

Available online at www.sciencedirect.com



Fisheries Research 70 (2004) 397-407



www.elsevier.com/locate/fishres

# Remedies for pseudoreplication

Russell B. Millar\*, Marti J. Anderson

Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand

# Abstract

Pseudoreplication is the failure of a statistical analysis to properly incorporate the true structure of randomness present in the data. It has been well documented and studied in the ecological literature but has received little attention in the fisheries literature. Avoiding pseudoreplication in analyses of fisheries data can be difficult due to the complexity of the statistical procedures required. However, recent developments in statistical methodology are decreasing the extent to which pseudoreplication has to be tolerated. Seven examples are given here, beginning with simple design-based remedies and progressing to more challenging examples including the model-based remedies of mixed-effects modelling, generalized linear mixed models, state-space models, and geostatistics.

© 2004 Elsevier B.V. All rights reserved.

*Keywords:* Generalized linear mixed models; Geostatistics; Maximum likelihood; Mixed-effects; Nonparametric MANOVA; Overdispersion; Pseudoreplication; Random effects; Sequential population analysis; State-space models

# 1. Introduction

Pseudoreplication is the incorrect modelling of randomness, and is a notoriously rampant affliction in ecological field experiments (Hurlbert, 1984; Heffner et al., 1996). However, it is not widely recognized that pseudoreplication is also a widespread affliction in the analysis of fisheries data. This may be because the terminology of manipulative field experiments is not always terribly meaningful in the fisheries context, and because the complex and nonlinear structure of many fisheries models requires a different kind of remedy for pseudoreplication.

\* Corresponding author. Tel.: +64 9 373 7599; fax: +64 9 373 7000.

E-mail address: r.millar@auckland.ac.nz (R.B. Millar).

The seven examples herein demonstrate how pseudoreplication can be eliminated (or at least ameliorated) by appropriate consideration of the structure of the randomness inherent in the data. The common feature of these examples is lack of independence in the data, in the sense that the observations cannot be truly considered a simple random sample from the population of interest. The first two are simple design-based remedies of a generic nature and pseudoreplication is avoided by basing inference on groupings of the data that can be considered a simple random sample. This simple approach is insufficient to handle the more complex structure of the data presented in the remaining five examples. In particular, Examples 3-6 employ the concept of random effects as a remedy for pseudoreplication, albeit in very different ways.

 $<sup>0165\</sup>text{-}7836/\$$  – see front matter 2004 Elsevier B.V. All rights reserved. doi:10.1016/j.fishres.2004.08.016

A mixed-effects model is one that includes both fixed and random effects. The so-called fixed effects could correspond to the effect of treatments, but more generally are parameters associated with the system under study (e.g. average density, growth rate, instantaneous rate of natural mortality). Mixed-effects models use the concept of random effects to emulate the randomness inherent in the data. State-space models also use random effects, but in the context of time series data. The complex details of fitting mixed-effects and state-space models are not covered here, and can be found in the referenced literature.

Mixed-effects and state-space models have been in use for several decades, but until recently have been largely restricted to the context of normal linear models due to the computational difficulty of applying them more generally. Thus, these models have not previously been of great practical relevance to the field of quantitative fisheries modelling, where it has been much easier to turn a blind eye to pseudoreplication or at best to use an ad-hoc fix. However, since the early 1990s, the understanding of such models and the software for fitting them has improved to the point where they can now be applied to many fisheries models. This is especially true of fisheries models that are fitted under the Bayesian paradigm (Punt and Hilborn, 1997) because the Bayesian software uses Monte Carlo computer methods that do not require simplifying assumptions for tractability (e.g. Meyer and Millar, 1999). Indeed, there are many historical data-sets in the statistical literature that have received comprehensive mixed-effects analysis within the Bayesian paradigm, but have yet proved too difficult for the equivalent classical mixedeffects analysis (Meyer and Millar, 1999; Millar, 2004).

Hurlbert (1984) defined pseudoreplication as 'the use of inferential statistics to test for treatment effects with data from experiments where either treatments are not replicated (though samples may be) or experimental units are not statistically independent'. Hurlbert identified three common forms of pseudoreplication, simple, sacrificial, and a third which he called temporal but here will be called simple-temporal to distinguish it from more general forms of temporal pseudoreplication. Simple pseudoreplication occurs when an analysis fails to acknowledge that multiple observations have been taken on a single replicate of a treatment. Similarly, simple-temporal pseudoreplication is the failure to acknowledge the sequential measurement of multiple observations on the same treatment replicate. Many of Hurlbert's examples of simple pseudoreplication arise due to observations not being independent in space (due to clustering, say), and this will be denoted here as simple-spatial pseudoreplication. Simple pseudoreplication, in its various forms, is extremely common and its consequence is effectively an inappropriate inflation of the 'effective sample size', and this can lead to underestimates of standard errors and spurious statistical significance (i.e. an inflated Type I error rate). Sacrificial pseudoreplication works in the opposite direction—it sacrifices statistical power (i.e. inflates the Type II error rate) because it fails to recognize pairing or grouping structure in the data.

Hurlbert's definitions of pseudoreplication are somewhat limited in the context of fisheries models. Many fisheries experiments are mensurative (i.e. nonmanipulative) and complex, and cannot be expressed using the vernacular of experimental design (i.e. using terms such as 'treatment' and 'experimental unit'). For this reason, the terms temporal pseudoreplication and spatial pseudoreplication will be used here in a general sense, without assuming them to be special cases of simple pseudoreplication. Here, the working definition of pseudoreplication will be the use of inferential statistics when, at some level of the design or model, independence is incorrectly assumed.

In the first two examples, the experimental design is relatively straightforward and pseudoreplication can be avoided simply by identifying the independent experimental units and applying design-based inference solely to these. In particular, in the second example (two-stage sampling), spatial pseudoreplication is remedied by considering transects as the basic experimental unit rather than the evenly-spaced quadrats on the transects. The third example presents a design-based random-effects remedy to spatial pseudoreplication when the data are collected at different spatial scales. In this example, the magnitude of variability in multivariate community assemblage data at different spatial scales was of direct interest. Example 4 looks at sequential population analysis (SPA). It includes a temporal random effect and is an example of a state-space model. Examples 5 and 6 use generalized linear mixed models, and the final example briefly motivates geostatistics as a remedy to avoid pseudoreplication in spatial data.

Example 6 is covered in greatest depth and includes analyses of data from a size selectivity experiment. It demonstrates that simple pseudoreplication may not necessarily be detected by naive goodness of fit tests, and that incorrect inference would result.

The sequential population analysis in Example 4 is a particularly potent reminder of the unfortunate consequences of pseudoreplication. Indeed, the following excerpt from Myers and Cadigan (1995a) shows that pseudoreplication compromised the ability of SPA to provide sound inference about the abundance of cod on the Grand Bank of Newfoundland. This was once the largest cod fishery in the world, but collapsed in the early 1990s.

An analysis of traditional catch-at-age data in conjunction with research surveys, which assumed that research survey estimation errors of abundance by age and year were independent, led assessment biologists to the conclusion that the collapse was caused by an increase in natural mortality in the first half of 1991. We constructed a statistical model to test this hypothesis. The results do not support the hypothesis .... We also demonstrate that the usual assumption that estimation errors from research trawl surveys are independent is not valid, and can lead to invalid inference and unreasonable estimates of abundance.

# 2. Examples

The first example is generic and is not based on any particular application. The remaining six examples are from the fisheries related literature and their primary purpose is to demonstrate that fisheries designs and models commonly make questionable assumptions of independence, and to show how the analysis can be implemented to eliminate or ameliorate pseudoreplication.

The seven examples are:

- 1. Paired data (sacrificial).
- 2. Two-stage intertidal sampling of shellfish (spatial).
- Community assemblage analysis at different spatial scales (simple-spatial).
- 4. Sequential population analysis (simple, temporal).
- 5. Comparison of fish density inside and outside a marine reserve (simple).

- 6. Size selectivity analysis (simple).
- 7. Abundance surveys (spatial).

#### 2.1. Example 1: paired t-test

One of the standard traps for undergraduate students is the mistake of analyzing paired data as though they were unpaired. In the case of normally distributed data, this would be the use of a two-sample *t*-test instead of a paired *t*-test. The two-sample *t*-test is not valid because the pair of observations made on an experimental unit (i.e. subject) are not independent. Moreover, the two-sample *t*-test sacrifices statistical power because differences between the experimental units will inflate the estimates of the standard error within each treatment group and thus may obscure between-treatment effects. Indeed, in the terminology of Hurlbert (1984), this is an example of sacrificial pseudoreplication.

The paired *t*-test assumes that the within-pair differences in measurements are a simple random sample and applies a one-sample *t*-test to these differences. This design-based remedy is particularly elegant, because it avoids the need to model the actual data (the pairs of measurements from each experimental unit) and therefore eliminates any need to explicitly model between-subject variability. The model is simply that the differences are an i.i.d. sample from a normal distribution.

# 2.2. Example 2: two-stage sampling of shellfish

Morrison et al. (1999) implemented a two-stage sampling scheme for intertidal shellfish and applied it to several beaches in the Auckland region of New Zealand. At each beach, the first stage was the random allocation of transects and the second stage was the digging of four  $0.1 \text{ m}^2$  quadrats at 10 m intervals along each transect. (The first stage of sampling was stratified according to prior knowledge of shellfish density, but this will not be considered here.)

The quadrats are clearly not a simple random sample from the beach, because the inclusion of a quadrat in the sample guarantees that there will be at least one other quadrat within a 10 m radius. Morrison et al. (1999) appropriately noted that the quadrats are pseudoreplicates and that the transects are the basic experimental unit. Their inference was based on the average density within the transect. This also mitigates criticisms about the non-randomness of the systematic sampling of quadrats within each transect because, for the purpose of inference, the 'observations' consist of a simple random sample of measured transect densities.

#### 2.3. Example 3: underwater visual counts of fish

Anderson and Millar (2004) used a multilevel experimental design to survey fish abundance on inshore reefs off the north-eastern coast of New Zealand. The primary objectives were to assess differences in species assemblages between urchin-grazed barrens and kelp habitats at different spatial scales, and thus the habitats can be considered treatments. Four locations, separated by hundreds of kilometers, were sampled (Fig. 1). Within each location, fish were counted at each of four different randomly located sites, separated by hundreds to thousands of metres. At each site, the two habitats were sampled: kelp forests (i.e. areas characterised by relatively dense cover of the kelp Ecklonia radiata) and 'barrens' (i.e. areas characterised by little or no macro-algal cover and dominated by the grazing urchin Evechinus chloroticus). Within each habitat, divers on SCUBA did a visual survey by swimming ten (haphazardly chosen) 25 m transects and identifying and recording the number of each species of fish observed within a distance of 2.5 m on either side of the transect.

The sampled transects are clearly not a simple random sample of all possible 25 m transects that could have been swum over kelp or barren habitat off the NE coast of New Zealand. Instead, sampling takes place at three levels (i) firstly, a sample of locations, (ii) within each location a sample of sites, and (iii) within each site a sample of transects (stratified by habitat type). Quantification of the relative magnitude of the spatial variability inherent in these three levels of sampling, and any interaction between habitat effect and spatial scale, were both fundamental research objectives of Anderson and Millar (2004). Hence the analysis needed to formally incorporate the three levels of spatial variability, and the effects of habitat and interactions. An analysis that ignored the design (say, by simply combining the species assemblage data across all transects for each habitat) would not only be guilty of simple pseudoreplication but would also be failing to utilize information the data contain about spatial variability in assemblages.

For approximately normally distributed data, the above design can be formally analyzed using a multiway mixed-effects ANOVA with location, site, and transect as appropriately specified random-effects and habitat as a fixed effect, and with appropriate interaction terms. However, the primary focus of Anderson and Millar (2004) was to model the multivariate assemblage data recorded from each transect. A mixedeffects ANOVA-type model was deployed, but using design-based permutation to partition the variability in a dissimilarity matrix, rather than using a model based on assumptions of multivariate normality (Anderson, 2001a, 2001b). They found that variability in species assemblages was highest at the scale of individual transects, and that variability from site to site and from location to location was comparable. Moreover, the effect of habitat was not uniform and it varied between locations and between sites within locations.

# 2.4. Example 4: sequential population analysis

If  $N_{a,y}$  is the number of fish of age *a* at the start of year *y* and  $C_{a,y}$  is the catch (assumed known) taken during year *y*, then a typical sequential population dynamics model is of the form

$$N_{a_0,y} = R_y \tag{1}$$

$$N_{a,y} = N_{a-1,y-1} e^{-m} - C_{a-1,y-1} e^{-m/2}, \quad a > a_0$$
(2)

where *m* is the natural mortality,  $a_0$  is the age of recruitment, and  $R_y$  is the recruitment in year *y*. Eq. (2) is derived from assuming that fishing occurs in the middle of the year.

The data are a rectangular table of relative numbersat-age,  $I_{a,y}$ . For example, Table 1 of Myers and Cadigan (1995a) lists the average catch-per-surveytow of Grand Banks cod for age-classes a = 3, ..., 12and survey years y = 1978, ..., 1993. For simplicity, it will be assumed here that the survey takes place at the start of the year (this is not a crucial assumption, see Myers and Cadigan (1995a)). Then, the data are typically modelled as lognormally distributed according to the equation

$$I_{a,y} = QN_{a,y} e^{\epsilon_{a,y}}$$
(3)



Fig. 1. Map of northeastern New Zealand, showing the locations and sites sampled by Anderson and Millar (2004). [Reprinted with permission from Elsevier.]

where parameter Q denotes the 'relative catchability' of the survey gear, and  $\epsilon_{a,y}$  are assumed to be i.i.d. normal.

The sequential population analysis is modelling a table of relative numbers-at-age, with age and year indexing the rows and columns of the table. In this sense, age and year can be considered pre-assigned row and column treatments, with the effects of these treatments modelled via Eqs. (1) and (2). Equivalently, it may be more natural to consider cohort and year as the treatments, because each additional year adds another level to both the year 'treatment' and cohort 'treatment'.

It can therefore be argued that a SPA suffers simple pseudoreplication because, for year (i.e. treatment) y, it is the same set of survey tows that is used to calculate the average catch-per-survey-tow for each age-class, ages 3–12 in the case of the Grand Banks cod. Consequently, the data,  $I_{a,y}$ , within each year will be positively correlated. Myers and Cadigan (1995b) confronted this pseudoreplication by explicitly incorporating a random year effect into Eq. (3)

$$I_{a,y} = QN_{a,y} e^{\epsilon_{a,y} + \epsilon_y} \tag{4}$$

and they provided a computational algorithm for obtaining maximum likelihood estimates of the model parameters.

Temporal pseudoreplication arises because it is the same cohort (i.e. treatment) that is sampled over a number of years. Eq. (2) is a structural formula describing the expected decline in size of a cohort due to catch and natural mortality, but process variability will cause the actual cohort size to deviate from that expected. Moreover, the effect of any deviation in (2) necessarily impacts on the size of the cohort in subsequent years, which in turn will cause temporal correlation in the data  $I_{a,y}$ .

The temporal correlation can be modelled by including random error terms in the population dynamics equation (2). This form of model is called a state-space model. Maximum likelihood fits of state-space models are challenging, but if sufficient simplifying assumptions are made then they can be fitted using the Kalman filter (e.g. Sullivan, 1992; Freeman and Kirkwood, 1995; Kimura et al., 1996). Bayesian stock assessment modelling does not require the simplifying assumptions, notwithstanding the difficulty of obtaining a reliable posterior sample if the model is overly complex. For example, Meyer and Millar (1999) describe methodology for fitting nonlinear and nonnormal state-space surplus production models, while Millar and Meyer (2000a) provide more detail of the underlying theory. Millar and Meyer (2000b) fit a Bayesian state-space sequential population analysis to the cod data of Myers and Cadigan (1995a) and show how to incorporate the random year effect in Eq. (4).

# 2.5. Example 5: comparison of fish density inside and outside a marine reserve

Willis and Babcock (1997) used the catch-per-uniteffort (CPUE) of volunteer recreational fishers as a measure of the relative density of snapper (*Pagrus auratus*) within and adjacent to the Leigh Marine Reserve on the northeast coast of New Zealand. The objective of this study was to estimate the ratio of snapper density between the reserve and adjacent non-reserve areas (Millar and Willis, 1999).

Fishing from small boats was conducted on the 4 days of 15 June, 29 June, 7 and 15 December in 1996. The region of interest was partitioned into ten areas, with areas 1 and 2 being outside of the reserve in the

northwest direction, areas 3–8 being inside the reserve, and areas 9 and 10 being outside of the reserve in the southeast direction. Each boat present was assigned to fish in a specified area in the morning, and assigned to a different area in the afternoon. The skippers were told to choose sites 'haphazardly' within the assigned area, and to fish from that location for 30 min. This permitted a maximum of six sites within an area to be fished by a boat in any given morning or afternoon session.

A total of 22 boats were involved, but the actual boats participating on each of the 4 days was never exactly the same. Only one boat participated throughout the entire study and 15 boats fished on only 1 day. Boats fishing on a single day therefore fished in only two areas, with one exception where the skipper inadvertently crossed an area boundary and fished a third area.

The CPUE data cannot be considered a simple random sample. Rather, it could be said that a 'random' selection of recreational boats was made, and used on four 'randomly' chosen days. Had the data been balanced with respect to allocation of boats over treatments (reserve and non-reserve) and dates then it may have been possible to devise an elegant reprieve from pseudoreplication (e.g. Example 1), but the data are extremely unbalanced here. Thus, Millar and Willis (1999) used a generalized linear mixed model (GLMM) to model the CPUE data as (overdispersed) Poisson with the four dates and 22 boats treated as random effects. That is, the four dates were considered a random sample of possible dates and the boats were considered a random sample of all recreational fishing boats.

Maximum likelihood fits of GLMM's are extremely difficult because calculation of the likelihood function typically requires high dimensional numerical integration (McCulloch and Searle, 2001). For this reason, Millar and Willis (1999) used the SAS macro GLIM-MIX (Littell et al., 1996) to fit the GLMM using a form of pseudolikelihood (Wolfinger and O'Connell, 1993). Software is now becoming available that can do true maximum likelihood fits of simpler forms of GLMM's (see Example 6 below).

#### 2.6. Example 6: size selectivity data

Experiments to estimate the size-selectivity of a fishing gear typically involve multiple deployments of two or more variants of the gear. In the case of trawls or

Table 1

traps this might be an experimental gear and a cover or control gear. For gillnets, several nets of differing mesh size are typically fished simultaneously. The data consist of the lengths of all fish caught (or a subsample if the number caught is large) in each of the gear variants from each deployment.

Analysis of size-selectivity data generally proceeds by fitting some form of selection curve using maximum likelihood. For example, the symmetric logistic curve has become the default standard for trawl selectivity, notwithstanding the need to verify goodness of fit. The asymmetric Richards curve is typically employed if the logistic indicates lack of fit. The likelihood function for the curves is derived using the assumption that the length frequency data are Poisson distributed, which is implicitly assuming independence of fish. The SELECT methodology of Millar (1992) showed that maximum likelihood fits can most easily be achieved by (loosely speaking) modelling the difference in catch between the gear variants, rather than needing to model the catches themselves.

Selectivity studies typically suffer from simple pseudoreplication on at least two levels. The first arises because the treatment (deployment of the gear) is not replicated at the level of individual fish. The Poisson assumption inherent in the likelihood is implicitly saying that the fate of each fish (i.e. capture or escape) is independent of every other fish. One possible violation of this independence assumption arises from the schooling behaviour of fish. This form of pseudoreplication is very common in analyses of count data, and results in extra-Poisson variation. Fortunately, the magnitude of the extra-Poisson variation can be estimated from goodness of fit diagnostics, and an overdispersion correction applied (e.g. Millar and Fryer, 1999).

In the majority of selectivity analyses, the length frequency data from all deployments of the gears are summed and these total length frequencies are then modelled. This is the cause of the second level of simple pseudoreplication. It has been observed in many studies that selectivity can vary immensely from deployment to deployment of fishing gear, despite all possible attempts to replicate experimental conditions. This form of pseudoreplication cannot be adequately ameliorated by applying the standard overdispersion correction. For example, Table 1 shows carapace length frequency data of school prawn (*Metapenaeus macleayi*). These data were summed over 19 replicate deployments of an ex-

School prawn catches su	ummed over 19	replicate de	eployments	of a
covered codend				

Carapace length (mm)	Codend catch	Cover catch
6	0	3
7	0	0
8	0	2
9	0	7
10	2	10
11	6	48
12	10	56
13	31	95
14	74	175
15	313	478
16	904	970
17	1227	824
18	1090	450
19	567	136
20	186	19
21	48	0
22	15	0
23	2	0

perimental square-mesh stow net that was deployed with a fine mesh cover to capture all prawns that escaped through the codend (Macbeth et al., in press). The resulting retention proportions were well fitted by an asymmetric Richards selection curve (Fig. 2, Table 2) and, with just two exceptions, the deviance residuals were less than unity in magnitude. No symptoms of pseudoreplication are evident. The fit gives an estimated carapace length of 50% retention,  $l_{50}$ , of 16.2 mm with a standard deviation of just 0.06 mm.

A very different story becomes apparent if the catches from the individual deployments are examined. The fitted selection curve and estimated standard deviations are necessarily the same as those obtained from the fit to the summed data, however it is now clear that massive extra-Poisson variation is present (Fig. 2, Table 2). Indeed, the deviance residuals exceeded 5

Table 2

Goodness of fit statistics for the Richards selection curve fitted by maximum likelihood to the school prawn data

	Summed data	Individual hauls data
Pearson $\chi^2$	3.0	995.2
d.o.f.	7	99
Overdispersion	n.s.	10.1

Only length classes with expected catch exceeding 3 were included. n.s. denotes not significantly different from unity.



Fig. 2. Selectivity curve fitted by maximum likelihood to the observed retention proportions, and the resulting residual plot. Left column: fit to the retention proportions obtained from summing the catch data over the 19 replicate deployments. Right column: fit to the retention proportions observed in each deployment. Note that the fitted selectivity curve is the same in both cases.

in magnitude on 13 occasions and the overdispersion factor, calculated as the ratio of the goodness-of-fit  $\chi^2$  statistic divided by degrees of freedom, is estimated to be 10.1. Millar et al. (2004) used the squareroot of this overdispersion factor as an ad-hoc correction to the estimated standard deviations. Thus, the estimated standard deviation of  $l_{50}$  was increased by a multiplicative factor of 3.2. That is, from 0.06 to 0.2 mm.

The form of the between-deployment variability can be inspected by performing individual selectivity fits to the catch data from each deployment. Fig. 3 shows the estimated  $l_{50}$ 's and selection ranges (SR, the difference between the lengths of 25% and 75% retention probability) from individual selection curve fits to the 19 replicate deployments. If the treatment effect (i.e. selectivity) was the same in each deployment then, because they each contain the true unknown selectivity parameter with a priori probability 0.95, the 95% confidence intervals should overlap (with perhaps one or two exceptions). This appears plausible for SR, but is clearly not the case for the  $l_{50}$  parameter. Mixed-effects modelling of between-haul variation is addressed by Fryer (1991) and Millar et al. (2004). Both of these mixed-effects approaches allow selectivity parameters (e.g. length of 50% retention) to vary randomly from haul to haul, but they differ in their implementation. Fryer (1991) applied linear normal mixed models to the estimated selectivity parameters obtained from individual fits from each deployment. In contrast, Millar et al. (2004) uses the SAS procedure NLMIXED (SAS Institute Inc., 1999) to fit a GLMM directly to all the individual haul data.

# 2.7. Example 7: biomass surveys

Some biomass surveys use randomized sampling locations, and are often stratified by a relevant environmental variable, such as depth if a trawl survey (e.g. Doubleday, 1981). This design-based approach ensures that the catches at each sampling location are independent because the sampling locations are random (Brus and de Gruitjer, 1997). Thus, the sample mean (by strata



Fig. 3. Estimated selection curve parameters, and 95% confidence intervals, from 19 replicate deployments of a stow net in a school prawn fishery.

if applicable) of the observed densities provides an unbiased estimator of the biomass density, and the sample variance (by strata if applicable) is an appropriate estimator of the true sampling variability of this mean. Several model-based extensions to this approach exist and these typically involve using latitude, longitude and other relevant environmental information to introduce a structural model for the expected density over the survey region (e.g. Smith, 1990; Evans et al., 2000) for the purpose of obtaining a more statistically efficient estimate of biomass.

Random surveys can be criticized as being inefficient because they do not guarantee 'even' coverage over the survey region. Systematic biomass surveys offer an alternative whereby sampling locations are (typically) arranged on regularly spaced and pre-determined grid points. This type of design can also be advantageous in terms of implementation. However, the observations at each sampling location are now no longer a random sample from the population of all possible sampling locations, and hence are not independent. Lack of randomness also applies to survey data that are collected over a continuous track, as is often the case with acoustic data. Analyzing such data using design-based inference would be pseudoreplication.

Systematic biomass surveys can be analyzed in many ad-hoc ways, a number of which appear to work quite well (Wolter, 1984). The more formal approach is to use model-based analysis. This form of spatial analysis was primarily developed within the field of geological science, and hence is routinely known as geostatistics. The central notion of geostatistics is that of an underlying spatial supermodel, and to regard the actual population being sampled as a single realization from that spatial model. Kriging (Cressie, 1991) is possibly the most widely used geostatistical method in fisheries. More recently, the transitive method (Petitgas, 1993; Bez, 2002) has been demonstrated as an alternative that is applicable to samples taken on regularly spaced grids.

# 3. Discussion

Examples 3–6 introduced mixed-effects models (including state-space models) for capturing the structure of randomness in a wide variety of fisheries data, and Example 7 briefly mentioned techniques for coping with non-random biomass surveys. These examples are typical of quantitative fisheries analysis, and share the property that the data are not an independent sample from the population of interest. A naive analysis of such data would therefore commit pseudoreplication.

Fitting mixed-effects models can be challenging in all but the simple case of linear models and normal errors (McCulloch and Searle, 2001). However, software is increasingly becoming available for more complex mixed-effects models. This includes the SAS procedure NLMIXED (SAS Institute Inc., 1999) and the SAS macros GLIMMIX and NLIN-MIX (Littell et al., 1996). The NLMIXED procedure is preferred because it fits the model using a numerical approximation to the true likelihood, whereas GLIMMIX and NLINMIX use a modified likelihood method that can have poor properties in some situations (Millar and Willis, 1999). However, NLMIXED is restricted to models with just one random effect. The freely available R language (Anonymous, 2003; http://www.r-project.org/) has user-provided software packages that include functions glmmPQL, nlme, and GLMM. Functions glmmPQL and nlme are similar in implementation to GLIMMIX and NLINMIX, respectively, and function GLMM is similar in implementation to NLMIXED. The freely available WinBUGS software (Spiegelhalter et al., 1999; http://www.mrcbsu.cam.ac.uk/bugs/) can be used to fit a wide variety of Bayesian mixed-effects models, and many such examples are included with its documentation.

One form of simple pseudoreplication that mixedeffects models cannot remedy is extrapolation of inference to a population greater than the one sampled. For example, Millar and Willis (1999) estimated the density of snapper inside the Leigh Marine Reserve to be 11 times that in areas adjacent to the reserve (Example 5). This ratio cannot be used to infer an effect of marine reservation, because the experiment was not replicated, as only the Leigh Marine Reserve was used, nor were any similar experiments conducted prior to establishment of the Leigh Marine Reserve. Similarly, sizeselectivity studies are typically conducted with a single gear from a single boat, in a limited area, over a short period of time. The deployments are therefore a long way from being representative of the (hypothetical) population of all possible deployments of the gear in the entire fishery. Therefore, if the estimated selection curve is used in stock assessments or discard studies of the fishery then pseudoreplication has been committed.

The reader who has made it thus far may be wondering whether pseudoreplication is a sufficiently great sin as to warrant the extra effort of attempting the somewhat complicated remedies herein. This will depend on the particular application and whether there are sufficient data to permit a successful fit. For example, our experience with size-selectivity data suggests that mixed-effects models can usually be used provided that sufficient fish are caught in each deployment. The catch is usually sufficiently large in trawl experiments, but seldom so in gillnet or hook experiments. In such cases, ad-hoc remedies may exist, including those demonstrated in Example 6, and the variance formulae of Wolter (1984) as a simpler alternative to geostatistics.

The properties observed in Example 6 are typical of simple pseudoreplication. Pooling of the catch data (Table 1) explicitly ignores the possibility of variability in replicate deployments of the treatment (i.e. trawl codend). The total number of fish measured may number in the thousands, and this results in implausibly small estimates of standard error and possible spurious statistical significance if two or more gear variants are being compared. In these experiments, the total number of deployments typically numbers in the tens, and the replication at this level needs to be considered for valid statistical inference.

Clearly, careful thought is needed in the analysis of fisheries data to avoid pseudoreplication when lack of independence occurs at some level of the design or model. We have demonstrated a variety of ways that pseudoreplication can arise and suggested methodology for remedying it. This work is just the tip of an iceberg. Our hope is that it will have made the reader aware of pseudoreplication in analyses of fisheries data, and prompted her to keep a wary eye out for it. Pseudoreplication appears in far more guises than demonstrated here.

# References

- Anderson, M.J., 2001a. A new method for non-parametric multivariate analysis of variance. Aust. Ecol. 26, 32–46.
- Anderson, M.J., 2001b. Permutation tests for univariate or multivariate analysis of variance and regression. Can. J. Fish. Aquat. Sci. 58, 626–639.
- Anderson, M.J., Millar, R.B., 2004. Spatial variation and effects of habitat on temperate reef fish assemblages in north eastern New Zealand. J. Exp. Mar. Biol. Ecol. 305, 191–221.
- Anonymous, 2003. R: A Language And Environment For Statistical Computing. R Development Core Team, Vienna.
- Bez, N., 2002. Global fish abundance estimation from regular sampling: the geo-statistical transitive method. Can. J. Fish. Aquat. Sci. 59, 1921–1931.
- Brus, D.J., de Gruitjer, J.J., 1997. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil. Geoderma 80, 1–44 (with discussion).
- Cressie, N.A.C., 1991. Statistics for Spatial Data. Wiley, New York. Doubleday, W.G. (Ed.), 1981. Manual on Groundfish Surveys in the
- Northwest Atlantic, NAFO Sci. Coun. Studies, vol. 2. NAFO. Evans, G.T., Parsons, D.G., Veitch, P.J., Orr, D.C., 2000. A localinfluence method of estimating biomass from trawl surveys, with Monte Carlo confidence intervals. J. Northw. Atl. Fish. Sci. 27,
- 133–138.
  Freeman, S.N., Kirkwood, G.P., 1995. On a structural time series method for estimating stock biomass and recruitment from catch and effort data. Fish. Res. 22, 77–98.
- Fryer, R.J., 1991. A model of between-haul variation in selectivity. ICES J. Mar. Sci. 48, 281–290.
- Heffner, R.A., Butler, M.J., Reilly, C.K., 1996. Pseudoreplication revisited. Ecology 77, 2558–2562.
- Hurlbert, S.H., 1984. Pseudoreplication and the design of ecological field experiments. Ecol. Monogr. 54, 187–211.

- Kimura, D.K., Balsiger, J.W., Ito, D.H., 1996. Kalman filtering the delay difference equation: practical approaches and simulations. Fish. Bull. US 94, 678–691.
- Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., 1996. SAS System for Mixed Models. SAS Institute Inc., North Carolina.
- Macbeth, W.G., Broadhurst, M.K., Millar, R.B., in press. Coveredcodend experiments to assess selectivity in an Australian penaeid prawn stow net fishery. Bull. Mar. Sci.
- McCulloch, C.E., Searle, S.R., 2001. Generalized, Linear, and Mixed Models. Wiley, New York.
- Meyer, R., Millar, R.B., 1999. BUGS in Bayesian stock assessment. Can. J. Fish. Aquat. Sci. 56, 1078–1087.
- Millar, R.B., 1992. Estimating the size-selectivity of fishing gear by conditioning on the total catch. J. Am. Statist. Assoc. 87, 962–968.
- Millar, R.B., 2004. Simulated maximum likelihood in practice. Aust. NZ J. Statist.
- Millar, R.B., Broadhurst, M.K., MacBeth, W.G., 2004. Modelling between-haul variability in the size selectivity of trawls. Fish. Res. 67, 171–181.
- Millar, R.B., Fryer, R.J., 1999. Estimating size-selection curves of trawls, traps, gillnets, and hooks. Rev. Fish Biol. Fish. 9, 89–116.
- Millar, R.B., Meyer, R., 2000a. Non-linear state space modelling of fisheries biomass dynamics by using Metropolis-Hastings within-Gibbs sampling. Appl. Statist. 49, 327–342.
- Millar, R.B., Meyer, R., 2000b. Bayesian state-space modeling of age-structured data: fitting a model is just the beginning. Can. J. Fish. Aquat. Sci. 57, 43–50.
- Millar, R.B., Willis, T.J., 1999. Estimating the relative density of snapper in and around a marine reserve using a log-linear mixed effects model. Aust. NZ J. Statist. 41, 383–394.

- Morrison, M.A., Pawley, M.D.M., Browne, G.N., 1999. Intertidal surveys of shellfish populations in the Auckland region 1997–1998 and associated yield estimates. NZ Fish. Ass. Res. Doc. 99/25.
- Myers, R.A., Cadigan, N.G., 1995a. Was an increase in natural mortality responsible for the collapse of northern cod? Can. J. Fish. Aquat. Sci. 52, 1274–1285.
- Myers, R.A., Cadigan, N.G., 1995b. Statistical analysis of catchat-age data with correlated errors. Can. J. Fish. Aquat. Sci. 52, 1265–1273.
- Petitgas, P., 1993. Geostatistics for fish stock assessments: a review and an acoustic application. ICES J. Mar. Sci. 50, 285–298.
- Punt, A.E., Hilborn, R., 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. Rev. Fish Biol. Fish. 7, 35–63.
- SAS Institute Inc., 1999. SAS/STAT<sup>®</sup> User's Guide, Version 8. SAS Institute Inc., Cary, NC.
- Smith, S.J., 1990. Use of statistical models for the estimation of abundance from groundfish trawl survey data. Can. J. Fish. Aquat. Sci. 47, 894–903.
- Spiegelhalter, D.J., Thomas, A., Best, N.G., 1999. WinBUGS, Version 1.2, User Manual. MRC Biostatistics Unit, Cambridge.
- Sullivan, P.J., 1992. A Kalman filter approach to catch-at-length analysis. Biometrics 48, 237–257.
- Willis, T.J., Babcock, R.C., 1997. Investigation of Methods for Assessing Reef Fish Populations and the Effects of Marine Reserve Protection. Department of Conservation, New Zealand.
- Wolfinger, R., O'Connell, M., 1993. Generalized linear mixed models: a pseudo-likelihood approach. J. Statist. Comput. Simul. 48, 233–243.
- Wolter, K.M., 1984. An investigation of some estimators of variance for systematic sampling. J. Am. Statist. Assoc. 79, 781–790.