Journal of Statistical Planning and Inference 144 (2014) 69–80

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Design and analysis of experiments testing for biodiversity effects in ecology

R.A. Bailey [a,*], Julia Reiss [b]

[a] School of Mathematical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK
[b] Department of Life Sciences, Whitelands College, University of Roehampton, London SW15 4JD, UK

ARTICLE INFO

Available online 28 September 2012

Keywords:
Biodiversity
Design of experiments
Family of models
Hasse diagram

ABSTRACT

It is now widely believed that biological diversity is good for the natural environment. One way that ecologists test this is to place random collections of species in mini-environments and then measure some outcome. Statisticians have been working with fresh-water ecologists to improve this in two ways. The first is that the subsets of species are carefully chosen, not random. The second is that a nested family of plausible models is fitted. The results of three experiments suggest that biodiversity can have no effect at all, but that there are other plausible underlying mechanisms.

Implications for the design of such experiments, the understanding of the family of models, and the analysis of the data are discussed.

1. Introduction

There are many experiments in ecology whose results seem to suggest that biodiversity is generally a good thing. Often a large collection of different species is considered, and random subsets of these species are used as treatments and put into some artificial set-up that mimics nature. The measured response is some eco-desirable outcome. Very often, the conclusion is that the greater the number of different species, the better the outcome.

For example, Bell et al. (2005) used random subsets from a collection of 72 bacterial species. These were grown on sterile leaf discs in a sterile fluid to resemble bacterial assemblages found on decomposing beech leaves. The authors found that bacteria were "more active" when species richness was high (they showed higher respiration); as a consequence species-rich assemblages will be able to decompose leaves faster than species-poor ones. Another example comes from Cardinale (2011), who showed that biodiversity of stream algae improves stream water quality because diverse algae assemblages can take greater advantage of differences in their environment (so-called *niche partitioning*); that is, they can grow better when species richness is high and take up more pollutants as a consequence. He used a laboratory set-up consisting of cultured algae that were added to artificial "mini-streams".

The experiment by Cardinale was very much in the tradition of those plant ecologists who were the first to design "biodiversity and ecosystem functioning" (B-EF) experiments in the 1990s. These experiments specifically addressed the effects of biodiversity (species richness) on particular "ecosystem processes", such as nutrient uptake (see Loreau et al., 2002). Because of global change and the species loss it causes, B-EF experiments are now a major research topic in Ecology. These experiments all vary in terms of the species used (often they use either plant or animal species and do not mix

* Corresponding author.
E-mail addresses: r.a.bailey@qmul.ac.uk (R.A. Bailey), Julia.Reiss@roehampton.ac.uk (J. Reiss).
URLS: http://www.maths.qmul.ac.uk/~rab (R.A. Bailey), http://www.roehampton.ac.uk/Staff/JuliaReiss (J. Reiss).

0378-3758/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2012.09.008

them), response variables measured (for example, how productive a plant assemblage is or how good an animal assemblage is at using food resources) and in terms of their statistical analysis and experimental design (for example, short term studies where species do not reproduce or long term studies where species grow and reproduce). Many studies have "failed" to show that biodiversity is important for ecosystem processes (for example, McKie et al., 2008; Perkins et al., 2010). In general, the statistical analysis does not seem to address the mechanisms that could govern biodiversity effects or explain why no biodiversity effects can be observed.

The experiments we describe here were specifically designed to address how mixtures of animals "perform" with regard to how efficiently they consume food and generate fine particular matter (that can be used by other organisms as food) in a "short term" experiment. In addition, two of the experiments described addressed not only the role of species richness, but also that of size within species.

In Section 2 we describe three experiments in which the subsets of species were carefully and deliberately chosen. Section 3 gives the models that we fitted to the data. These included not only the 'biodiversity' model where the number of different species is considered to be a quantitative factor, but also models that would be more familiar to people running experiments on mixtures of different ingredients—see Cornell (2002). We found that Hasse diagrams helped the biologists to understand the relationships between the different models.

Section 4 briefly summarizes our conclusions from the data analysis, and suggests a new graphical method of summarizing the analysis-of-variance table. Finally, Section 5 considers some questions about how such experiments should be designed.

2. The experiments

The first two experiments are described by Reiss et al. (2011). Six types of freshwater organisms called invertebrate detritivore shredders were used. These types were three species, with two size classes within each species. For simplicity, they are referred to here as *A*, *B*, …, *F*. The experimental unit was a jar. Twelve organisms were put into each jar. The treatments were thus the combinations of types put into each jar.

Table 1 shows the treatments used in the first experiment. There were six treatments called *monocultures* where all 12 organisms in the jar were of the same type. There were 15 further treatments called *dicultures*: in these, there were six organisms of one type and six of another. Finally, there were 20 treatments called *tricultures* in which three different types of organism were used, four of each. Thus there were 41 treatments altogether.

There was also a 'control' treatment with no organisms. It was expected that it would perform very differently from the other 41 treatments, and this is indeed what happened. Thus all the models discussed in Section 3 need to be augmented with an extra parameter for the control: compare Bailey (2008, Sections 3.2 and 5.9). For simplicity, we ignore the control treatment here, as it makes no difference to our conclusions from the three experiments described or to our discussion of the planning, description and analysis of such experiments.

The experiment was carried out in four blocks of 41 jars. The blocks were transparent containers, each holding a two-dimensional array of jars. All jars in each block (total area of about one square metre) shared the same light source and were submerged in the same stream water (a mesh net at the top of the jars allowed water to flow in but kept the animals inside the jar). Carefully measured amounts of alder leaf litter were put into each jar. Then one treatment was added to each jar, in such a way that each treatment occurred in exactly one jar in each block. The jars were left for 28 days: then the amount of leaf litter eaten was measured. This was converted into a daily leaf decomposition rate. A secondary measure was the quantity of fine particulate organic matter (FPOM) in the jar after 28 days.

It is important to note that each measured response had a single value for each jar. As in many biodiversity experiments on fauna, it was impossible to subdivide the response and attribute parts to different types of organism. This contrasts with experiments on biodiversity of flora, where it is often possible to measure the yield of each species in the pot or plot.

The first experiment was designed to find out not only if the responses were affected by the number of types of organism present but also if it mattered which combinations of types were present. The second experiment took this a little further by replacing the tricultures by so-called *uneven dicultures*, in which there were eight organisms of one type and four of the other. Thus there were the 51 treatments shown in Table 2. This experiment was run in three blocks of 51 jars. The same two responses were measured as in the first experiment.

The third experiment is described by Reiss et al. (2010). It concerned a slightly different ecological setting, and different organisms, but was otherwise similar to the first experiment, in that the treatments were monocultures, dicultures and tricultures. However, this time the jars in each block did not share the same water as they were not submerged in a water

Table 1
The 41 treatments in the first experiment.

Number	Treatment	Name	Example	Richness Level
6	*A*, …, *F*	Monoculture	12 of type *A*	1
15	*AB*, …, *EF*	Diculture	6 of *A*, 6 of *B*	2
20	*ABC*, …, *DEF*	Triculture	4 of *A*, 4 of *B*, 4 of *C*	3