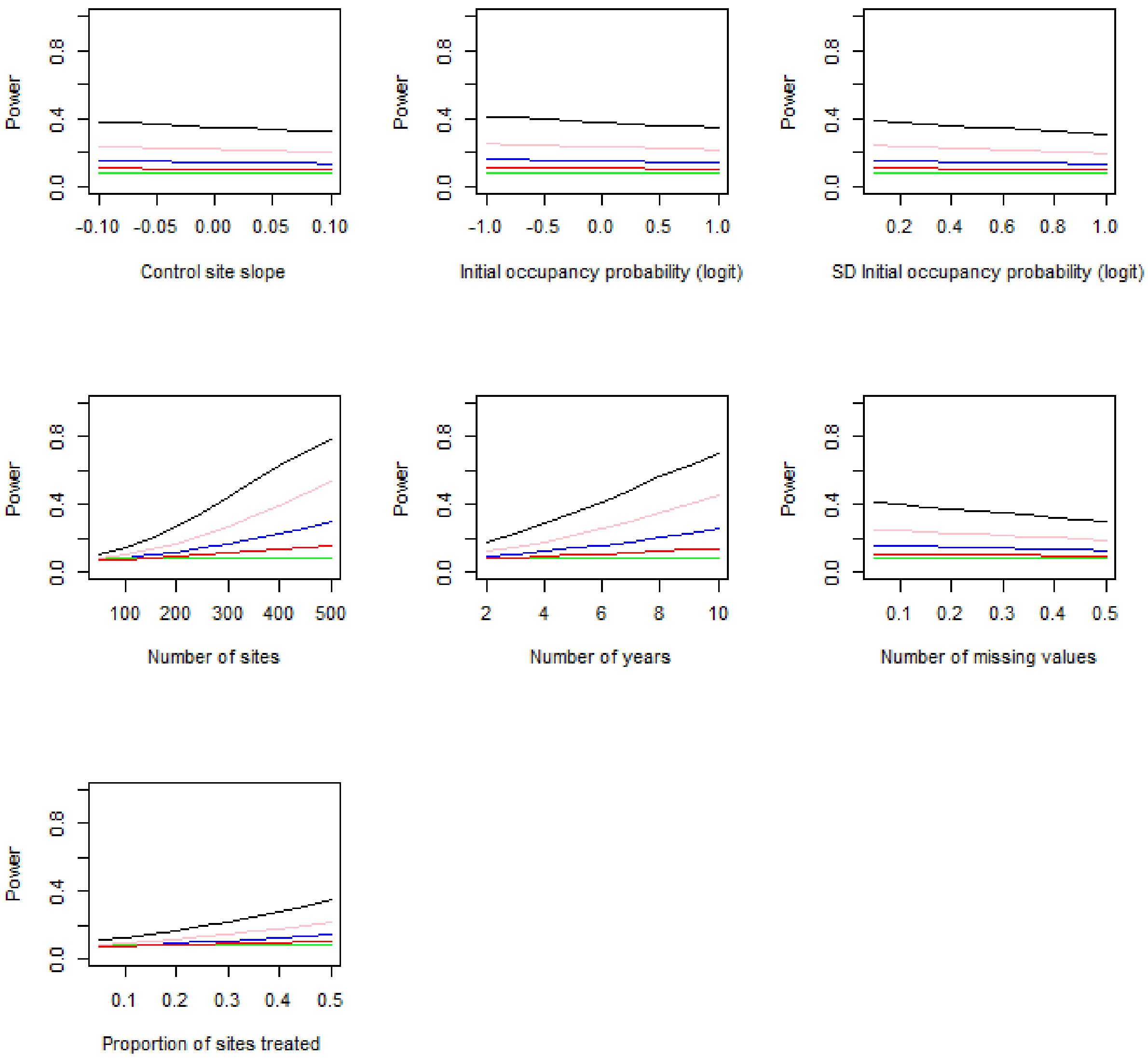
Control slope	0.0
Logit Initial occupancy probability (mean)	0.8
Logit Initial occupancy probability (SD)	0.5
No. sites	250
No. years	5
Proportion of visits missed	0.3
Proportion of sites treated	0.5

Figure 9 (a)



For the Binomial model, used for binary presence absence type data, there was lower power achievable than in the equivalent Poisson case. This was as one would expect as the count data analysed in the Poisson case contained more information than a simple binary indicator. This additional information proved to be vital as in the Binomial case here, reasonable power was only achieved by significantly increasing the number of sites or the length of monitoring. Other parameters had very little effect under this specific scenario.

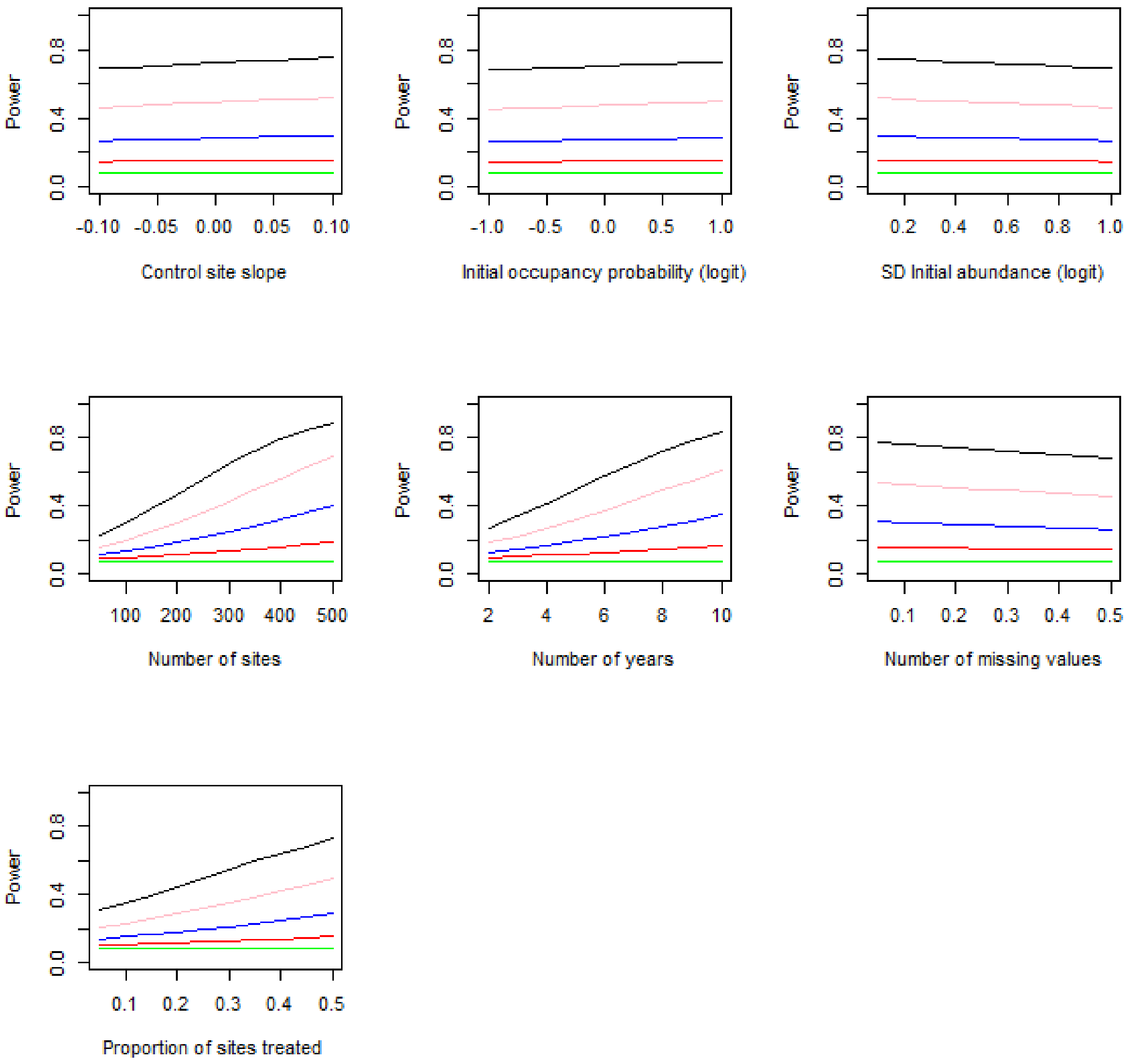
Figure 9 (b)



Similar to the Poisson case, the Binomial model for binary data also showed the mirroring we would anticipate for when α is a negative or positive influencing factor.

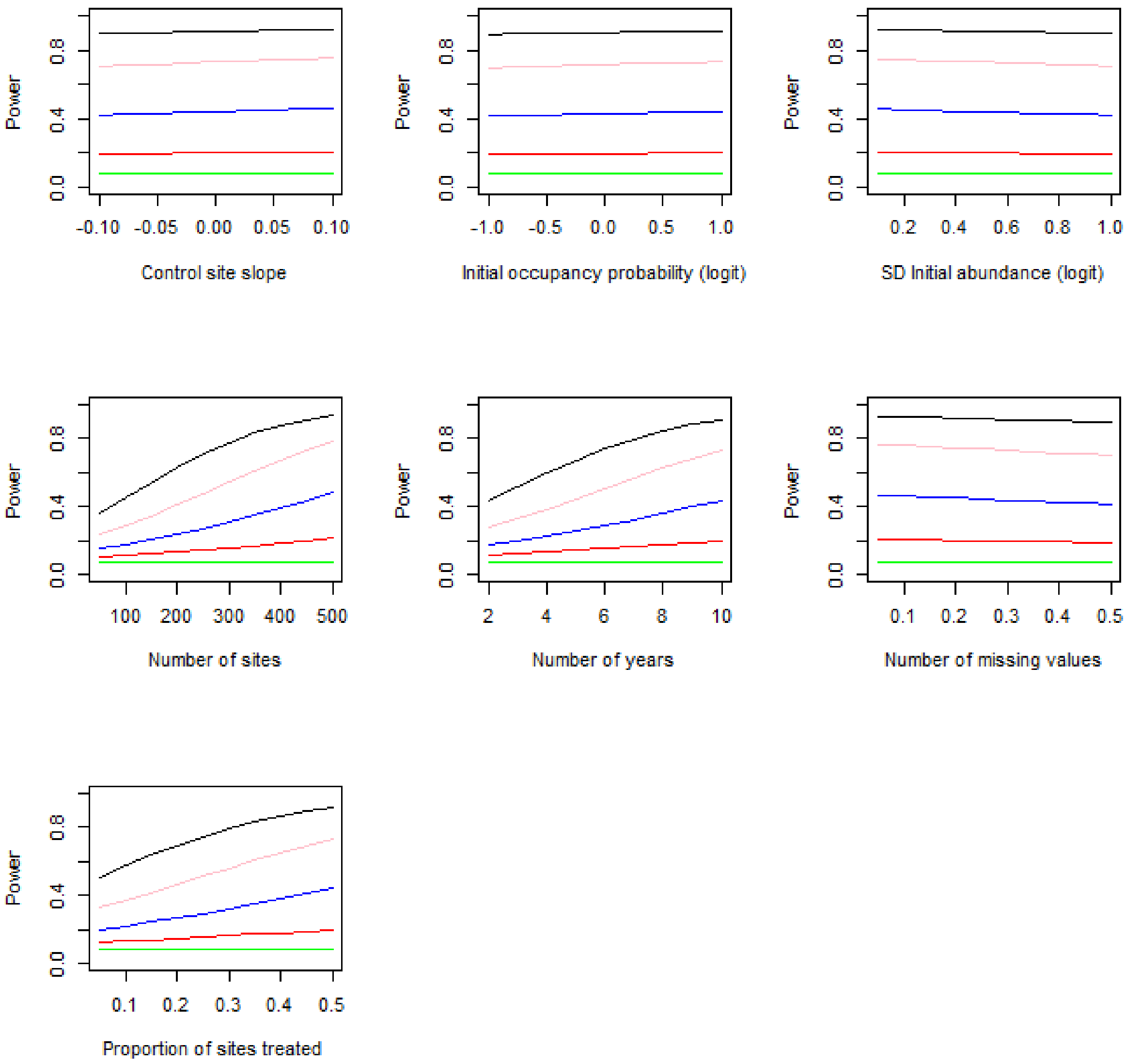
Figure 10: Power curves with more sites/years (where appropriate) compared to Figure 9. Numbers of sites and years were respectively 350 and 8 (a) and 450 and 10 (b). All other parameters as Figure 9; negative values of α only were shown from -0.1 (black) to zero (green), also as Figure 9.

Figure 10 (a)



Here the number of sites and the length of monitoring had been greatly increases and we could immediately see that large magnitudes of α (α=-0.1, black line or α=-0.075, pink line) could be detected with sufficient statistical power. What was also noticeable is how the proportion of sites treated could affect power in this case. Even with a large number of sites and a long time series, if the proportion of sites treated fell below 0.3, power no greater than approximately 50% could be achieved. This was far more influential on the binary data here than for the count data analysed in the Poisson model.

Figure 10 (b)



When the number of sites and length of monitoring was increased even further, it was clear that the actual value of α had the greatest influence on power. Large effects could be estimated with reasonable power, but small effects (α=-0.025, red line), were not captured with any more than approximately 20% power in all cases.

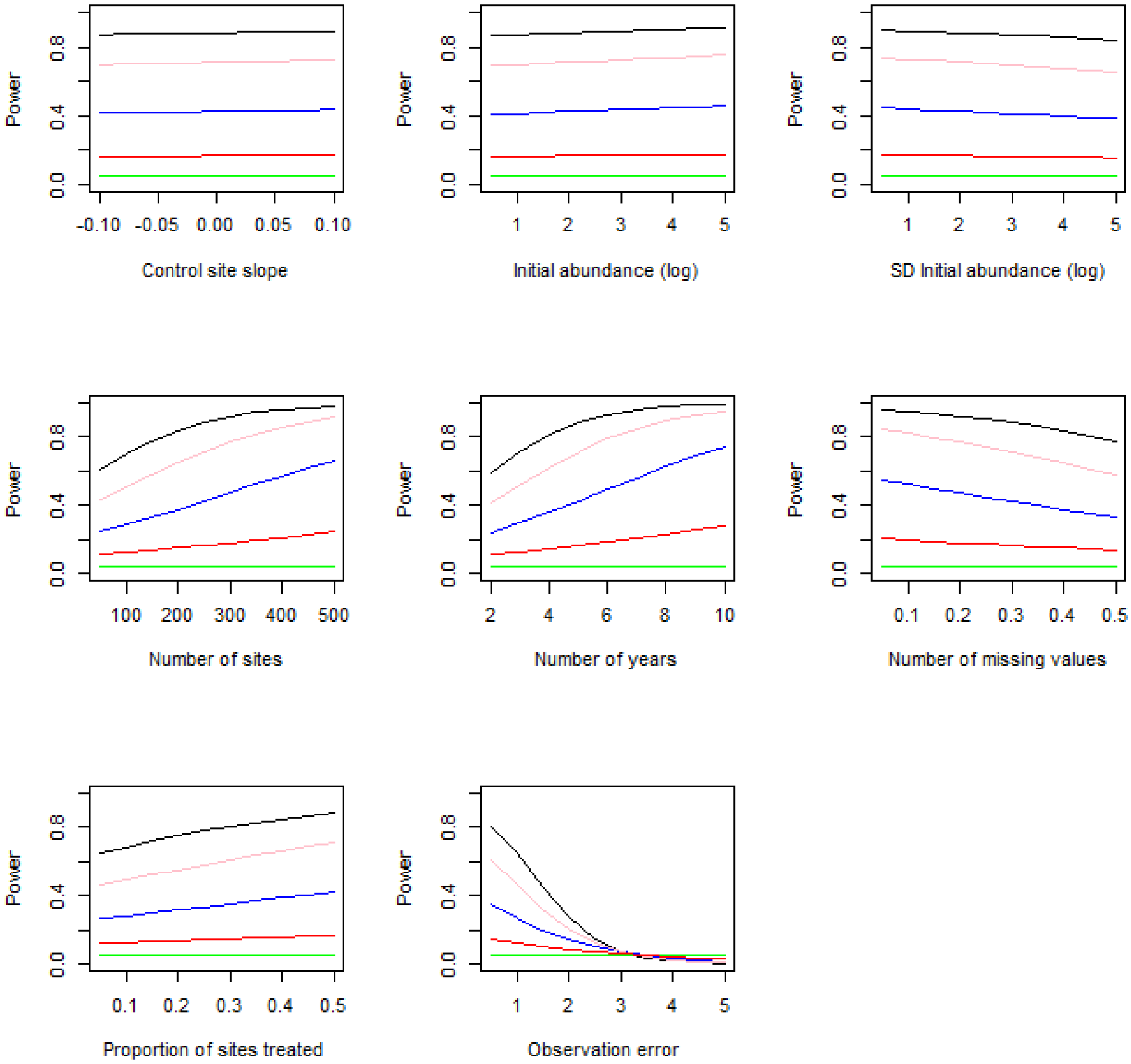
Figure 11: Power of Normal model to detect specified negative (a) or positive (b) treatment effects (α, in the notation of Freeman and Newson, 2008) versus assorted population- or scheme-specific parameters. All parameters were held constant, other than α and those varying along the x-axis in each diagram, these constant values being as shown below. Curves in (a) represent α = -0.1 (black), -0.075

(pink), -0.05 (blue), -0.025 (red) and 0 (no effect; green). Curves in (b) represent α = 0.1 (black), 0.075 (pink), 0.05 (blue), 0.025 (red) and 0 (no effect; green)

Other parameter values, where fixed:

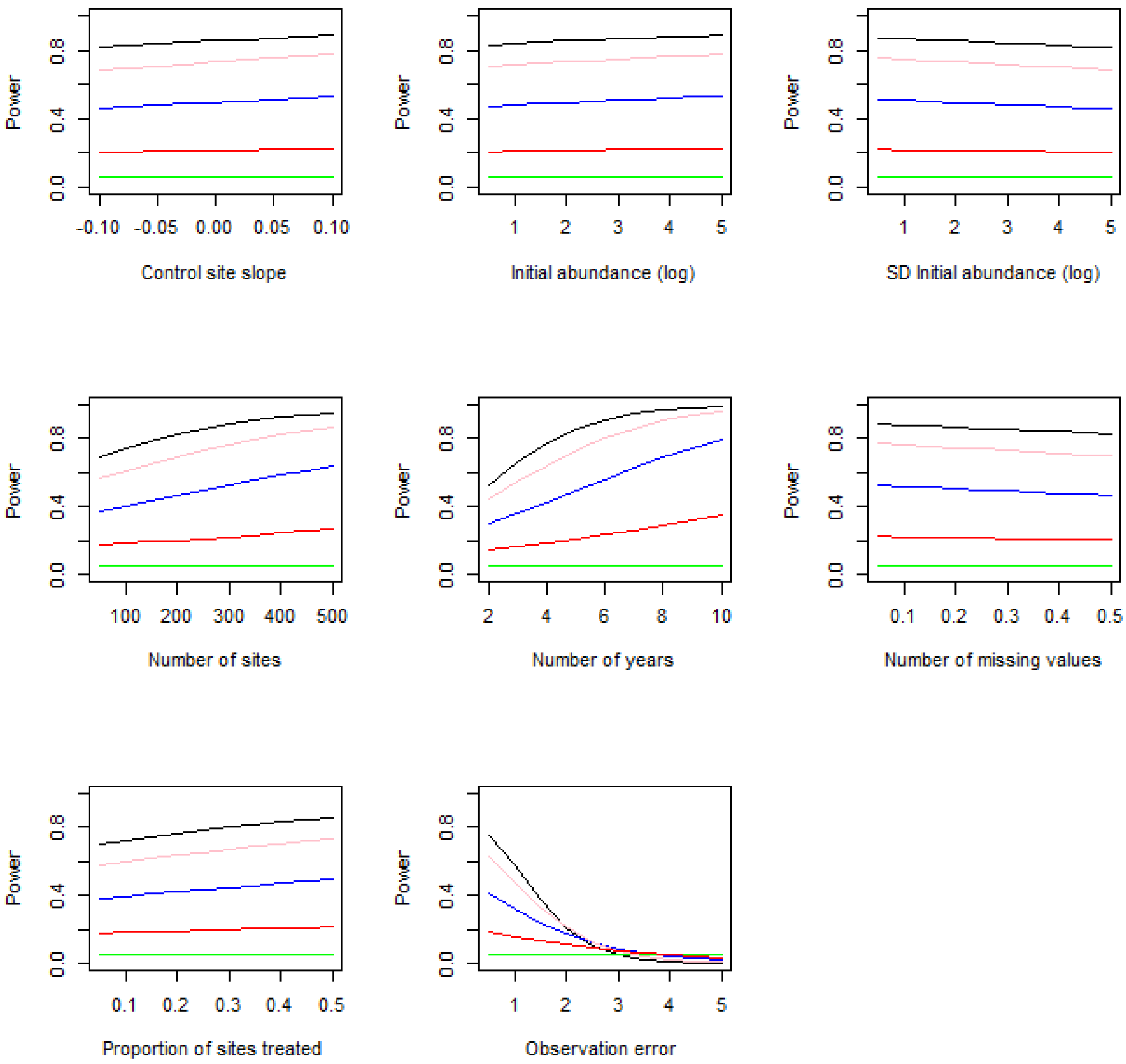
Control slope	0.0
Log Initial count (mean)	2
Logit Initial count (SD)	2
No. sites	250
No. years	5
Proportion of visits missed	0.3
Proportion of sites treated	0.5
Observation error standard deviation	0.1

Figure 11 (a)



For continuous data, analysed using the Normal distribution, we could see that the observation error associated with the sample data had the biggest influence on power. As the error increased, the power to detect an effect, no matter how big, significantly reduced. This parameter had an even greater influence than the number of sites surveyed or the length of the monitoring activity. This suggested that ESNs should do all they can to ensure that observation error is kept to an absolute minimum, perhaps by quality control procedures or greater training of surveyors.

Figure 11 (b)



For positive values of α, the same relationships were observed with the observation error again being the most dominant factor in influencing power.

Figure 12: Power curves with greater observation error and initial count (where appropriate) compared to Figure 11. Observation error SD and log initial count are respectively 0.8 and 1 (a) and 2 and 0.5 (b). All other parameters as Figure 11; negative values of α only are shown from -0.1 (black) to zero (green), also as Figure 11.

Figure 12 (a)

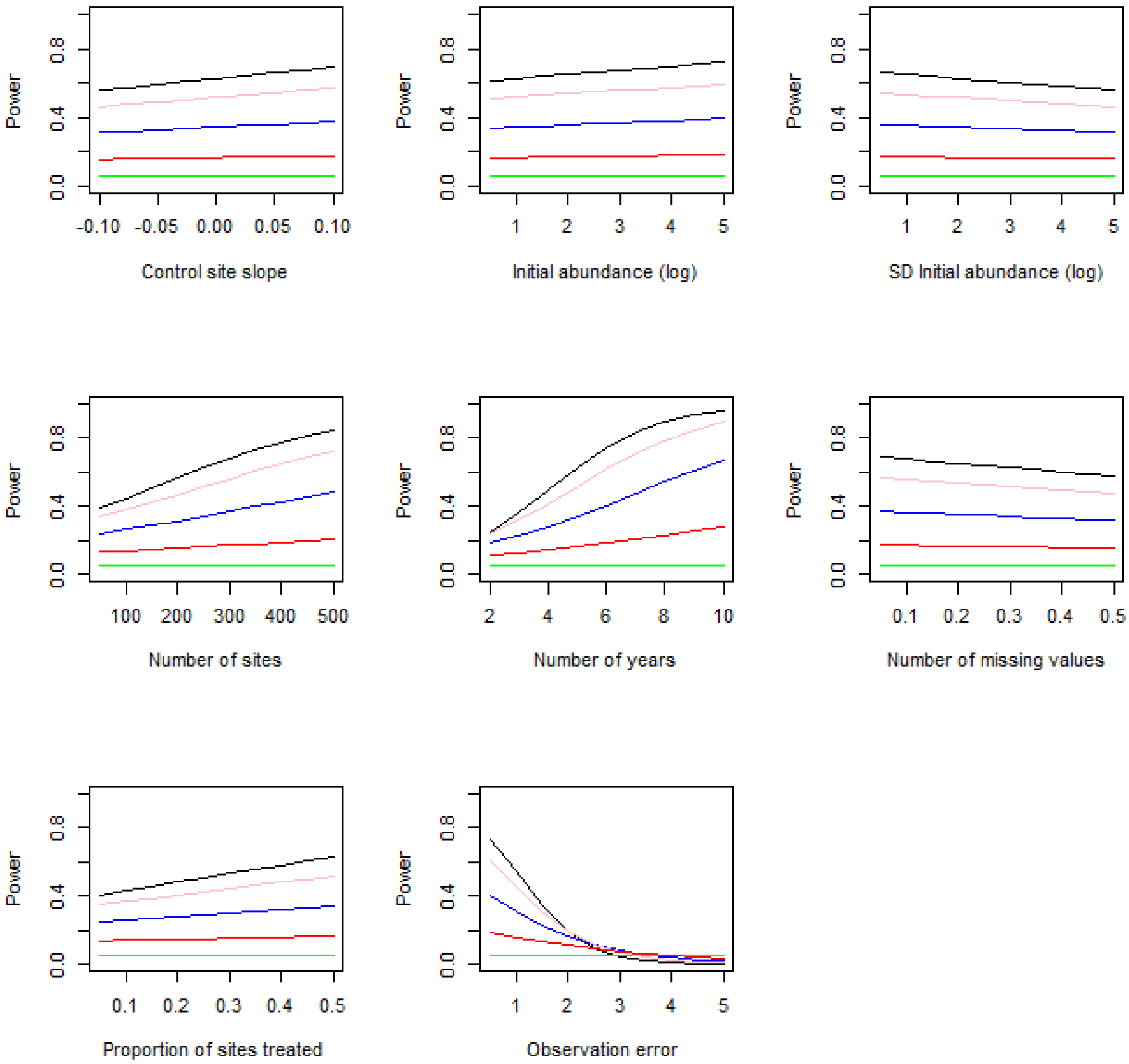
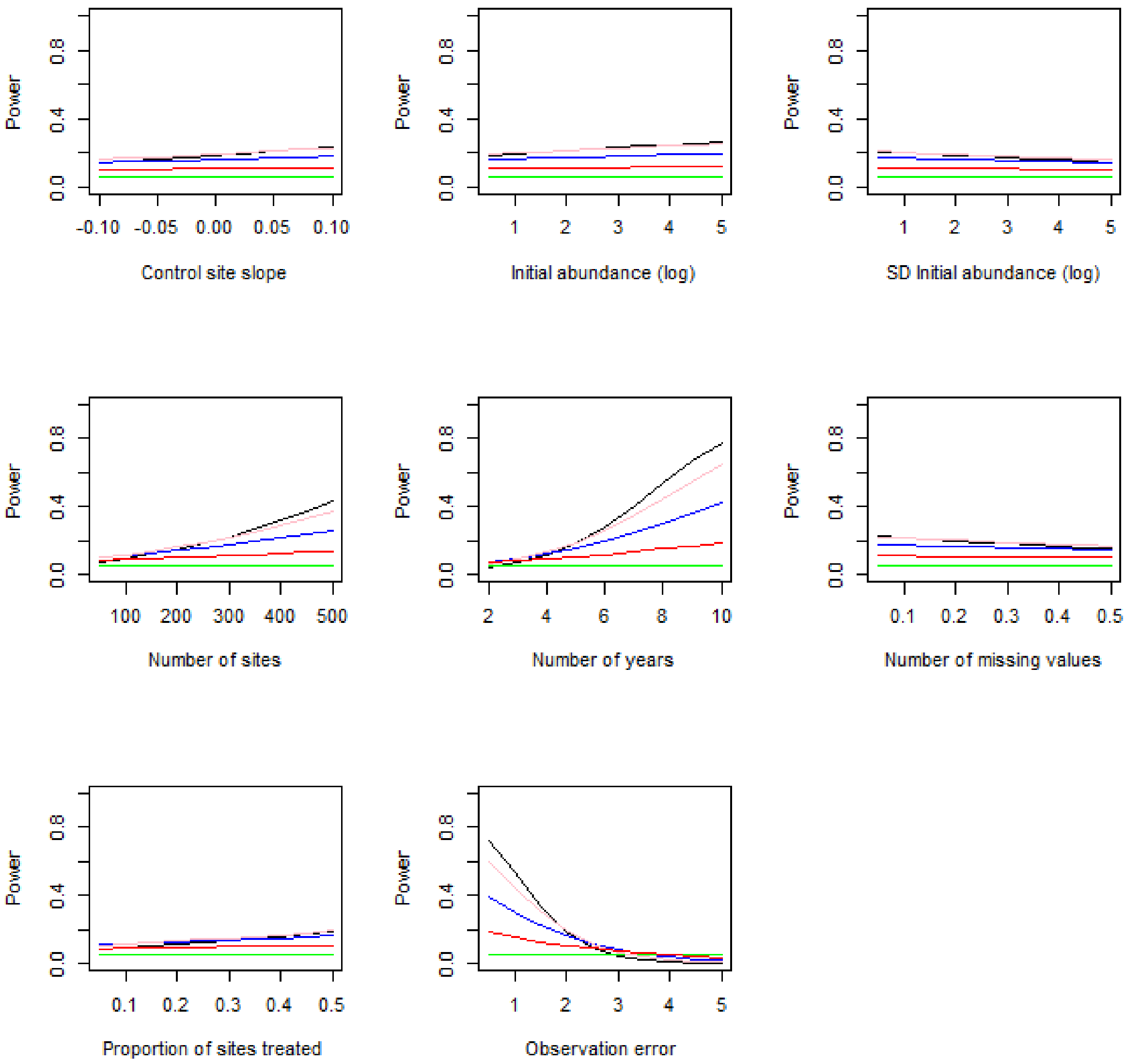


Figure 12 (b)



As the observation error increased even further, the plots showed that only a very long time series, with a large effect of α could result in sufficient power (number of years=10, α=-0.1, black line). None of the other parameters could achieve power greater than 40%.

Figure 13: Power curves with a selection of values for no. of sites/years, observation error and initial count (where appropriate) compared to Figure 11. Observation error SD and log initial count are respectively 0.8 and 1 (a) and 2 and 0.5 (b). In each case 350 sites were visited in 8 years. All other parameters as Figure 11; negative values of α only are shown from -0.1 (black) to zero (green), also as Figure 11.

Figure 13
(a)

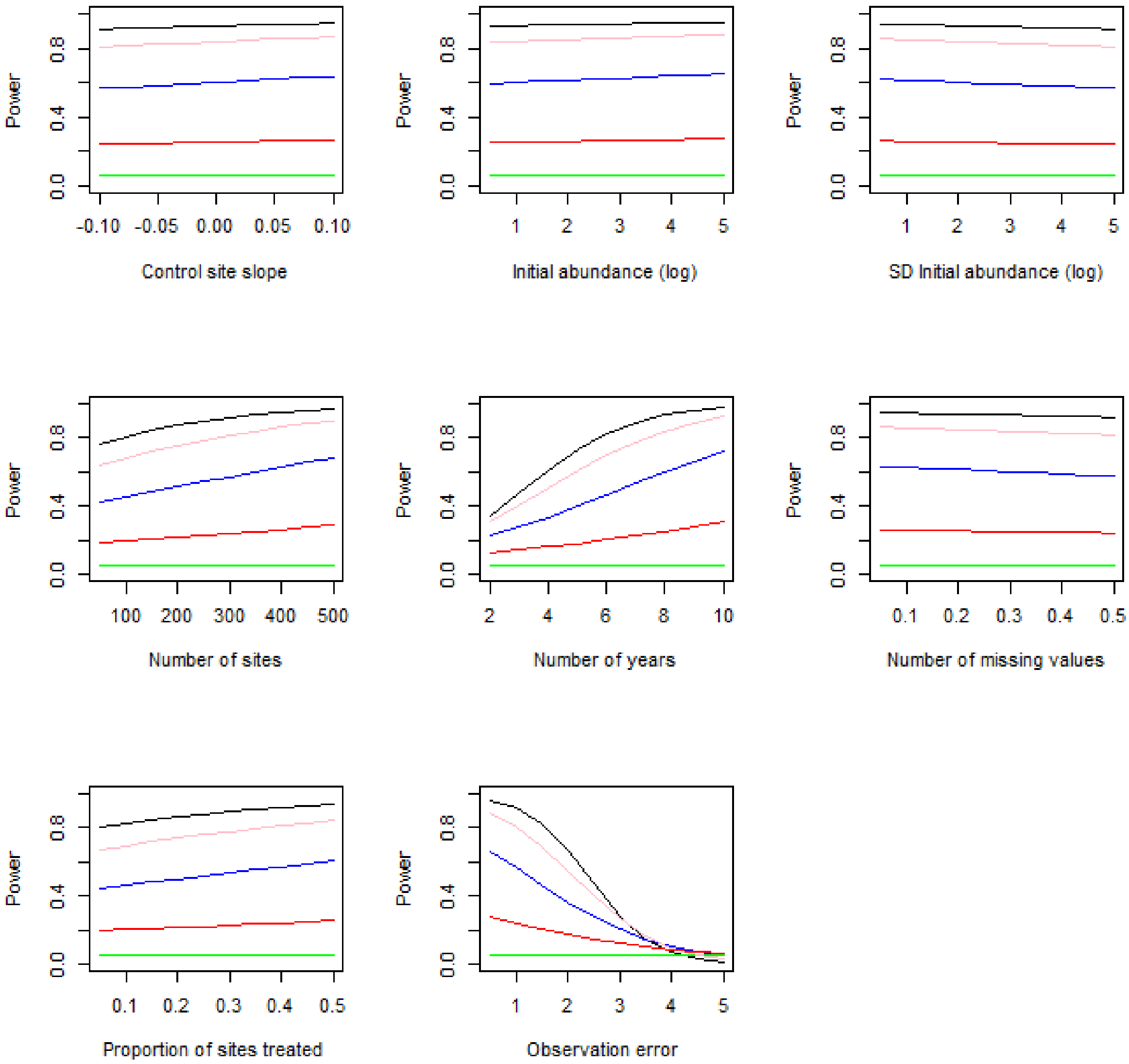
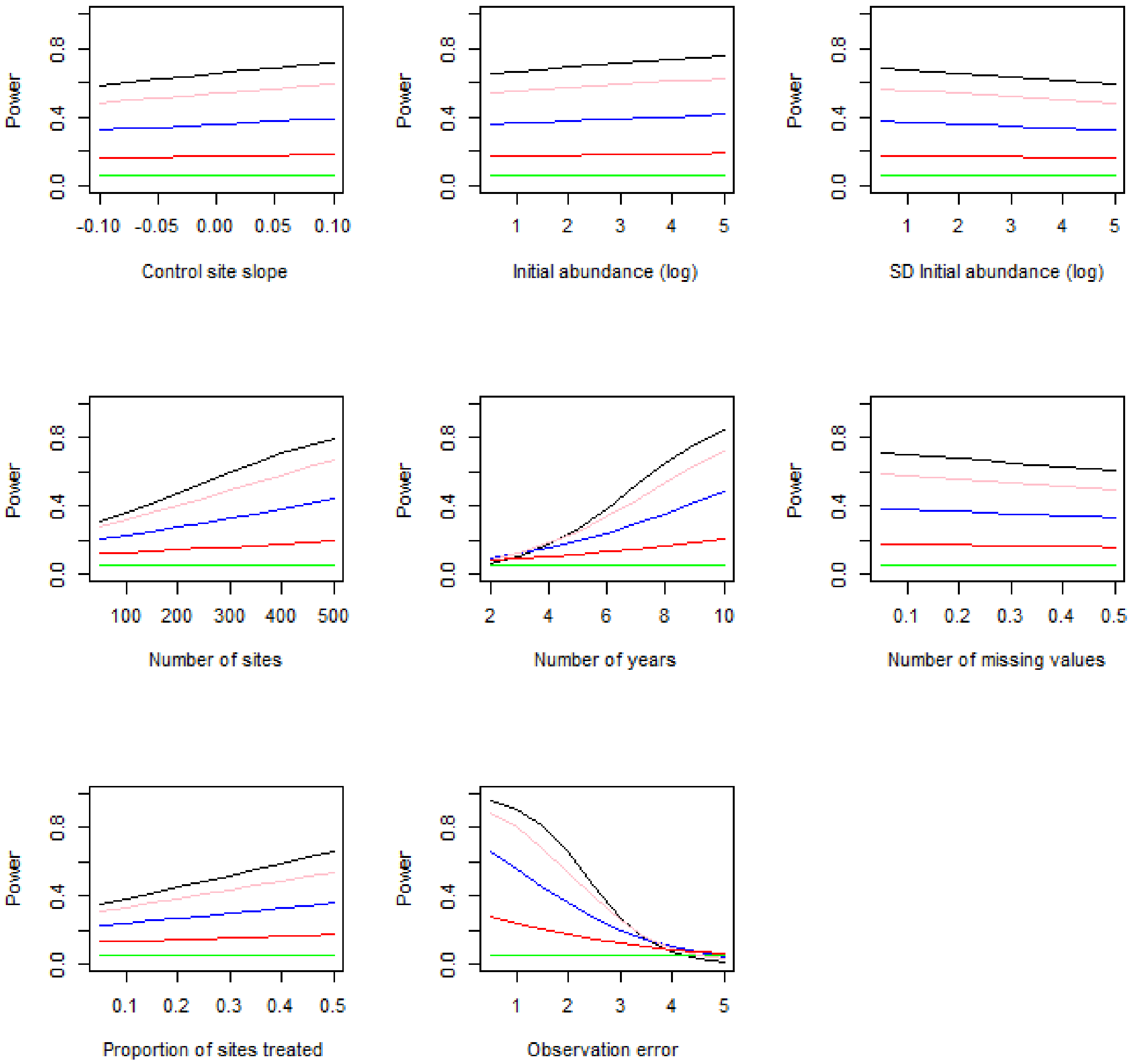


Figure 13 (b)

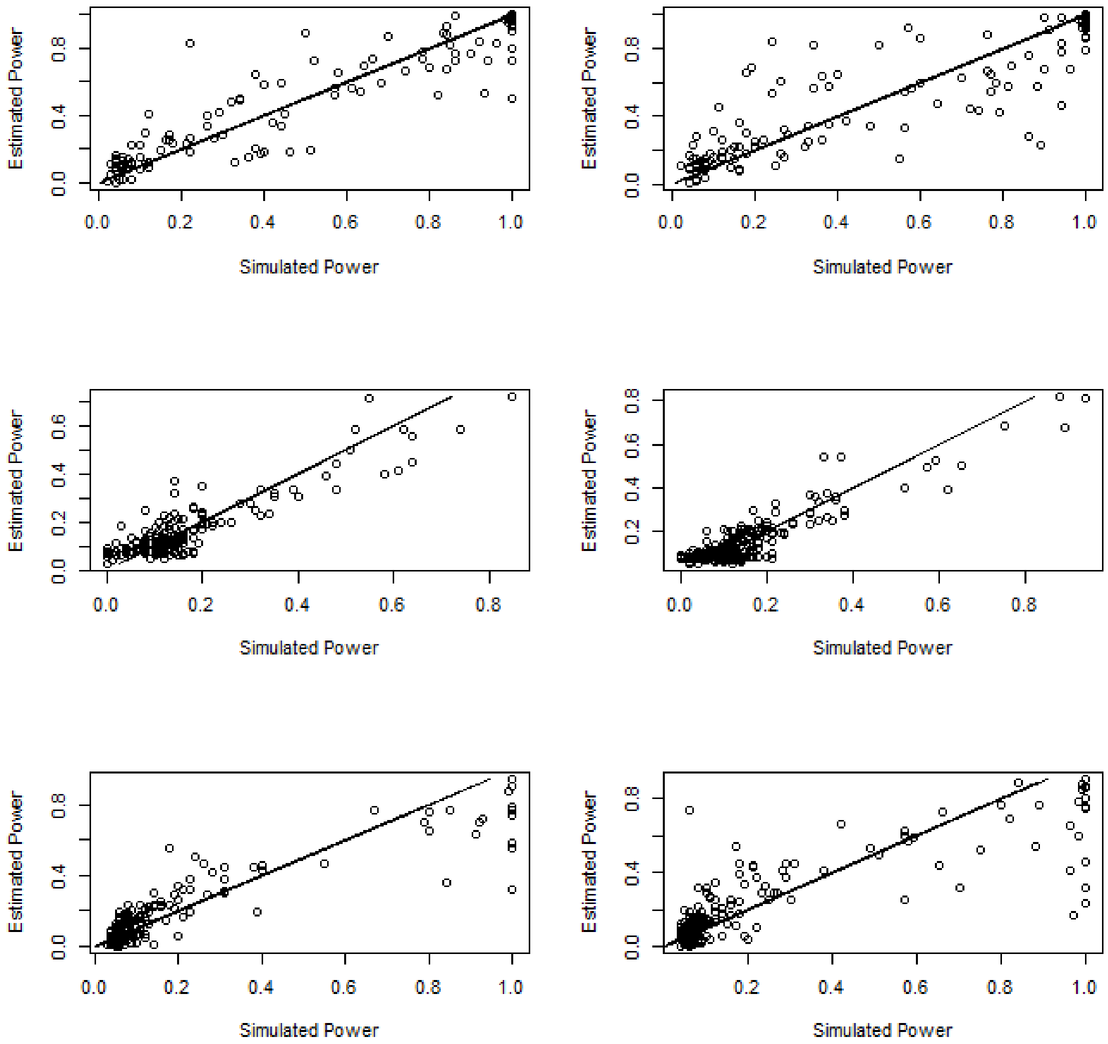


As the number of sites and the length of monitoring was increased, we saw that the biggest influencing factor on the power to detect an effect was the size of the effect itself. There was clear separation between the power curves for different levels of α, with low effect sizes (e.g. red line) never produced power above 20%, whereas high effect sizes (e.g. black line) often produced power in excess of 60%.

4.2.3. Generic Equation

The generic equation as defined in Equation 2 was estimated separately for the Binomial, Poisson and Normal cases and also, in light of issues around fitting, separately for positive and negatives values of α. As power was constrained to lie between 0 and 1 a logit link function was used in a Binomial glm evaluated in R. The resulting goodness of fit plots for each of the 6 models are shown below in Figure 14 and show that the generic equations did a reasonable job of estimating power. Having examined the fits, we could be confident in the shapes of the relationships estimated between the predictor variables, $X_1 - X_9$, and power. Although there was error associated with the power estimation, the speed and efficiency of the equations provided a useful tool and guide for power estimation in this context. As the generic equation is based on a model of power that is linear in the fixed effects, performance and better model fit could potentially be achieved by investigating non-linear associations. Further work would be needed to consider this in detail.

Figure 14: Power estimated under the appropriate generic equation (y axis) versus that observed from a number of repeated simulations (x axis): Poisson(top), Binomial (centre) and Normal (bottom) models, with α < 0 (left hand panels) or α>0 (right hand panels).



The coefficients of the resulting power equations are given in Table 16. As predictors, we used the strength of the effect α, the quadratic term α^2 and first-order interactions between α and the remaining variables. Doing this ensured that the predicted power was the same for α=0, irrespective of the values of the further parameters. Under the theory, this value (by definition, the probability of rejecting the null hypothesis of no treatment effect even though it is true) should be equal to 0.05, as testing was carried out at the 5% significance level. Though it was possible to constrain the power to exactly equal this value, we had rather left it as an estimable intercept, thereby providing an additional means of model verification by comparing this estimated value with the 'ideal' value, 0.05. As the models were given on logit scales, the intercept should be equal to -2.95. Appendix 6 provides a clear guide on how to use these models in practice to achieve estimates of power.

Table 16: Coefficients in linear equations used to approximate power in the models of this report.

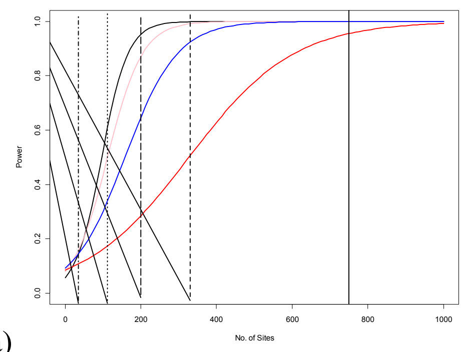
Predictor	Poisson	Binomial	Normal

	α<0	α>0	α<0	α>0	α<0	α>0
Intercept	-2.67	-2.19	-2.44	-2.39	-3.01	-2.84
α	92.08	-90.33	21.50	-21.21	-24.69	31.69
$α^2$	-178.36	-132.34	77.99	115.21	-73.48	-204.6
α:Slope	-92.24	16.77	-16.29	-13.22	-9.19	29.27
α:Abundance_Mean	-27.24	21.27	-1.17	-1.45	-0.95	1.08
α:Abundance_Var	-28.17	19.79	3.13	-3.92	1.16	-0.96
α:Duration	-13.62	11.80	-3.25	2.93	-5.51	5.49
α:Missed	61.88	-19.40	10.09	-10.64	40.65	-0.11
α:N_Sites*(1-Treated)	-0.029	0.075	-0.016	0.01	-0.016	0.0078
α:N_Sites*Treated	-0.55	0.44	-0.13	0.14	-0.14	0.089
α:Sigma					15.82	-0.16

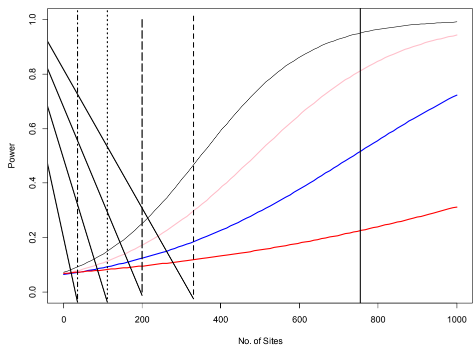
Notes: coefficients of models given on logit scales.

4.2.4. Mapping real data onto the power curves.

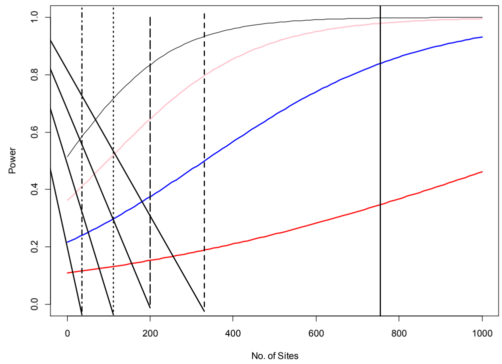
The plots shown in Figures 6-13 demonstrated multiple scenarios fed into the generic power equations and the resulting power relationships with individual factors. The same approach could be taken to investigate the power that an existing scheme had to detect a change in a given indicator over a set period of time. As an example of mapping some existing networks onto the derived power curves, we chose 5 networks that meet the idealised standard (see Section 3). These were: Vigie-Nature, Bats; Dragonfly Monitoring Scheme (The Netherlands); Common species census (Slovakia); Census of wintering and staging wildfowl and geese (Estonia); and the Butterfly monitoring scheme (The Netherlands). As information on specific indicators was unknown for these sites (we do not have detailed information on the mean value and variance of a specific species), a number of assumptions had to be made in order to produce power curves. Assumptions were made for parameters for the Poisson, Binomial and Normal cases in line with those values assumed in Figures 6a, 9a and 11a respectively. Only data on number of sites was taken from the inventory to map these networks onto the power curves. Power was plotted as a function of sample size in each of the three (Poisson, Binomial and Normal) cases so that the values observed for the individual networks could be added. The resulting power curves are shown in Figure 15 and demonstrated that, given a number of conditions, some of the networks show the potential for high statistical power. In the Poisson case (a) and Normal case (c) reasonable power under the hypothetical scenario was achieved for all networks apart from the Census of wintering and staging wildfowl and geese when α is large (black curve). In the Binomial case (b), only the Butterfly monitoring scheme, with its large number of sites, achieved reasonable power for high levels of α. For low levels of α, only the Butterfly monitoring scheme in the Poisson case, achieved high power. Both the Binomial and Normal cases failed to achieve reasonable power, even for the longest and biggest networks, when α was low.



a)



b)



c)

Figure 15: Power against sample size plotted for: a) the Poisson case - parameters as defined in Figure 6a; b) the Binomial case – parameters as defined in Figure 9a; and c) the Normal case – parameters as defined in Figure 11a. Existing monitoring networks are superimposed showing where their current power lies given their sample size. Solid line: Butterfly monitoring scheme, Long dashed line: Vigie-Nature, Short dashed line: Dragonfly Monitoring Scheme, Dotted line: Common species census, Dot Dash line: Census of wintering and staging wildfowl and geese.

As Figure 15 above exemplifies one specific scenario, the task was repeated using different values for the initial abundance mean and initial abundance variation, these were halved in all cases. The results from the same exercise under this new scenario are presented in Figure 16. We can see that this has not made too much difference in each of these cases presented. This reflects what is shown in Figures 6a, 9a and 11a, where initial abundance did not have too much influence on the overall power.

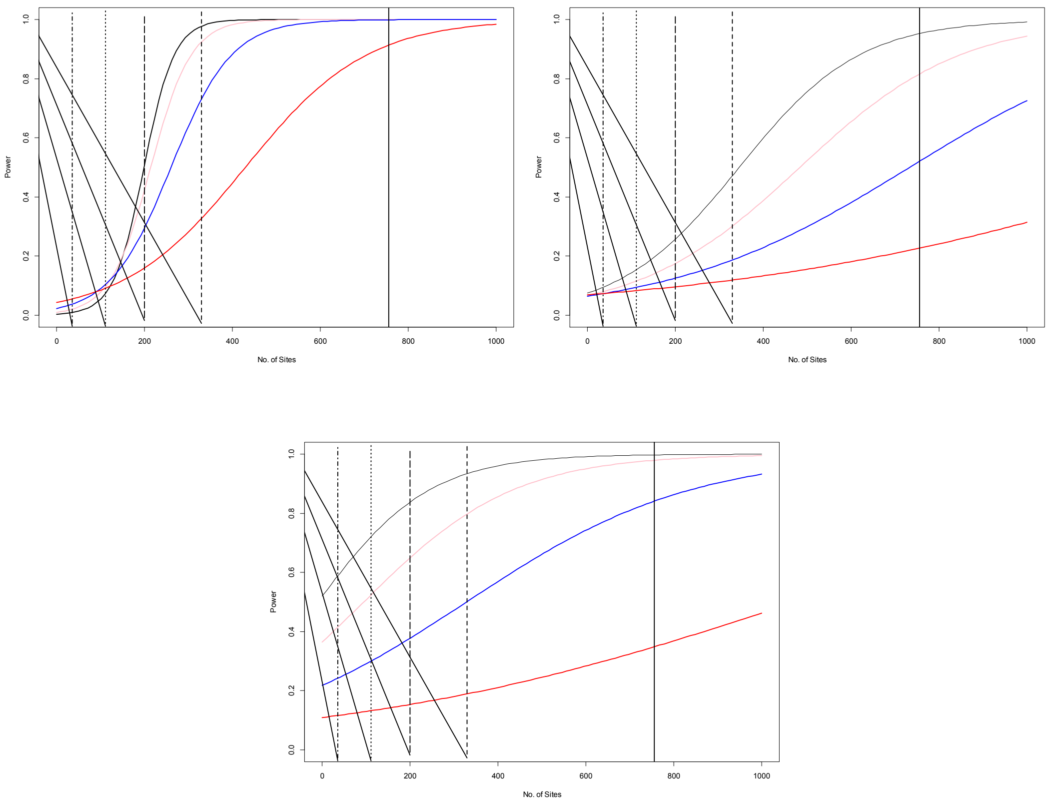


Figure 16: Power against sample size plotted for: a) the Poisson case - parameters as defined in Figure 6a but with log initial abundance mean and sd halved; b) the Binomial case – parameters as defined in Figure 9a but with log initial abundance mean and sd halved; and c) the Normal case – parameters as defined in Figure 11a but with log initial abundance mean and sd halved. Existing monitoring networks are superimposed showing where their current power lies given their sample size. Solid line: Butterfly monitoring scheme, Long dashed line: Vigie-Nature, Short dashed line: Dragonfly Monitoring Scheme, Dotted line: Common species census, Dot Dash line: Census of wintering and staging wildfowl and geese.

4.3. Networks Suitable for PMEM

It is important that consideration is not only given to the statistical properties of any network prior to its use for PMEM, but also to its reporting properties, geographic properties and the protocol adopted. This is especially important when consideration for involvement in PMEM is based on existing networks without any alteration or amendment to their current data collection and reporting strategy is planned or possible.

In collating the inventory of networks, a set of criteria and their properties were identified that would define an idealised network (see Section 3). In practice this is very difficult to define and often subjective. However, there are some elements that are important to have information about if a network is to be useful for EU-wide PMEM and it recognized that an effort to identify the networks that go some way to achieve these standards is important.

In identifying the different facets of networks that contribute to an idealised contributor to PMEM and from the work undertaken in compiling the inventory (Section 3), it is clear that difficulties remain in bringing together all the relevant information due to inconsistencies in reporting and archiving of important supplementary information. Broader scale umbrella organisations provide an excellent foundation for pulling together this information across networks and achieve some element of consistency. These broad scale initiatives have been identified here as the "Suitable EU-wide networks" and highlight the importance of EU-wide programmes that bring together information that would be crucial for PMEM.

Table 17: Summary of the key attributes that make up the different network classes considered, together with some examples for these standard networks that were identified within the inventory.

Network Class	Criteria	Example ESNs that meet the criteria
Suitable EU-wide Networks	European-wide	Pan-European Common Bird Monitoring Scheme, PECBMS
	Multisite, even distribution	Butterfly Monitoring Scheme, BMS
	Observations at least annually	European Topic Centre on Water, Eionet Water
	Standard protocol	Constant Effort Sites Ringing
	Protocol well documented	The Biodiversity Information System for Europe (BISE)
	Trained surveyors	EEA, Air quality
	Validated data	
	Analysis method well documented	
	Access to raw data	
Suitable national networks	Member state	Base de Données Analyse des Terres
	Multisite, even distribution	Suivi des Analyses de Terre

	Observations at least annually	British and Irish Ringing Scheme
	Standard protocol	Broedvogelmonitoring
	Protocol well documented	Landelijk meetnet vlinders
	Trained surveyors	Zoogdierenvereniging
	Validated data	The UK Butterfly Monitoring Scheme
	Analysis method well documented	Breeding Bird Survey
	Access to raw data	Seguimiento de Anfibios y Reptiles de España
Possible networks – more information needed	Member state	Inventaire, Gestion et Conservation des Sols
	Multisite	Riksskogstaxeringen
	Standard protocol	United Kingdom Cereal Pathogen Virulence Survey
	Trained surveyors	Base de Donnees Carbone France
	Validated data	Suivi des Analyses de Terre
	Analysis method well documented	Meetnet Reptielen
		ECN data centre
		Nationell Inventering av Landscapet i Sverige
		Watervogeltelling
		Haudelinnustiku punktloendused

4.4. Conclusions

Having examined the statistical power to detect a treatment effect under a range of scenarios and produced an equation to easily estimate power under differing conditions / hypothesised changes, we could use this to make recommendations for environmental networks and their use for PMEM of regulated products in the future. In particular, the power estimates provide useful information when considering any potential changes to networks for their specific use in PMEM and provide some key context when interpreting non-significant effects.

4.4.1. Sample size

It is clear from the derived generic equation that sample size is one of the main contributing factors in determining the power of any network to detect an effect of a release to the environment of a product. Sample size, unlike many other factors such as magnitude of the effect, is something that, to a certain extent, the networks themselves can control (some volunteer networks have little or no control over this however). Therefore sample size is something that could be investigated and recommendations could be made on a more appropriate number of sites that are required to increase statistical usefulness of data obtained from these networks. For any specific network, the generic equations, given in Section 4.2, can easily be used to look at the sample size to power relationship, as shown in the power curves for the case studies investigated. There are, however, some important aspects to consider, which are described below.

4.4.1.1. Combining networks

There are some clear examples of networks in the inventory that collect the same information on the same environmental indicator but across different geographic regions. Each of these networks is analysed individually and the power to detect an effect is related to each specific network. Given such scenarios where the same data are being collected, it would seem obvious to attempt to combine the data and analyse the pooled resource with increased power rather than separately analysing the data from each network, where each of these networks has low sample size and therefore low power. These small separate networks may give different results, especially if some locations across this broad geographical scale have not had the same exposure to the product whose post-market effects are being assessed. The variation among sites may then hide the effect of the product that was significant at some of the sites. Combining results for different networks, therefore, is perhaps not as straightforward as it may seem, as there may be important covariates influencing the response variable across different geographic regions and different elements of variability from each constituent data supplier.

Let us consider an example of the Pan-European Common Bird Monitoring Scheme, which is an umbrella group of multiple networks each operating at Member State level. If we make some assumptions about the parameters in the generic equation, we can investigate where each member organisation lies on the respective power curves, given its sample size and length of survey. This was done only for the Poisson case as birds are typically recorded as counts. Figure 17 shows the resulting power curves for 4 member networks of the Pan-European Common bird Monitoring scheme where the parameter input values have been fixed according to:

Control slope	0.0
Log Initial abundance (mean)	0.1
Log Initial abundance (SD)	2
Proportion of visits missed	0.3
Proportion of sites treated	0.5

The different lines represent different levels of α {solid line = -0.1, dashed line = -0.075, dotted line = -0.05, dot dash line = -0.025}. The figure shows that generally the longer running schemes with the highest number of sites have the greatest power. The differing shape of the power curves show that, for some networks (e.g. Denmark), there is little benefit to be gained from adding additional sites, whereas for others (e.g. Slovakia) there can be large gain in power with little extra effort.

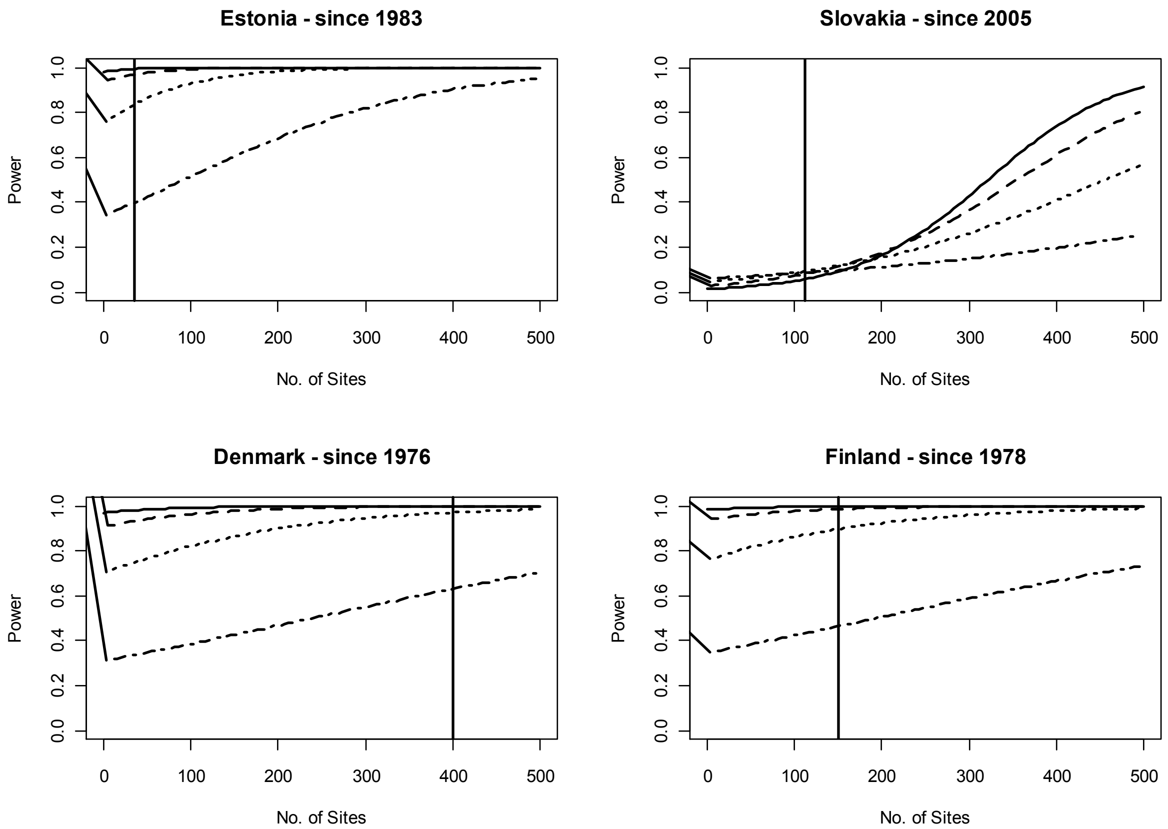


Figure 17: Example: Poisson power curves against sample size for 4 networks, with their current sample size indicated. Lines indicate different levels of α: solid line α=-0.1; dashed line α=-0.075; dotted line α=-0.05; dot-dash line α=-0.025.

When attempting to combine data across networks we had to a) take the most conservative value across the individual networks for variables such as length of time the scheme had been running and b) to make assumptions about any additional hierarchical variance component. This gave a very approximate idea about power, but more clearly provided information on the potential gains that could be made when data is analysed in this way. The resulting power curve is shown in Figure 18, where the combined sample size across the 4 networks is indicated by the vertical black line. For the Slovakian network, a major gain in power could be achieved when an additional number of sites were added, whereas for the other networks power is reduced as the length of time series has been reduced to match that of Slovakia. This does, however, assume that the control slope and background noise in the population is equivalent in each of the different networks. This of course may not necessarily be true and under certain circumstances power can be lost when combining networks because of added variability in the data. The results from the simulation study showed that in most scenarios, power is independent of the slope in the control population, but high variation in the data can dramatically reduce power. More complex hierarchical models would be needed to fully investigate the advantages and disadvantages of combining data across networks.

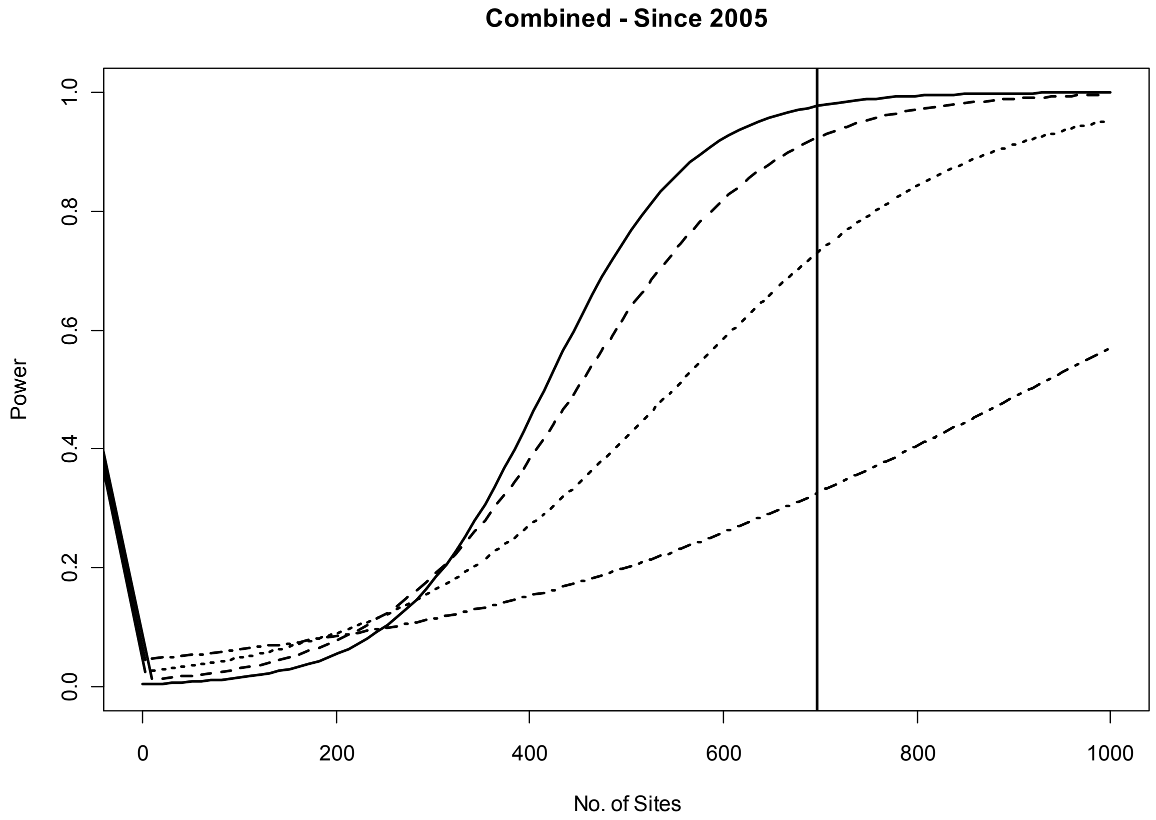


Figure 18: Poisson power curve of combined data across the 4 networks, taking most conservative values across constituent networks for length of time. Combined sample size is indicated by the vertical black line. Lines indicate different levels of α: solid line α=-0.1; dashed line α=-0.075; dotted line α=-0.05; dot-dash line α=-0.025.

Van Strien et al. (2010) analysed data post trend analysis and there may be value in taking that approach rather than pooling raw data and analysing it as shown above. There may be some merit into looking into the Van Strien method using pooled data, in parallel to the power analysis presented here. This is because in some circumstances there may be additional power gained if all the data for a particular network can be used to estimate its trend then these trend estimates are modelled together, whereas for other networks more value could be gained by pooling raw data as opposed to derived metrics.

To investigate combining data from networks further and in sufficient detail, one would need to conduct further simulation studies that include additional levels of variation in the model hierarchy and allow for different protocols in differing groups. Such an investigation would also need to take into account the different lengths of time that each constituent scheme has been running and the different numbers of "treated" and missed sites as defined in Table 16. As some schemes are volunteer based and some professionally surveyed, differences in the observation error also need to be accounted for. However, this would need large scale, complex simulation studies that are extremely difficult and time–consuming. Furthermore, these analyses would probably need to be run on a case-by-case basis. Therefore there is potential for further simulation studies that, in collaboration with the generic approach to power estimation presented here, seek to further understand the advantages in pooling data from different sources.

4.4.1.2. Associated Costs of increasing sample size

The simulation study showed that increasing the sample size of any monitoring network or survey activity has a positive effect on the power to detect any treatment effect. However, increasing sample sizes, in most cases the number of sites surveyed, comes at a cost. The associated costs of increasing sample sizes depend on many different factors and are specific for each network and the type of data collection they carry out, which makes it very difficult to consider general implications. We therefore examined two specific networks run by the Centre for Ecology and Hydrology: the Wider Countryside Butterfly Survey (WCBS) and the Countryside Survey (CS). These two schemes operate very differently and the costs involved in increasing the sample size provide some context when considering this option.

The UK Wider Countryside Butterfly Survey, set up in 2009 to compliment the UK Butterfly Monitoring Scheme, is a network of standardised fixed route walks (transects) established at multiple sites and surveyed for butterflies each year on a regular basis under reasonable weather conditions from the beginning of April until the end of September. Transects are located in a stratified-random sample of 1 km squares across the UK and are mainly surveyed by volunteers, though a small number of sites are surveyed by professionals. Approximately 700 WCBS sites were sampled in 2010.

There are two possible approaches for increasing sample sizes within the Wider Countryside Butterfly Survey: professional and trained volunteer. Here we focus on increasing the sample size by using trained volunteers as it provides a contrast to the purely professional approach adopted by the Countryside Survey. Very approximate calculations suggest that for an extra 100 squares (an almost 15% increase in sample size) a one-off recruitment cost of £30k and thereafter an annual cost of £10k would be appropriate. Over a 10 year period the associated costs would therefore be approximately £120k for an extra 100 squares. Using the generic equation we can estimate the increase in power if 800 sites were surveyed over 10 years against 700 sites over 10 years. In this case the generic equation reveals that for the £120k spent an estimated potential increase of 6% in power could be achieved.

Countryside Survey (CS) is a stratified random sample of 591 1km squares over Great Britain (GB) conducted entirely by professionals with extensive Quality Assurance and Quality Control procedures in place to ensure optimum quality and efficiency of the data. Within each 1km square, detailed data on a high variety of biophysical measurements are taken, including extensive botanical surveys, soil measurements, water quality and habitat condition. The design of the survey ensures that it is spatially representative of GB and unbiased in terms of sampling effort. However, as it is only conducted approximately every 8 years, the temporal representation is weak and the time taken to detect any effects is therefore affected. Long time periods between surveys would miss short term effects caused by small releases of products subject to PMEM and it would be very difficult to prove cause and effect.

To improve the suitability of CS for PMEM, there are two options: the first is to increase the number of squares; the second is to conduct the survey at more frequent, regular time intervals, thus enabling higher resolution temporal analyses and changes overtime to be examined.

With previous experience of CS, the average cost per square across the UK is approximately £7k. Again, we can use the generic equation to look at how this cost translates into power gained. Rather than 600 sites every 8 years, we can look at the power to detect an effect against cost if 100 sites were surveyed every year or if 800 sites were surveyed every 8 years. In this case, the generic equation reveals that for the extra £1.46M spent, it would be better to sample every year (potential increases in power of up to 42%) as opposed to surveying 800 sites once every 8 years (potential increases in power of up to 11%).

The comparison of the WCBS and CS in terms of the cost to power gained is interesting due to the differences between the two schemes in terms of volunteer and professional recording. Whilst the use of volunteers can offer significant increases in sample size for relatively little extra cost, it often induces further sources of variability. Furthermore, one has to recognise that there is a limit to the sample size achievable as the pool of available volunteer surveyors is not inexhaustible. The use of professionals is more expensive, but the ability to have more control over sample location, effort, consistency and observer quality can result in higher returns in power for the same number of additional sites. Cost and sample size, along with other facets of the survey in question should be translated into power before making any comparison. It is therefore important when consideration is given to changing, or adding to, an existing scheme that the cost versus power relationship and the contributing factors to power is fully understood.

4.4.1.3. New surveys

Alongside looking at how sample size of existing surveys may be increased to attain greater statistical power, it is also important to consider any new schemes that may be expected to grow as the network itself matures. This is especially true of citizen science surveys whereby volunteers conduct the vast majority of the recording, because adoption and uptake of the scheme amongst the volunteer community may be slow. Box 4.1 provides an example of a new survey started only in the last couple of years that provides key information on environmental indicators and has the potential to expand over the coming years. With the advent of modern technologies such as smart phones and electronic data capture, there are many more examples of recently formed citizen science schemes setup to fill a gap in the existing knowledge base.

<u>Box 4.1 UK Farm Pollinator survey</u>

As an example of a new scheme we consider the recent UK Farm Pollinator Survey. The scheme started in 2012 as a pilot and saw 550 volunteers taking part across 36 unique farms throughout the UK. The scheme is run as a citizen science effort in collaboration with the

Linking Environment And Farming (LEAF) initiative and sees volunteers recording the number of pollinator insects of different taxonomic classes they identify during each visit to a farm.

As interest and awareness in the scheme grows and sufficient backing from the appropriate sponsors is maintained, one might expect the sample size to increase significantly over the next few years. Even if sample size remains constant, the longer the scheme runs for, the more power it will have to detect a change. For relatively very little resource and maintenance costs, there is potential for the scheme to see a significant increase in power over the next decade. The UK farm pollinator survey may also prove to be a rich resource for certain types of PMEM because it specifically targets farms as opposed to attempting to be representative of a wider population. Such targeting could be an advantage if the aim is to detect change in agro ecosystems.

Any conclusions drawn on recently started schemes should therefore keep this in mind. Power may be low for such startup schemes in their present state, but, perhaps even without interjection, the scheme may expand and grow into a robust, data rich resource. Startup schemes also offer greater flexibility to change and adapt as opposed to longer term schemes that may be compromised if any changes were sought. So startup schemes should always be considered as a useful resource for PMEM and efforts made to achieve the appropriate number of samples to increase statistical power rather than being ignored because of any current misgivings in terms of statistical power.

4.4.2. Uptake

As the analysis considered throughout this report concerns comparing a treatment effect (some change in the agro ecosystem) versus a control effect (the response of the remaining population), it is important not just to have an overall sufficient sample size, but also sufficient sample size in each of the two (treatment / control) groups. This was shown in the power curve relationships by the "Proportion of sites Treated", which can lead to significant increases in power. This term encapsulates two key but very different components. Those being: uptake and sample coincidence. Uptake of a product is obviously important in terms of viewing its effectiveness though is something that no network has any control over. If uptake is very low, it will be extremely difficult for any monitoring network to infer anything with sufficient power, although any impact is also likely to be low, and therefore any environmental damage limited. However, the monitoring networks themselves do dictate where their sites are and hence any overlap between an area of uptake and an area under survey.

The coincidence of survey sites with uptake area is something that if one was to design a survey to monitor the effect of a specific change in the agro ecosystem may ensure that this happens. This obviously would only work for that specific change and as PMEM is part of general surveillance, it would *a priori* not necessarily be clear what the influencing factor (treatment) would be. PMEM is designed to detect unexpected effects and so tailoring a scheme for one specific change could be a poor use of resource. Thus there will inevitably be a trade-off between power for any specific change, and generality. For clarity, if one knew that uptake amongst the community was reasonable, yet many monitoring schemes showed low sample sizes in the treatment group, one may be tempted to recommend that the scheme be modified, adapted or added to in order to encompass these sites. There is obviously a trade-off between encompassing these sites yet retaining the statistically robustness of a schemes original design, its report requirements and detection of any other effect resulting from change in the agro ecosystem.

It is important to recognise, however, that when there is no bias in the uptake of a product, regionally or otherwise, one may expect a well-designed scheme to achieve sufficient overlap. Hence the criteria on sample size distribution placed on the definition of suitable networks. The issue will be most prominent when there is low uptake of a product with a particular bias in the spatial and temporal distribution of the product uptake.

4.4.3. Reporting guidelines

As demonstrated in the results when investigating optimal networks, it is the umbrella organisations that show the greatest consistency and greatest availability of information. There is therefore clearly some benefit in the existence of such groups. Hence there is a potential role for the authorities to play in collating the information and ensuring that all relevant contextual data is available when presenting results from individual monitoring schemes.

Obtaining information from networks themselves has proven to be very difficult and yet the vast majority of them report to some governing body or are funded, at least in part, by a governing sponsor. There may be some merit, at the point of reporting to the respective body, for the authority in question to collate the information required as per the inventory and archive the required documents or links in one place.

4.4.4. Additional Factors

This work does not take account of additional factors that may affect the environmental outcome under consideration. In what we have presented we have modelled a control effect. This “control” group could in practice, however, have a high level of variability as it effectively contains all of the effects of other factors not included in the model. For example if we were investigating the effect of a particular product released to the environment on bumblebee abundance and we knew that length of hedgerow also affected the abundance of bumblebees, then we should include this covariate in the model to separate out hedgerow influence before looking at the impact of the product in question itself.

If these other factors were known and could be included, then that would lead to an increase in power as some of the unexplained variability in the control group would be accounted for. It would therefore be reasonable to suggest that networks themselves should endeavour to collect information on any known factor affecting the environmental outcome so that this could be included in a statistical model suitable for PMEM.

5. Comments and Recommendations

5.1. Comments

Gathering information on statistical analyses that are being used to identify changes in agro ecosystems and which could potentially be used to assess the post market impact of regulated products proved to be very difficult. In the published literature insufficient information in titles and abstracts made systematic literature searches impractical, while web sites from Ecological Surveillance Networks and projects did not give sufficient information on the methods for data collection, the number and location of sites being used, or how they dealt with the data to produce graphs of trends over time. If a web site or report did show trends there was no statistical assessment of whether these changes were within the expected Normal fluctuations or did represent a significant change in the ecosystem. There are, however, some clear examples of where a network has used and provided the reference to a stand-alone tool specifically designed for such a purpose, for example the use of the TRIM software package (Pannekoek J and Van Strien A, 2005), which has excellent documentation. However detailed information on how the TRIM software was applied to specific data sets collected was rarely available. This type of referencing and documentation is inconsistent amongst networks.

The collated inventory of networks we have provided has shown that there are existing monitoring schemes collecting data across a wide range of indicators that would be of potential use in PMEM. No obvious gaps in the types of data being recorded were identified. A full list of indicator categories covered by the networks in the inventory is given below.

- Air Quality
- Animal health
- Biodiversity - amphibians
- Biodiversity - birds
- Biodiversity - fish
- Biodiversity - fungi
- Biodiversity - insects
- Biodiversity - mammals
- Biodiversity – other arthropods
- Biodiversity - plants
- Biodiversity – reptiles
- Human health
- Plant health
- Soil Function
- Sustainable agriculture
- Water quality

There are a number of existing monitoring schemes that meet all the criteria that we believe could define a suitable EU-wide network, and many more may be achieving the same standard though operating at Member State level as opposed to EU level. These schemes are characterised by having multiple site locations with an even/representative spatial distribution; observations made at least once per year; data collected according to some well-defined, well-documented protocol by trained surveyors; information that the data itself is validated and accessible; and well documented information on data collection methods and data analysis. Examples of such schemes are provided in Table 17 and can be identified from the inventory.

There are a number of existing monitoring networks that, as part of an umbrella group, collect the same data according to very similar protocols. The key difference often being that the individual networks have been operating for different lengths of time and at a different number of sites in different countries and regions. Some of the networks will have relatively low power to detect an environmental effect due to utilizing a small number of sites over a limited time period, whereas others could have very high statistical power. There is therefore much value to be gained in combining data across geographic regions to enhance the power of the statistical analysis. Even when the most conservative of assumptions are made on parameters input into a generic equation (such as length of time the scheme has run for e.g. Figure 18), there is potential value to be gained from pooling data across a region. Not only can this lead to enhancing the ability to detect ecosystem changes, and to increase in power for the detection, but also gain a better understanding of the factors that influence the data and a gain in confidence when interpreting results that are significant at EU scale, rather than within specific geographic regions.

In order to utilize the surveillance network data, our consultation with statisticians found that most would recommend some variant of the generalised linear model (note that this includes many modelling approaches often referred to in different ways), when posed with the task of identifying some form of change in an agro ecosystem. The externally refereed statistical section of this report that we have provided gives a clear, brief description, together with key references, of the methodologies ecological statisticians believe most useful for PMEM.

In Section 1.4, we mention some existing "off the shelf" statistical methods that are available in many different software packages. These should provide sufficient power to detect trends in environmental indicators. Some methods proved to be more powerful than others and to select the most appropriate method one must take into account the question to be answered and the specific data set, data type and the distribution of data available. Generally it was found that the statistical model fitting technique made little difference to the results, but inclusion of complex error structures in the model could have a significant impact on the outcome and conclusions. Our sample analyses show that one should not fit complex parametric models to small quantities of sample data, and non-parametric methods would be more appropriate. In addition the data should be tested for auto-correlation and heterogeneity of variance. As found with Abrahantes et al. (2004) the number of hierarchies in the variance structure can have significant impact on the power of the analysis and therefore it is essential to make sure that this is taken into account when deciding on which models to fit to the data.

The power for detecting change from environmental monitoring networks has been demonstrated for a number of scenarios / indicators by use of the generic equations provided. The demonstration shown in Figure 17 indicates that this is simple to do in practice and also shows that, under certain conditions, a number of exemplar schemes that are collecting data do show sufficient statistical power to detect a change in an agro ecosystem.

5.2. Recommendations

One of the key criteria for suitable networks is that the data is collected either by professional surveyors or trained volunteers who follow clearly defined and documented data collection methods. This is to ensure quality and comparability of the data being used in any analysis. Such schemes, however, are likely to be more expensive and resource demanding to organize. Volunteer-run schemes offer the best value for money and relatively little resource, and have the potential to obtain large quantities of data. If control over and specific training of professional and volunteer surveyors is impractical, then it is important that the schemes incorporate a rigorous data and methods quality control component in order to reach the suitable standard. This hopefully will ensure that the data collected in the field is of sufficient quality to be utilized in any statistical analyses. There should therefore be an emphasis on developing quality control tools such as the Indicia (http://www.biodiverseit.co.uk/indicia.html) platform, which ensures quality of data by using controlled electronic data capture methods, which can be utilised across multiple similar networks.

The use of statistical modelling tools can demonstrate significant correlations between changes in the agro ecosystem management, such as the use of pesticides or herbicides, or release of genetically modified organisms, and environmental changes. However, such correlative studies and analyses cannot specify a causal link between the two. This is due to a number of reasons, the most obvious being the potential of missing covariates in the model and correlated factors that are not known or not considered. Any significant correlation discovered will however suggest potential links and causal pathways that will provide specific hypotheses that should then be verified under controlled experimental conditions. There are therefore clear links between ongoing monitoring activities and the development of targeted small and large scale studies. Determining the ecological mechanisms and potential cause and effect of substances subjected to PMEM will benefit from co-ordinated activity across a wide range of stakeholders

The types of monitoring activity that has/is being conducted in different countries are shown in Table 12 in Section 3. Many countries show a good range of monitoring activity across environmental indicators (noted above). Biodiversity of birds and insects appear to have the most monitoring activity across the different countries likely as a result of the interest from volunteers and professionals working with endangered and charismatic species, while human and animal health, soil and air quality show the least monitoring activity. There are some countries with greater coverage across environmental indicators than others and some indicators that are only monitored in a small number of countries. This may, however, be an issue of "network visibility" as opposed to a genuine lack of information. This too though is an important issue to consider. This information (Table 12, Section 3), in conjunction with the inventory itself, should be used to guide any additional monitoring.

The inventory of environmental networks developed as part of this project shows that the majority of activity is ongoing and that plans for continued monitoring and increased citizen science activities are already in place. It is therefore important, to be useful for guiding PMEM, that the inventory of networks is maintained and kept up to date. This should be managed and not just left to the initiatives of ESNs and ESPs to enter information, as it is essential to ensure quality of the information. However, the effort required for this should not be underestimated. Guidance and suggestions could be offered to the network managers on how to improve the web sites and documentation of the work that they do and data they collect in order to increase the usefulness of their activities.

The required characteristics of a suitable EU-wide network are demonstrated in Table 17 i.e. a network that conducts its monitoring in a way that will make it useful for PMEM and which provides important meta-data and ancillary contextual information needed for accurate interpretation of any statistical analysis. These criteria, coupled with sufficient (>70%) statistical power to detect an effect for a particular indicator, will form the basis of a scheme suitable for PMEM. The characteristics of a suitable network clearly indicate how a design could be of maximum policy relevance, guidelines which EFSA could publish for new schemes to follow.

Further research on the way that networks and the data they collect could be combined to increase spatial coverage and data comparability will enhance our understanding of the drivers of environmental change. Such research will also further inform the assumptions we have made here during our data simulations to assess the data needs for statistical analysis. This will then provide further evidence towards the potential use of environmental surveillance network data. Analysis of hierarchical variance structures and different sampling regimes will unveil the true potential behind the possibilities of pooling data.

The use of data collected by environmental surveillance networks, both professionally and volunteer-led, has great potential to increase our investigations into PMEM to assess the unforeseen consequences of some change in the agro ecosystem on the environment and, coupled with additional investigation and experiments, to potentially enhance our ability to avoid such consequences. In order for this to be successful, increased standardization of data collection, availability of documentation, collaboration across networks, supporting established networks, developing new networks and keeping track of all networks will significantly enhance the potential for useful PMEM activities and assessments.

ACKNOWLEDGEMENTS

We are grateful to the people that responded to our request for request for suggestion on key words to use in a literature search, or useful publications that might be useful for our efforts to find information on the statistics that are used in environmental and species monitoring programs and that might be useful to determine the impact of products that are subject to post market monitoring. The following people provided information and also agreed that their names could be given in these acknowledgments. Note that the people that we have acknowledged here have not seen this report, and naming them does not indicate that they agree with the contents. Danny Hoofman, Helen Roy and Michael Pocock (CEH Wallingford), Christopher Andrews, Adam Vanbergen and Jan Dick (CEH Edinburgh), Terry Parr and Simon Smart (CEH Lancaster), Leslie Firbank (University of Leeds), Staffan Eklöf (The Swedish Board of Agriculture), Mike O'Driscoll and John Joseph (Department for Environment, Food and Rural Affairs, UK), Alexander Stein (International Food Policy Research Institute, USA), Katherine Bainbridge (Science and Technology Committee, British Parliament), Joe Perry (formerly at Rothamsted Experimental Station), Wiebke Züghart and Mathias Otto (Federal Agency for Nature Conservation (BfN) / Germany) Esther Esteban Rodrigo (DG Agricultural Production and Markets, Madrid). We also gratefully acknowledge the comments made by 7 external, anonymous statisticians who contributed to the section in this report on the description of the appropriate statistical methods for analysing ESN data.

REFERENCES

Abrahantes, J. C., Molenberghs, G., Burzykowski, T., Shkedy, Z., Abad, A. A., & Renard, D. (2004). Choice of units of analysis and modeling strategies in multilevel hierarchical models. Computational statistics & data analysis, 47(3), 537-563.

Benjamini, Y. (2010). Discovering the false discovery rate. Journal of the Royal Statistical Society B, 72: 405–416.

Besag, J. (1975) Statistical Analysis of Non-Lattice Data. The Statistician, 24: 179–195.

Breslow, N.E. and Clayton, D.G. (1993) Approximate Inference in Generalized Linear Mixed Models. Journal of the American Statistical Association, 88: 9–25.

Carey, P.D.; Wallis, S.; Chamberlain, P.M.; Cooper, A.; Emmett, B.A.; Maskell, L.C.; McCann, T.; Murphy, J.; Norton, L.R.; Reynolds, B.; Scott, W.A.; Simpson, I.C.; Smart, S.M.; Ullyett, J.M.. 2008 Countryside Survey: UK Results from 2007. NERC/Centre for Ecology & Hydrology, 105pp. (CEH Project Number: C03259).

Chandler, R. and Scott, M. (2011) Statistical Methods for Trend Detection and Analysis in the Environmental Science. Wiley (Statistics in Practice series)

Cochran, W. G. (1952) The ¬χ2 test of goodness of fit. Annals of Mathematical Statistics, 25: 315

Commission implementing decision of 21 December 2011 concerning the adoption of a financing decision to support voluntary surveillance studies on honeybee colony losses (notified under document C(2011) 9597) (2011/881/EU). OJ L343, 23.12.2011, p.119-120.

Commission implementing decision of 4 July 2012 concerning a financial contribution by the Union to certain Member States to support voluntary surveillance studies on honeybee colony losses (notified under document C(2012) 4396) (2012/362/EU). OJ L176, 6.7.2012, p.65-69.Commission Regulation No 87/2011 of 2 February 2011 designating the EU reference laboratory for bee health, laying down additional responsibilities and tasks for that laboratory and amending Annex VII to Regulation (EC) No 882/2004 of the European Parliament and of the Council. OJ L29, 3.2.2011, p.1-4.

Commission Regulation (EU) No 87/2011 of 2 February 2011 designating the EU reference laboratory for bee health, laying down additional responsibilities and tasks for that laboratory and amending Annex VII to Regulation (EC) No 882/2004 of the European Parliament and of the Council. OJ L29, 3.2.2011, p.1-4.

Congdon, P. (2006) Bayesian Statistical Modelling, 2nd edition. Wiley

Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora (Habitats directive). OJ L206, 22.7.1992, p.7–50.

Council Directive 2000/29/EC of 8 May 2000 on protective measures against the introduction into the Community of organisms harmful to plants or plant products and against their spread within the Community. OJ L169, 10.7.2000, p.1-112.

Cox, D.R. (1975) Partial Likelihood. Biometrika, 62: 269-276

Cressie, N. and Wikle, C.K. (2011) Statistics for Spatio-Temporal Data. Wiley

Davison, A.C. and Hinkley, D.V. (1997) Bootstrap Methods and their Application. Cambridge University Press

Diggle, P.J., Kaimi, I. and Abellana, R. (2010) Partial-Likelihood Analysis of Spatio-Temporal Point-Process Data. Biometrics 66: 347-354.

Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy (Water directive). OJ L327, 22.12.2000, p.1-72.

Directive 2001/18/EC of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC. OJ L106, 17.4.2001, p.1-38.

Directive 2009/128/EC of the European Parliament and of the Council of 21 October 2009 establishing a framework for Community action to achieve the sustainable use of pesticides. OJ L309, 24.11.2009, p.71-86.

Directive 2009/147/EC of the European Parliament and of the Council of 30 November 2009 on the conservation of wild birds (Birds directive). OJ L20, 26.1.2010, p.7–25.

De Haan, L. and Ferreira, A. (2006) Extreme value theory: An introduction. Springer.

De'Ath, G. (2002). Multivariate regression trees: a new technique for modeling species-environment relationships. Ecology 83: 1105-1117.

Edwards, A.W.F. (1972) Likelihood. Cambridge University Press, Cambridge (expanded edition, 1992, Johns Hopkins University Press, Baltimore).

Efron, B. (1982) The jackknife, the bootstrap, and other resampling plans. Society for Industrial and Applied Mathematics, Philadelphia.

Efron, B. (1983) Estimating the error rate of a prediction rule: improvement on cross-validation. Journal of the American Statistical Association, 78: 316-331.

Efron, B. and Tibshirani, R. J. (1994) An Introduction to the Bootstrap. Chapman & Hall/CRC (Monographs on Statistics & Applied Probability).

Efron, B. (2010). Large-Scale Inference. Cambridge University Press.

EFSA (2010) Application of systematic review methodology to food and feed safety assessments to support decision making. *EFSA Journal,* **8,** 90.

Fewster, R. M., Buckland, S. T., Siriwardena, G. M., Baillie, S. R. and Wilson, J. D. (2000) Analysis of population trends for farmland birds using generalized additive models. Ecology, 81: 1970-1984.

Fisher, R. A. (1954) Statistical Methods for Research Workers. 12th Edition. Oliver and Boyd.

Freeman, S.N., N. David G. Noble, Stuart E. Newson and Stephen R. Baillie (2007) Modelling population changes using data from different surveys: the Common Birds Census and the Breeding Bird Survey British Trust for Ornithology, The Nunnery, Thetford, Norfolk, IP24 2PU, UK

Freeman, S.N. and Newson, S.E. (2008) On a log-linear approach to detecting ecological interactions in monitored populations. Ibis, 150, 2, 250-258.

Friedman, M. (1937) The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association, 32: 675–701

Green, P.J. and Silverman, B.W. (1994) Nonparametric Regression and Generalized Linear Models. Chapman and Hall.

Hails et al. 2012. Determining and increasing the sensitivity of existing environmental surveillance monitoring networks to detect unanticipated effects that may occur in the environment in response to the cultivation of genetically modified crops. SFFS0001: Final Report

Hastie, T. J. and Tibshirani, R. J. (1990) Generalized Additive Models. Chapman & Hall/CRC.

Hastie, T., Tibshirani, R. and Friedman, J. (2009) The Elements of Statistical Learning – Data Mining, Inference, and Prediction, 2nd ed. Springer.

Higgins, J.P.T, and Green S. (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Accessed online from www.cochrane-handbook.org. During February 2013

Hosmer, D. W., Lemeshow, S. and May, S. (2011) Applied survival analysis: regression modeling of time to event data. Wiley-Interscience.

Khangura, S., Konnyu, K., Cushman, R., Grimshaw, J., Moher, D., (2012). Evidence summaries: the evolution of a rapid review approach. *Syst Rev* 1, 10.

Kleijn, D., Berendse, F., Smit, R., and Gilissen, N. (2001) Agri-environment schemes do not effectively protect biodiversity in Dutch agricultural landscapes. Nature, 413: 723-725.

Kruskal, J. B. (1964). Nonmetric multidimensional scaling: a numerical method. Psychometrika 29: 115-130.

Kruskal, J.B. and Wallis, W.A. (1952) Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association 47: 583–621

Kolmogorov, A. N. (1933) Sulla Determinazione Empirica di una Leggi di Distributione.

Giornale dell'Istituto Italiano degli Attuari, 4: 83-91.

Lee, Y. and Nelder, J.A. (2001) Hierarchical generalized linear models: a synthesis of generalised linear models, random-effect models and structured dispersions. Biometrika 88: 987-1006.

Liang, K.Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. Biometrika 73: 13-22.

Lin, X. and Zhang, D. (1999) Inference in generalized additive mixed models by using smoothing splines. Journal of the Royal Statistical Society B, 61: 381-400.

Manly, B. F. J. (1992) Randomization and Monte Carlo methods in biology. Chapman and Hall, New York

McCullagh, P. and Nelder, J. (1989) Generalized Linear Models, Second Edition. Boca Raton: Chapman and Hall/CRC.

Nelder, J. and Wedderburn, R. W. M. (1972) Generalized Linear Models. J. R. Statist. Soc. A, 135: 370-384.

O'Connor, A.M., Lovei, G.L., Eales, J., Frampton, G.K., .Glanville, J., Pullin, A.S., Sargeant, J. (2012) Implementation of systematic reviews in EFSA scientific outputs workflow. (CFT/EFSA/AMU/2010/01)

Pannekoek J and Van Strien A, 2005. TRIM 3 Manual (TRends & Indices for Monitoring data). CBS Statistics Netherlands. 58pp.

Patterson, H.D. and Thompson, R. (1971) Recovery of inter-block information when block sizes are unequal. Biometrika 58: 545-554.

Pinheiro, J. and Bates, D.M. (2000) Mixed-Effects Models in S and S-PLUS. Springer.

Pitman, E.J.G. (1937) Significance tests which may be applied to samples from any population. Royal Statistical Society Supplement 4: 119-130 and 225-32 (parts I and II).

Priestley, M. B. (1981) Spectral analysis and time series (Volumes 1 & 2). Academic Press.

Regulation (EC) No 1185/2009 of the European Parliament and of the Council of 25 November 2009 concerning statistics on pesticides. OJ L324, 10.12.2009, p.1–22.

Roy, H.E., Pocock, M.J.O., Preston, C.D., Roy, D.B., Savage, J., Tweddle, J.C. & Robinson, L.D. (2012) Understanding Citizen Science & Environmental Monitoring. Final Report on behalf of UK-EOF. NERC Centre for Ecology & Hydrology and Natural History Museum.

Rue, H., Martino, S. and Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society B, 71: 319-392.

Rundlöf, M. and Smith, H.G. (2006) The effect of organic farming on butterfly diversity depends on landscape context. Journal of Applied Ecology, 43: 1121–1127.

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003) Semiparametric Regression. Cambridge University Press

Smirnov, N. V. (1939) On the Estimation of the Discrepency between Empirical Curves of Distribution for Two Independent Samples, Bulletin of Moscow, 2: 3-16.

Stone, M. (1974) Cross-validatory choice and assessment of statistical predictions. J. R. Statist. Soc. B, 36: 111-147.

ter Braak, C.J. (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology, 67: 1167-1179.

Turner, H. and Firth, D. (2012) Generalized nonlinear models in R: An overview of the gnm package. Cran-R Project.

van den Wollenberg, A. L. (1977). Redundancy analysis: an alternative for canonical correlation analysis. Psychometrika, 42: 207-219.

van Strien, A. J., Pannekoek, J. E. R. O. E. N., & Gibbons, D. W. (2001). Indexing European bird population trends using results of national monitoring schemes: a trial of a new method. Bird Study, 48(2), 200-213.

van Strien, A., Pannekoek, J., Hagemeijer, W. and Verstrael, T. (2004) A loglinear Poisson regression method to analyse bird monitoring data. Bird Census News, 13: 33-39.

Venables, W.N. and Ripley, B.D. (2002) Modern Applied Statistics with S. Fourth edition. Springer.

Vinzi, V., Chin, W.W., Henseler, J. et al., eds. (2010) Handbook of Partial Least Squares. Springer.

Wackernagel, H. (2003) Multivariate geostatistics. Third edition. Springer-Verlag.

Walker, K. J., Critchley, C. N. R., Sherwood, A. J., Large, R., Nuttall, P., Hulmes, S., Rose, R. and Mountford, J. O. (2007) The conservation of arable plants on cereal field margins: an assessment of new agri-environment scheme options in England, UK. Biological Conservation, 136: 260-270.

Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. Biometrika, 61: 439–447.

Wilcoxon, F. (1945) Individual comparisons by ranking methods. Biometrics Bulletin, 1: 80–83.

Wood, S.N. (2000) Modelling and Smoothing Parameter Estimation with Multiple Quadratic Penalties. J. R. Statist. Soc. B, 62: 413-428.

Wood, S. N. (2006) Generalized Additive Models: An Introduction with R. Chapman & Hall/CRC.

Mei, Y. (2010) Efficient scalable schemes for monitoring a large number of data streams. Biometrika, 97: 419-433.

Zhu, H. and Wang, M.-C. (2012) Analysing bivariate survival data with interval sampling and application to cancer epidemiology. Biometrika, 99: 345-361.

Zuur, A., Ieno, E.N. and Smith, G.M. (2007) Analyzing Ecological Data. Springer.

Zuur, A., Ieno, E.N., Walker, N., Saveliev, A. and Smith, G.M. (2009) Mixed Effects Models and Extensions in Ecology with R. Springer.

Zuur, A., Saveliev, A. and Ieno, E.N. (2012) Zero Inflated Models and Generalized Linear Mixed Models with R. Springer.

ADDITIONAL REFERENCES SUGGESTED BY EXTERNAL STATISTICIAN BUT NOT CITED WITHIN THE TEXT HERE

Anderson, M.J. and Thompson, A.A. (2004) Multivariate control charts for ecological and environmental monitoring. Ecological Applications, 14: 1921-1935.

Anderson, M.J. (2008) Animal-sediment relationships revisited: characterising species' distributions along an environmental gradient using canonical analysis and quantile regression splines. Journal of Experimental Marine Biology and Ecology, 366: 16-27.

Besbeas, P. and Morgan, B.J.T. (2012) Kalman filter initialisation for integrated population modelling. Journal of the Royal Statistical Society: Series C (Applied Statistics), 61: 151-162.

Fox, R., Warren, M.S., Brereton, T.M., Roy, D.B. and Robinson, A. (2011) A new Red List of British butterflies. Insect Conservation and Diversity, 4: 159-172.

Hoef, J.M.V. and Boveng, P.L. (2007) Quasi-Poisson vs. Negative Binomial Regression: How Should We Model Overdispersed Count Data? Ecology, 88: 2766-2772.

Kery, M., Gardner, B., Stoeckle, T., Weber, D. and Royle, J.A. (2011) Use of Spatial Capture-Recapture Modeling and DNA Data to Estimate Densities of Elusive Animals. Conservation Biology, 25: 356-364.

Kulldorf, M. (1997) A spatial scan statistic. Communications in Statistics - Theory and Methods, 26: 1481-1496.

McCrea, R.S., Morgan, B.J.T., Gimenez, O., Besbeas, P.T., Lebreton, J.D. and Bregnballe, T. (2010) Multi-site integrated population modelling. Journal of Agricultural, Biological and Environmental Statistics, 15: 539-561.

MacKenzie, D.I., Nichols, J.D., Sutton, N., Kawanishi, K. and Bailey, L.L. (2005) Improving inferences in population studies of rare species that are detected imperfectly. Ecology, 86: 1101-1113.

Newman, K.B., Fernández, C., Thomas, L. and Buckland, S.T. (2009) Monte Carlo inference for state-space models of wild animal populations. Biometrics, 65: 572-583.

Roy, D.B., Rothery, P., Moss, D., Pollard, E. and Thomas, J.A. (2001) Butterfly numbers and weather: predicting historical trends in abundance and the future effects of climate change. Journal of Animal Ecology, 70: 201-217.

Toms, J.D. and Lesperance, M.L. (2003) Piecewise regression: a tool for identifying ecological thresholds. Ecology, 84: 2034-2041.

van Strien, A.J., Termaat, T., Kalkman, V., Prins, M., De Knijf, G., Gourmand, A.L. Houard, X., Nelson, B., Plate, C., Prentice, S., Regan, E., Smallshire, D., Vanappelghem, C. and Vanreusel, W. (2013) Occupancy modelling as a new approach to assess supranational trends using opportunistic data: a pilot study for the damselfly Calopteryx splendens. Biodiversity and Conservation, 22: 673-686.

Wheeler, D.C. and Waller, L.A. (2008) Mountains, valleys, and rivers: The transmission of raccoon rabies over a heterogeneous landscape. Journal of agricultural, biological, and environmental statistics, 13: 388-406.

APPENDICES

1. SEARCH TERMS AND WILD CARDS USED DURING TESTING AND PRODUCTION OF ENDNOTE LIBRARIES.

Search terms with wild cards to show how wild cards were used to cover variations in spelling and pleural terms

Search terms for habitat or ecosystem		Search terms for organism that may be recorded in an ESN		Terms that might indicate statistical analysis carried out	
Search term with wild card	**Wild card includes**	**Search term with wild card**	**Wild card includes**	**Search term with wild card**	**Wild card includes**
agriculture		Alga*	Algae, algal	abundance*	abundances
agri-environment	Agri-environments	amphibian*	Amphibians	air-borne	
alpine		animal*	Animals	annual	
bog*	bogs	arthropod*	Arthropods	biodiversity	
connectivity		bats*	Bats	census*	censuses
dune*	dunes	bee*	Bees	collection*	collections
ecosystem*	ecosystems	beetle*	Beetles	Collection*time *	Collection-times, collection-time
environment*	environments	bird*	Birds	"collection*-*time*"	Collection - time , collection - times
estuary*	estuaries	breeding-bird*	breeding-birds	comparison*	comparisons
farm*	farms, farmland	bug*	Bugs	correlation*	correlations
		bumblebee*	Bumblebees	daily	
forest*	Forests, forestry	butterf*	Butterflies	data-mining	
geographical		carabid*	carabids	decline*	declines
grassland*	grasslands	caterpillar*	caterpillars	Decreas*	Decreases, decreasing
habitat*	habitats	deer		demographic	
heathland*	heathlands	diatom*	Diatoms	demography	
"land-use"		earthworm*	earthworms	densities	
lake*	lakes	fish*	Fishes	density	
Landscape*	landscapes	frog*	Frogs	detection	

Local*	localities	fungi		distribution*	distributions
Long		hedgehog*	Hedgehogs	diversit*	Diversity, diversities
long-term		herb*	Herbs	driver*	drivers
lowland*	lowlands	herbaceous		drought*	droughts
marine		herbivore*	Herbivores	duration	
moorland*	moorlands	hoverfl*	Hoverfly, hoverflies	effect*	effects
mountain*	mountains	insect*	insects	flood*	floods
national		invertebrate*	Invertebrates	historical	
network*	networks	ladybird*	Ladybirds	hourly	
niche		Lepidoptera*	Lepidopteron	impact*	impacts
peat		Macro-invertebrate*	Macro-invertebrates	Increase*	Increases, increasing
Peatland*	peatlands	macroinvertebrate*	macroinvertebrates	Index*	indexes
Peat*-*land*	Peat – land, peat - lands	macrophyte*	macrophytes	indicator*	indicators
Peat*land*	Peat-land, peat-lands	mammal*	mammals	indices	
Plot*	plots	microb*	Microbes, microbial	Intensity*	Intensity, intensities
regional		microorganism*	microorganisms	inter-annual	
riparian		microorganizim*	microorganizims	likelihood	likelihoods
Saltmarsh*	saltmarshs	mollusc*	Molluscs	Measur*	Measurements measuring
salt*-*marsh*	Salt - marsh or salt – marshes	moss*	mosses	Method*	Method, methods
salt*marsh	salt-marsh or salt-marshes	moth*	moths	monitoring	
sea*	seas	mycological		monthly	
semi-natural		passerine*	passerines	mortalit*	Mortality, mortalities
terrestrial		plankton		multi-annual	
Transect*	transects	plant*	Plants	pattern*	patterns
upland*	uplands	predator*	Predators	population*	populations
vegetation		reptile*	Reptiles	prediction*	predictions
freshwater*	freshwaters	shrub*	Shrubs	probabilit*	Probability, probabilities
wetland*	wetlands	snake*	Snakes	quantit*	Quantity quantities
wildland*	wildlands	soil*	Soils	quantitifying	
woodland*	woodlands	species		Regulation	

		toad*	Toads	"Remote*sensing"	Remote sensing, remote-sensing
		tree*	Trees	"Remote*-*sensing"	Remote-sensing, remote – sensing, remote-sensing, remote - sensing
		wader*	Waders	response*	responses
		waterfowl		satellite	
		weed*	weeds	scale*	scales
		zooplankton		seasonal	
				seasonality	
				sensing	
				series	
				severity	
				spatial	
				"species* richness"	Species richness, speciesrichness
				"species *-*richness"	Species-richness, species – richness, species - richness
				spread	
				status	
				surveillance	
				survey*	surveys
				survival	
				survivorship	
				temporal	
				time*	times
				trap*	traps
				trend*	trends
				variation*	variations
				wind*	winds
				yearly	

2. Initial systematic searches using a wide variety of key words that were suggested by (or the same type) as those suggested by experts or obtained from the ecological literature (These searches were not saved).

Using search terms suggested by experts restricted only by research Domain Science and Technology in WoK

Number of references 5702960 Time span 2002-2012 Search 22 February 2013 Changed moth* to moth and moths to avoid mother and bee* to bee and bees to avoid been and still got 4715231 2 April	
WoK search filters	Search terms used
Research Domains=(SCIENCE TECHNOLOGY)	Topic=((agriculture or agrienvironment or agri*environment* or agri*-*environment* or alpine or bog* or connectivity or dune* or ecosystem* or environment* or estuary* or farm* or forest* or geographical or grassland* or grass*land* or grass*-*land* or habitat* or heathland* or heath*land* or heath*-*land* or "land-use" or "land*use" or "land*-*use*"or lake* or landscape* or land*scape or land*-*scape or local* or long or long*-*term or lowland* or low-land * or low*-*land or marine or moorland* or mountain* or national or network* or niche or peat or peatland* or peat*land* or peat*-*land* or plot* regional or riparian or salt marsh* or salt*-*marsh* or salt*marsh or sea* or seminatural* or semi*natural or semi*-*natural or terrestrial or transect* or upland* or vegetation or freshwater* or fresh*-*water* or fresh*-*water* or wetland* or wet*land or wet*-*land* or wildland* woodland*)) AND Topic=((abundance* or air-borne or air*-*borne or air*borne or annual or biodiversity or census* or collection* or Collection-time or comparison* or correlation* or daily or data-mining or decline* or decreas* or demographic or demography or densities or density or detection or distribution* or diversit* or driver* or drought* or duration or effect* or flood* or historical or hourly or impact* or increase* or Index* or indicator* or indices or Intensity* or inter-annual or likelihood or lidar or Measur* or method or monitoring or monthly or mortalit* or multi-annual or pattern* or population* or prediction* or probabilit* or quantit* or quantitifying or Regulation or remote or response* or satellite or scale* or seasonal or seasonality or sensing or series or severity or spatial or species-richness or spread or status or survey* or survival or survivorship or temporal or time* or trap* or trend* or variation* or wind* or yearly)) AND Topic=((algae or algal or amphibian* or animal* or arthropod* or bats* or bee* or beetle* or bird* or breeding-bird* or bug* or bumblebee* or butterf* or carabid* or caterpillar* or deer or diatom* or earthworm* or fish* or frog* or fungi or hedgehog* or herb* or herbaceous or herbivore* or hoverfl* or insect* or invertebrate* or ladybird* or Lepidoptera* or Macro-invertebrate* or macroinvertebrate* or macro*invertebrate* or macro*-*invertebrate or macrophyte* or mammal* or microb* or microorganism* or microorganizim* or micro*organism* or micro*-*organism* or micro*organizm* or micro*-

	organizm or mollusc* or moss* or moth* or mycological or passerine* or plankton or plant* or predator* or reptile* or shrub* or snake* or soil* or species or toad* or tree* or wader* or waterfowl or zooplankton))

Using Expert suggested search terms with Research Area filters in WoS

Number of references 109303 time span 2002-2012 Search 22 February 2013	
WoK search filters	Search terms used
Research Domains=(SCIENCE TECHNOLOGY) AND Research Areas=(ENVIRONMENTAL SCIENCES ECOLOGY OR MARINE FRESHWATER BIOLOGY OR FISHERIES OR BIODIVERSITY CONSERVATION OR AGRICULTURE OR BEHAVIORAL SCIENCES OR GEOGRAPHY OR DEMOGRAPHY OR ZOOLOGY OR PHYSICAL GEOGRAPHY OR PLANT SCIENCES OR COMPUTER SCIENCE OR OCEANOGRAPHY OR PHYSIOLOGY OR MICROBIOLOGY OR WATER RESOURCES OR FORESTRY OR MYCOLOGY OR URBAN STUDIES OR MATHEMATICS OR ENTOMOLOGY or OR MATHEMATICS OR WATER RESOURCES OR AGRICULTURE OR OCEANOGRAPHY OR MYCOLOGY OR FORESTRY OR FISHERIES OR BIODIVERSITY CONSERVATION OR MATHEMATICAL COMPUTATIONAL BIOLOGY OR REMOTE SENSING OR PLANT SCIENCES OR MICROBIOLOGY OR DEMOGRAPHY)	Topic=((agriculture or agrienvironment or agri*environment* or agri*-*environment* or alpine or bog* or connectivity or dune* or ecosystem* or environment* or estuary* or farm* or forest* or geographical or grassland* or grass*land* or grass*-*land* or habitat* or heathland* or heath*land* or heath*-*land* or “land-use” or “land*use” or “land*-*use*”or lake* or landscape* or land*scape or land*-*scape or local* or long or long*-*term or lowland* or low-land * or low*-*land or marine or moorland* or mountain* or national or network* or niche or peat or peatland* or peat*land* or peat*-*land* or plot* regional or riparian or salt marsh* or salt*-*marsh* or salt*marsh or sea* or seminatural* or semi*natural or semi*-*natural or terrestrial or transect* or upland* or vegetation or freshwater* or fresh*-*water* or fresh*-*water* or wetland* or wet*land or wet*-*land* or wildland* woodland*)) AND Topic=((abundance* or air-borne or air*-*borne or air*borne or annual or biodiversity or census* or collection* or Collection-time or comparison* or correlation* or daily or data-mining or decline* or decreas* or demographic or demography or densities or density or detection or distribution* or diversit* or driver* or drought* or duration or effect* or flood* or historical or hourly or impact* or increase* or Index* or indicator* or indices or Intensity* or inter-annual or likelihood or lidar or Measur* or method or monitoring or monthly or mortalit* or multi-annual or pattern* or population* or prediction* or probabilit* or quantit* or quantitifying or Regulation or remote or response* or satellite or scale* or seasonal or seasonality or sensing or series or severity or spatial or species-richness or spread or status or survey* or survival or survivorship or temporal or time* or trap* or trend* or variation* or wind* or yearly)) AND Topic=((algae or algal or amphibian* or animal* or arthropod* or bats* or bee* or beetle* or bird* or breeding-bird* or bug* or bumblebee* or butterf* or carabid* or caterpillar* or deer or diatom* or earthworm* or fish* or frog* or fungi or hedgehog* or herb* or herbaceous or herbivore* or hoverfl* or insect* or invertebrate* or ladybird* or Lepidoptera* or Macro-invertebrate* or macroinvertebrate* or macro*invertebrate* or macro*-*invertebrate or macrophyte* or mammal* or microb* or microorganism* or microorganizim* or micro*organism* or micro*-*organism* or micro*organizm* or micro*-*organizm* or mollusc* or moss* or moth* or mycological or passerine* or plankton or plant* or predator* or reptile* or shrub* or snake* or soil* or species or toad* or tree* or wader* or waterfowl or

	zooplankton))

Comprehensive search terms, filtered by Research Domain, Research Areas and European countries.

Number of references 317699 time span 2002-2012 Search 22 February 2013	
WoK search filters	Search terms used
Research Domains=(SCIENCE TECHNOLOGY) AND Research Areas=(ENVIRONMENTAL SCIENCES ECOLOGY OR MARINE FRESHWATER BIOLOGY OR FISHERIES OR BIODIVERSITY CONSERVATION OR AGRICULTURE OR BEHAVIORAL SCIENCES OR GEOGRAPHY OR DEMOGRAPHY OR ZOOLOGY OR PHYSICAL GEOGRAPHY OR PLANT SCIENCES OR COMPUTER SCIENCE OR OCEANOGRAPHY OR PHYSIOLOGY OR MICROBIOLOGY OR WATER RESOURCES OR FORESTRY OR MYCOLOGY OR URBAN STUDIES OR MATHEMATICS OR ENTOMOLOGY or OR MATHEMATICS OR WATER RESOURCES OR AGRICULTURE OR OCEANOGRAPHY OR MYCOLOGY OR FORESTRY OR FISHERIES OR BIODIVERSITY CONSERVATION OR MATHEMATICAL COMPUTATIONAL BIOLOGY OR REMOTE SENSING OR PLANT SCIENCES OR MICROBIOLOGY OR DEMOGRAPHY) AND Countries/Territories=(SCOTLAND OR GERMANY OR HUNGARY OR IRELAND OR UK OR SPAIN OR FRANCE OR ITALY OR ESTONIA OR NETHERLANDS OR ENGLAND OR SLOVAKIA OR SWEDEN OR POLAND OR SLOVENIA OR ROMANIA OR LATVIA OR FINLAND OR WALES OR NORWAY OR BULGARIA OR BELGIUM OR LUXEMBOURG OR DENMARK OR PORTUGAL OR CZECH REPUBLIC OR AUSTRIA OR GREECE)	Topic=((agriculture or agrienvironment or agri*environment* or agri*-*environment* or alpine or bog* or connectivity or dune* or ecosystem* or environment* or estuary* or farm* or forest* or geographical or grassland* or grass*land* or grass*-*land* or habitat* or heathland* or heath*land* or heath*-*land* or “land-use” or “land*use” or “land*-*use*”or lake* or landscape* or land*scape or land*-*scape or local* or long or long*-*term or lowland* or low-land * or low*-*land or marine or moorland* or mountain* or national or network* or niche or peat or peatland* or peat*land* or peat*-*land* or plot* regional or riparian or salt marsh* or salt*-*marsh* or salt*marsh or sea* or seminatural* or semi*natural or semi*-*natural or terrestrial or transect* or upland* or vegetation or freshwater* or fresh*-*water* or fresh*-*water* or wetland* or wet*land or wet*-*land* or wildland* woodland*)) AND Topic=((abundance* or air-borne or air*-*borne or air*borne or annual or biodiversity or census* or collection* or Collection-time or comparison* or correlation* or daily or data-mining or decline* or decreas* or demographic or demography or densities or density or detection or distribution* or diversit* or driver* or drought* or duration or effect* or flood* or historical or hourly or impact* or increase* or Index* or indicator* or indices or Intensity* or inter-annual or likelihood or lidar or Measur* or method or monitoring or monthly or mortalit* or multi-annual or pattern* or population* or prediction* or probabilit* or quantit* or quantitifying or Regulation or remote or response* or satellite or scale* or seasonal or seasonality or sensing or series or severity or spatial or species-richness or spread or status or survey* or survival or survivorship or temporal or time* or trap* or trend* or variation* or wind* or yearly)) AND Topic=((algae or algal or amphibian* or animal* or arthropod* or bats* or bee* or beetle* or bird* or breeding-bird* or bug* or bumblebee* or butterf* or carabid* or caterpillar* or deer or diatom* or earthworm* or fish* or frog* or fungi or hedgehog* or herb* or herbaceous or herbivore* or hoverfl* or insect* or invertebrate* or ladybird* or Lepidoptera* or Macro-invertebrate* or macroinvertebrate* or macro*invertebrate* or macro*-*invertebrate or macrophyte* or mammal* or microb* or microorganism* or microorganizim* or micro*organism* or micro*-*organism* or micro*organizm* or micro*-*organizm* or mollusc* or moss* or moth* or mycological or passerine* or plankton or plant* or predator* or reptile* or shrub* or snake* or soil* or species or toad* or tree* or wader* or waterfowl or zooplankton))

3. Details of the literature search terms and results obtained that were incorporated into EndNote libraries and saved for this project. Including results for systematic and non systematic searches

Name	Search terms	Exclusions	Number of refs
amphibian	Topic=(amphibian or amphibians) AND Topic=(increas* or decreas* or variation* or difference* or trend* or reduc* or expan* or decline*) AND Topic=(statistic* or meta*-*analysis or meta*analysis)	NOT Topic=(dna or taxonom* or gen*)	583
bat	Topic=(bat or bats) AND Topic=(ecology or environment* or ecosystem*) AND Topic=(increas* or decreas* or variation* or difference* or trend* or reduc* or expan* or decline*)	NOT Topic=(dna or taxonom* or gen*) Refined by: Research Areas=(ENVIRONMENTAL SCIENCES ECOLOGY OR ZOOLOGY OR MATHEMATICAL COMPUTATIONAL BIOLOGY OR PHYSIOLOGY OR BEHAVIORAL SCIENCES OR BIODIVERSITY CONSERVATION OR GEOGRAPHY OR PLANT SCIENCES OR MATHEMATICS OR AGRICULTURE OR COMPUTER SCIENCE OR	628

		FORESTRY) AND Research Domains=(SCIENCE TECHNOLOGY)	
Bee insect moth	Topic=(bee or bees or insect or insects or moth or moths or athropod*) AND Topic=(increas* or decreas* or variation* or difference* or trend* or reduc* or expan* or decline*) AND Topic=(statistic* or meta*-*analysis or meta*analysis) Search language=Auto	NOT Topic=(dna or taxonom* or gen*) Refined by: Research Areas=(ENTOMOLOGY OR ZOOLOGY OR ENVIRONMENTAL SCIENCES ECOLOGY OR MATHEMATICAL COMPUTATIONAL BIOLOGY OR MATHEMATICS OR PLANT SCIENCES OR AGRICULTURE OR DEMOGRAPHY OR FORESTRY OR MICROBIOLOGY OR BIODIVERSITY CONSERVATION) AND Research Domains=(SCIENCE TECHNOLOGY) Note cannot use bee* because this brings “been” and not moth* because get mother	1351
Bird to include for example waterbird and birds	Topic=(*bird* or passerine*) AND Topic=(ecology or environment* or ecosystem*) AND Topic=(increas* or decreas* or variation* or difference* or trend* or reduc* or	NOT Topic=(dna or taxonom* or gen*) Refined by: Research Domains=(SCIENCE TECHNOLOGY)	544

	expan* or decline*) AND Topic=(statistic* or meta*-*analysis or meta*analysis)		
Butterflies	Topic=(butterfly or butterflies or lepidopt*) AND Topic=(ecology or environment* or ecosystem*) AND Topic=(increas* or decreas* or variation* or difference* or trend* or reduc* or expan* or decline*) AND Topic=(statistic* or meta*-*analysis or meta*analysis)	NOT Topic=(dna or taxonom* or gen*)	69
Plant species	Topic=(("plant species")) AND Topic=(ecology or environment* or ecosystem*) AND Topic=(increas* or decreas* or variation* or difference* or trend* or reduc* or expan* or decline*)	NOT Topic=(dna or taxonom* or gen*)	4635
Reptiles	Topic=(("reptile or reptiles")) AND Topic=(ecology or environment* or ecosystem*) AND Topic=(increas* or decreas* or variation* or difference* or trend* or reduc* or expan* or decline*)	NOT Topic=(dna or taxonom* or gen*)	87
			7810

Systematic search on Web of Knowledge for network publications by network name

Name and home website	Search terms	Exclusions and reasons	results
United Kingdom			
Agri-environment scheme monitoring in England			211
Uk Butterfly monitoring network http://www.ukbms.org/	Topic=(butterfly monitoring scheme or UKBMS)		72
Bee, Wasps and Ants recording society UK			0
Botanical Society of the British Isles http://www.bsbi.org.uk/			7
BirdTrack http://www.bto.org/volunteer-surveys/birdtrack/about			4
British Arachnological Society http://www.britishspiders.org.uk/			3
British Bryological Society http://www.britishbryologicalsociety.org.uk/			4
British Trust for Ornithology http://www.bto.org/	Topic=("british trust for ornithology")	Cannot use BTO because of number of non-relevant refs	68
Butterfly monitoring scheme http://www.ukbms.org/			72

Butterfly Conservation http://butterfly-conservation.org/			123
Cranefly, ground beetle, hoverfly, ladybird, Opiliones, Pseudoscorpions, Staphylinidae, stenini and spider Recording Schemes http://www.ladybird-survey.org/recording.aspx http://www.hoverfly.org.uk/portal.php http://srs.britishspiders.org.uk/ http://www.brc.ac.uk/recording_schemes.asp		not sorted by organism	19
Countryside survey http://www.countrysidesurvey.org.uk/			36
British Dragonfly Society http://www.british-dragonflies.org.uk/			1
Buglife http://www.buglife.org.uk/			1
Breeding Bird Survey http://www.bto.org/			380
British Trust for Ornithology and *bird* http://www.bto.org/			68
Environmental change network (ECN)	Topic=("environmental change network")	ECN gave many references about physics electronics telecommunications etc.	24
Forest Inventory network NFI Could cover several countries	Topic=("forest inventory" or "national forest inventory")	Cannot use NFI because of number	587

		of irrelevant refs	
Garden BirdWatch http://www.bto.org/volunteer-surveys/gbw			1
Lesser Horseshoe bat Summer Roost Survey http://www.bats.org.uk/pages/about_us.html			0
Mammal Society http://www.mammal.org.uk/			7
Royal society for the protection of birds"	Topic=(rspb or "royal society for the protection of birds")		54
national amphibian reptile recording scheme http://narrs.org.uk/index.php	Topic=(narrs or national amphibian reptile recording scheme)		0
National Bat Monitoring Programme http://www.bats.org.uk/pages/batmonitoring.html			4
National Moth Recording Scheme http://www.mothscount.org/text/27/national_moth_recording_scheme.html			4
Natural England http://www.naturalengland.org.uk			14
Plantlife http://www.plantlife.org.uk/			7
Rothamsted Insect trapping network http://www.rothamsted.ac.uk/			0
Royal Society for the protection of birds			54

http://www.rspb.org.uk/ourwork/science/publications.aspx			
Small mammal Monitoring small.mammals@mammal.org.uk			7
Spider recording scheme http://srs.britishspiders.org.uk/			0
The Mammal Society monitoring programmes http://www.mammal.org.uk/mmm			0
Tracking mammals partnership http://jncc.defra.gov.uk/page-1757			2
UK Phenology Network http://naturestimeline.com/			2
Wildlife Trust http://www.wildlifetrusts.org/			23
Czechoslovakia			
"Czech Society for Ornithology" or "Česká společnost ornitologická" http://www.birdlife.cz/index.php?a=cat.53	Topic=("Czech Society for Ornithology" or "Česká společnost ornitologická" or CSO) AND Topic=(*birds*)	Used *bird* to avoid multiple references related to waste water (all 6 not useful)	6
French networks Multiple sites searched for together	Topic=("The French National Agency for Water and Aquatic Environments" or "L'Office national de l'eau et des milieux aquatiques - Eaufrance"		1

	or "Soil Scientific Interest Group" or "Groupement d'Intérêt Scientifique Sol" or "National water and aquatic environments - Hydrobiological and Fish Network" or "Informations sur les Milieux Aquatiques pour la Gestion Environnementale" or "Office national de l'eau et des milieux aquatiques" or "National inventory of the natural heritage" or "Muséum national d'Histoire naturelle" or "Conservatoires Botaniques Nationaux" or "Research Centre for Ringing Bird Populations" or "Centre de Recherches par le Baguage des Populations d'Oiseaux" or "French Odonatology Society" or "Société française d'Odonatologie" or "Ministry of Food, Agriculture and Forests - Environment" or "Ministère de L'Agriculture de l'Agroalimentaire et de la Forêt - Environnement")		
Netherlands			
	Topic=("SOVON Bird Research		Only I

SOVON Bird Research Netherlands	Netherlands" or sovon or Sovon or "SOVON Vogelonderzoek Nederland" or "Bird Research Netherlands" or "Vogelonderzoek Nederland")		article in English
Netherlands sites Searches of multiple sites together	Topic=("Reptile, Amphibian & Fish Conservation the Netherlands" or "Reptielen Amfibieën Vissen Onderzoek Nederland" or "TINEA Foundation – Small Butterflies" or "Stichting TINEA – Kleine Vlinders" or "Foundation fauna and flora research" or "Stichting Veld Onderzoek Flora en Fauna" or "Netherlands Entomological Society" or "Nederlandse Entomologische Vereniging" or "Bryology and Lichenology Working Group" or "De Bryologische en Lichenologische Werkgroep" or "European Invertebrate Survey Netherlands" or "Stichting European Invertebrate Survey (EIS) - Nederland" or "Dutch Butterfly Conservation" or Vlinderstichting or "Dutch Mammal Society" or	references not sorted by site No publications automatically available	36

	Zoogdierenvereniging" or "Plant Research International (Bees)" or "Foundation Dutch Flora Research" or "Stichting Floristisch Onderzoek Nederland" or "TINEA Foundation – Small Butterflies" or "Stichting TINEA – Kleine Vlinders")		
Germany			
German butterfly network Tagfalter-de http://www.tagfalter-monitoring.de/			14
Germany Multiple sites searched for together	Topic=("German Soil Monitoring" or "Boden-Dauerbeobachtungsflächen" or "The Federation of German Avifaunisten" or "Dachverband Deutscher Avifaunisten" or "German Bee monitoring" or "Deutsches Bienenmonitoring" or "Society of the German Bee Research Institutes" or "Arbeitsgemeinschaft der Institute für Bienenforschung" or "German Butterfly Monitoring" or "Tagfalter-Monitoring Deutschland" or "The core environmental indicator system of the Federal Environment Agency" or "Umwelt-Kernindikatorensystem des Umweltbundesamtes")		5

Portugal			
Portuguese Society for the Study of Birds http://www.spea.pt/en/	Topic=("Portuguese Society for the Study of Birds" or "Sociedade Portuguesa para o Estudo das Aves)	Can not include SPEA as many references that are nothing to do with ecology. No scientific references recommended AIRO - SPEA publication	0
European general			
European general networks searched for together	Topic=("BirdLife International European division" or "Butterfly Conservation Europe" or "Co-ordinated bird ringing throughout Europe" or "European Red List" or "ENVironmental ASsessment of Soil for mOnitoring" or "The European Birds Census Council")		22
			1943

Results for Non systematic search – visits to web sites of selected ESN’s

Name and home website	Website address of publications	Description of publications	availability
United Kingdom			
Agricultural-Supply chain initiative on modified agricultural crops	none		0

http://www.scimac.org.uk/who.html			
Amphibian - National amphibian reptile recording scheme http://narrs.org.uk/index.php	http://narrs.org.uk/news.php	One report downloaded Found 13 refs in the report	2
Bee, Wasps and Ants recording society UK http://www.bwars.com/	http://www.bwars.com/index.php?q=downloads	About 95 publications mainly maps of distributions, notes on taxonomy and new records of species observations in different places.	1 looked statistical in English
Biodiversity- The Biodiversity Impacts of Climate Change Observation Network	http://bicco-net.org/	75	32 selected
British trust for Ornithologists http://www.bto.org/	http://www.bto.org/research-data-services/publications this has links to each year of publications http://www.bto.org/research-data-services/publications/publications-list http://www.bto.org/research-data-services/publications/research-reports http://www.bto.org/research-data-services/publications/bird-study http://www.bto.org/volunteer-surveys/ringing/publications/ringing-migration	Several places for publications. Searched through lists many not available	56
Biodiversity- The Biodiversity Impacts of Climate Change	http://bicco-net.org/climate-change-publications	75 publications on page (since 2009)	32 selected

Observation Network http://bicco-net.org/			
Butterfly- Uk Butterfly monitoring network http://www.ukbms.org/		100+ papers working on downloading	
Environmental change network	http://www.ecn.ac.uk/	Looked at 193	44 selected
Rothamsted Insect trapping network http://www.rothamsted.ac.uk/	LIGHT-TRAP NEWSLETTER No. 36. DECEMBER 2011 http://www.rothamsted.ac.uk/insect-survey/LightTrapNewsletters/LightTrap36.pdf LIGHT-TRAP NEWSLETTER No. 35. 2010 http://www.rothamsted.ac.uk/insect-survey/LightTrapNewsletters/LightTrap35.pdf LIGHT-TRAP NEWSLETTER No. 34. 2009 http://www.rothamsted.ac.uk/insect-survey/LightTrapNewsletters/LightTrap34.pdf LIGHT-TRAP NEWSLETTER No. 33. 2008 http://www.rothamsted.ac.uk/insect-survey/LightTrapNewsletters/LightTrap33.pdf LIGHT-TRAP NEWSLETTER No. 32. 2007 http://www.rothamsted.ac.uk/insect-survey/LightTrapNewsletters/LightTrap32.pdf LIGHT-TRAP NEWSLETTER No. 31. 2006 http://www.rothamsted.ac.uk/insect-survey/LightTrapNewsletters/LightTrap31.pdf LIGHT-TRAP NEWSLETTER No. 30. 2005	Most publications listed in annual newsletters not directly on site. Looked at lists in annual newsletters (2012 not available). All have to find manually.	9 Selected from reports-issues: 2011 1/11 2010 0/11f 2009 3/22 2008 0/31 2007 0/17 2006 1/17 2005 0/18 2004 2/10 2003 1/16 2002 1/14

	http://www.rothamsted.ac.uk/insect-survey/LightTrapNewsletters/LightTrap31.pdf LIGHT-TRAP NEWSLETTER No. 29. 2004 http://www.rothamsted.ac.uk/insect-survey/LightTrapNewsletters/LightTrap29.pdf LIGHT-TRAP NEWSLETTER No. 28. 2003 http://www.rothamsted.ac.uk/insect-survey/LightTrapNewsletters/LightTrap28.pdf LIGHT-TRAP NEWSLETTER No. 27. 2002 http://www.rothamsted.ac.uk/insect-survey/LightTrapNewsletters/LightTrap27.pdf		
Belgium			
ARABEL - Arachnologia Belgica http://www.arabel.ugent.be/nl.php	http://www.arabel.ugent.be/nl.php?page=nb05	Most related to taxonomy or new records of occurrence Most recent 2005 has 5 papers in Dutch – 2004 11 papers and 2003 6 papers. No publications in English.	0
Belgion conservation Biologie de l'evolution et de la conservation http://www.bionat.ulg.ac.be/index.php?mn=0&pg=0	Page of publications http://www.bionat.ulg.ac.be/index.php?mn=1&pg=130&tri=all http://www.bionat.ulg.ac.be/index.php?mn=3&pg=330	About 130 pubs from 2002 – 2012 most taxonomic or evolutionary or records -not useful for statistics	0
Czechoslovakia			
Bee research institute	Publications listed by person only	Only listed until 2001 last in 2001	0

Výzkumný ústav včelařský v Dole http://www.beedol.cz/bee-research-institute/	eg director Dr. Ing. František Kamler http://beedol.cz/files/kamler_bibl.pdf Ing. Vladimír Veselý, CSc. http://beedol.cz/files/vesely_bibl.pdf	and few in any year in English	
Mapping and protection of Czech butterflies Mapování a ochrana motýlů České republiky http://www.lepidoptera.cz/	http://www.lepidoptera.cz/publikace/clanky	Gives two publications in English not available	2
http://www.birdlife.cz/index.php?a=cat.53 "Czech Society for Ornithology" or "Česká společnost ornitologická"		If use CSO all seem to be about waste water Website does not have list of publications just details of the Sylvia journal with many publications from Czech work and elsewhere. – Sylvia not listed in web of Knowledge	0
Alternative for "Czech Society for Ornitology" or "Česká společnost ornitologická" http://www.birdlife.cz/index.php?ID=2402		Link to Journal Sylvia abstracts only for 2012	0
Alternative for Czech Bird publications http://jpsp.birds.cz/vysledky.php?menu=publications		None give need to send request for specific publications	0
Germany			

tagfalter-monitoring.de (*Jill need translation*) http://www.tagfalter-monitoring.de/	Link to http://www.lepiforum.de/lepidopterenforum/forum/literatur_tagfalter.htm For example http://www.bluehende-landschaft.de/ for beekeepers	Downloaded an saved an example report gives literaturesfor butterfles moths small butterflies and caterpillars but all are pictures and guides and no scientific studies Other links to web sites - need to check	2
Portugal			
Portuguese Society for the Study of Birds http://www.spea.pt/en/	(http://www.spea.pt/pt/publicacoes/anuario-ornitologico/ http://www.spea.pt/pt/publicacoes/airo/ http://www.spea.pt/pt/publicacoes/airo/ http://www.spea.pt/en/publications/pardela/	Pardella is a magazine with a sample copy 2011 and 2012 available – required a subscription Advertise Airo scientific publication – need to purchase- listed in Web of Knowledge	0
Slovakia			
Slovakia- Lepidoptera of Slovak butterflies web site http://www.lepidoptera.sk/uvod		No publications apart from maps, pictures and guides to butterfly species	0
Slovensko Birds of Slovakia web site	http://vtaky.sk/stranka/56-Publikacie.html http://www.tichodroma.sk/en/index.html	Publication Tichodroma Magazine [ISSN 1337-026X] not in Web of Knowledge need subscription	2 out of 10 in English selected

Sweden			
Swedish Ornithological Society www.sofnet.org Swedish Ornithological Society Sveriges Ornitologiska Förening Also link from Lund University http://www4.lu.se/o.o.i.s/15056	Looked for www.sofnet.org Selected research tab got http://www.sofnet.org/sveriges-ornitologiska-forening/forskning/ then to environmental monitoring http://www.sofnet.org/sveriges-ornitologiska-forening/forskning/miljoovervakning/ then to Swedish bird survey http://www.zoo.ekol.lu.se/birdmonitoring/ then to publications Publications was sent to http://www.zoo.ekol.lu.se/birdmonitoring/Publikationer.htm	62 publication listed on pubs page 14 in English	6 selected
			182

4. RESULTS OF STATISTICAL METHOD COMPARISON SIMULATION STUDY

Table A4.1 Data types generated in the simulation run, together with an Id to enable cross-comparison with results tables and figures.

Data Type Id	Response Type	No. of Sites	Start Value	Error	Length of Monitoring	Samples per Year	Change %	No. of Variables to Simulate
1	Count	10	100	NA	5	12	0.1	10
2	Count	50	100	NA	10	1	0.1	10
3	Count	1	100	NA	5	52	0.1	10
4	Count	10	100	NA	25	1	0.1	10
5	Count	1	100	NA	25	12	0.1	10
6	Count	1	100	NA	10	1	0.1	10
7	Count	10	10	NA	5	12	1	10
8	Count	50	10	NA	10	1	1	10
9	Count	1	10	NA	5	52	1	10
10	Count	10	10	NA	25	1	1	10
11	Count	1	10	NA	25	12	1	10
12	Count	1	10	NA	10	1	1	10
13	Continuous	10	100	25	5	12	0.1	10
14	Continuous	50	100	25	10	1	0.1	10
15	Continuous	1	100	25	5	52	0.1	10
16	Continuous	10	100	25	25	1	0.1	10
17	Continuous	1	100	25	25	12	0.1	10
18	Continuous	1	100	25	10	1	0.1	10
19	Continuous	10	100	5	5	12	1	10
20	Continuous	50	100	5	10	1	1	10
21	Continuous	1	100	5	5	52	1	10
22	Continuous	10	100	5	25	1	1	10
23	Continuous	1	100	5	25	12	1	10
24	Continuous	1	100	5	10	1	1	10
25	Presence/Absence	10	0.5	NA	5	12	0.1	10
26	Presence/Absence	50	0.5	NA	10	1	0.1	10
27	Presence/Absence	1	0.5	NA	5	52	0.1	10
28	Presence/Absence	10	0.5	NA	25	1	0.1	10
29	Presence/Absence	1	0.5	NA	25	12	0.1	10
30	Presence/Absence	1	0.5	NA	10	1	0.1	10
31	Presence/Absence	10	0.9	NA	5	12	1	10
32	Presence/Absence	50	0.9	NA	10	1	1	10
33	Presence/Absence	1	0.9	NA	5	52	1	10
34	Presence/Absence	10	0.9	NA	25	1	1	10
35	Presence/Absence	1	0.9	NA	25	12	1	10
36	Presence/Absence	1	0.9	NA	10	1	1	10

Table A4.2 List of methods used to analyse the data within the simulation runs, together with an Id for cross-comparison with results tables and figures.

Method	Method Type Id
Generalised Linear Model (known family, eg Poisson)	1
Generalised Linear Model (Gamma distribution assumed)	2
Generalised Additive Model (known family, eg Poisson)	3
Generalised Additive Model (Gamma distribution assumed)	4
Generalised Linear Mixed Model (Maximum Likelihood fit) | site = random	5
Generalised Linear Mixed Model (REML fit) | site = random	6
Generalised Linear Mixed Model (MCMC fit) | site = random	7
Generalised Linear Mixed Model (PQL fit) | site = random	8
Generalised Linear Mixed Model (PQL fit, AR (1) process incl) | site = random	9
Generalised Linear Mixed Model (REML fit, AR(1) process incl) | site = random	10
Generalised Estimating Equations (known family) | site = random	11
Generalised Estimating Equations (Gamma distribution assumed) | site = random	12
Bootstrap based resampling	13
Wilcox Test	14
Kruskal Wallis test	15
Redundancy Analysis	16
CUSUMS	17

5.

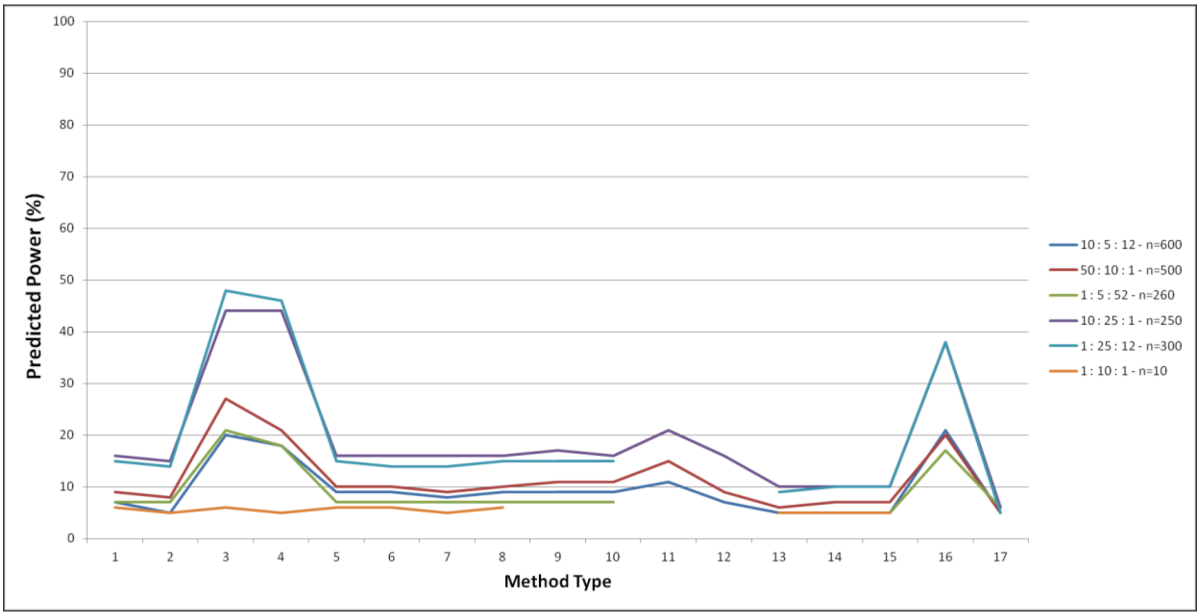
Count data with a hypothetical 0.1% Annual Change

Table showing relative power of each method to detect a 0.1% annual change for each of the corresponding data types as given in table A4.1.

	Method Type																
Data Type	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	7	5	20	18	9	9	8	9	9	9	11	7	5	5	5	21	5
2	9	8	27	21	10	10	9	10	11	11	15	9	6	7	7	20	5
3	7	7	21	18	7	7	7	7	7	7			5	5	5	17	6
4	16	15	44	44	16	16	16	16	17	16	21	16	10	10	10	38	6
5	15	14	48	46	15	14	14	15	15	15			9	10	10	38	5
6	6	5	6	5	6	6	5	6					5	5	5		

Relative power of each method plotted as a graph.

Legend indicate the no. of sites: no. of years : no. per year – n=total sample size



6.

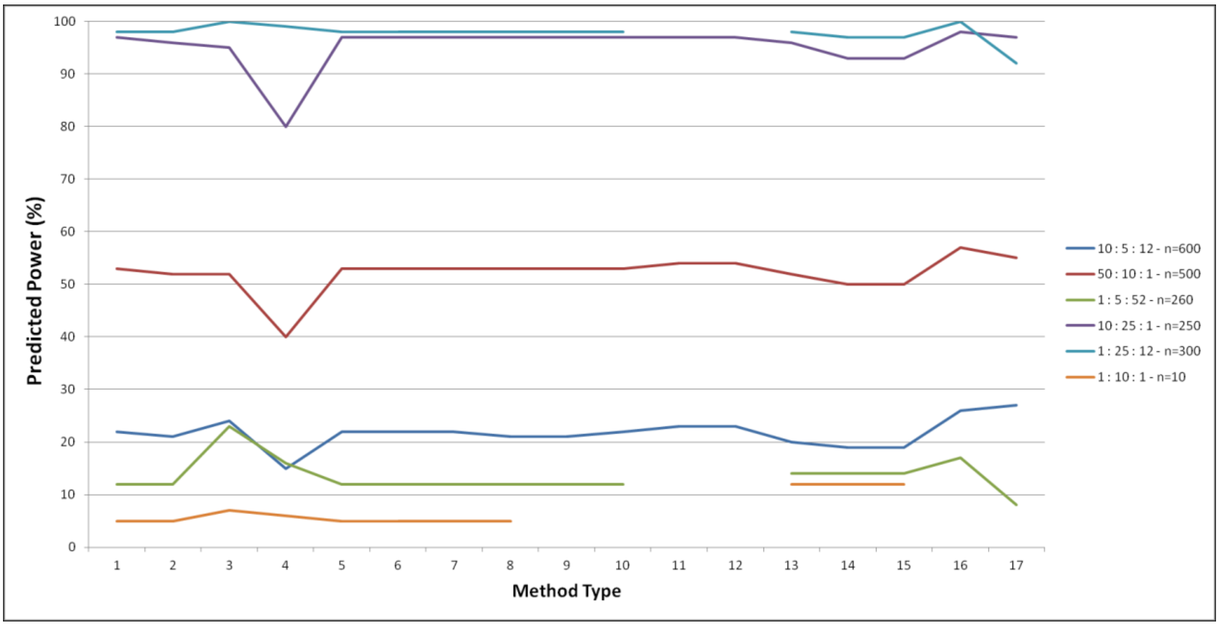
Count data with a hypothetical 1% Annual Change

Table showing relative power of each method to detect a 1% annual change for each of the corresponding data types as given in table A4.1.

	Method Type																
Data Type	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
7	22	21	24	15	22	22	22	21	21	22	23	23	20	19	19	26	27
8	53	52	52	40	53	53	53	53	53	53	54	54	52	50	50	57	55
9	12	12	23	16	12	12	12	12	12	12			14	14	14	17	8
10	97	96	95	80	97	97	97	97	97	97	97	97	96	93	93	98	97
11	98	98	100	99	98	98	98	98	98	98			98	97	97	100	92
12	5	5	7	6	5	5	5	5					12	12	12		

Relative power of each method plotted as a graph.

Legend indicate the no. of sites: no. of years : no. per year – n=total sample size



7.

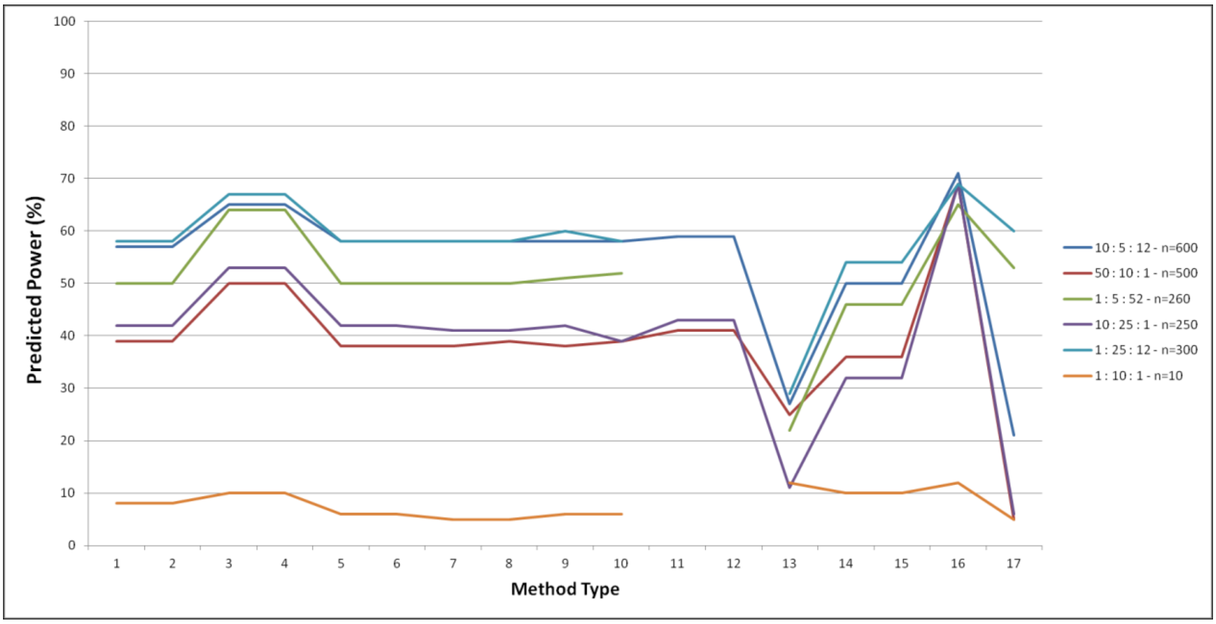
Continuous data with a hypothetical 0.1% Annual Change

Table showing relative power of each method to detect a 0.1% annual change for each of the corresponding data types as given in table A4.1.

	Method Type																
Data Type	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
13	57	57	65	65	58	58	58	58	58	58	59	59	27	50	50	71	21
14	39	39	50	50	38	38	38	39	38	39	41	41	25	36	36	69	5
15	50	50	64	64	50	50	50	50	51	52			22	46	46	65	53
16	42	42	53	53	42	42	41	41	42	39	43	43	11	32	32	69	6
17	58	58	67	67	58	58	58	58	60	58			29	54	54	69	60
18	8	8	10	10	6	6	5	5	6	6			12	10	10	12	5

Relative power of each method plotted as a graph.

Legend indicate the no. of sites: no. of years : no. per year – n=total sample size



8.

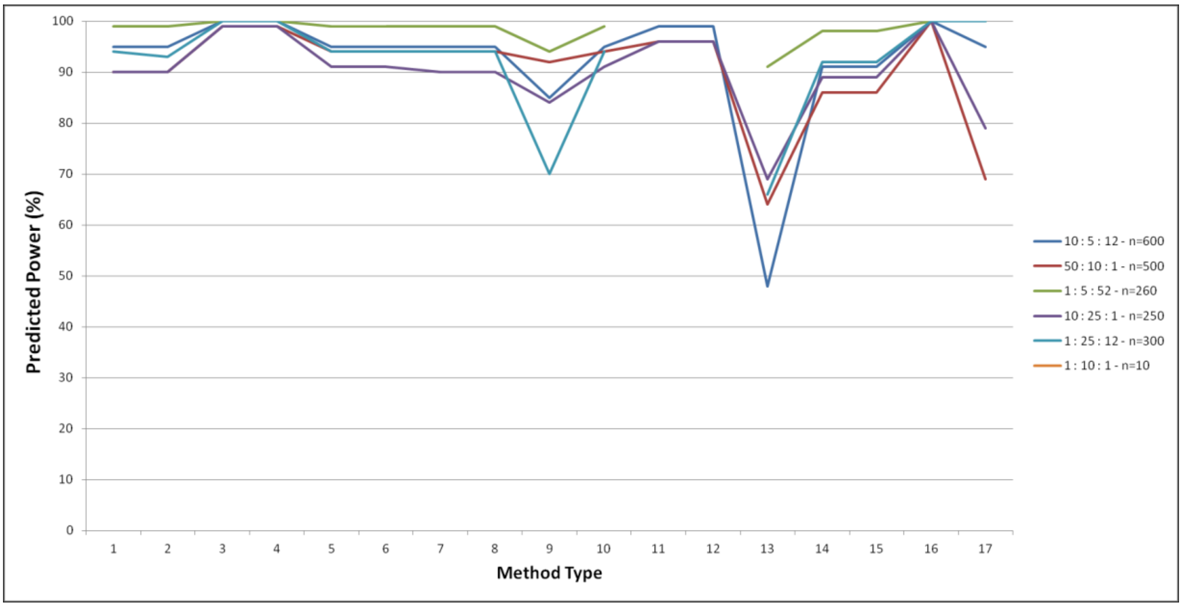
Continuous data with a hypothetical 1% Annual Change

Table showing relative power of each method to detect a 1% annual change for each of the corresponding data types as given in table A4.1.

	Method Type																
Data Type	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
19	95	95	100	100	95	95	95	95	85	95	99	99	48	91	91	100	95
20	90	90	99	99	94	94	94	94	92	94	96	96	64	86	86	100	69
21	99	99	100	100	99	99	99	99	94	99			91	98	98	100	100
22	90	90	99	99	91	91	90	90	84	91	96	96	69	89	89	100	79
23	94	93	100	100	94	94	94	94	70	94			66	92	92	100	100
24																	

Relative power of each method plotted as a graph.

Legend indicate the no. of sites: no. of years : no. per year – n=total sample size



9.

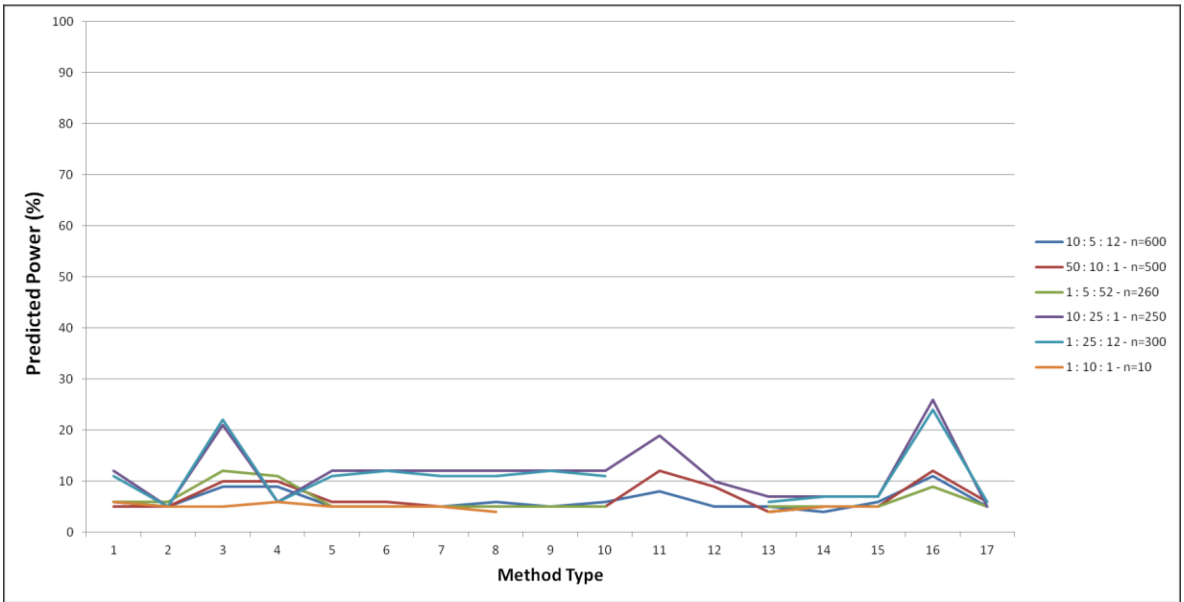
Presence Absence data with a hypothetical 0.1% Annual Change

Table showing relative power of each method to detect a 0.1% annual change for each of the corresponding data types as given in table A4.1.

	Method Type																
Data Type	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
25	6	5	9	9	5	5	5	6	5	6	8	5	5	4	6	11	5
26	5	5	10	10	6	6	5	5	5	5	12	9	4	5	5	12	6
27	6	6	12	11	5	5	5	5	5	5			5	5	5	9	5
28	12	5	21	6	12	12	12	12	12	12	19	10	7	7	7	26	5
29	11	5	22	6	11	12	11	11	12	11			6	7	7	24	6
30	6	5	5	6	5	5	5	4					4	5	5		

Relative power of each method plotted as a graph.

Legend indicate the no. of sites: no. of years : no. per year – n=total sample size



10.

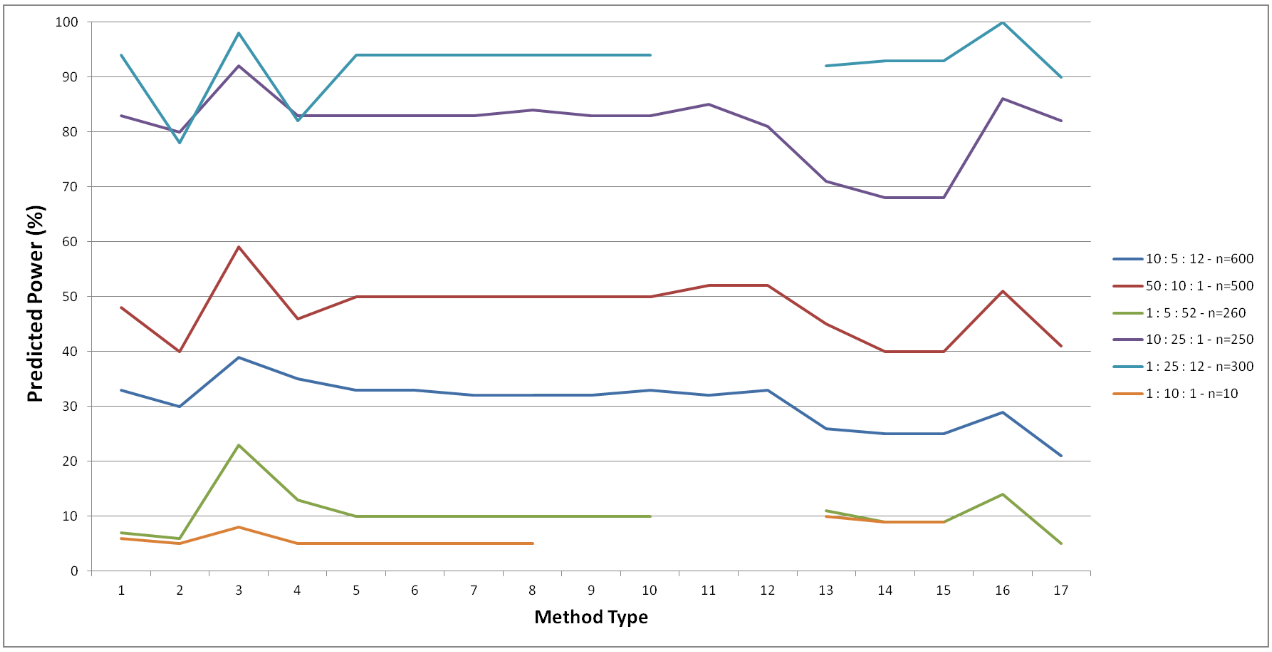
Presence Absence data with a hypothetical 1% Annual Change

Table showing relative power of each method to detect a 1% annual change for each of the corresponding data types as given in table A4.1.

	Method Type																
Data Type	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
31	33	30	39	35	33	33	32	32	32	33	32	33	26	25	25	29	21
32	48	40	59	46	50	50	50	50	50	50	52	52	45	40	40	51	41
33	7	6	23	13	10	10	10	10	10	10			11	9	9	14	5
34	83	80	92	83	83	83	83	84	83	83	85	81	71	68	68	86	82
35	94	78	98	82	94	94	94	94	94	94			92	93	93	100	90
36	6	5	8	5	5	5	5	5					10	9	9		

Relative power of each method plotted as a graph.

Legend indicate the no. of sites: no. of years : no. per year – n=total sample size



5. GUIDANCE NOTES FOR READING AND INTERPRETING THE PYTHIA DATABASE OF ESNS.

Data on the organisations (ESNs) and programmes (ESPS) ware collected in different sheets. The relation between the two can be found either by looking at column D on the “programmes” sheet (for ESPs belonging to which ESN) or on the separate excel table “link organization-programme” (for the ESNs having which ESPs).

Data are collected as found on the website or in manuals and/or reports. Possible values for each item were either a selected number of predetermined values or were free to fill in as described in Tables 10 and 11 in the report.

If a certain data type was not readily available it was indicated as “unknown”. “Blanks” are items that were not looked for, as is the case for most EuMon derived ESNs/ESPs. In those cases only the relevant information was taken from the EuMon and not further elaborated. A link to the EuMon website is provided for more information.

Sample size is interpreted in 2 ways: either as numbers of sites monitored or as the size of 1 sample.

When “Depending on species/parameters” was used the number of posiible values was too high to fit the table.

N.a. means “not applicable”.

6. EXAMPLE OF USING THE GENERIC POWER EQUATION

To use the generic power equations developed in Section 4 to produce an estimate of power, one must be aware that the coefficients are given on the logit scale. Let us take as an example the coefficients from the model fitted to count data, where the effect we wish to investigate the power of is a negative one. So, the model to use would be the Poisson model for $\alpha<0$, which is provided below.

Predictor	**Poisson**
	$\alpha<0$
Intercept	-2.67
α	92.08
α^2	-178.36
α:Slope	-92.24
α:Abundance_Mean	-27.24
α:Abundance_Var	-28.17
α:Duration	-13.62
α:Missed	61.88
α:N_Sites*(1-Treated)	-0.029
α:N_Sites*Treated	-0.55

If data on the nine predictor variables were available, power could easily be estimated. Let's take the nine predictor variables as defined for the first Poisson case in Section 4.2.

Predictor	**Poisson**
α	-0.1
Slope	0
Abundance_Mean	0.1
Abundance_Var	2
N_Sites	250
Duration	5
Missed	0.3
Treated	0.5

The power would then be estimated by the following steps:

g_power = ((-2.67) + (92.08)-0.1 +*
(-178.36)-0.1*-0.1 +*
(-92.24)-0.1*0 +*
(-27.24) -0.1*0.1 +*
(-28.17) -0.1*2 +*
(-13.62) -0.1*5 +*
(61.88) -0.1*0.3 +*
(-0.029) -0.1*250*(1-0.5) +*
(-0.55) -0.1*250*(0.5))*

power = exp(g_power) / (1 + exp(g_power))

which in this case returns a value of 0.988.

Glossary [and/or] abbreviations

ESN: environmental surveillance networks

ESP: environmental surveillance programme

PMEM: post-market environmental monitoring