Monitoring stoat *Mustela erminea* control operations: power analysis and design

J.A. Brown

Biomathematics Research Centre
Department of Mathematics and Statistics
Private Bag 4800
University of Canterbury
Christchurch

C.J. Miller

Department of Conservation
Private Bag 701
Hokitika

No. 164 April, 1998

Monitoring stoat (*Mustela erminea*) control operations: power analysis and design.

J.A. Brown[1], C.J. Miller[2].

[1]Biomathematics Research Centre
Department of Mathematics and Statistics
Private Bag 4800
University of Canterbury.

[2]Department of Conservation
Private Bag 701
Hokitika

Abstract

The aim of this report is to look at design of tracking tunnel monitoring for stoat control operations. We use tracking tunnel data from a study in North Okarito forest to develop a stoat-monitoring design and to evaluate the success of different survey protocols. We discuss the importance of understanding statistical power. Power is a measure of the likelihood of reaching the correct conclusion about the success of a monitoring programme. Power is one of the crucial factors that should be considered in designing a monitoring programme. The relative costs of falsely concluding a control operation was successful when it was not, or of falsely concluding a control operation was not successful when it was, need to be carefully evaluated.

We recommend for monitoring stoats with tracking tunnels that tunnel-stations be spaced 1 km apart and multiple tunnels be used at each station. The actual survey design for monitoring will depend on the acceptable error rates, the desired level of power, the target percentage kill, and the pre-control stoat density. We have constructed a model to estimate the number of stations and inspections for combinations of type I error rates, power, target percentage kill and pre-control stoat density.

1. Introduction

1.1 Monitoring stoat control operations

Monitoring is an essential component of conservation management. The Department of Conservation conducts *operational* monitoring to assess whether its control targets (e.g. percentage kill of a pest animal) have been achieved, and *performance* monitoring to assess how well a management action has protected a specific resource (e.g. reduction in predation of eggs and chicks).

The success or failure of a monitoring programme can be measured by its ability to detect a biologically significant population change. This ability depends on the way the monitoring is designed (Green 1979; Norton 1996). Poorly designed monitoring programmes can increase the risk of making either of two potentially costly mistakes: concluding that an operation was successful in reducing pest numbers to a target level, when in fact it was not; or concluding the control operation was not successful, when in fact it was. The first mistake may result in the loss of the conservation resource that was thought to have been protected, and the second may result in a manager repeating a control operation when it was not necessary.

There are an increasing number of stoat control operations being carried out by the Department of Conservation. The monitoring component of these operations must be appropriately designed. This is particularly important when control operations are conducted over large areas to protect widely dispersed or wide ranging species, e.g. kiwi in North Okarito forest, or a suite of species as in a mainland island situation. King & Edgar (1977, and see also King 1994) recommended that tracking rates from ink-print tunnels were the best means to monitor stoats in New Zealand and Moors (1978) first used tracking tunnels at Kaikoura to monitor the seasonal and annual changes in a mustelid population. To date however there has not been a standardised protocol developed for the spatial layout of tracking tunnels to monitor stoat control operations.

1.2 Aim

The aim of this report is to look at design of tracking tunnel monitoring for stoat control operations. We used tracking tunnel data from a study in North Okarito forest (Miller & Elliott 1997) and data from a current study on stoat home ranges in podocarp forest (Miller *et al.*, unpubl. data) to develop a stoat-monitoring design and to evaluate the success of different survey protocols.

1.3 Two important concepts for environmental monitoring

1.3.1 Independent data

Ideally, monitoring should be designed so the data are independent to maximise the information gained from the field surveys. Independent data means that every tracking tunnel-station or every station inspection, is informative. If two tracking stations are placed close together and the same stoat tracks both stations there is little gained by way of new information from the second tracking station, because the presence of stoats has been detected in the first station. Similarly, if a repeat inspections of tracking tunnels is conducted soon after the first inspection, and in that time there is little change in the stoat population, most of the information on the tracking rates will have been collected during the first inspection.

In both these examples the survey effort would be better used in checking stations that were further apart. In the first example, the stations should be placed further apart on the ground (spatially), and in the second example, the station inspections should be further apart in time (temporally). With adequate spatial and temporal spacing, for a given population size, whether one tunnel-station is tracked or not has no influence on whether the neighbouring tunnel-station will be tracked. This is known as statistical independence.

1.3.1 Statistical power

Statistical power is a measure of the likelihood of a monitoring programme reaching the correct conclusion. If a stoat population has declined after a control operation a monitoring programme with sufficient power should be able to detect this change. Power is one of the crucial factors that should be considered in designing a monitoring programme (Taylor and Gerrodette, 1993, Fairweather 1991).

Consider the following example. If a stoat population has been reduced by 50% after a control operation, and the population were monitored pre- and post-control, the correct change in population size would be estimated if the tracking rates had reduced by 50%. However, if the tracking rates reduced by a smaller amount, or not at all, and the statistical test does not provide sufficient evidence to say with confidence that the population has been reduced the manager may decide that the control operation was not successful. The manager may then repeat the control operation unnecessarily, or, erroneously conclude that their control strategy was not effective and change strategies. Statistically this is known as committing a type II error, β (Peterman 1990, Steidl *et al.* 1997). A type II error is when the statistical test fails to detect a change.

With stoat control the cost of making this type of error, where the success of the control operation is not detected, can be considered "long term", that is, unnecessary resources may be expended, or feedback on the effectiveness of control strategies for refinement may be delayed. In the short term kiwi will still be protected because the control operation will be repeated, but at extra and unnecessary cost.

The alternative "error" scenario is that the stoat control operation was not effective, i.e. the population did not reduce but data from the monitoring meant the manager concluded there had been a reduction. This is quite plausible because, by chance, on one survey the tracking rates may be high, and by chance, on a post-control survey the rates may be low, even with no real change in the stoat population. If the tracking rates reduced by 50%, the manager would infer that the control operation was successful and fail to take necessary remedial action to protect kiwi from predation. Statistically this is known as committing a type I error, α, (Peterman 1990, Steidl *et al.* 1997). A type I error is when the statistical test is used to falsely detect a change. There is an obvious immediate and high cost in making this type I error because no further protection may be given to the kiwis even though stoat numbers remain high.

These type I and II errors can be estimated, and are usually defined as being a fraction, or a probability, e.g., 0.05 is commonly used as the type I error rate, and 0.2 are often considered acceptable as type II error rates. It is important to note that these two values are not "written in stone", and that the choice of acceptable probabilities, α and β, should be relevant to the situation.

Statistical power is estimated from the type II error rate. Consider the example above where a population has been reduced with a successful control operation. If the probability of making a type II error is 0.2 ($\beta = 0.2$), then the probability of making the correct decision is 0.8. This is calculated as $1- \beta$. In this example, having statistical power of 0.8 means there is an 80% chance that the statistical test will lead to the correct decision that the population has been reduced.

Monitoring programmes should be designed to achieve high power. There are three ways to achieve this. One way is to increase the survey effort, e.g. using more tracking stations. With greater effort the tracking rates are likely to be more reliable and are more likely to accurately reflect the true situation. Another way to improve statistical power is to have high tracking rates and larger reductions in tracking rates. A third way to improve power is to change the type I error rate, e.g. from 0.05 to 0.1, because the type I error rate is positively related to power. When the type I error is high, the type II error rate is low and power is high (Peterman 1990). Choosing which combination of these is acceptable becomes a management decision, based on factors such as resources, and the consequences of making the wrong decision.

2. Method

2.1 Data collection

Tracking tunnel data were collected during an experiment to test the effectiveness of poisoned eggs for controlling stoats (Miller & Elliott 1997). The study was conducted in North Okarito forest at two sites known as River and Loop. The sites were approximately 4 km apart. The same grid layout of 29 stations, with one tracking tunnel at each station was used at both sites (Fig. 1). Tunnels, either 500 m or 250 m apart were baited with a small piece of beef and inspected approximately weekly for stoat tracks. The tracking tunnels were used to monitor the relative abundance of stoats, and to estimate the percentage kill achieved. The difference between pre- and post-control tracking rates was used as an estimate of the percentage reduction in the stoat population.

Data were collected from 21 tunnel inspections at Loop site over a period of 5 months (22 August 1996 to 25 March 1997), and at River site, from 20 tunnel inspections over a similar period (21 August 1996 to 26 March 1997). Hens eggs poisoned with 1080 were placed in the River site after the 12th tunnel inspection, and in the Loop site after the 19th tunnel inspection.

2.2 Data analysis

Definition of tracking rate

We defined the tracking rate as the average proportion of tunnel-stations found tracked on an inspection. For example, if 20 stations are inspected five times and the number of stations found to be tracked were 2, 3, 2, 1, and 2, the tracking rate would be $\{(2 + 3 + 2 + 1 + 2)/5\}/20 = 0.1$. If, after the control operation, the number of stations tracked were 1, 0, 1, 2, and 1, the post-control tracking rate would be $\{(1 + 0 + 1 + 2 + 1)/5\}/20 = 0.05$. The estimate of the kill rate was calculated from the difference between the pre- and post-control tracking data: $(0.1 - 0.05)/0.1 = 0.5$, or 50% control.

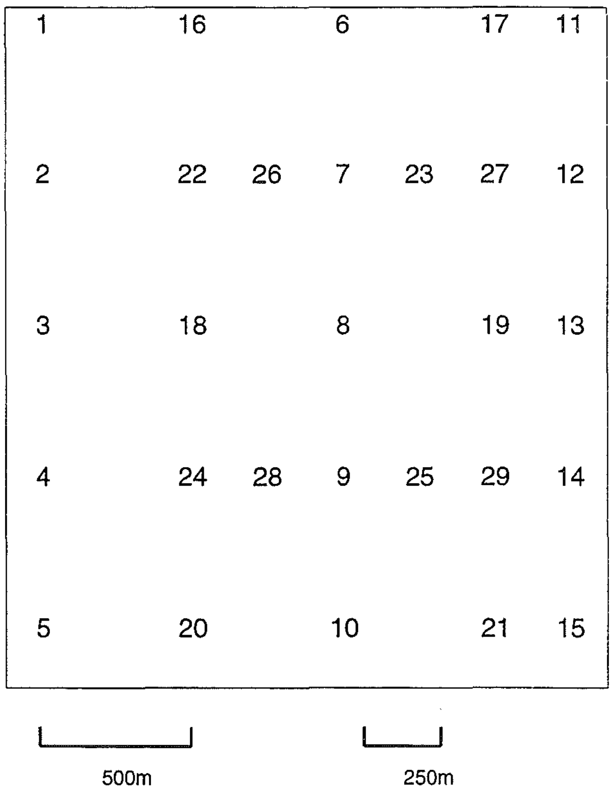


Fig. 1. Tunnel-station layout, North Okarito forest. There were 29 tunnel-stations in total, spaced either 500 m or 250 m apart.

Independence of tunnel-station counts

We calculated the distances between all pairs of stations and the correlations between pre-control station counts to see if there was any relationship between the spatial distance between stations and the differences in station counts. Station pairs were divided into 500m intervals of inter-station distances. The station count was the number of times the station was found tracked during the pre-control inspections. If there were spatial correlation then stations located close to each other would have similar counts.

The relationship between the temporal distance between stations and the differences in station counts was examined with a Mantel randomisation test (Manly 1997, p. 174). Station counts in this temporal-analysis were the number of tunnels that were found tracked at an inspection. If there were temporal correlation inspections spaced 1 or 2 weeks apart would have similar counts. To estimate the significance level of any observed correlation, the correlation between the matrix of the station count differences and the matrix of the inverse of the spatial differences between inspections were calculated and compared to 9999 correlations where one of the matrices was randomly shuffled.

Estimation of power

The power to detect a difference in station counts pre- and post-control was estimated by simulation (Beier and Cunningham 1996) using a binomial model. The post-control probability of a station being tracked was calculated as a proportional reduction of the pre-control probability. Individual station counts were the number of inspections the station was tracked and the observed difference in the sum of the station counts between pre- and post-control was calculated. The observed difference was compared to differences when the pre- and post- station counts were pooled and the pool sampled, with replacement, 999 times. The proportion of simulated samples where the difference was at least as small as the observed difference was used as the 1-tailed significance level.

The generation of station counts, and simulated resampling, was repeated 1000 times and power was estimated as the proportion of the significance levels less than the type I error rate, α. The factors varied in the simulation were: the number of stations, the number of station inspections in the survey (the same number of inspections were used for pre- and post-control surveys), the initial probability of a station being tracked and the percentage decrease in this probability, and, α.

Factor levels were:

- pre-control probability of a station being tracked: 0.08, 0.15, and 0.30
- % control: 25%, 33.33%, 50%, and 75%
- number of station inspections (for either pre- or post-survey): 5, 10, and 15
- number of stations: 10, 25, and 50
- α: 0.05, 0.1, and 0.2

The initial probability levels were chosen based on the data from North Okarito forest. The average tracking rate over 19 inspections at Loop site was 1.5517, or 0.07 of the tunnels on an individual inspection were tracked. At River site the average tunnel count over 12 inspections was 1.0690, or 0.09 for an individual inspection.

The simulation method assumes that the probability of a station being tracked on inspection is independent of the probability of it being tracked on a previous, or subsequent inspection, and independent of the tracking rate at neighbouring stations.

3. Results

3.1 Station spacing - spatially

The number of the tunnels found to be tracked at an inspection ranged from 0 to 12 (Fig. 2). On average, the overall tracking rate was 0.08, i.e., 8% of the tunnels were tracked at any individual inspection. Some of the tunnels were never tracked during the monitoring (Fig. 3), e.g., tunnel 16 was never tracked, while tunnel number 6, located 500 m away (Fig. 1) was tracked on five inspections at Loop site, and on 3 inspections at River site.

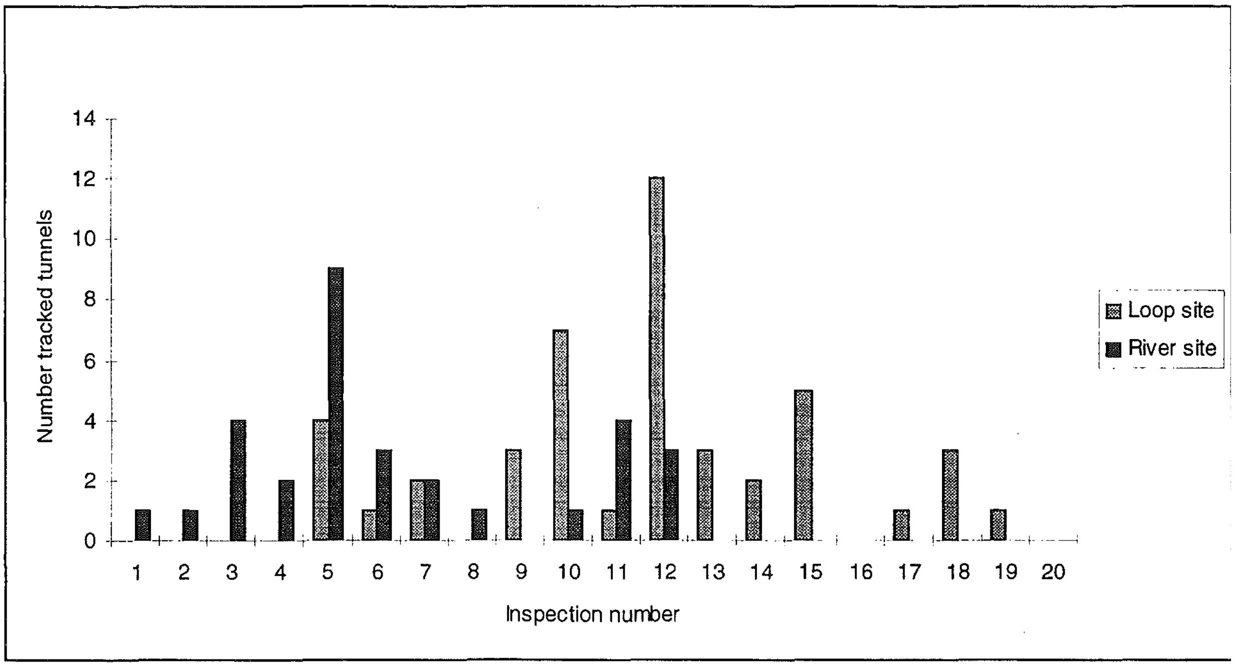


Fig. 2. Tunnel counts – the number of tunnels tracked at each inspection.

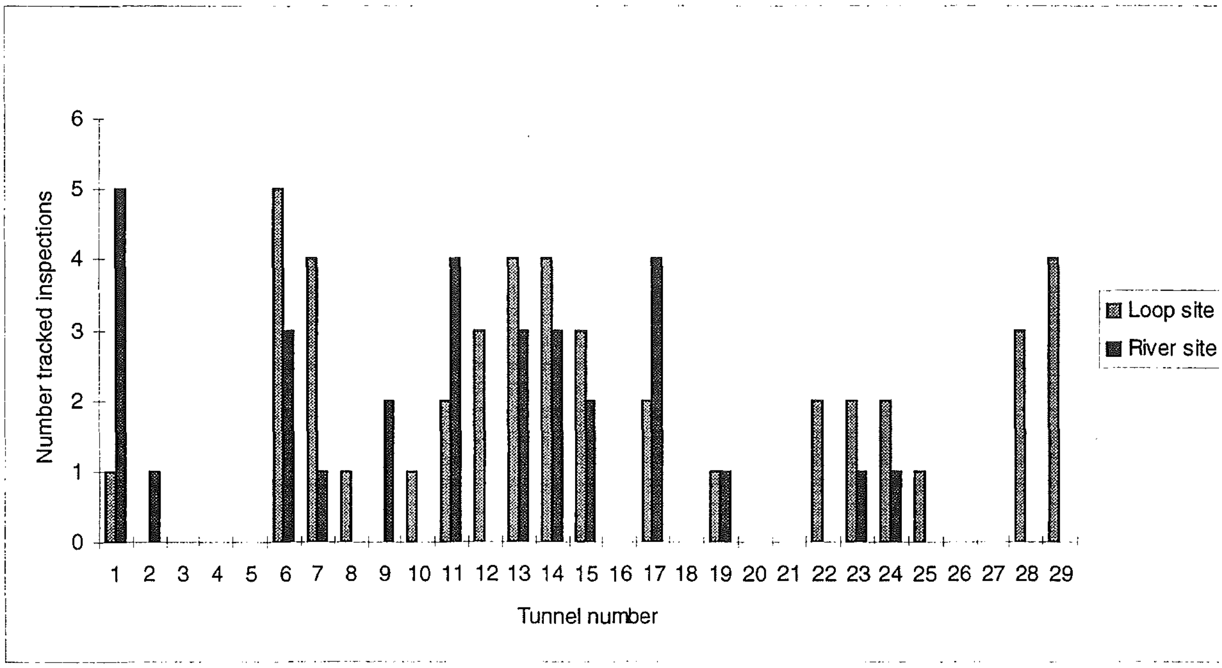


Fig. 3. Tunnel counts – the number of inspections the tunnels were tracked.

Counts from stations over 2 km apart were negatively correlated, that is, when one station tunnel count was high the other was low (Fig. 4). This suggests that within the study site stoats were not uniform in density, but were patchily distributed at a scale greater than 2 km. Similar trends were observed at both sites. Stations that were 500m apart had similar counts (positive correlation). Positive correlation would be expected if the same stoats were running through adjacent tunnels. Stoat home ranges for podocarp forest have been reported to be 27 ha to 190 ha for females, and 145 to

326 ha for males (Miller *et al.*, unpubl. data). For these home ranges, with a station spacing of 500m, more than one tunnel would be within a home range and tracking rates of adjacent stations cold be correlated.

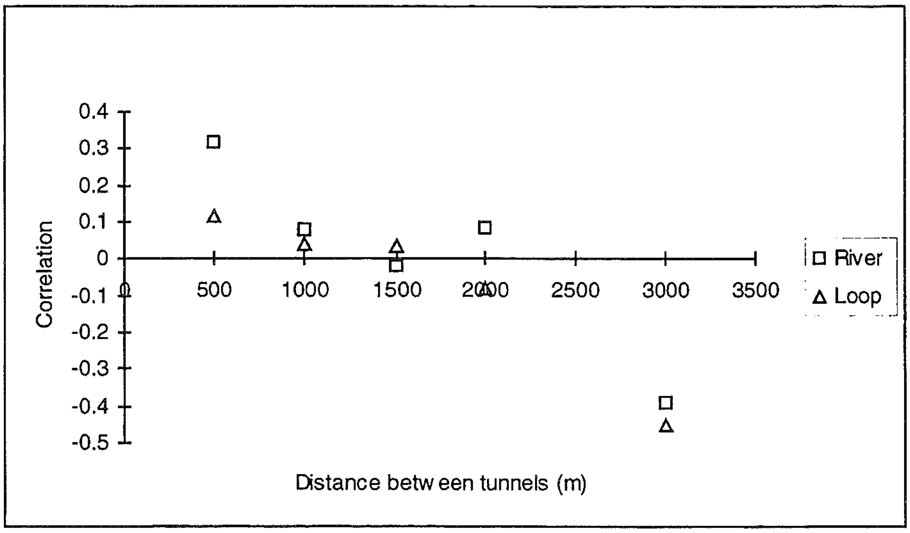


Fig. 4. Spatial correlation between tunnel counts.

3.2 Station spacing - temporally

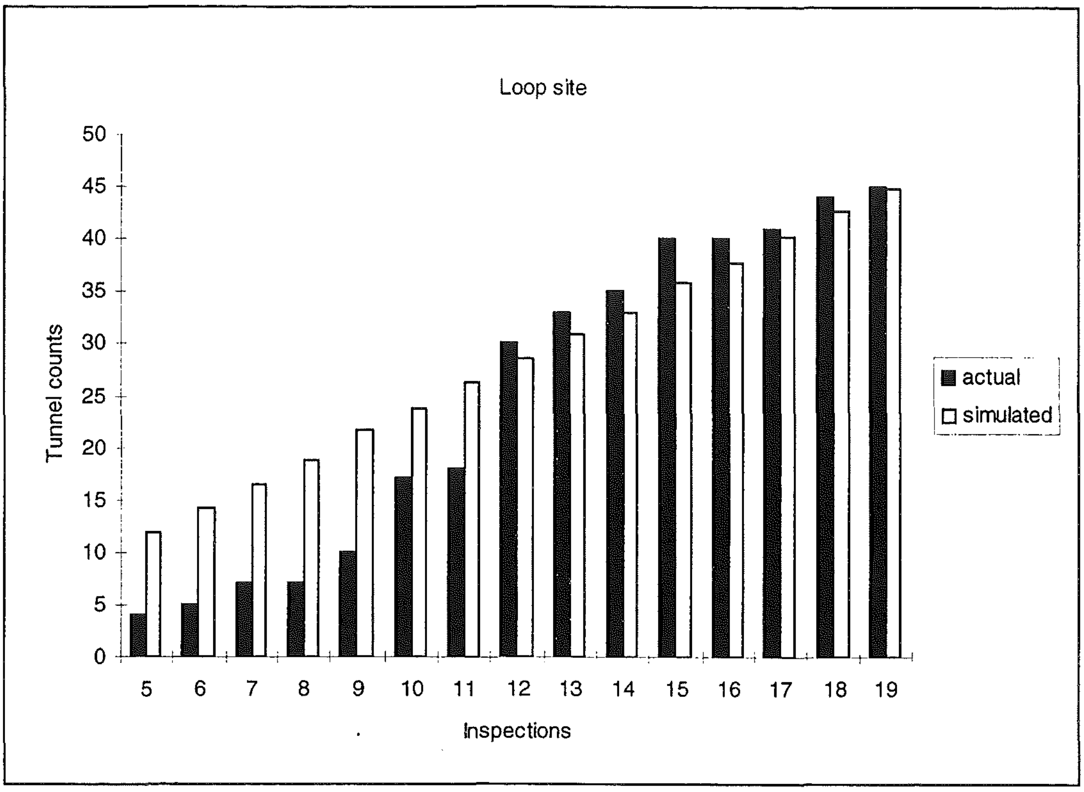
Station counts varied among inspections (Fig. 2). For example, at Loop site, 12 of the 29 tunnels were recorded as tracked in the inspection on the 12th week while in week 13 only 3 were tracked. The variation in weekly counts differed between the two sites. Week 12 had the highest count at Loop site, and at River site week 5 had the highest count.

In the first 4 weeks at Loop site no tunnel was tracked, and at River site the first 2 weeks had low counts. These results are consistent with behavioural response where an initial period is spent learning to run through the tunnels.

There was no evidence that the counts were correlated between the weekly inspections at either Loop site ($r = 0.06$, $P = 0.79$), or River site ($r = 0.04$, $P = 0.62$).

3.3 Estimates of power to detect control rates

The binomial model fitted the study data reasonably well and was a useful tool for approximating the effect of changes in survey design on statistical power. A comparison of the expected counts from the binomial model with the actual counts show that in the initial period of the survey the model slightly overestimated the station counts and underestimated the counts in the later period for Loop site (Fig. 5). This trend was the opposite of the trend at River site.



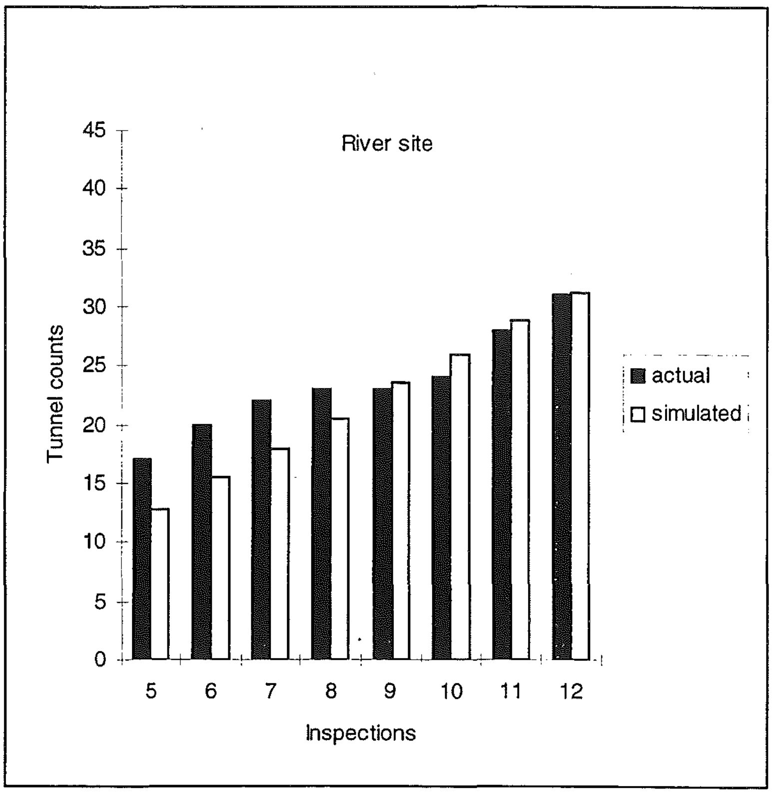


Fig. 5. Comparison of actual and simulated counts for Loop and River sites.

Survey effort

The results from the model suggest that the ability to detect the true control rate (power) for stoats improves with more survey effort (Fig. 6). With five weekly inspections the chance of detecting that a control operation achieved a 50% kill was estimated to be 40% when there were 25 stations, the initial probability of a station being tracked between inspections is 0.08, and $\alpha = 0.05$. With twice as many stations, 50, the chance increased to 60%. Extra survey effort can be either by way of more stations, or more inspections. For example, we estimated to have an 80% chance of detecting a 50% control we estimate that either 50 stations should be used and inspected weekly over 10 weeks, or 30 stations inspected over 15 weeks.

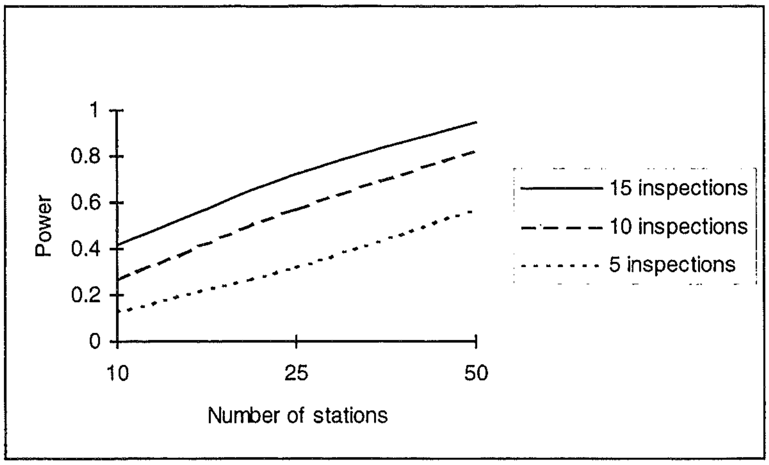


Fig. 6. Estimated power with different number of inspections for detecting a 50% kill, with a pre-control tracking rate of 0.08.

Control rate

Our model shows that larger reductions in stoat populations can be more reliably detected than smaller kill rates. With 5 weekly inspections the chance of detecting a 50% control with 25 stations was estimated to be 25% but the chance of detecting a 75% control was estimated to be 65% (Fig. 7). With more survey effort large kill rates can be even more reliably detected. With 10 inspections and 25 stations the chance of detecting a 50% control was estimated to be 50% and the chance of detecting a 75% control was estimated to be 85%. With ten inspections and 50 stations the ability to detect a 75% control was estimated as 90%.

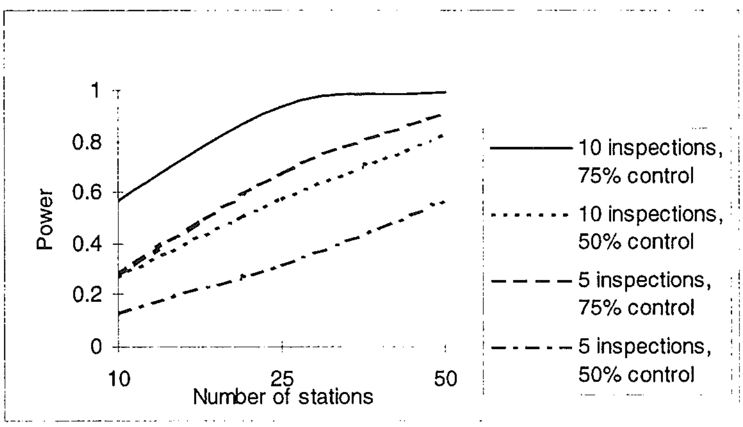


Fig. 7. Estimated power with different numbers of inspections and percentage kill, with a pre-control tracking rate of 0.08.

Tracking rate

Higher pre-control tracking rates improved the chance of detecting the percentage kill with our model. For example, with a pre-control tracking rate of 0.15 the ability to detect a 50% control with five weekly inspections, and 25 stations was 50%, but increased to 90% when the pre-control rate was 0.30 (Fig. 8). With a pre-control tracking rate of 0.08 even with 50 stations and five inspections the chance of detecting a 50% kill was only 50%, although with 15 inspections the power increased 90% (Fig. 8).

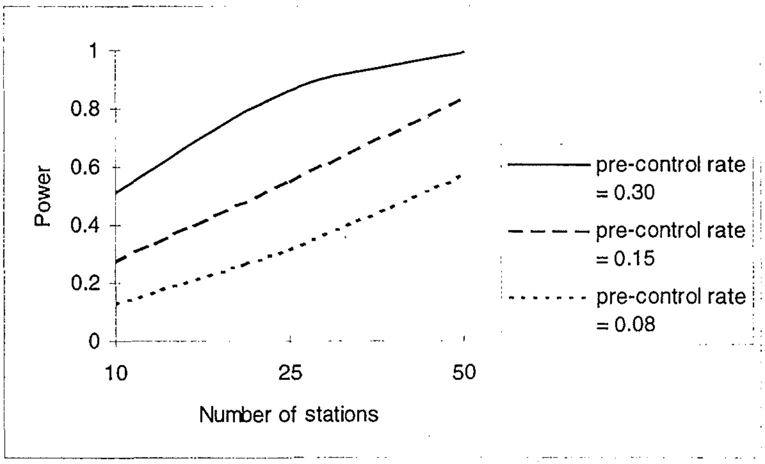


Fig. 8. Estimated power with different pre-control tracking rates for detecting 50% kill with five inspections.

Type I error rate

The ability to detect the true control rate was improved by accepting higher risks of falsely detecting a successful control operation when in fact there has been no reduction in the stoat population (type I error, α). For example with a type I error rate, α, of 0.05, and with 25 stations the chance of detecting a 50% kill was 50%, but when α = 10, the chance was 70%. Power improved further to 85% if α = 0.2 were used (Fig. 9).

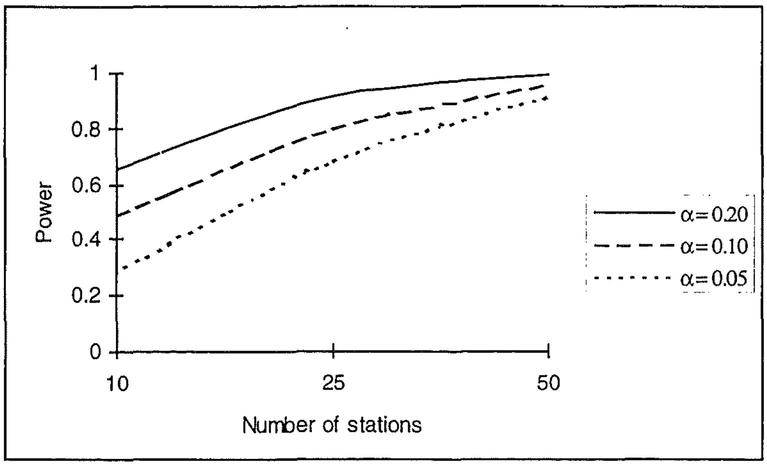


Fig. 9. Estimated power with different type I error rates, α, for detecting 50% kill with five inspections.

4. Discussion

Spacing of tunnel-stations
Tunnel-stations should be preferably at the scale of home ranges to reduce the chance of the same stoat running through more than one tunnel (Diefenbach *et al.* 1994). Since home ranges are likely to be elongated in shape (e.g., Miller *et al.*, unpubl data, Murphy and Dowding 1994) this would suggest wide spacing should be used.

In their study of stoat home ranges in beech forest Murphy and Dowding (1994) found the average range length to be 2.3 ± 0.3 km for females and 4.0 ± 0.9 km for males. Stations spaced more than 4 km apart would help ensure stations were in separate home ranges and that the counts of tracked stations were independent. However, this spacing is impractical for logistical reasons. Stoat densities are likely to be higher than what would be expected with these home ranges because home ranges can overlap and contract or expand between years and within a year, between seasons (Murphy and Dowding, 1995, Miller *et al.*, unpubl data). In addition, there may be other non-territorial animals such as juveniles and transients in the area. The presence of these other animals confounds the independence of tunnel-stations because even at 4 km spacing the same stoat could be in transit and running through many stations.

There may be no one ideal spacing, because even at very wide spacing stations may not be independent. On balance, we recommend a spacing of 1 km. In this study, there was little correlation between station counts and it is a practical and feasible spacing.

Station inspections
There was no evidence in our study that station counts spaced a week apart were correlated, and for future monitoring we recommend continuing with weekly counts. There are potential problems with this design in that the same stoat may be tracking the same station each week (Kendall *et al.* 1992).

We did observe that the initial station counts were low, consistent with other tracking tunnel studies. When tracking data are used to estimate control rates, initial low station counts will cause an underestimate of control rates. Therefore we recommend that the initial 3 weeks of data be excluded from the analysis. Alternatively, the stations could be laid out for 3 weeks prior to any inspection but undertaking these initial inspections is preferable to help identify any initial tunnel aversion.

Multiple tunnels at a station
The ability to detect a reduction in the stoat population with our model improved with higher tracking rates. One way to achieve higher tracking rates is to use multiple tunnels at each station. This will reduce the bias from not detecting a stoat given that they are in the area (Zeilinski and Stauffer 1996, Diefenbach *et al.* 1994) because, if there are stoats near the station there is more chance of at least one tunnel in the station being tracked.

With multiple tunnels at each station, the data collected at an inspection would be whether the station is tracked or not, that is, whether at least one tunnel is tracked during inspections. The station is considered the sample unit, rather than the individual tunnels.

At each station the tunnels should be clustered closely together with the stations 1 km apart. The optimum number of tunnels within a station needs further research and field validation, but as an estimate we recommend three tunnels, with the three tunnels spaced 50 m apart.

An alternative method for increasing the tracking rate is to use a interval longer than 1 week between station inspections. The risk with this strategy is that the total time interval between the start and finish of the monitoring survey will be long, and stoat behaviour may change and confound the tracking rates. For example, as alternative food sources become available in the forest the stoat's attraction to the tunnel bait will change. After a control operation stoats may begin re-invading the control area, and if the post-control monitoring extends into when re-invasion begins to occur control rates will be underestimated.

Estimation of control rates

Estimates of control rates calculated using the difference between pre- and post-control tracking rates should use the proportion of stations that had at least one tunnel tracked in an inspection, averaged over inspections. The sample size for this analysis is the number of stations and therefore stations should be spaced apart by a reasonable distance, e.g., 1km, to be independent.

The layout of multiple tunnels at widely spaced stations may seem counter-intuitive. Since 1 km has to be walked between stations why not have some intermediate tunnels along the path, e.g. 100m apart, which can be inspected for little extra cost? This is best answered by an example. Consider a design with a line of 10 stations of 3 tunnels where the stations were 1 km apart, and another with a line of 30 tunnels 100 m apart. Both designs use 30 tunnels. In the first design the sample size is 10 since there are 10 stations, and in the later the sample size is not 30 as would be expected but 1 since the tunnels are not independent.

The tunnels within each station are not independent and the number of tunnels is not the size of the sample. It is erroneous to use the number of tunnels as the sample size - in our proposed design, or in other monitoring design - when tunnels are not independent and with no adjustment for this bias. Typically pest monitoring uses multiple lines of stations (e.g., lines of tunnels or traps). If the stations are not independent but the lines are, the size of the sample is the number of lines and there may be too few lines to ensure large enough sample sizes to achieve sufficient power to detect the true control rate. With the proposed design of widely spaced tunnel-stations, each station contributes to the sample size and adequate sample sizes may be achieved without excessive effort. If lines of non-independent tunnels are used the only way to increase the sample size would be to increase the number of lines which can be costly.

Another design to avoid the problem of lack of independence of station counts is to randomly locate the stations in the study area (van der Meer 1997, Lesica and Steele 1996). However, this can be an impractical design for sites such as North Okarito where access is very difficult between stations.

These results from our analysis are based on grids of stations, but more commonly stations are laid out along a line and this may give some variation between our results and other studies (Brown *et al.* 1996). When stations are systematically laid out along the survey line, the stations should still be 1 km apart. The first station should be randomly located within the first km and subsequent stations at 1 km intervals.

Survey effort

The level of survey effort (the number of stations, and the number of inspections) for monitoring can be estimated by using our model and Fig. 6 to 9, provided the pre-control tracking rate can be estimated. If this is not available previous year's data, or data from similar sites can be used. Any estimate should be conservative, i.e., low; to ensure monitoring is as least as effective as predicted. Our estimate of the ability to detect the true kill rate improved with more survey effort, either by having more inspections or more stations. For example, we estimated a 50% kill could be detected with 60% power with either: 20 stations and 15 inspections; 30 stations and 10 inspections; or 55 stations and five inspections (Fig. 6). In most studies it would be more cost efficient to increase the number of stations rather than the number of inspections to improve power.

Our model can also be used to estimate power retrospectively once the monitoring has been conducted (Thomas 1997). If a control operation was thought to have not been effective, retrospective power analysis can be used to estimate the likelihood of being able to detect the target percentage kill given the monitoring effort. If the likelihood was low, e.g., 20% the manager should have little confidence in the monitoring results, and their survey protocol for monitoring should be refined.

The survey design used will depend on the required amount of power and the acceptable error rates. The ability to detect the actual control-rate is a compromise between making errors in reporting a control as being successful when it was not, and failing to detect a successful control (type I and II errors respectively). The ratio between the costs of these two errors can be used to help decide on appropriate levels of power rather than choosing an arbitrary value (Mapstone 1995).

We suggest that the ratio of the cost of a type II error to a type I error is 0.5, i.e., the immediate short term loss of falsely concluding a control was successful when it was not is twice the longer term cost of failing to detect a successful control. This is because if a control were not successful managers would need to undertake extra protection work for the vulnerable kiwis. If the unsuccessful control operation were thought to be successful extra work may not be carried out.

With this cost ratio, the ratio between the type I and II error rates should be 0.5, e.g., if the type I error rate were chosen to be 0.1 then the type II error rate should be set at 0.2, which will give power of 0.8. If a lower type I error rate were used, e.g., 0.05, then the type II error rate would be 0.1, with power of 0.9. With this procedure of setting power based on relative error costs, managers need only decide on the acceptable type I error rate.

Other design considerations
The ability to detect a control-rate may be improved by using a stratified design, where additional survey effort is focused on areas within the study site of either lower stoat numbers, or higher variability among stations (Zeilinski and Stauffer 1996). In the more usual stratified design extra effort is allocated to areas of higher abundance, but for monitoring stoats more power would be achieved by putting more effort into the areas of the low tracking rates. Extra effort can be either via more station inspections, or with more tunnels in each station. We observed that for Loop site some groups of tunnels were tracked more than others and this would suggest that within this site the area could be stratified.

Future Research
There are a number of questions to be answered, e.g., how many tunnels should be used at each station, and what spacing should be used for the tunnels and for the stations? Field validation of the proposed method of multiple tunnel-stations spaced 1 km apart should be undertaken.

6. Recommendations – Management and Research

Management

We recommend for monitoring stoats with tracking tunnels that:

- Tunnel-stations be spaced 1 km apart,
- Stations be inspected weekly,
- Multiple tunnels are used at each station. We recommend stations of three tunnels spaced 50 m apart,
- Data from the first 3 weeks after tunnels are first laid out and baited should not be used in estimating control rates.

The actual survey design for monitoring and the amount of survey effort; the number of stations and inspections, will depend on the acceptable error rates, the desired level of power, the target percentage kill, and the pre-control stoat density.

The decision on what are acceptable error rates should be evaluated carefully by considering the consequences of making such an error. The relative cost of the type I and type II errors can be used to estimate the type II error rate, and the desired level of statistical power. We recommend that the ratio, $cost_{II}/cost_{I} = 0.5$ be used as a preliminary estimate, and with a type I error of 0.1, the type II error should be 0.2, which will result in 80% power

Research

- Field trials should be conducted, and results analysed for statistical power, to more accurately define optimum tunnel layout and to field validate our results.
- A computer model should be produced for managers to use in planning monitoring and deciding on the number of stations and inspections. We have compiled the mathematical part of such a model for combinations of type I error rates, power, target percentage kill and pre-control stoat density but with further funding the programme could be developed to make it more user-friendly and accessible to managers. With some modifications the model also could be used for planning monitoring of other pest species, e.g., possums, rats, mice, and for use with traps

and wax blocks. Such a model will be useful for evaluating future monitoring strategies. Retrospectively it can be used to check that there was sufficient statistical power in a monitoring operation to be able to detect the true kill rate.

7. Acknowledgements

We thank Colin O'Donnell and Malcolm Thomas for comments and advice.

8. References

Beier, P. Cunningham, S.C. 1996. Power of track surveys to detect changes in cougar populations. *Wildlife Society Bulletin* 24:540-546.

Brown, K.P., Moller, H., Innes, J., Alterio, N. 1996. Calibration of tunnel tracking rates to estimate relative abundance of ship rates (*Rattus rattus*) and mice (*Mus Musculus*) in a New Zealand forest. *New Zealand Journal of Ecology* 20(2):217-275.

Diefenbach, D.D., Conroy, M.J., Warren, R.J., James, W.E., Baker, L.A., Hon, T. 1994. A test of the scent-station survey technique for bobcats. *Journal of Wildlife Management* 58:10-17.

Fairweather. P.G. 1991. Statistical power and design requirements for environmental monitoring. *Australian Journal of Marine and Freshwater Research* 42: 555-567.

Green, R.H. 1979. Sampling design and statistical methods for environmental biologists. Wiley, New York.

Kendall, K.C., Metzgar, L.H., Patterson, D.A., Steele, B.M. 1992. Power of sign surveys to monitor population trends. *Ecological Applications* 2:422-430.

King, C.M. & Edgar, R.L. 1977. Techniques for tracking and trapping stoats (Mustela erminea) a review and a new system. New Zealand Journal of Zoology 4: 193-212.

Lesica, P., Steele, B.M., 1996. A method for monitoring long-term population trends: an example using rare arctic-alpine plants. *Ecological Applications* 6:879-887.

Mapstone, B.D. 1995. Scaleable decision rules for environmental impact studies: effect size, type I, and type II errors. *Ecological Applications* 5:401-410.

Manly, B.F.J. 1997. Randomisation, bootstrap and Monte Carlo methods in biology. Chapman and Hall, London.

Miller, C.J. & Elliott, M. 1997. Experimental stoat control North Okarito forest, West Coast. Pp. 19-23 In Joanna Sim & Alan Saunders (eds.) Predator workshop 1997. Proceedings of a workshop held 21 - 24 April 1997, St Arnaud, Nelson Lakes.

Miller, C.J., Elliott, M., & Alterio, N. (unpubl. data).

Moors, P.J. 1978. Methods for studying predators and their effects on forest birds. Pp. 47-57 in: The ecology and control of rodents in New Zealand nature reserves. Department of Lands and Survey, Information Series No. 4.

Murphy, E.C., Dowding, J.E. 1994. Range and diet of stoats (*Mustela erminea*) in a New Zealand beech forest. *New Zealand Journal of Ecology* 18:11-18.

Murphy, E.C., Dowding, J.E. 1995. Ecology of the stoat in *Nothofagus* forest: home range, habitat use and diet at different stages of the beech mast cycle. *New Zealand Journal of Ecology* 19:97-109.

Norton, D.A. 1996. Monitoring biodiversity in New Zealand's terrestrial ecosystems. In: McFadgen, B. & Simpson, P. (comp.) Biodiversity: papers from a seminar series on biodiversity. Science & Research Division, Department of Conservation, Wellington. Pp. 19-41.

O'Donnell, C.F.J., Dilks, P.J., & Elliott, G.P. 1996. Control of stoat (*Mustela erminea*) population irruption to enhance mohoua (*Mohoua ochrocephala*) breeding success in New Zealand. *New Zealand Journal of Zoology* 23: 279-286.

Peterman, R.M. 1990. Statistical power analysis can improve fisheries research and management. *Canadian Journal of Fisheries and Aquatic Sciences* 47:2-15.

Steidl, R.J., Hayes, J.P., Schauber, E. 1997. Statistical power analysis in wildlife research. *Journal of Wildlife Management* 61:270-279.

Taylor, B.L., Gerrodette, T. 1993. The uses of statistical power in conservation biology: the vaquita and northern spotted owl. *Conservation Biology* 7:489-500.

van der Meer, J. 1997. Sampling design of monitoring programmes for marine benthos: a comparison between the use of fixed versus randomly selected stations. *Journal of Sea Research* 37: 167-179.

Thomas, L. 1997. Retrospective power analysis. *Conservation Biology* 11: 276-280.

Zeilinski, W.J., Stauffer, H.B. 1996. Monitoring *Martes* populations in California: survey design and power analysis. *Ecological Applications* 6:1254-1267.