Fisheries and Oceans Canada / Pêches et Océans Canada
Science / Sciences

C S A S | **S C C S**

Canadian Science Advisory Secretariat | **Secrétariat canadien de consultation scientifique**

Research Document 2008/064 | **Document de recherche 2008/064**

Not to be cited without
permission of the authors *

Ne pas citer sans
autorisation des auteurs *

An evaluation of Steller sea lion (*Eumetopias jubatus*) pup counts from 35mm oblique images

Évaluation des dénombrements de petits de l'otarie de Steller (*Eumetopias jubatus*) à partir d'images 35 mm prises à angle oblique

P. F. Olesiuk[1], D. G. Calkins[2,3], K. W. Pitcher[3],
W. L. Perryman[4], C. Stinchcomb[4] and M. Lynn[4]

[1]Fisheries and Oceans Canada, Pacific Biological Station, Nanaimo, B.C., V9T 6N7
[2]Alaska Department of Fish and Game, 333 Raspberry Road, Anchorage, AK 99518
[3]Alaska SeaLife Center, 301 Railway Avenue, P.O. Box 1329, Seward, AK 99664
[4]National Marine Fisheries Service, Southwest Fisheries Science Centre
P.O. Box 271, La Jolla, CA 9203

ISSN 1499-3848 (Printed / Imprimé)

Canada

ABSTRACT

The precision and accuracy of Steller sea lion (*Eumetopias jubatus*) pup counts made from oblique 35mm aerial slides was assessed by comparing them to ground drive-counts and counts from aerial vertical medium-format images. DFO flew surveys using oblique 35mm photography within 2 days of ADF&G's ground drive-counts at Forrester Island, Alaska, in 1994, 1995, 1997 and 1998. In 1998 and 2002, aerial surveys were conducted at rookeries in B.C. and Forrester Island using both 35mm oblique and vertical medium-format photography, the latter of which was shown to provide pup counts statistically equivalent to ground drive-counts (Snyder et al. 2001). As expected, ground drive-counts provided the most precise pup counts (CV=0.047) and are widely regarded as the most accurate method against which other techniques are generally validated. In an earlier study, Snyder et al. (2001) showed that vertical medium-format images provided as good precision (CV=0.048) as ground counts on Alaskan rookeries. However, our initial counts of medium-format images for the B.C. survey in 1998 were less precise (CV=0.094) because one reader obtained counts that were consistently (8 of 10 sites) and significantly ($0.0001<P<0.0110$) greater than the other reader. The precision was improved (CV=0.056) by replacing the lower of the initial two reader's counts with those of a third reader, whose counts agreed closely with the higher of the initial two readers. There were no significant differences among readers and fairly good precision (CV=0.063) for the medium-format images of B.C. rookeries in 2002, leading to an overall CV=0.060 for the two medium-format surveys. Counts made from oblique 35mm slides were reproducible among readers (CV=0.085), and were similar for surveys replicated on different dates (CV=0.102), but appeared to be slightly biased.

Comparison of pup counts from oblique 35mm slides to ground drive-counts at Forrester Island indicated that the 35mm counts tended to be significantly lower (P<0.0001), with about 80% of the pups seen on the ground evident in the 35mm slides. The degree of bias seemed relatively constant on a site-by-site basis (slope=0.797-0.813; 95% *CI* of 0.738-0.893) and between the 4 years (mean 79.7%; range 76-85%). Similarly, comparison of 35mm slide counts with medium-format counts at Forrester Island also indicated that only about 80% of pups seen in medium-format images were evident in the 35mm slides (slope=0.797-0.813; 95% *CI* of 0.693-0.891). The 35mm counts appeared to be less biased on B.C. rookeries, with about 96% of pups in the medium-format images being evident in the 35mm slides. We suspect that the difference in accuracy of 35mm pup counts between Forrrester Island and B.C. rookeries may be due to differences in the size and topography of rookeries. We therefore recommend that a correction factor of 1.25 (95% CI of 1.12-1.44) be applied to pup counts made from 35mm slides at Forrester Island, and a correction factor of 1.05 (95% CI of 1.018-1.075) be applied to pup counts made from 35mm slides at B.C. rookeries, to account for pups that are obscured in photographs taken at oblique angles.

In 2006, we began a transition from 35mm slides to digital photography. To insure consistency of the survey time-series, all rookeries in B.C. where photographed using both a film and digital SLR camera. There was close agreement between the digital and film counts for both pups ($0.9942<r^2<0.9954$) and non-pups ($0.9994<r^2=0.9994$), and in both cases the intercepts passed through the origin ($0.423<P<0.742$) and slopes were not significantly different from unity ($0.503<P<0.849$) indicating the counts were statistically indistinguishable from one another.

RÉSUMÉ

La précision et l'exactitude des dénombrements de petits de l'otarie de Steller (*Eumetopias jubatus*) effectués à partir de diapositives 35 mm prises à un angle oblique à partir d'un vion a été évaluée à l'aide d'une comparaison avec des dénombrements effectués sur le sol et es dénombrements faits à partir d'images de format moyen prises à la verticale, également epuis un avion. Le MPO a mené desrelevés aériens à l'aide de photographies 35 mm prises à angle oblique dans les deux jours qui ont suivi des dénombrements menés au sol par DF&G à l'île Forrester, en Alaska, en 1994, en 1995, en 1997 et en 1998. En 1998 et en 2002, on a également mené des relevés aériens à des roqueries de la C.-B. et à l'île Forrester à l'aide de photographies 35 mm prises à angle oblique et de photographies de format moyen prises à la verticale, ce qui a permis de démontrer que la deuxième méthode a fourni des dénombrements des petits équivalents sur le plan statistique aux dénombrements effectués sur le terrain (Snyder *et al*., 2001). Comme on s'y attendait, les dénombrements effectués sur le sol ont fourni les dénombrements de petits les plus précis (*CV* = 0,047) et sont largement considérés comme étant la méthode la plus précise pour effectuer la validation d'autres techniques en général. Dans une étude antérieure, Snyder *et al*. (2001) ont démontré que les images de format moyen prises à la verticale affichaient une précision (*CV* = 0,048) aussi bonne que celles des dénombrements effectués au sol dans les roqueries de l'Alaska. Toutefois, nos dénombrements initiaux faits à partir d'images de format moyen pour le relevé de 1998 en C.-B. ont été moins précis (*CV* = 0,094) du fait que le lecteur donnait des dénombrements qui étaient constamment (à 8 des 10 sites) et de façon significative (0,0001 < *P* < 0,0110) supérieurs à ceux obtenus avec l'ancien lecteur. La précision a été améliorée (*CV* = 0,056) lorsque l'on a remplacé les deux dénombrements les moins élevés des lecteurs par ceux d'un troisième lecteur, dont les dénombrements étaient pratiquement identiques aux dénombrements les plus élevés des deux lecteurs initiaux. Aucune différence significative n'a été observée entre les lecteurs et la précision a été relativement bonne (*CV* = 0,063) pour les images de format moyen des roqueries de la C.-B. en 2002, ce qui nous a donné un coefficient de variation global de 0,060 pour les deux relevés effectués avec des photographies de format moyen. Les dénombrements faits à l'aide des diapositives 35 mm prises à angle oblique ont pu être reproduits d'un lecteur à l'autre (*CV* = 0,085) et ont été similaires pour les relevés qui ont été réeffectués à de nouvelles dates (*CV* = 0,102), mais ont semblé légèrement biaisés.

La comparaison des dénombrements de petits effectués à partir de diapositives 35 mm prises à angle oblique et les dénombrements effectués au sol à l'île Forrester a révélé que les dénombrements provenant des diapositives avaient tendance à être considérablement inférieurs (P < 0,0001) et qu'environ 80 % des petits observés sur le sol étaient visibles sur les diapositives 35 mm. Le degré de biais a semblé relativement constant d'un site à l'autre (pente = 0,797-0,813; IC de 95 % de 0,738-0,893) et entre les quatre années (moyenne de 79,7 %; fourchette de 76 à 85 %). De la même façon, la comparaison des dénombrements à partir de diapositives 35 mm et des dénombrements à partir de photographies de format moyen à l'île Forrester a également révélé que seulement 80 % des petits observés sur les photographies de format moyen étaient visibles sur les diapositives 35 mm (pente = 0,797-0,813; IC de 95 % de 0,693-0,891). Les dénombrements faits à partir des diapositives 35 mm ont semblé être moins biaisés dans le cas des roqueries de la C.-B. du fait qu'environ 96 % des petits qui apparaissaient sur les photographies de format moyen étaient visibles sur les diapositives 35 mm. Nous pensons que la différence dans la précision des dénombrements effectués avec les diapositives 35 mm entre l'île Forrester et les roqueries de la C.-B. peut être attribuable àdes différences au chapitre de la taille et de la topographie des roqueries. Nous recommandons par conséquent qu'un facteur de correction de 1,25 (IC de 95 % de 1,12- 1,44) soit appliqué aux

dénombrements de petits faits à partir de diapositives à l'île Forrester et qu'un facteur de correction de 1,05 (IC de 95 % de 1,018-1,075) soit appliqué aux dénombrements de petits effectués à partir de diapositives 35 mm aux roqueries de la C.-B. afin de tenir compte des petits qui n'apparaissent pas clairement dans les photographies prises à angle oblique.

En 2006, nous avons commencé à délaisser les diapositives 35 mm au profit de photographies numériques. Afin d'assurer l'uniformité de la série chronologique des relevés, nous avons photographié toutes les roqueries de la C.-B. en utilisant un appareil à pellicule et un appareil SLR numérique. Nous avons observé une forte corrélation entre les dénombrements faits avec les appareils numériques et à pellicule tant chez les petits ($0,9942 < r_2 < 0,9954$) que chez les juvéniles et les adultes ($0,9994 < r_2 = 0,9994$) et, dans les deux cas, les points d'interception ont croisé les origines ($0,423 < P < 0,742$) et les pentes n'ont pas différé de façon significative par rapport à l'unité ($0,503 < P < 0,849$), ce qui indique que les dénombrements ne peuvent être distingués statistiquement l'un de l'autre.

INTRODUCTION

Steller sea lions (*Eumetopias jubatus*) breed along the North Pacific Rim from the Kuril Islands and Kamchatka Peninsula, west through the Aleutian and Pribilof Islands into the Gulf of Alaska, and south along the continental shelf as far as central California. Three stocks are recognized based on genetic differences (Bickham et al. 1996; Baker et al. 2005) and phylogeographic patterns (Loughlin 1997). The Asian stock breeds in Russia west of the Commander Islands and the Western stock breeds in the Commander and Aleutian Islands and Gulf of Alaska west of Cape Suckling (144°W). The Eastern stock breeds in Southeast Alaska, British Columbia, Oregon and north-central California. The western stock, having declined by about 80% since the 1970s (Merrick et al. 1987; Loughlin et al. 1992; Trites and Larkin 1996; Loughlin 1998, NMFS 2007) was listed as *endangered* under the U.S. *Endangered Species Act,* has been the focus of much research in recent years (NMFS 2007). In contrast, the eastern stock appears to be stable or increasing over much of its range (Calkins et al. 1999; Brown and Riemer 1992; Olesiuk 2007; Pitcher et al. 2007), but it was nevertheless listed as *threatened* in the U.S. due to concerns the declines – which were first observed in eastern Aleutian Islands and spread to the Gulf of Alaska (Braham et al. 1980) – may continue spreading to the Eastern stock, and some uncertainty at the time regarding the genetic division of stocks. In Canada, *COSEWIC* originally concluded in 1987 that the species was *not at risk,* but more recently in 2003 recommended the species be designated as *special concern* under the *Species at Risk Act.* The re-designation was based primarily on the unexplained declines that had occurred in western Alaska, the species is sensitive to disturbances while on land, and the limited number (3) of breeding sites in Canadian waters. Given the widespread distribution of Steller sea lions which spans several state and federal jurisdictions, it is important that survey procedures be co-ordinated and calibrated amongst the various agencies responsible for monitoring populations in different parts of their range.

Two basic approaches have been used to monitor the abundance of pinnipeds. For species that tend to be widely scattered with breeding activities dispersed, such as harbour seals, population estimates have been derived by surveying all sex- and age-classes at haulout sites and applying corrections for animals at sea and hence missed during surveys. Corrections are generally based on haulout patterns as they relate to survey conditions, or on the proportion of radio-tagged animals hauled out during surveys (Harvey 1987; Huber 1995; Withrow and Loughlin 1997; Olesiuk 1999). For Steller sea lions, behavioural patterns vary widely among individuals and with reproductive status, and differ among sex- and age-classes (Merrick and Loughlin 1997; Swain 1996). The proportion of animals visible during surveys also varies seasonally and with environmental variables such as ocean swell height, time-of-day, tide height, and time relative to low tide (Mathison and Lopp 1963; Calkins et al. 1999). Loughlin et al. (1992) reviewed existing data and concluded that they were insufficient for estimating the proportion of animals at sea and missed during surveys, but he indirectly estimated that 75% of all animals were counted based on the number of animals observed during surveys relative to the expected actual population size (estimated from pup production – see below). Using similar techniques and slightly more comprehensive survey data, Sease et al. (unpublished manuscript) estimated that 82% of animals were counted during surveys.

The second approach for estimating abundance, particularly for species that congregate to breed, is to measure pup production and estimate total population size from the expected ratio of pups: non-pups (Berksen and DeMaster 1995). For Steller sea lions, virtually all pups are born at traditional breeding areas, and the pupping season is fairly synchronized throughout the range with births generally occurring from late May to early July

(Bigg 1985; but see Pitcher et al. unpubl.). Newborn pups are not competent swimmers and are thus confined to rookeries for about the first month of life (Sandegren 1970). Surveys of rookeries timed to coincide with the end of the pupping season, by which time most pups have been born and few have begun to disperse, can thus provide a fairly accurate estimate of total annual pup production. Based on life tables, Calkins and Pitcher (1982) calculated there would be a total 4.5 animals (including pups) in the population for each pup born. Trites and Larkin (1996), using the same life table data, derived a ratio of 4.6 and suggested that 10% of pups were missed in surveys, resulting in a multiplier of 5.1.

While pup production theoretically provides the best estimate of total sea lion abundance, in practice pups are more difficult to census than other age-classes. Pups are small and black and tend to blend in with the substrate. They are often positioned close to their mothers or form tight pods with other pups, and are thus more difficult to see from the air or vantage-points on land, especially at oblique angles. For these reasons, pups have traditionally been surveyed by ground drive (spook) counts, where non-pups are driven off the rookery and counters walk through the rookery tallying the number of pups. While this is widely regarded as the most accurate method, it can be disruptive. In the only study that has attempted to quantify the impact, Lewis (1987) reported that 30% fewer mothers maintained contact with pups (49% versus 71% respectively) in a year in which drive-counts were conducted compared to a control year with no drive-count.

Recently, researchers at the Southwest Fisheries Science Center (SWFSC) developed aerial census techniques and procedures for pinnipeds using motion-stabilized vertical medium-format photography. This medium-format system has been used to obtain more precise and accurate pup counts for several species of pinniped (Lowry and Perryman 1992; Lowry et al. 1996; Lowry 1999). In an evaluation for Steller sea lions at Año Nuevo, California, Westlake et al. (1997) concluded that pup counts from vertical medium-format photograph were more precise (less variable) than those made with binoculars and spotting scopes from various vantage points along the periphery of the rookery. In a more recent evaluation at ten Alaskan rookeries, pup counts made from medium-format images were found to be statistically indistinguishable from ground drive-counts, and it was concluded that vertical aerial photography was a viable method for surveying Steller sea lion pups (Snyder 1998; Snyder et al. 2001).

In British Columbia, Steller sea lion rookeries are situated on small isolated islands, and good vantage points for counting pups do not exist or are not easily accessible. The rookeries have been designated as Ecological Reserves or are protected in a National Park Reserve, which prohibits regular disturbances. Since the early 1970s, Department Fisheries and Oceans (DFO) has surveyed Steller sea lions in B.C., including pups, using oblique 35mm aerial photography (Bigg 1985). While other researchers have reported little success with this method for counting pups (Mathisen and Loop 1963; Withrow 1982; Brown and Riemer 1992; Snyder 1998), these evaluations appear to have been based on black and white photographs, or on colour slides taken for the purpose of counting non-pups. For the DFO surveys, somewhat specialized photographic procedures have been developed specifically for counting pups, but they have never been evaluated.

In this report, we evaluate the oblique 35mm slide counts made by DFO by comparing them with concurrent ground drive-counts and vertical medium-format images. During 1994-95 and 1997-98 , DFO conducted aerial pup surveys using oblique 35mm photography at Forrester Island, Alaska, within 1-2 days of Alaska Department of Fish and Game's (ADF&G) ground drive- counts. In 1998 and 2002, Forrester Island and all B.C. rookeries were surveyed using both vertical medium-format photography and 35mm oblique photography. Our objectives here are to evaluate whether the oblique 35mm pup counts represent: 1) a precise index of pup abundance, as indicated by a higher degree of correlation with ground and medium-format counts; and 2) an accurate estimate of pup production, as indicated by a 1:1 relationship with ground counts and medium-format images. As part of a transition from film to digital photography, in 2006 we photographed all B.C. rookeries using both a film and a digital 35mm camera. The pup and non-pup counts from slides and digital images are compared to evaluate whether the switch to digital photography can be made without comprising the consistency of the Steller sea lion survey time-series.

METHODS

Survey Procedures:

35mm Oblique Photography:

The 35mm oblique aerial surveys were flown in a *DeHavilland* Beaver floatplane. This type of aircraft provided an ideal platform for such work as it has an opening front window for taking photographs, can fly slow with superior stalling characteristics, and is capable of making steep bank turns up to 60°. These features were considered essential for manoeuvring around rookeries where downdrafts and our prop-wash were commonly encountered while making circles tight enough to obtain good photographs. All of the 35mm surveys used in this assessment were flown by the senior author, who has been conducting such surveys for DFO since the early 1980s.

Special effort was taken to insure 35mm slides were of suitable quality for counting pups (compared to non-pups). Animals were photographed with a *Pentax* or *Nikon* SLR autofocus camera equipped with a fast, high-quality 70-200mm zoom lens using *Extachrome* 400 ASA or, when light levels allowed, preferably *Kodachrome* 200 ASA film. Light readings were taken from the ocean surface away from land to prevent distortion by the reflection of breaking surf. We tended to slightly over-expose film which enhanced the visibility of the dark pups, especially on darker substrates. Several passes were made over each group of animals using steep bank turns at an altitude of 150-200 metres, and we shot a mosaic of overlapping photographs, usually at the highest magnification allowed by the zoom lens.

The 35mm slides were counted by projecting the image onto white paper using a *Prado Leitz* projector, which provided superior optics. We began by previewing all passes and selecting the highest quality images. Groups of animals were usually counted from the same pass, so we could use both individual animals and physical features to delineate boundaries between overlapping slides. We generally tried to make counts from the centre of overlapping frames, where optical distortion was minimal. Pups were distinguished on the basis of colour and small size and marked on the paper with felt pen, and the marks tallied once the count was complete. We adopted a “balance-of-probability” approach, rather than counting only images that could positively be identified as pups (which would lead to an underestimate) or all images that could possibly have been pups (which would lead to an overestimate). We began by quickly going over the slide and marking those that were very clearly pups, and then carefully deliberating over each of those for which there was some uncertainty. It is worth noting that this was a very tedious and time-consuming process, and required about an order of magnitude more effort and far more slides than would typically be necessary for Steller non-pup surveys. For instance, for the province-wide survey in 1998, the author utilized a total of 127 slides to count 2,071 pups on B.C. rookeries, and another 87 slides to count 2,364 pups on Forrester Island.

In 2006, we evaluated using 35mm digital photography instead of film. All rookeries in B.C. were photographed using both digital and film SLR cameras within a few minutes of each other. Digital images were taken with a 10.2 mega-pixel Nikon D200 single-lens reflex camera equipped with the same 80-200mm f2.8 Nikor lens used on the film camera[1]. Im Digital images and PhotoShop psd files were managed (and if necessary adjusted) in an Aperture library in OS X on a MacPro computer, and counted in PhotoShop CS3 using a Reindeer Graphics Image Processing Tool Kit. Following the methodology developed by Withrow (pers. comm.) and adapted by Olesiuk (2006) for harbour seal surveys, separate layers were created for pup counts, non-pup counts, and demarcation lines and notes. Animals were marked on each layer with colour-coded symbols using the brush tool, and tallied using the Reindeer Graphics Count tool. Photographic slides were processed as outlined above, and the two counts compared to insure consistency of the survey time-series.

Medium-Format Vertical Photography:

The medium-format surveys were flown by SWFSC personnel who have with extensive experience using this technique. The surveys were conducted from a twin-engine *Aero Commander* specially equipped with a belly mounted camera portal. Photographs were taken with a medium-format (126mm) KA-76 camera that had originally been developed for military reconnaissance. The camera was equipped with forward motion compensation, such that film was advanced at the same speed as the plane. Photographs were taken with about 80% overlap on *Kodak Aerochrome* HS SO-359 film developed as 126x91mm colour transparencies. Transects were flown over all occupied areas of the rookery at an altitude of 195 – 210 meters, and film slightly overexposed.

Medium-format images were counted on a light table equipped with a movable, variable-power microscope. A clear sheet of acetate was placed over the frame being counted, and each pup marked with a fine-tip felt pens while being tallied with a hand counter. The same criteria and degree of subjectivity was used as has been used for the 35mm slides, at least for the counts made by the senior authors. Once the frame had been counted, the acetate sheet was transferred to the preceding or proceeding frame, which had been taken from a slightly different position as the aircraft advanced along the transect. Each mark was checked, and usually some additional pups became evident that had been missed in the first count. The process was repeated with other frames covering the same area, such that each count was based on the cumulative tally in 3-4 images. The larger medium-format images covered a broader area (80% of altitude) than 35mm slides, and it was generally possible to count the entire rocks, except for large islands or long beaches. This required many fewer images – for example, in the 1998 survey the senior author used a third as many medium-format images as 35mm slides (41 versus 127) to count the B.C. rookeries, and was thus a much less tedious process.

[1]Given the smaller sensor size on the digital camera, the 80-200mm lens provided the equivalent of a 120-300mm lens on the digital camera. However, since most images taken near the middle of the zoom range, this had little effect on the magnification of the images.

Ground Drive Counts:

Ground counts at Forrester Island were conducted by various Alaska Department of Fish and Game personnel, all having extensive experience working with Steller sea lions. A team of usually two persons were landed on the rookery by small boat, which would drive most non-pups into the water. Young pups generally stayed on land and gathered in groups, and were quite approachable, and the team would walk through the rookery together and independently tally the number of pups present on hand counters. Occasionally fairly large groups of older pups would enter the water, and approximations were made of the size of these groups.

Comparisons:

For the ground and 35mm aerial comparisons at Forrester Island, we attempted to fly the 35mm aerial surveys just before the rookeries were disturbed by the ground drive-counts. Aerial surveys were conducted on 01-July-94, 22-June-95, 28-June-97, and 03-July-98. In 1997, the ground count was made the day following the aerial survey, and in 1995 and 1998 the ground counts were made 2 days after the aerial survey. In 1994, the aerial survey was delayed due to weather, and was flown the day following the ground drive-count.

For the medium-format and 35mm comparison, we attempted to fly the surveys (neither of which disturbed the animals) at about the same time. The medium format survey of Forrester Island and all B.C. rookeries was flown on 21-June-98. For the 35mm surveys, the Scott Islands, which account for about 70% of total pup production in B.C., was flown two days later on 23-June-98. The 35mm survey at North Danger Rocks was conducted on 03-July-98, and at Cape St. James and Forrester Island on 03-July-98, which was 11 and 12 days after the medium-format survey respectively. In order to evaluate whether pup numbers may have changed over this period, a second 35mm survey was conducted of the Scott Islands on 04-July-98, which was 13 days after the medium-format survey and 11 days after the first 35mm survey. In 2002, the medium-format survey at Forrester Island was flown on 05-July-02, and at B.C. rookeries on 06-July-02, and the 35mm survey flown 2-3 days earlier. However, while the medium-format survey was being conducted on the Scott Islands, we simultaneously flew a second 35mm survey. Since these two surveys were conducted within an hour of each other, conditions were identical and for the most part the exact same animals were being photographed, so we used this series to compare non-pup counts as well as pup counts.

Statistical Analysis:

As a measure of the *precision* (reproducibility) of the counts associated with each survey method, we calculated a small-sample coefficient of variation:

[1] $CV = (S/\overline{X})(1+1/4n)$

where S denotes the standard deviation, $\overline{X}$ the mean, and n the sample size (as per Snyder 1998). The first term is the classic *CV*, expressed as a proportion of the mean, and the second term a small-sample size correction (Sokal and Braumann 1980). We calculated *CV*s for each site, and averaged them as an indicator of the typical precision of each method.

Survey methods were compared by examining the linear relationship between counts by each method on a site-by-site basis. We were interested in two distinct aspects of the relationship: its strength and its position. As a measure of *strength,* we calculated the correlation coefficient:

[2] $r = \Sigma xy / (\Sigma x^2 \Sigma y^2)^{1/2}$

where if X and Y were to denote the counts obtained by the two methods being compared, Σx^2 and Σy^2 denote their sum of squares. $(1-r^2)$ provides a measure of the residual variation that was unexplained by a linear model. It should also be noted that the correlation coefficient is scale invariant: adding or subtracting a constant from either or both variables, and or multiplying either or both by a constant, has no effect on its value. The correlation coefficient thus provides a pure measure of the strength of the linear relationship, but conveys no information about its position. A high degree of correlation only indicates that the two variables are proportional to one another, but not whether one variable represents, say, half or the same or twice the other variable. A high correlation coefficient merely indicates that the two methods being compared provide a good *relative index* of one another.

The second aspect of the relationship of interest was its position as indicated by its slope and intercept:

[3] $Y = \beta_0 + \beta_1 X$

which at first glance, appears to be a simple exercise in linear regression. However, in this case there are measurement errors associated with both variables – two counters reading the same image or walking through the same rookery will get slightly different counts – which violates a fundamental assumption of linear regression that the independent variable is measured without error. A structural (errors-in-variables regression) model was therefore considered to be more appropriate:

[4] $\hat{Y} = \beta_0 + \beta_1 \hat{X}$ or $(Y + \delta 1) = \beta_0 + \beta_1 (X + \delta 2)$

where the true number of animals present at the time of each survey, X and Y, are not known, and the observed counts, $\hat{Y} = (Y+\delta 1)$ and $\hat{X} = (X+\delta 2)$, are each subject to measurement error, $\delta 1$ and $\delta 2$ respectively. Unfortunately, it is impossible to fit equation [4] to a series of observations without providing some ancillary data. There are several statistical approaches to this problem, but the most common and straightforward is to specify the ratio of the measurement error variances:

[5] $\lambda = \sigma_{\delta 1}^2 / \sigma_{\delta 2}^2$

which we estimated from the ratios of the *CV*s associated with counts for each survey method (equation [1]). In cases where the same methods were being compared (i.e. comparing two observer's counts, or digital images versus photographic slides), λ was set to unity. Given λ, an estimate of the slope of the structural relationship, B_1, can be obtained by either maximum likelihood (Kendall and Stuart 1967) or least squares (Sprent 1969) methods:

[6] $B_1 = \Sigma y^2 - \lambda \Sigma x^2 + ((\Sigma y^2 - \lambda \Sigma x^2)^2 + 4\lambda(\Sigma xy)^2)^{1/2} / 2\Sigma xy$

and its intercept subsequently obtained from $B_0 = \bar{Y} - B_1 \bar{X}$. We excluded sites containing no or few (<10) pups from the calculations so as to prevent them from artificially forcing the relationship through the origin.

Heuristically, λ can be thought of as a weighting factor that dictates the relative emphasis given to minimizing vertical departures and horizontal departures from the underlying relationship. As λ approaches 0, the emphasis in on minimizing vertical departures and it can be shown that the structural relationship converges on a classical *Y* on *X* regression. As λ approaches ∞, the emphasis is on horizontal departures, and the structural relationship converges on a classical *X* on *Y* regression.

Several important differences between the structural relationship and standard linear regression are worth noting. First, the structural relationship is symmetrical, so it does not matter which variable is designated as being dependent or independent – the slopes will be the exact reciprocal of one another. Second, the structural relationship provides a pure indication of the position of the linear relation and is unaffected by its strength, and measurement error or noise can be added or removed from either variable without affecting the slope and intercept estimates (so long as the ratio of measurement errors, λ, is maintained or the parameter adjusted). Consequently, an intercept of, say, 0 and slope of 1 in a structural relation merely indicates there is no consistent bias between the two methods, but reveals nothing about their relative precision. This differs from standard linear regression, where adding measurement error or noise diminishes the slope. In essence, we have separated the issue of precision and accuracy, with the correlation coefficient (r^2) providing a pure measure of precision, and the parameters of the structural relationship a pure measure of the relative accuracy of the two methods. In standard regression analysis, these two concepts are inter-twined (testing Ho:β=0 in standard regression is equivalent to testing Ho: r^2=0). Thus, in addressing the basic question as to whether 35mm counts are as reliable as medium-format and ground drive-counts, we need to consider both the relative accuracy (intercept approaches 0 and slope approaches 1) and precision (r^2 approaches 1).

As would be expected for count data, variance tended to increase with the mean for all three counting methods. We therefore applied a variance-stabilzing transformation to counts, and determined the most appropriate transformation using Taylor's Power Law (Taylor 1961). In all cases, the exponent of the power relationship was greater than 1 and less than 2 (range 1.48 to 1.87), indicating that the most suitable transformation fell somewhere between a square root and logarithmic transformation. None of the exponents were significantly different from an intermediate value of 1.5, so we applied a fourth-root transformation, which has been found to be generally applicable to counts in other applications (Downing 1979; France et al. 1995).

RESULTS

Oblique 35mm Slides:

For the 1998 survey, the 35mm slides for all B.C. rookeries (except, due to time constraints, Cape St. James) were read by two experienced readers (Table 1; Table 2a). There was a high degree of correlation and very significant linear relationship between the two readers for both the original counts:

[7] $GE = 9.23 + 1.062 \times PO$ (r^2=0.9924; $F_{(1,7)}$=778.6; P<0.0001)

and the fourth-root transformed counts:

[8] $GE = 0.03 + 1.019 \times PO$ (r^2=0.9828; $F_{(1,7)}$=342.1; P<0.0001)

(Figure 1). Neither of the slopes differed significantly from one ($F_{(1,6)}$=2.67; P=0.1463 and $F_{(1,6)}$=0.12; P=0.7400 respectively) or intercepts from zero ($F_{(1,6)}$=0.93 P=0.3718 and $F_{(1,6)}$=0.03; P=0.8724 respectively). Both equations thus simplified to:

[9] $GE = PO$

which we regarded as indicating that the two reader's counts were statistically equivalent to one another. Where available, their average (*CV*=0.085) was used in subsequent comparisons.

The consistency of 35mm counts between the two readers could not be attributed to the fact the two readers were merely looking at the same images, or because the quality of images were particularly good for that survey. Several passes were made at most sites, and each counter usually selected a different set of slides to count (only 29 of the 198 slides counted for the comparison were selected by both counters). The 35mm survey results also appeared to be reproducible (Tables 2a and 2b). For the Scott Islands, which were surveyed on 22-June and again on 04-July in 1998, there was good agreement both between the raw counts:

[10] $PO(Jul4) = 14.53 + 0.931 \times PO(Jun22)$ (r^2=0.9945; $F_{(1,6)}$=910.0; P<0.0001)

and transformed counts:

[11] $PO(Jul4) = 0.253 + 0.937 \times PO(Jun22)$ (r^2=0.9935; $F_{(1,6)}$=764.2; P<0.0001)

(Figure 2). Similarly, there was good agreement between the Scott Island surveys replicated on 02-July and again on 06-July in 2002:

[12] $PO(Jul6) = -7.36 + 1.086 \times PO(Jul2)$ (r^2=0.9827; $F_{(1,6)}$=284.8; P<0.0001)

and transformed counts:

[13] $PO(Jul6) = -0.105 + 1.039 \times PO(Jul2)$ (r^2=0.9824; $F_{(1,6)}$=279.7; P<0.0001)

(Figure 2). For all the relationships, none of the intercepts were statistically different from zero, and none of the slopes significantly different from one ($F_{(1,5)}$=3.19 P=0.1342; $F_{(1,5)}$=5.02, P=0.0662; $F_{(1,5)}$=4.21, P=0.0954; $F_{(1,5)}$=3.43 P=0.1136; $F_{(1,5)}$=0.05 P=0.8299; $F_{(1,5)}$=1.83, P=0.2244; $F_{(1,5)}$=0.19, P=0.6828; $F_{(1,5)}$=0.40, P=0.5500 respectively), indicating that counts from each of the two duplicated surveys could be regarded as being statistically equivalent. The variability of the replicated surveys (*CV*=0.101) was only slightly greater than the variability among readers for the same survey.

Ground Drive-Counts:

Ground counts were conducted by various Alaska Department of Fish and Game biologists, all with extensive experience working with sea lions (Table 1). In total, there were three different teams of counters: *DC* and *DM* in 1994 and 1995, *DM/BP* and *US/DS* in 1997, and *BP* and *KP* in 1998 (Figure 3). There was a high degree of correlation and significant linear relationship between each pair of observers for both the original counts:

[14] $DC = 2.38 + 0.998 \times DM$ (r^2=0.9842; F=622.4; P<0.0001)

[15] $(DM/BP) = -51.11 + 1.222 \times (US/DS)$ (r^2=0.9966; F=1,481.2; P<0.0001)

[16] $BP = 1.39 + 0.994 \times KP$ (r^2=0.9997; F=8,699.0; P<0.0001)

and the transformed counts:

[17] $DC = -0.014 + 1.004 \times DM$ (r^2=0.9912 F=1,132.3; P<0.0001)

[18] $(DM/BP) = -0.414\ 1.110 \times (US/DS)$ (r^2=0.9921 F=629.4; P<0.0001)

[19] $BP = 0.068 + 0.986 \times KP$ (r^2=0.9995; F=5,863.2; P<0.0001)

For the first and third team of observers, the intercepts were not significantly different from zero on either the original and transformed scale ($F_{(1,8)}$=0.01; P=0.9197 and $F_{(1,8)}$=0.01; P=0.9202; $F_{(1,4)}$=0.04, P=0.8623 and $F_{(1,4)}$=1.31, P=0.3351) and the slopes not significantly different than one ($F_{(1,8)}$=0.003; P=0.9559 and $F_{(1,8)}$=0.02; P=0.8990; $F_{(1,4)}$=0.37, P=0.5775 and $F_{(1,4)}$=1.27, P=0.3236), indicating that the counts could be regarded as statistically equivalent to one another. For the second team of observers, however, the intercept was significantly greater than zero for the original counts ($F_{(1,5)}$=10.20, P=0.0242) and marginally significant for transformed counts ($F_{(1,5)}$=4.84, P=0.0791), and the slope significantly less than one for both the original counts ($F_{(1,5)}$=49.16, P=0.0004) and transformed counts ($F_{(1,5)}$=6.18, P=0.0474). The latter departure from equivalency was due almost entirely to an usually large discrepancy in one count at a particularly large site (1,383 versus 1,166 for North Rocks), which accounted for over 90% of the difference in the total count for that year. The ground counts by various observers can, with the odd exception, be regarded as being statistically equivalent to one another. Their average (*CV*=0.047) was used in subsequent comparisons.

Vertical Medium Format Images:

The medium-format images for the B.C. survey conducted in 1998 were initially counted by two readers. There was a high degree of correlation and significant relationship between the counts on both the original scale:

[20] $GS = 11.39 + 0.8195 \times PO$ (r^2=0.9958; F=1,674.3; P<0.0001)

and transformed scale:

[21] $GS = 0.253 + 0.9039 \times PO$ (r^2=0.9927; F=952.4; P<0.0001)

(Figure 4). However, the slope was significantly different from one for both the raw counts ($F_{(1,7)}$=82.0, P<0.0001) and transformed counts ($F_{(1,7)}$=10.8, P=0.0110). The intercept was not quite significantly different from zero for the raw counts ($F_{(1,8)}$=3.55; P=0.1016), but was marginally significant for the transformed counts ($F_{(1,8)}$=5.27, P=0.0554). The counts, which in total differed by about 15.4%, could thus not be regarded as being statistically equivalent. The discrepancy between the two readers inflated the variability (*CV*=0.094), which was less precise than replicate counts made both on the ground and from 35mm slides.

The discrepancy in counts between the two readers was not due to a few outliers; one reader (*GS*) consistently (8 of 10 sites) got lower counts than the other (*PO*). Nor could the disparity be attributed to the inherent variability in counting the medium format images. One of the readers (*PO*) counted each images twice (Figure 5), and obtained a high degree of consistency on both the original scale:

[22] $PO(2^{nd}) = -2.542 + 1.0037 \times PO(1^{st})$ (r^2=0.9982; F=8,647.4; P<0.0001)

and transformed scale:

[23] $PO(2^{nd}) = -0.055 + 1.0104 \times PO(1^{st})$ (r^2=0.9969; F=5,070.7; P<0.0001)

with none of the parameter estimates differing significantly from that expected under equivalency ($F_{(1,16)}$=2.83, P=0.1120; $F_{(1,16)}$=0.12, P=0.7361; $F_{(1,16)}$=1.60, P=0.2236; and $F_{(1,16)}$=0.54, P=0.4724). In an earlier assessment, the other reader (*GS*) had also double-counted medium format images for Alaskan rookeries, and reported a high degree of precision (*CV*=0.035 for *PO* on B.C. rookeries in 1998 and *CV*=0.052 for *GS* on Alaskan rookeries in 1997).

The disparity in pup counts between the two readers was not due to errors in matching sites or mismatching images. Counts of non-pups (Figure 6), which were generally very easy to discern in the images, were nearly identical for the two readers, both on the original scale:

[24] $GS(non\text{-}pups) = -1.21 + 0.991 \times PO(non\text{-}pups)$

(r^2=0.9983; F=5,896.1; P<0.0001)

and transformed scale:

[25] $GS(non\text{-}pups) = 0.054 + 0.986 \times PO(non\text{-}pups)$

(r^2=0.9988; F=8,133.1; P<0.0001)

and again none of the parameter estimates differed from that expected given equivalency ($F_{(1,10)}$=0.02, P=0.8800; $F_{(1,10)}$=0.47, P=0.5082; $F_{(1,10)}$=1.17, P=0.3054; and $F_{(1,10)}$=1.63 P=0.2279).

Given the disparity between the first two readers, we had the 1998 images re-examined by a third experienced reader (*CS*). The resulting counts were nearly identical to the higher of the initial two readers (*PO*) on both the original:

[26] $CS = -3.185 + 0.9938 \times PO$ (r^2=0.9977; F=2,591.1; P<0.0001)

and transformed scale:

$$[27]\quad CS = 0.085 + 0.9744 \times PO \qquad (r^2=0.9931;\ F=869.8;\ P<0.0001)$$

(Figure 4) and again none of the parameters differed from the expected values given equivalency ($F_{(1,10)}$=0.24, P=0.6432; $F_{(1,6)}$=0.10, P=0.7612; $F_{(1,6)}$=0.45, P=0.5289; and $F_{(1,6)}$=0.61 P=0.4615). In contrast, the third reader's counts once again tended be higher than the lower of the initial two readers (*GS*) on the original scale:

$$[28]\quad GS = 12.290 + 0.8362 \times CS \qquad (r^2=0.9955;\ F=1{,}337.6;\ P<0.0001)$$

and transformed scale:

$$[29]\quad GS = 0.141 + 0.9372 \times CS \qquad (r^2=0.9914;\ F=692.1;\ P<0.0001)$$

(Figure 4). As was the case in comparing *GS*'s counts to *PO*'s, the slope was significantly less than one for the raw ($F_{(1,6)}$=51.5, P=0.0002), but not quite significant for the transformed counts ($F_{(1,6)}$=3.14, P=0.1198). The intercepts were not significantly different from zero for the raw or transformed counts ($F_{(1,6)}$=2.70; P=0.1517); $F_{(1,6)}$=1.06, P=0.3437).

Given the close agreement between two of the three readers (*PO* and *CS*), both of which got counts that were significantly higher than the other reader (*GS*), we opted to discard the latter's counts, and used the average of the two high counts in subsequent analyses. This reduced the variability of the 1998 counts from medium-format images by about half, resulting in a CV=0.056.

For the 2002 survey, the medium-format images were counted by 3 experienced readers (2 experienced readers for two sites). In this case, there was a high degree of consistency among all three readers on both the original scale:

$$[30]\quad CS = 2.86 + 0.9716 \times LX \qquad (r^2=0.9992;\ F=9{,}011.9;\ P<0.0001)$$

$$[31]\quad CS = 5.23 + 0.9647 \times ML \qquad (r^2=0.9995;\ F=9{,}650.1;\ P<0.0001)$$

$$[32]\quad LX = -1.92 + 0.9973 \times ML \qquad (r^2=0.9991;\ F=5{,}304.5;\ P<0.0001)$$

and transformed scale:

$$[33]\quad CS = 0.119 + 0.9690 \times LX \qquad (r^2=0.9966;\ F=2{,}078.7;\ P<0.0001)$$

$$[34]\quad CS = 0.007 + 0.9934 \times ML \qquad (r^2=0.9984;\ F=3{,}184.8;\ P<0.0001)$$

$$[35]\quad LX = -0.183 + 1.0363 \times ML \qquad (r^2=0.9971;\ F=1{,}733.0;\ P<0.0001)$$

None of the parameter estimates differed significantly from the values expected given equivalency ($F_{(1,7)}$=0.26, P=0.6231; $F_{(1,8)}$=4.10, P=0.0775; $F_{(1,5)}$=0.75, P=0.4273; $F_{(1,5)}$=5.79, P=0.0528; $F_{(1,5)}$=0.05, P=0.8292; $F_{(1,5)}$=0.04, P=0.8483; $F_{(1,7)}$=1.91, P=0.2096; $F_{(1,7)}$=2.14, P=0.1815; $F_{(1,5)}$=0.01, P=0.9314; $F_{(1,5)}$=0.14, P=0.7211; $F_{(1,5)}$=2.99, P=0.1445; $F_{(1,5)}$=2.13, P=0.1948). The medium-format counts were therefore averaged in subsequent analyses. The level of variability for the 2002 counts (CV=0.063) was similar to the variability between the two high readers for the 1998 counts (*CV*=0.056), giving an overall *CV*=0.060 for replicate counts from the medium-format images.

35mm Slides versus Ground Counts:

In all four years that concurrent oblique 35mm aerial surveys and ground-drive counts were conducted at Forrester Island, there was a high degree of correlation and significant linear relation between the two techniques on both the original scale ($0.8475<r^2<0.9901$; $0.0033>P>0.0001$) and transformed scale ($0.7365<r^2<0.9830$; $0.0135<P<0.0001$) (Figure 7). An *ANOVA* indicated no significant year effect (F=0.088 and P=0.7694 for the raw counts, and F=0.87 and P=0.3592 for transformed counts), indicating that data for all years could be pooled. The resulting overall relationship was:

[36] $35mm = 35.0 + 0.7125 \times Ground$ (r^2=0.9107; F=254.9; P<0.0001)

for the raw counts. Since its intercept was not significantly different from zero ($F_{(1,25)}$=2.25, P=0.1460) but its slope significantly less than one ($F_{(1,25)}$=43.3, P<0.0001), the relationship was forced through the origin and simplified to:

[37] $35mm = 0.7970 \times Ground$ (r^2=0.9640; F=696.2; P<0.0001)

with the slope having a *SE*=0.029 (95% *CI* of 0.738-0.856).

For the transformed counts, the relationship was:

[38] $35mm = 0.20 + 0.9018 \times Ground$ (r^2=0.9176; F=278.5; P<0.0001)

and again the intercept was again not significantly different from zero ($F_{(1,25)}$=0.78, P=0.3847) but the slope was significantly different from one ($F_{(1,25)}$=21.2 P=0.0001), so the relationship was forced through the origin and simplified to:

[39] $35mm = 0.9494 \times Ground$ (r^2=0.9965; F=7,404.1; P<0.0001)

with *SE*=0.011 (95% *CI* of 0.927-0.972). Back-transformed to the original scale (by raising the slope to the power of 4) gives:

[40] $35mm = 0.8125 \times Ground$

(95% *CI* of 0.738-0.893). Both equations [37] and [40] indicate that 35mm slide counts are systematically biased and represent about 80% of ground counts. The bias appeared to be fairly consistent among sites, as indicated by the strong linear relationships between 35mm and ground counts on a site-by-site basis (r^2>0.90). The bias also appeared to be fairly constant between years, which for the four annual surveys ranged from 75.6-85.1% (mean 79.7%) (Table 3). We therefore conclude from this comparison that about 20% of pups counted during ground drive-counts at Forrester Island were missed in the 35mm slides, and that a correction of 1.24 (95% CI of 1.12-1.36 based on widest interval on raw and transformed scale) should be applied to adjust for this bias.

35mm Slides versus Medium-Format Images:

For B.C. rookeries, there was a high degree of correlation and significant linear relationships between 35mm slide counts and medium-format image counts on both the original scale:

[41] *35mm(BC)*= 12.78 + 0.9357 × *Medium(BC)*

(r^2=0.9954; F=3,237.9; P<0.0001)

and transformed scale:

[42] *35mm(BC)* = 0.2942 + 0.9289 × *Medium(BC)*

(r^2=0.9926; F=1,001.2; P<0.0001)

(Figure 8 – top panel). On the original scale, the intercept was not significantly from zero ($F_{(1,15)}$=2.93, P=0.1075), but the slope was significantly different than one ($F_{(1,15)}$=15.3, P=0.0012). We therefore forced the regression through the origin, simplifying it to:

[43] *35mm(BC)* = 0.9564 × *Medium* (BC)

(r^2=0.9973; F=5,884.8; P<0.0001)

which indicated that about 4.4% (*SE*=1.24%) fewer pups were evident in the 35mm slides compared to the medium-format images for B.C. rookeries, suggesting that a correction of 1.046 (95% CI of 1.018-1.075) should be applied to 35mm pup counts to account for missed animals.

The results were more complicated on the transformed scale, as both the intercept was significantly different than zero ($F_{(1,15)}$=6.47, P=0.0225) and the slope significantly different than one $F_{(1,15)}$=5.94, P=0.0268), suggesting that on transformed scale the bias increased with group size (the bias was negative up to 4.14 on transformed scale, which equated to 292 animals on the original scale). Based on the sizes of groups counted during the B.C. surveys, the medium-format counts would have been expected to be 1.010 times the 35mm count in 1998 and 1.044 times the 35mm count in 2002, both of which fell within the 95% confidence interval of the slope for the raw counts. The actual total medium-format pup count was 1.048 times the 35mm count in 1998, and 1.012 times the medium-format count in 2002, both of which again fell within the 95% confidence interval. Overall, these analysis indicate that for B.C. rookeries a small correction factor of 1.046 (95% CI of 1.018-1.075) applied to counts made from 35mm slides would account for missed pups.

For Forrester Island, Alaska, there was also a high degree of correlation and significant linear relationships between 35mm slide counts and medium-format image counts on both the original scale:

[44] *35mm(AK)* = 13.34 + 0.7614 × *Medium(AK)* (r^2=0.9325; F=165.8; P<0.0001)

and transformed scale:

[45] $35mm(AK) = 0.0213 + 0.9367 \times Medium(AK)$ (r^2=0.9740; F=448.9; P<0.0001)

(Figure 8 – bottom panel). On the original scale, the intercept was not significantly from zero ($F_{(1,12)}$=0.10, P=0.9248), but the slope was significantly different than one ($F_{(1,12)}$=17.0, P<0.0001). The regression was thus forced through the origin, reducing it to:

[46] $35mm(AK) = 0.7891 \times Medium(AK)$ (r^2=0.9809; F=666.7; P<0.0001)

with SE=0.044 and 95% CI of 0.693-0.885. On the transformed scale, the intercept was not significantly different from zero ($F_{(1,12)}$=0.01, P=0.9921), but when forced through the origin the slope was significantly different from one ($F_{(1,13)}$=37.2, P<0.0001). The regression thus simplified to:

[47] $35mm(AK) = 0.9508 \times Medium(AK)$ (r^2=0.9994; F=11,702; P<0.0001)

with SE=0.010, which back-transformed to the original scale (by raising to the power of 4), translates to:

[48] $35mm(AK) = 0.8172 \times Medium(AK)$

with 95% CI of 0.748-0.891. Both equations [46] and [48] indicate that about 20% fewer pups (21.1% for original counts and 18.3% for transformed counts) were evident in the 35mm slides compared to the medium-format images. This suggests a correction of 1.25 (95% CI of 1.130-1.442 for original counts and 1.122-1.337 for transformed counts) should be applied to 35mm pup counts made at Forrester Island to account for missed animals. The correction for missed pups on Forrester Island was significantly higher than the correction on B.C. rookeries for both the original and transformed counts. However, the corrections for Forrester Island based on a comparison of 35mm slides to medium-format images (1.22-1.27; equations [46] and [48]) were consistent with and not significantly different than the corrections for Forrester derived based on a comparison of 35mm slides to ground drive-counts (1.23-1.25; equations [37] and [40]).

For the Scott Islands in 2002, where the oblique 35mm and vertical medium-format surveys were flown within an hour of each another, we also compared non-pup counts. Unlike young pups which are confined to land, non-pups are constantly arriving at and departing from rookeries on feeding trips, so numbers fluctuate diurnally and between days, which normally complicates a direct comparison. There was very good agreement between the two methods on both the original scale:

[49] $35mm(Non\text{-}Pup) = 11.70 + 0.9847 \times Medium(Non\text{-}Pup)$

(r^2=0.9937; F=786.3; P<0.0001)

and transformed scale:

[50] $35mm(Non\text{-}Pup) = 0.215 + 0.9603 \times Medium(Non\text{-}Pup)$

(r^2=0.9953; F=1,055.2; P<0.0001)

(Figure 9). And none of the parameters were significantly different than would be expected given statistical equivalency ($F_{(1,5)}$=0.20, P=0.6702; $F_{(1,5)}$=0.19; P=0.6772; $F_{(1,5)}$=2.59; P=0.1687; $F_{(1,5)}$=1.82; P=0.2264). The total number of non-pups counted in the 35mm and medium-format images (3,787 and 3,762) differed by less than 1%.

Medium Format Images versus Ground Counts:

Although we made no direct comparison between medium-format and ground drive-counts as part of our assessment, Snyder et al. (2001) made such a comparison for Alaskan rookeries surveyed in 1997 and 1998. For sake of completeness and consistency, we re-analyzed the data given in their Table 1 (and incorporated a few additional counts made by *CS*) using the same statistical procedures employed for the other comparisons in this report.

There was very strong relationship between ground drive-counts and medium-format images on both the original scale:

[51] *Medium* = 9.34 + 0.9844 × *Ground* (r^2=0.9344; F=199.5; P<0.0001)

and following a fourth-root transformation:

[52] *Medium* = -0.090 + 1.0179 × *Ground* (r^2=0.9719; F=485.1; P<0.0001)

(Figure 10). None of the parameters differed significantly from those expected given statistical equivalency ($F_{(1,14)}$=0.05, P=0.8286; $F_{(1,14)}$=0.05, P=0.8202; $F_{(1,14)}$=0.16, P=0.6939; $F_{(1,14)}$=0.15, P=0.6999), so both simplified to:

[53] *Medium-format = Ground Count*

indicating the two methods provide counts that were statistically equivalent to one another. Although our study did not directly compare ground counts to medium-format counts, we did compare both methods to counts made from oblique 35mm slides at Forrester Island. The similarity of the correction factors derived by both methods (1.23-1.25 for ground drive-counts and 1.22-1.27 for counts from medium-format images) also suggests the two methods can be regarded as being statistically equivalent.

Digital 35mm Images versus 35mm Photographic Slides:

For the 2006 survey, all B.C. rookeries were photographed using both film and digital 35mm SLR cameras (Table 5). There was a high degree of correlation and very significant linear relationship between pup and non-pup counts made from film and digital images on both the original scale:

[54] *Pups(Digital)* = 7.2 + 0.990 × *Pups(Film)*

(r^2=0.9954; $F_{(1,14)}$=2,619; P<0.0001)

[55] Non-*Pups(Digital)* = 9.86 + 0.980 × *Non-Pups(Film)*

(r^2=0.9994; $F_{(1,17)}$=24,250; P<0.0001) and the fourth-root transformed counts:

[56] *Pups(Digital)* = -0.0025 + 0.9959 × *Pups(Film)*

(r^2=0.9942; $F_{(1,14)}$=2,053; P<0.0001)

[57] Non-*Pups(Digital)* = 0.016 + 0.9977 × *Non-Pups(Film)*

(r^2=0.9994; $F_{(1,17)}$=25,373; P<0.0001)

(Figure 11). For pups, neither of the slopes differed significantly from one ($F_{(1,13)}$=0.285; P=0.603 and $F_{(1,13)}$=0.038; P=0.849 respectively) or intercepts from zero ($F_{(1,13)}$=0.689 P=0.423 and $F_{(1,13)}$=0.19; P=0.672 respectively). Similarly, for non-pups, neither of the slopes differed significantly from one ($F_{(1,16)}$=0.47; P=0.503 and $F_{(1,16)}$=0.37; P=0.553 respectively) or intercepts from zero ($F_{(1,16)}$=0.112 P=0.742 and $F_{(1,16)}$=0.48; P=0.501 respectively). Both equations thus simplified to:

[58] *Pups(Digital) = Pups(Film)*

[59] *Non-Pups(Digital) = Non-Pups(Film)*

which indicated the counts from 35mm digital images and film were statistically equivalent to one another. Where available, their averages (*CV*=0.060 for pups and *CV*=0.012 for non-pups) were used in subsequent comparisons.

DISCUSSION

As expected, ground drive-counts provided the most precise index of pup production (*CV*=0.047), and for the most part the differences between various personnel conducting the counts were negligible. However, in one instance there was an appreciable discrepancy (16.6%) between observers on a particularly large rookery, which indicated that even this technique can occasionally be subject to considerable measurement error. Nevertheless, it remains the standard by which other techniques are evaluated.

In contrast to what has been reported by other investigators (Withrow 1982; Brown and Riemer 1992; Snyder 1998), we found that pup counts made from oblique 35mm slides provided a fairly precise index of pup production. The 35mm counts were fairly consistent when read independently by two observers (*CV*=0.085) and also in two cases where surveys were replicated 2-11 days later (*CV*=0.102). In contrast, Snyder (1998) concluded that 35mm photography was unsuitable for counting Steller sea lion pups. He described the ADF&G slides used in his evaluation as being “blurry, had low resolution, and their focus distorted due to angle of plane and the ground”. The photographs were apparently taken by an inexperienced photographer from a faster-moving twin-engine plane at oblique angles through a bubble window. Obviously there are many factors that can affect the quality of 35mm photographs that must be taken into account when evaluating survey procedures. It is therefore unlikely that our findings are applicable to 35mm photography in general, but rather apply to the specific procedures and photographers used in the B.C. surveys.

Precision of the medium-format pup counts varied. In an earlier assessment, Snyder et al. (2001) reported an overall level of precision for medium-format images (*CV*=0.048) that was comparable to ground counts (*CV*=0.047). One of us read the medium-format images taken during the survey in B.C. in 1998 twice, and the counts were highly reproducible (*CV*=0.035). Although the second reader did not conduct replicate counts for the B.C. survey, he did so for the 1997 and 1998 surveys of Alaskan rookeries and in both cases got very reproducible counts (*CV*=0.052 and *CV*=0.042 respectively). Surprisingly, however, for the B.C. survey in 1998, there was considerable variability in the counts between the initial two readers, with one getting consistently and significantly higher pup counts than the other, resulting in a loss of precision (*CV*=0.094). However, when the medium-format images were blindly counted by a third reader, the counts were almost identical to the original reader who tended to get the higher counts, but were once again consistently and significantly higher than the original reader who tended to get the lower counts. This led us to discard the lower of the two original counts, and use the average of the two higher counts, resulting in improved precision (*CV*=0.056). For the B.C. survey in 2002, there was good agreement between the three readers that counted the medium-format images, again resulting in good precision (*CV*=0.063). Not including the low counts that had been discarded, overall precision of the replicate pup counts from medium-format images for B.C. rookeries (*CV*=0.060), which was intermediate to that reported for ground drive-counts in Alaska (*CV*=0.047) and from 35mm oblique slides for B.C. rookeries (*CV*=0.085).

The reason for the discrepancy between the first two readers for the 1998 survey is unknown. In the opinion of the senior author, the medium-format images were generally very good quality and pups could be discerned with relatively little subjectivity. Snyder (1998) also reported a few relatively imprecise (*CV*>0.10) counts in his study, and suggested that lack of focus and fatigue may have contributed to the imprecision. Fatigue, which would likely lead to under-counting, may have been a factor in our study as well. For the reader that got the significantly lower counts that were ultimately discarded, the main focus was Alaskan rookeries, and the B.C. images were examined after several hundred Alaskan images had already been counted. While he double-counted all Alaskan sites, he did not have time to conduct a second read of the B.C. images, so it cannot be ascertained whether his precision may have waned. For the reader that tended to get higher counts, the main focus were the B.C. images, and they were counted first and in duplicate. This illustrates the importance of having independent replicate counts made of images, or at very least having duplicates made by the same counter.

One of our primary objectives was to evaluate the accuracy of pup counts from 35mm slides. While it was appeared that the 35 mm slide counts were relatively precise, and reproducible among different observers and between surveys conducted on different dates, comparison of 35mm counts with ground drive-counts at Forrester indicated quite clearly that they were biased. Overall, it appeared that about 20% of the pups seen on the ground were not evident in the oblique 35mm slides. The degree of bias seemed relatively consistent among sites (slope=0.797-0.813 with 95% CI of 0.738-0.893) and constant between the 4 years (mean 79.7%; range 75.6-85.1%). This suggests that a relatively precise and accurate estimate of pup abundance at Forrester Island can be obtained from 35mm counts by applying a correction factor of 1.24 (95% CI of 1.12-1.36) to account for missed pups.

In an earlier assessment at Alaskan rookeries, it had been concluded that pup counts made from medium-format images were statistically indistinguishable from ground drive-counts (Snyder 1998; Snyder et al. 2001), which was also apparent in our re-analysis of their data. This implies that medium-format photography can also be used to truth pup counts made from oblique 35mm slides. For Forrester Island, the results of such a comparison were consistent with those obtained in comparing pup counts made from 35mm slides for ground-drive counts. Once again, it appeared that about 20% of pups were missed in the oblique 35mm slides, suggesting that a relatively accurate estimate of pup abundance at Forrester Island can be obtained from 35mm counts by applying a correction factor of 1.25 (95% CI of 1.12 to 1.44) to account for missed pups.

For B.C. rookeries, pup counts made from oblique 35mm slides also tended to be significantly lower than those made from vertical medium-format images, but the magnitude of the difference was much less than at Forrester Island. Regression analysis indicated that on B.C. rookeries only about 4% fewer pups were evident in the 35mm slides compared to the medium-format images, suggesting that for B.C. rookeries a correction of 1.05 (95% CI of 1.018-1.075) should be applied to 35mm pup counts to account for missed animals.

Having been the only one involved in counting both the oblique 35mm slides and vertical medium-format images, it is the senior author's opinion that the two methods provide about the same level of resolution, and about the same degree of subjectivity was employed in identifying pups in the two types of images. The photographic systems both used a similar ED-based type of film, and they are flown at about the same altitude. While the medium-format photographs are taken along a transect from a faster-moving aircraft, the camera is programmed so that the film advances at about the same speed the plane is flying over ground. The oblique 35mm slides are taken from slower moving aircraft while they are circling, and distortion due to motion is not a factor for either type of survey. Thus, it would seem the main reason vertical medium-format images tend to give higher counts is that they are taken vertically as opposed to oblique angles. In counting the medium-format images, pup counts often increase by several percent when the image is superimposed on the preceding or proceeding frame, indicating that even the slight change in angle as the plane advances over the rookery can affect the number of pups that might be obscured. While it might be possible to reduce bias in 35mm pup counts by comparing numbers in adjacent frames, this is impractical because the angle and direction from which they're taken constantly changes as the plane circles.

The reason for the apparent discrepancy in the proportion of pups obscured in the 35mm slides between Forrester Island (20% missed) and B.C. rookeries (4% missed) is unknown. The difference could either imply that a greater proportion of pups are missed in the vertical medium-format images of B.C. rookeries, or alternatively that a lower proportion of pups are missed in the oblique 35mm slides of B.C. rookeries. We cannot think of any reason or offer any explanation as to why a higher proportion of pups might be missed at B.C. rookeries in the medium-format images. However, differences in the size and topography of rookeries could explain why fewer pups are missed in oblique 35mm slides. At Forrester Island, most rookeries are situated on large rocky islands with steep terrain. When photographing these sites (particularly North Rocks), it is often necessary to make broader circles, resulting in photographs being taken at more oblique angles (especially toward the inner portions of the rookery). Due to their size, there was also a tendency to zoom out to include a greater area in the frame, or at least a tendency to select such frames showing a greater area when piecing together the mosaic. Combined with the steep topography of these rocky sites, one might expect more pups to be hidden in crevices and behind outcropping in the oblique photographs. In contrast, the largest aggregations of pups in B.C. (SE Triangle Island) tend to be spread more linearly along flat rocky ledges or pebble beaches along the shore, where they can be photographed at steeper angles as the plane flies along the coast, and there are fewer obstacles to obscure pups. Although some of the rookeries in B.C. also occur on rocky sites with steep terrain (Sartine Island, North Danger Rocks), these tend to be smaller sites that can circled more tightly, resulting in photographs being taken at less oblique angles. Weather and lighting conditions may also play a factor. Forrester Island lies in an area more likely to be covered in low overcast cloud. Its also situated at one end of the end of the B.C. survey route, and thus tends to get surveyed early or late in the 10:00-18:00 counting window. In contrast, the major rookeries in B.C. at Cape St. James and especially on the Scott Islands tend to have clearer weather, and since they are toward the centre of the B.C. survey route, tend to be counted under more optimal lighting conditions near mid-day.

The general conclusion to be drawn from this study is that ground drive-counts, medium-format images, and 35mm slides all provided a relatively precise index of pup abundance. When comparing the various methods, correlation coefficients routinely accounted for in excess of 90-95% of the total variation, leaving less than 5-10% due to measurement error, which far exceeds the 70-75% generally considered as being required for making predictions.

There are numerous other factors that may also contribute to the inaccuracies in estimates of pup production, some of which were not included in our assessment. Although ground counts are generally considered to be the most accurate, the Forrester Island data indicate that even experienced observers may occasionally disagree by up to 15%. Moreover, since the two counters generally walk through the rookery together, some sources of error are not entirely independent, such as missing a portion of the site or judging the size or missing groups of pups that disperse into the water, and will not be captured in the *CV* calculations. One of the drawbacks of ground counts is that they involve travel by ship and thus tend to extend over a longer time-period that aerial surveys. For example, the ground pup counts in Alaska have occasionally begun as early as the third week of June (Sease et al., unpublished manuscript), at which point 10-30% of pups have yet to have been born (Bigg 1985), which could introduce a bias similar in magnitude to that inherent in oblique 35mm photographs.

One advantage of using a 35mm single-lens reflex camera compared to the belly-mounted medium-format camera is that the area being photographed can be seen through the lens. In contrast, the vertical images require that straight transects be flown while the plane is level, and an experienced operator is required to insure complete coverage. Even so, in several instances during the 1998 survey portions of B.C. rookeries were missed in the medium-format images, but which appeared in the 35mm slides. In one case, a small rock was missed completely, and in two other cases, parts of a beach and a small island were only partially photographed. Judging from the number of pups counted in these areas from the 35mm slides, the missed sites accounted for about 6% of the total number of pups on B.C. rookeries. It should be noted, however, that most of the error (67%) occurred on a beach at the base of a 700 foot cliff, characterized by strong downdrafts, which is extremely challenging to photograph from any type of aircraft, and the staff on medium-format plane were unfamiliar with B.C. rookeries. Although we were unable to directly check for this type of error during the 2002 survey (because a different set of readers read the 35mm and medium-format images), it is highly unlikely that any sites were missed, at least on the Scott Islands, as we got almost identical non-pup counts, which are very easy to distinguish in both types of photographs. The potential source of error could be circumvented by concurrently taking 35mm slides, or having someone with local knowledge onboard, or by flying overview transects at higher altitudes.

One major drawback of using 35mm slides was the effort required in piecing together the mosaic of many images required to obtained good pup counts. The process is laborious and prone to mistakes. For example, the senior author used 144 slides to count the pups in the 1998 survey, which represents an average of only 15 pups per slide. As the series of photographs is taken, the perspective is constantly changing as the aircraft circles, and the process subject to error, especially for the largest sites like those at Forrester Island that must be matched on all four sides. In assessing an unusually large discrepancy between the Forrester Island ground counts and 35mm slide counts, the author discovered that one site, representing 4% of the total pup count for Forrester Island that year, had inadvertently been counted twice from different perspectives in the 35mm slides. Such errors would normally go unnoticed. For very large sites such as North Rocks at Forrester Island, it becomes almost impossible to piece together a mosaic at maximum magnification of a 200mm zoom lens, and the tendency to pan out is probably one of the reasons the correction for missed pups tends to be greater at Forrester Island than B.C. This could have serious implications for using 35mm slides to monitor long-term trends if there are shifts in distribution or changes in the size of aggregations on rookeries. It would be difficult to discern between temporal changes in the proportion of pups being missed in 35mm slides from real changes in their abundance. In contrast, the vertical medium-format images are all taken from about the same vantage and the images cover a much larger area, which not only greatly reduces these types of errors and biases, but also provides much better information on the spatial distribution of pups.

Several advancements in digital photography have provided the opportunity to use digital photography instead of film. The *Nikon D200* camera used in this survey provided high-resolution (10.2 mega-pixel) images that provided resolution similar to high-speed film under good lighting conditions. Perhaps more importantly, the LCD sensor provided 12 bits of depth for each of the 3 colour channels (compared to 8 bits for JPEG compressed images), which greatly enhanced shadow detail and was especially important for distinguishing pups on dark substrates. The resulting RAW files were large (~16-17 Megabytes), but the write acceleration and large buffers in the camera and fast transfer rates provided by modern compact flash cards (>20 Megabytes per second) allowed shooting burst speeds of 22 frames in 5 seconds followed by one frame every 2 seconds. This was sufficient for photographing most of the rookeries in a single pass, although several of the largest sites required two passes. In the authors judgement, the digital images were comparable to the highest quality photographs that could be obtained using *Kodachrome* slide film. In support, the comparison of film and digital images for B.C. rookeries in 2006 indicated the counts were statistically equivalent (but still limited by the fact taken from oblique angles and pups obscured). The transition to digital 35mm photography will greatly simplify the handling and archiving of images, and its much easier to compare overlapping images on computer screens that allow multiple windows than it is by manually shifting between projected slides.

A source of error inherent to all techniques is missing pups that were born away from rookeries, or on rookeries not known to exist. However, breeding rookeries appear to be traditional and extremely stable. All known rookeries in B.C. have been known to exist since the first field studies were conducted in 1913, with no new rookeries having been established in recent years. This also seems to be the case between California and Oregon (there are no rookeries in Washington). However, in SE Alaska, the rookery at Forrester Island apparently first became established in the 1920s (Bigg 1985), and four additional rookeries have since been colonized at Hazy Island in 1980s, and White Sisters, Graves Rocks and Biali Rocks in the 1990s (Calkins et al. 1999; Pitcher et al. 2003). The entire coast is not searched during the B.C. sea lion surveys, and it is possible new rookeries may have gone unnoticed. Occasionally a few pups are born at non-breeding haulout sites, but these appear to account for less than 0.2% of total pup production (Olesiuk 2003), although it should be noted that haulout sites are generally not photographed in as much detail as rookeries, so the chances of underestimating or missing pups is greater. In SE Alaska, the new rookeries were formed at existing haulout sites, most if not all of which are checked during the B.C. surveys, so it is unlikely that new rookeries have been missed.

Another consideration in surveying pups is the impact of surveys on populations. Aerial surveys seemed to have no discernible impact, and we have never witnessed appreciable numbers of animals entering the water while circling (although this has occasionally been observed at haulout sites). In contrast, as the name implies, during ground drive-counts most non-pups are driven into the water, which could disrupt mother-pup bonds and territorial males. In the only study that has attempted to quantify the impact, Lewis (1987) reported that 30% fewer mothers maintained contact with pups (49% versus 71% respectively) in a year in which drive-counts were conducted compared to a control year with no drive count. Given the precarious status of the western stock of Steller sea lions, the potential impact of the repeated disturbance caused by ongoing pup surveys warrants closer examination.

Finally, we would remind readers that the focus of this study was to evaluate pup counts made from 35mm slides at rookeries in B.C. and neighbouring waters, and our conclusions and recommendations do not necessarily extend to other regions. The typography, substrate type and accessibility of rookeries, as well as local weather conditions, must be considered when planning surveys. For example, an important element in obtaining useable 35mm slides was conducting surveys in small single-engine aircraft that are slow and manoeuvrable that allowed steep bank turns. Such aircraft are simply not practical or safe in the more remote sections of the range of Steller sea lions. Indeed, prevailing weather conditions in some areas, such as persistent fog in the Aleutian Islands, may preclude aerial surveys altogether. Rookery size, topography and substrate type could also determine the best survey methodology. Most of the rookeries in B.C. and SE Alaska are situated on low-lying islands and islets, and can be approached by aircraft. The rookeries in B.C. and SE Alaska are generally characterized by flat rock shelves, or in the case of Triangle Island by a light grey pebble beach, which are ideal for counting pups from photographs. Even so, it appears the magnitude of bias in counts from 35mm slides varied with the size, topography or location of sites. Aerial photographs may be entirely inappropriate for some types of substrate, such as dark bolder beaches. It is likely that different survey techniques will be the most appropriate or practical in different regions over the expansive range of Steller sea lions, but insofar as possible we would encourage further development, comparison and calibration of techniques.

CONCLUSIONS & RECOMMENDATIONS

1. We conclude that pup counts made from oblique 35mm slides provide a good *relative index* of pup production when compared to ground drive-counts ($0.91<r^2<0.96$) and medium-format images ($0.93<r^2<0.99$). It is therefore recommended that DFO continue to use 35mm aerial photography to monitor trends in pup production at B.C. rookeries. Whenever possible, 35mm slides should be read independently by at least two different readers to provide a measure of the precision of counts. Survey procedures currently employed should not be modified, or any modifications should be assessed against existing procedures so that survey results can be calibrated. Furthermore, since the impact of ground drive-counts remains largely unknown, and since the major Steller sea lion rookeries in B.C. have been designated as Ecological Reserves or fall within a National Park Reserve where regular disturbances are discouraged, it is recommended that ground drive-counts or other activities causing similar levels of disturbance not be conducted or permitted.

2. Comparison of pup counts from oblique 35mm slides with ground drive-counts and vertical medium-format images at Forrester Island indicate that approximately 20% of pups are obscured in the oblique 35mm slides. The bias seems to be consistent among sites and between years. It is therefore recommended that a correction factor of 1.25 (95% CI of 1.12-1.44) be applied to account for pups obscured in the 35mm slides. Comparison of pup counts from oblique 35mm slides with vertical medium-format images on B.C. rookeries indicate that approximately 5% of pups are obscured in the oblique 35mm slides. Again the bias seems to be consistent among sites and between years. It is therefore recommended that a correction factor of 1.05 (95% CI of 1.018-1.075) be applied to account for pups that are obscured in the 35mm slides.

3. It is recommended that additional comparisons between oblique 35mm and vertical medium-format techniques be conducted on B.C. rookeries so that the results can be generalized beyond the two survey years examined to date, and to evaluate whether the correction factors might change with population status. In order to be feasible and cost-effective, such comparisons would have to be arranged to be done in collaboration with SWFSC and ADF&G.

4. The collection of 35mm slides taken during the province-wide over the past 31 years represents a valuable and irreplaceable set of reference material for training and calibrating counters, and it is recommended that the slides be properly archived.

5. Comparison of both pup and non-pup counts from oblique 35mm photographic slides and digital images at B.C. rookeries during the 2006 survey indicate the two methods provided counts that were statistically indistinguishable. Since digital images are more cost-effective, easier to handle and archive, and simplify the counting process, it is recommended that future surveys be conducted using digital photography.

ACKNOWLEDGEMENTS

Many experienced and dedicated people helped in collecting the data used in this analyses. We are especially grateful to the highly skilled pilots who flew safely in some of the remotest parts of the BC coast: Jim Kirk and Louis Rouleau flew the 35mm surveys in a *DeHavilland Beaver*, and Dave Weintraub, Ralph Aiken, Jr., and the late Tom Blaesing flew the medium-format surveys in an *Aero Commander*. We owe our gratitude to all those who helped us with counts and surveys, who collectively represent nearly two centuries of experience studying Steller sea lions: Graeme Ellis, Dennis McAllister, Boyd Porter, Gary Snyder, Denise Spraker and Una Swain. Funding for the 35mm surveys in B.C. and at Forrester Island was provided by the Department of Fisheries and Oceans, and funding for the medium-format surveys in Alaska was provided by the Alaska Department of Fish and Game. We owe our appreciation to Brian Reader of National Parks Service (Gwaii Haanas), Doug Bifford of Ecological Reserves, and the Southwest Fisheries Science Centre for contributing funding towards the medium-format survey of B.C. rookeries, and to Tom Loughlin for providing travel funds to La Jolla for one of us (PFO) to count the medium-format images. Dennis McAllister and Boyd Porter helped compile data for the Forrester Island comparison.

LITERATURE CITED

Baker, C. S., T. R. Loughlin, V. Burkanov, C. W. Matson, R. G. Trujillo, D. G. Calkins, J. K. Wickliffe, J. W. Bickham. 2005. Variation of mitochondrial control region sequences of Steller sea lions: the three-stock hypothesis. Journal of Mammology (86)6: 1075-1084.

Berkson, J.M. and D.P. DeMaster. 1985. Use of pup counts in indexing population changes in pinnipeds. Can. J. Fish. Aquat. Sci. 42: 873-879.

Bickham, J.W., J.C. Patton, and T.R. Loughlin. 1996. High variability for control-region sequences in a marine mammal – implications for conservation and biogeography of Steller sea lion (*Eumetopias jubatus*). Journal of Mammalogy 77: 95-108.

Bigg, M.A. 1985. Status of the Steller sea lion (*Eumetopias jubatus*) and California sea lion (*Zalophus californianus*) in British Columbia. Can. Spec. Publ. Fish. Aquat. Sci. No. 77. 20pp.

Braham, H.W., R.D. Everitt, and D. Rugh. 1980. Northern sea lion population decline in the eastern Aleutian Islands. J. Wildl. Manage. 44: 35-33.

Brown, R.F and S.D. Reimer. 1992. Steller sea lion counts in Oregon during June and July, 1975-1991. In-house Report, Nongame Wildlife Program, Oregon Department Fish and Wildlife, Newport, Oregon, 97365. 12pp.

Calkins, D.G., D.C. McAllister, K.W. Pitcher and G.W. Pendleton. 1999. Steller sea lion status and trend in Southeast Alaska: 1979-1997. Mar. Mamm. Sci. 152: 462-477.

Calkins, D.G. and K.W. Pitcher. 1982. Population assessment, ecology and trophic relationships of Steller sea lions in the Gulf of Alaska. Pages 447-546 *in* Environmental Assessment of the Alaskan Continental Shelf, U.S. Department Commerce and U.S. Department of Interior, Final Reports of Principal Investigators, Vol. 19.

Downing, J. A. 1979. Aggregation, transformation, and the design of benthos sampling programs. J. Fish. Res. Board Can. 36: 1454-1463.

Edie, A.G. 1977. Distribution and movements of Steller sea lion cows (*Eumetopias jubatus*) on a pupping colony. MSc. Thesis. University of British Columbia, Vancouver, British Columbia. 81pp.

France, R., N. Yan, P. Olesiuk, R. Nero, P.A. Del Giorgio, and W. Keller. 1995. Secondary analysis of relationships between pelagic invertebrate predators and phytoplankton abundance and water clarity. Fresh. Biol. 34: 255-261.

Harvey, J. T. 1987. Population dynamics, annual food consumption, movements and dive behaviors of harbor seals, *Phoca vitulina richardsi*, in Oregon. Unpubl. PhD. Thesis, Oregon State Univ. 177 pp.

Huber, H. R. 1995. The abundance of harbor seals (*Phoca vitulina richardsi*) in Washington, 1991-1993. MSc. Thesis, University of Washington, Seattle, Washington. 37 pp.

Kendall, M. and A. Stuart. 1967. The advanced theory of statistics. Vol. 2 Inference and relationship. MacMilland Publishing, New York.

Lewis, J.P. 1987. An evaluation of a census-related disturbance of Steller sea lions. M.Sc. Thesis, University of Alaska, Fairbanks, Alaska. 89pp.

Loughlin, T.R. 1997. Using the phylogeographic method to identify Steller sea lion stocks. Pages 159-171 In: A.E. Dizon, S.J. Chivers, and W.F. Perrin, editors. Molecular Genetics of Marine Mammals. Special Publication #3 of the Society for Marine Mammalogy.

Loughlin, T.R. 1998. The Steller sea lion: a declining species. Biosphere Conserv. 1(2): 91-98.

Loughlin, T.R., A.S. Perlov, and V.A. Vladimirov. 1992. Range-wide survey and estimation of total number of Steller sea lions in 1989. Marine Mammal Science 8: 220-239.

Lowry, M.S. 1999. Counts of California sea lion (*Zalophus californianus*) pups from aerial color photographs and from the ground: A comparison of two methods. Mar. Mamm. Sci.15: 143-158.

Lowry, M.S. and W.L. Perryman. 1992. Aerial photographic census of California sea lion pups at San Miguel Island, California for 1987-1990 and San Nicholas Island, California for 1990. Administrative Report LJ-92-19. Department of Commerce, National Marine Fisheries Service, Southwest Fisheries Science Center, La Jolla. CA 19pp.

Lowry, M.S., W.L. Perryman, M.S. Lynn, R.L. Westlake and F. Julian. 1996. Counts of northern elephant seals, *Mirounga angustirostris*, from large-format aerial photographs taken at rookeries in southern California during the breeding season. Fish. Bull. U.S. 94: 176-185.

Mathisen, O.A. and R.J. Lopp. 1963. Photographic census of the Steller sea lion herds in Alaska, 1956-58. U.S. Fish Wildl. Serv., Spec. Sci. Rep. Fish. No. 424, 20pp.

Merrick, R.L., T.R. Loughlin. 1997. Foraging behavior of adult female and young of the year Steller sea lions in Alaskan waters. Can. J. Zool. 776-786.

Merrick, R.L., T.R. Loughlin and D.G. Calkins. 1987. Decline in the abundance of the northern sea lion, *Eumetopias jubatus*, in Alaska, 1956-86. Fish. Bull. 85:351-365.

NMFS (National Marine Fisheries Service). 2007. Draft revised Steller sea lion recovery plan, eastern and western distinct population segments (*Eumetopias jubatus*). National Marine Fisheries Service, Silver Spring, MD. 305 p.

Olesiuk, P. F. 1999. An assessment of the status of harbour seals (*Phoca vitulina*) in British Columbia. Department of Fisheries and Oceans, Canadian Stock Assessment Secretariat, Res. Doc. 99/33.

Olesiuk, P. F. 2007. Recent trends in the abundance of Steller sea lions (*Eumetopias jubatus*) in British Columbia. Department of Fisheries and Oceans, Canadian Stock Assessment Secretariat, Res. Doc. 2008/63.

Pitcher, K.W., D.G. Calkins, B.J. LeBouef, R.L. Merrick, and G.W. Pendleton. Spatial and temporal variation in the timing of births of Steller sea lions. (Unpublished Manuscript)

Pitcher, K. W., P. F. Olesiuk, R. F. Brown, M. S. Lowry, S. J. Jeffries, J. L. Sease, W. L. Perryman, C. E. Stinchcomb, and L. F. Lowry. 2007. Abundance and distribution of the eastern North Pacific Steller sea lion (*Eumetopias jubatus*) population. Fish. Bull. 107: 102-115.

Sandegren, F.W. 1970. Breeding and maternal behaviour of the Steller sea *lion (Eumetopias jubatus*) in Alaska. MSc. Thesis, University of Alaska, Fairbanks, Alaska.

Sease, J.L., R.F. Brown, V.N. Burkanov, D.G. Calkins, P.F. Olesiuk and A.E. York. Range-wide survey of Steller sea lions in 1994. NMML, Unpublished Report.

Snyder, G.M. 1998. An evaluation of alternative methods for counting Steller sea lion pups. MSc. Thesis, University of Alaska, Anchorage, Alaska. 70pp.

Snyder, G.M., K.W. Pitcher, W.L. Perryman and M.S. Lynn. 2001. Counting Steller sea lion pups in Alaska: an evaluation of medium-format, color aerial photography. Mar. Mamm. Sci. 17(1): 136-146.

Sokal, R.R. and C.A. Braumann. 1980. Significance tests for coefficients of variation and variability profiles. Systematics Zoology 29: 50-66.

Sprent, P. 1969. Models in regression and related topics. Methuen & Co.

Swain, U.G. 1996. Foraging behaviour of female Steller sea lions in Southeast Alaska and the eastern Gulf of Alaska. Pp. 135-166 In: Steller sea lion recovery Investigations in Alaska, 1992-1994. Alaska Dept. Fish and Game. Wildl. Tech. Bull. No. 13.

Taylor, L. R. 1961. Aggregation, variance, and the mean. Nature (Lond.) 189: 732-735.

Trites, A.W. and P.A. Larkin. 1996. Changes in abundance of Steller sea lions (*Eumetopias jubatus*) in Alaska from 1956 to 1992: how many were there? Aquatic Mammals 22(3): 153-166.

Westlake, R.L., W.L. Perryman, and K.A. Ono. 1997. Comparison of vertical aerial photographic and ground censuses of Steller sea lions at Ano Nuevo Island, July 1990-1993. Mar. Mamm. Sci. 13: 207-218.

Withrow, D.E. 1982. Using aerial surveys, ground truth methodology, and haulout behavior to census Steller sea lions, *Eumetopias jubatus*. MSc. Thesis. University of Washington, Seattle, Washington. 102pp.

Withrow, D.E. and T.R. Loughlin. 1997. A correction factor estimate for the proportion of harbor seals missed on sand bar haulouts during molt census surveys in 1996 near Cordova, Alaska. Annual report to the MMPA Assessment Program, Office of Protected Resources, NMFS, NOAA, Silver Spring, MD.

Table 1. Summary of personnel that participated in ground drive-counts or reading photographic images.

Initials:	Counter:	Affiliation:
PO	Peter F. Olesiuk	Department of Fisheries and Oceans
GE	Graeme M. Ellis	Department of Fisheries and Oceans
DC	Donald G. Calkins	Alaska Department of Fish and Game
KP	Kenneth W. Pitcher	Alaska Department of Fish and Game
DM	Dennis C. McAllister	Alaska Department of Fish and Game
BP	Boyd Porter	Alaska Department of Fish and Game
US	Una G. Swain	Alaska Department of Fish and Game
DS	Denise Spraker	Alaska Department of Fish and Game
GS	Gary M. Snyder	Alaska Department of Fish and Game
ML	Morgan Lynn	Southwest Fisheries Science Centre
LX	Lisa (get last name)	Southwest Fisheries Science Centre
CS	Charlie Stinchcomb	Southwest Fisheries Science Centre

Table 2a. Counts used for the comparisons between oblique 35mm slides and vertical medium-format images for B.C rookeries in 1998. For the Scott Islands, the 35mm survey was conducted on 22-June survey (see Table 4), and for other sites on 3-4 July. Numbers in parentheses indicate the estimated number of animals missed in the medium-format survey.

Site	**35mm Slides**			**Medium-Format Images**			
British Columbia – 1998	***PO***	***GE***	**Mean:**	***GS***	***CS***	***PO***	**Mean:**
Beresford Island	0	0	0	0	0	0	0
Maggot Island	68	67	67.5	71	70	76.5	73.8
Sartine Island	143	157	150.0	132	129	129.0	130.5
Triangle Island - Site A	0	0	0	0	0	0	0
Triangle Island - Site B	54	56	55.0	36(+10)	43 (+10)	33.5 (+10)	34.8 (+10)
Triangle Island - Site C	156	190	173.0	416(+80)	487 (+80)	486.0 (+80)	451 (+80)
Triangle Island – Site D	359	375	367.0				
Triangle Island - Site E	5	3	4.0	5	6	7.5	6.3
Triangle Island - Site F	111	157	134.0	117	143	154.5	135.8
Triangle Island - Site G	543	589	566.0	485(+30)	571 (+30)	587.5(+30)	536.3 (+30)
Cape St. James – North	93	(93)	93.0	95	484	98.7	96.9
Cape St. James – South	395	(395)	395.0	325		380.5	352.8
North Danger Reef	144	158	151.0	183	179	198.5	190.8
Missed in medium format	-	-	-	120	120	120	120
Total (CV)	**2071**	**2240**	**2155.5 (0.085)**	**1985**	**2232**	**2272.2 (0.035)**	**2128.6 (0.092)[1] (0.056)[2]**

[1]Based on variability among all 3 readers.
[2]Based on variability among the 2 higher readers (*PO* and *CS*) after the significantly lower reader (*GS*) had been discarded.

Table 2b. Counts used for the comparisons between oblique 35mm slides and vertical medium-format images for B.C rookeries in 2002.

Site	35mm Slides			Medium-Format Images			
British Columbia – 2002	***2-3July***	***6-July***	**Mean:**	***CS***	***LX***	***ML***	**Mean:**
Beresford Island	0	2	1.0	0	0	0	0.0
Maggot Island	77	75	76.0	67	71	-	69.0
Sartine Island	132	160	146.0	137	150	-	143.5
Triangle Island - Site A	0	0	0	0	0	0	0.0
Triangle Island - Site B	50	69	59.5	55	58	63	58.7
Triangle Island - Site C & D	827	774	800.5	857	860	879	865.3
Triangle Island - Site E	42	27	34.5	20	13	20	17.7
Triangle Island - Site F	180	193	186.5	128	126	129	127.7
Triangle Island - Site G	1037	1198	1117.5	1139	1189	1179	1169.0
Cape St. James – North	173	-	173.0	646	650	668	654.7
Cape St. James – South	462	-	462.0				
North Danger Reef	219	-	219.0	213	214	194	207.0
Total (CV)	**3199**	**3352**	**3290.5 (0.102)**[1]	**3262**	**3331**	**3344.5**	**3312.5 (0.063)**

[1]Includes data from 35mm surveys replicated at the Scott Islands on 22 June and 4 July, 1998 (see Table 4).

Table 2c. Counts used for the comparisons between oblique 35mm slides and vertical medium-format images for Forrester Island in 1997, 1998 and 2002. In 1997 and 2002 the mean for the medium-format images are based on 3-5 replicate reads by several experienced readers , and in 1998 they are based on duplicate reads by *GMS*.

Site	**35mm Slides**	**Medium-Format Images**				
		Replicate Counts				**Mean**
Forrester Island – 1997						
	PO	***CS***	***ML***	***GS***	***WP***	
Lowrie Island	883	803	1016	880	752	862.8
North Rocks	729	1022	1147	1111	1107	1096.8
East Rocks	78	95	98	100	92	96.3
Cape Horn Rocks	232	392	377	701	689	695.0
Sea Lion Rocks	211	331	290			
Total	**2133**					**2750.9**
Forrester Island – 1998						
	PO	***GS – 1st***		***GS – 2nd***		**Mean**
Lowrie Island	845	1044		1111		1077.5
North Rocks	872	1130		1114		1122.0
East Rocks	44	68		61		64.5
Cape Horn Rocks	258	331		335		333.0
Sea Lion Rocks	345	421		420		420.5
Total	**2364**	**2994**		**3041**		**3017.5**
Forrester Island – 2002						
	PO	***CS***	***LX***	***ML***		**Mean**
Lowrie Island	842	1121	1052			1086.5
North Rocks	739	1058	1001			1029.5
East Rocks	111	133	138			135.5
Cape Horn Rocks	342	386	372	384		380.7
Sea Lion Rocks	364	430	430	411		423.7
Total (CV; 1997-2002 combined)	**2398**	**3128**	**2993**	**-**		**3055.9 (0.049)**

Table 3. Counts used for the comparisons between oblique 35mm slides and ground drive-counts at Forrester Island.

Site	35mm Slides	Ground Drive Counts		
Forrester Island – 1994	***PO***	***DC***	***DM***	**Mean**
Lowrie Island	660	862	841	851.5
North Rocks	849	960	1074	1017.0
East Rocks	67	124	139	131.5
Cape Horn Rocks	239	392	383	387.5
Sea Lion Rocks	258	349	353	351.0
Total	**2073**	**2687**	**2790**	**2738.5**
Forrester Island – 1995	***PO***	***DC***	***DM***	**Mean**
Lowrie Island	723	(896)	(896)	896.0
North Rocks	843	1232	1147	1189.5
East Rocks	88	(108)	(108)	108.0
Cape Horn Rocks	291	296	290	293.0
Sea Lion Rocks	413	396	368	382.0
Total (CV)	**2358**	**2928**	**2809**	**2868.5**
Forrester Island – 1997	***PO***	***BP/DM***	***DS/US***	**Mean**
Lowrie Island	883	714	667	690.5
North Rocks	729	1383	1166	1274.5
East Rocks	78	131	147	139.0
Cape Horn Rocks	232	322	333	327.5
Sea Lion Rocks	211	369	368	368.5
Total	**2133**	**2919**	**2681**	**2800.0**
Forrester Island – 1998	***PO***	***BP***	***KP***	**Mean**
Lowrie Island	845	1079	1083	1081.0
North Rocks	872	995	999	997.0
East Rocks	44	60	57	58.5
Cape Horn Rocks	258	265	256	260.5
Sea Lion Rocks	345	374	389	381.5
Total **(CV)**	**2364**	**2773**	**2784**	**2778.5** **(0.047)**

Table 4. Counts used for comparison of 35mm pup counts for the Scott Islands on 22-June-98 and 11 days later on 04-July-98. The very close correspondence is likely somewhat fortuitous; it would have been expected that the second count would have been about 10% higher due to the number of births expected between the surveys based on the timing of pupping given in Edie (1977).

Site	22-June-98	04-July-98	Mean:
Beresford Island	0	0	0
Maggot Island	68	72	70.0
Sartine Island	143	148	145.5
Triangle Island - Site A	0	0	0
Triangle Island - Site B	54	57	55.5
Triangle Island - Site C	156	171	163.5
Triangle Island - Site D	359	328	343.5
Triangle Island - Site E& F	116	137	126.5
Triangle Island - Site G	543	528	535.5
Total **(CV)**	**1439**	**1441**	**1440** **(0.059)**

Table 5. Counts used for the comparisons between oblique 35mm photographic slides and 35mm oblique digital images for B.C rookeries in 2006.

Site	**Pup Counts**			**Non-Pup Counts**		
British Columbia – 2006	***Film***	***Digital***	**Mean:**	***Film***	***Digital***	**Mean:**
Maggot Island	59	65	62.0	604	601	602.5
Sartine Island	171	185	178.0	374	385	379.5
Triangle Island – Site A1	116	121	118.5	161	161	161.0
Triangle Island – Site A2	205	212	208.5	243	241	242.0
Triangle Island – Site B	0	0	0	159	156	157.5
Triangle Island – Site C	27	29	28.0	106	106	106.0
Triangle Island – Site D	134	118	126.0	300	312	306.0
Triangle Island – Site E	1398	1365	1381.5	1687	1654	1670.5
Triangle Island – Site F	534	560	547.0	540	530	535.0
Triangle Island – Site G	258	271	264.5	365	366	365.5
Triangle Island – Misc.	0	0	0	31	32	31.5
Cape St. James – North	538	603	570.5	618	632	625.0
Cape St. James – South	162	143	152.5	463	474	468.5
North Danger Reef – North	330	304	317.0	635	629	632.0
North Danger Reef - South	79	93	86.0	366	377	371.5
Virgin Rocks	50	52	51.0	509	523	516.0
Total **(CV)**	**4061**	**4121**	**4091.0** **(0.060)**	**7161**	**7179**	**7170.0** **(0.012)**

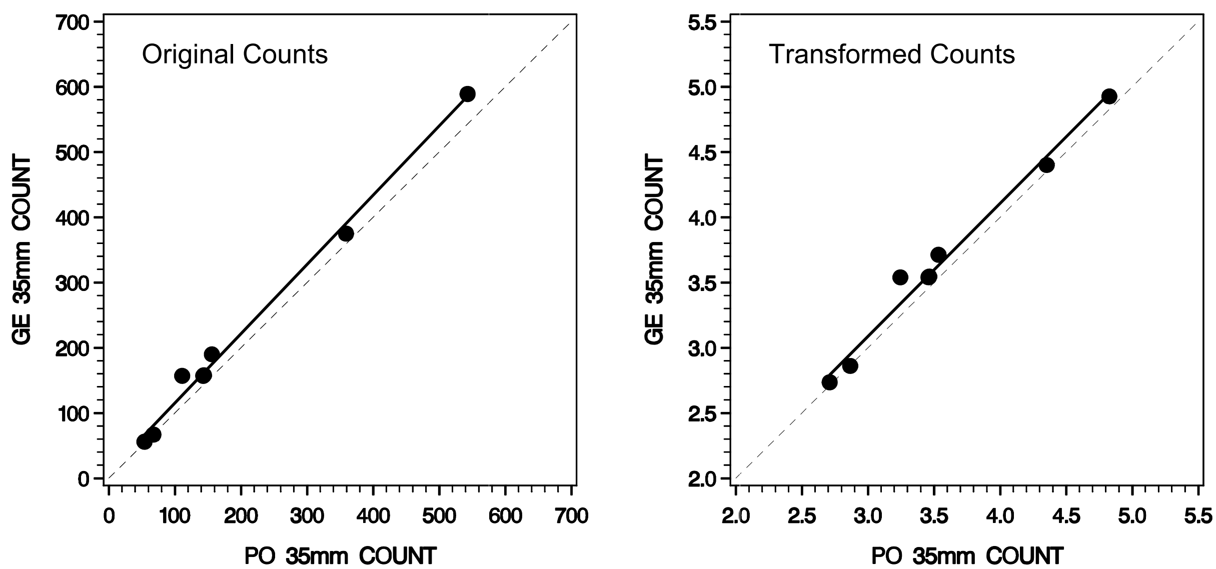

Figure 1. Comparison of 35mm slide counts for British Columbia rookeries in 1998 by two readers (*GE* and *PO*). The solid line represents the best fitting functional relationship, and the dashed line denotes 1:1 agreement.

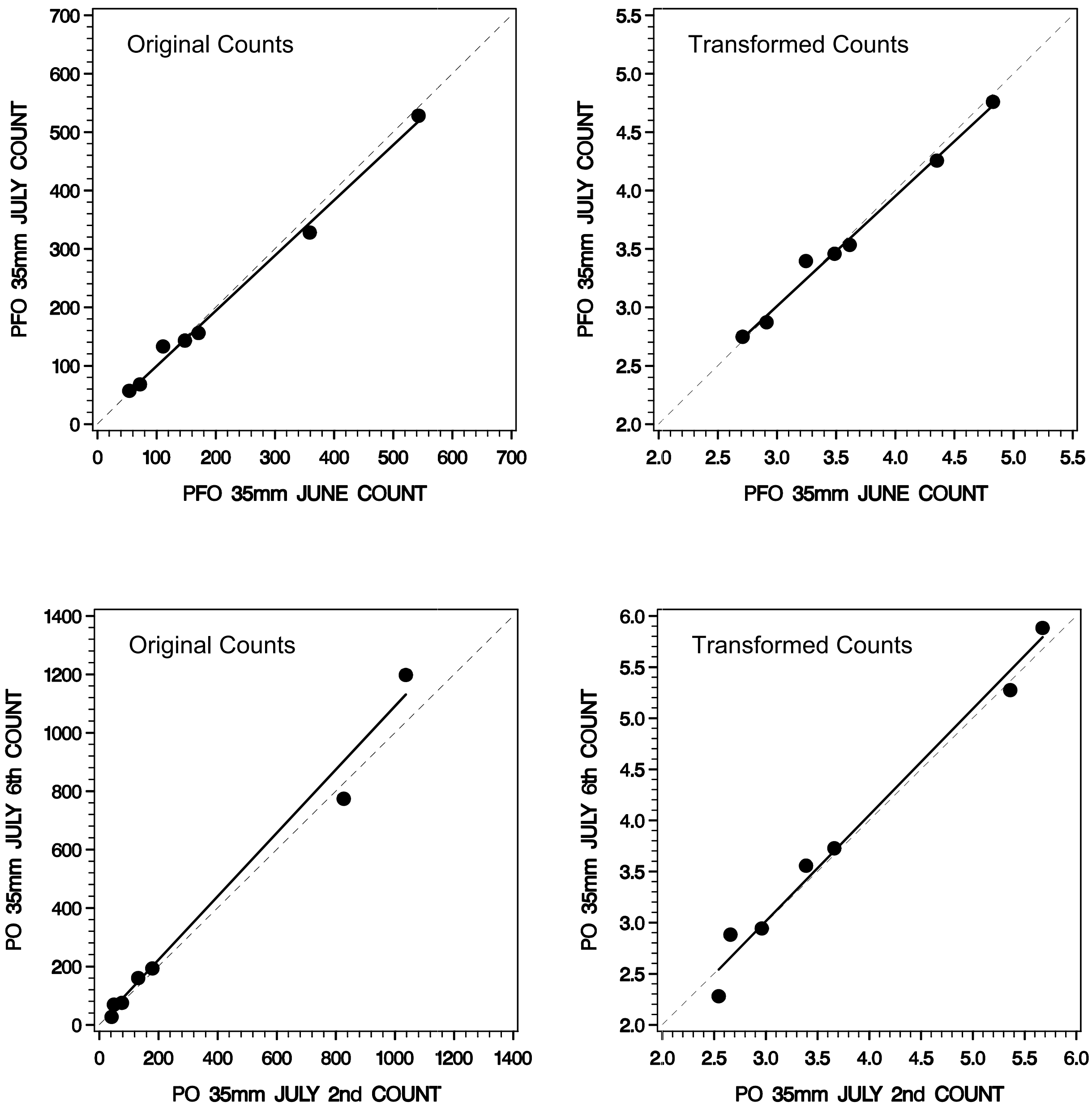

Figure 2. Comparison of 35mm slide counts for Scott Island survey conducted on 23-June-98 and replicated on 04-July-98 (top panel) and on 02-July-02 and 06-July-02 (bottom panel) . The solid lines represent the best fitting structural relationships, and the dashed lines denotes 1:1 agreement.

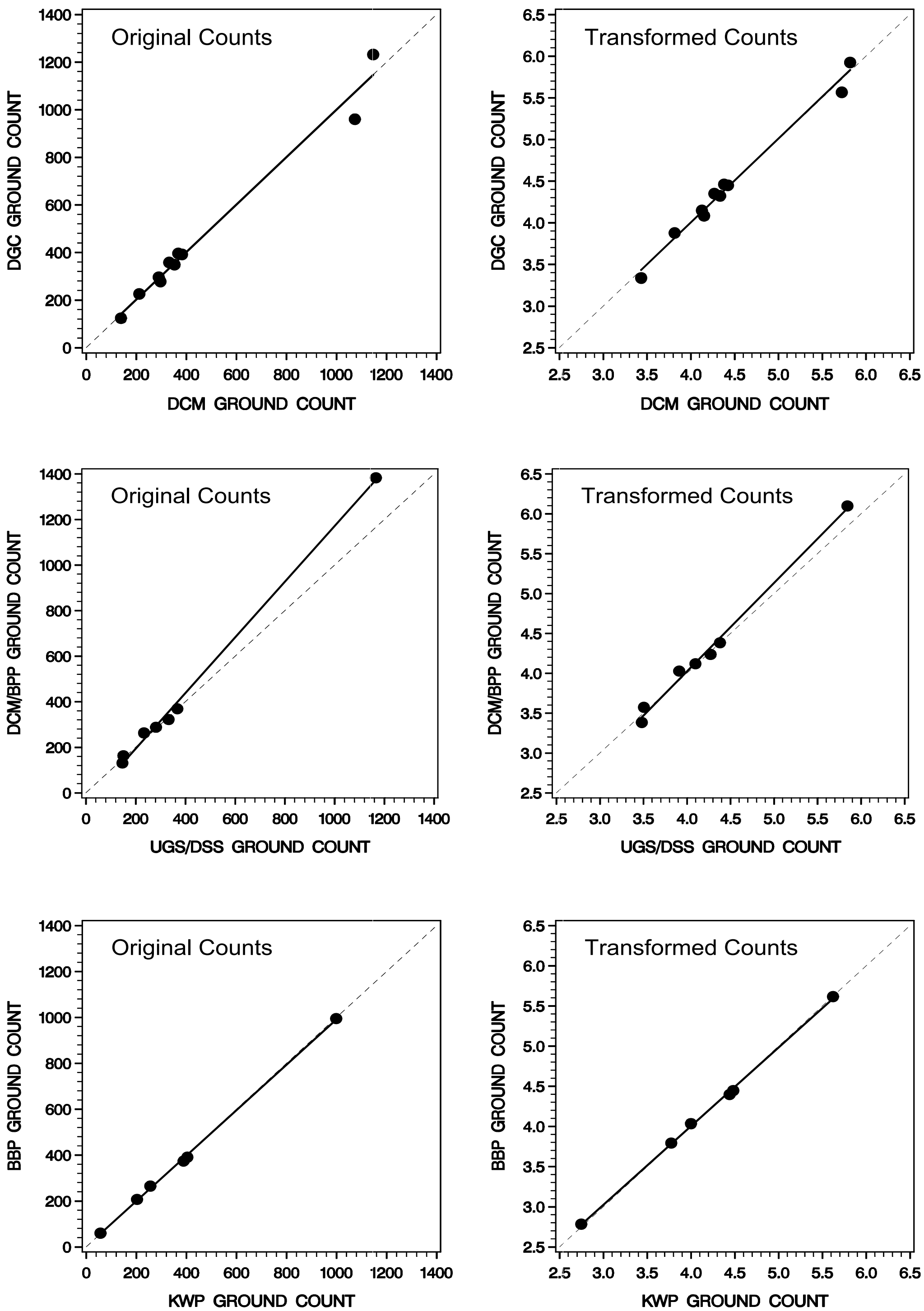

Figure 3. Comparison of ground drive-counts at Forester Island for each team of observers. The solid lines represent the best fitting structural relationship, and the dashed line denotes 1:1 agreement.

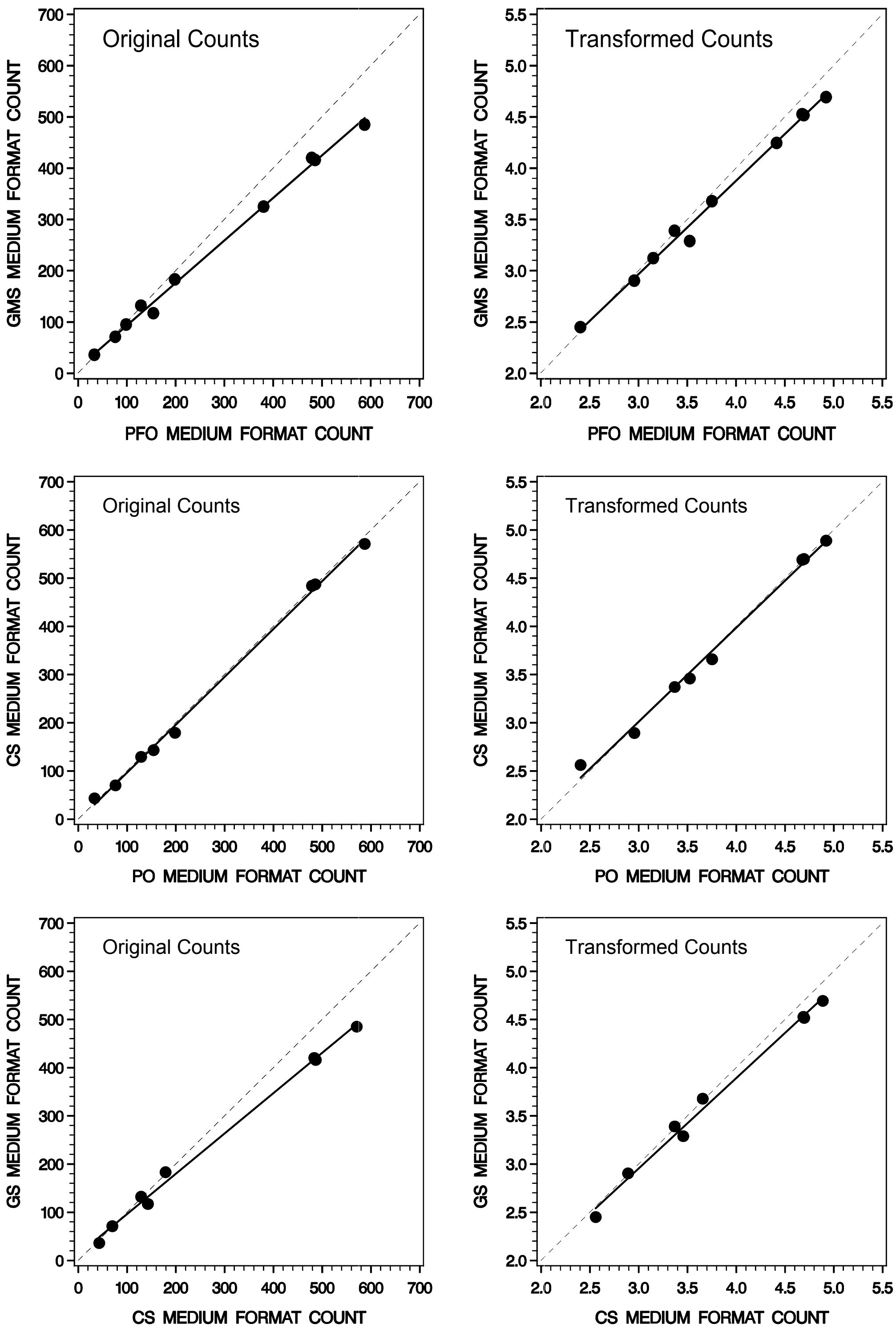

Figure 4. Comparison of medium-format images for British Columbia rookeries in 1998 read independently by three readers (*PO, GS and CS*). The solid lines represent the best fitting structural relationship, and the dashed line the 1:1 agreement.

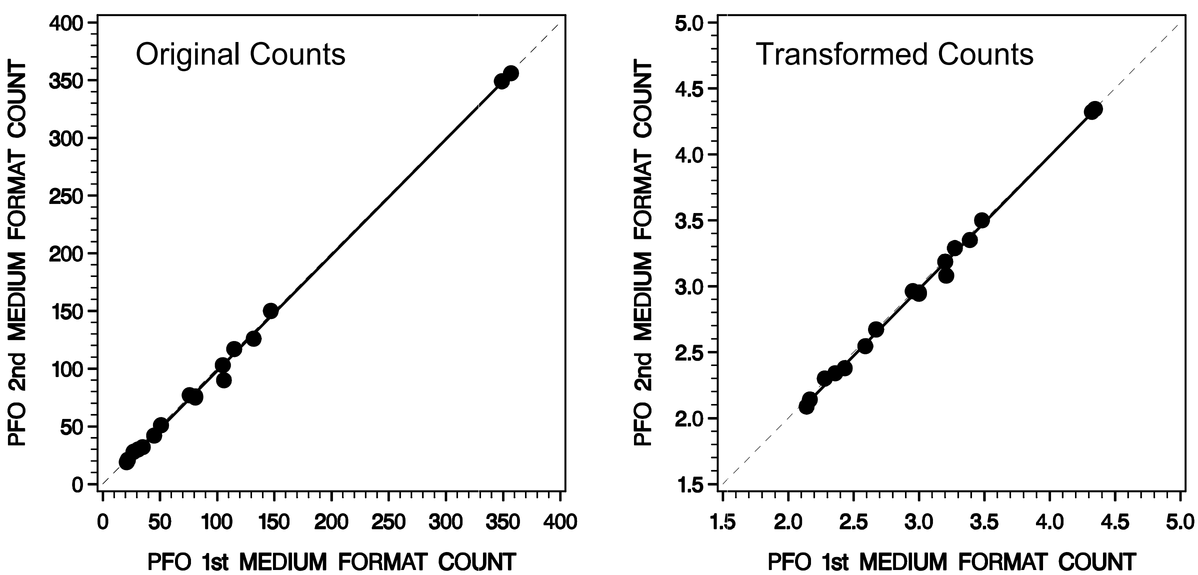

Figure 5. Comparison of 1st and 2nd reads of medium-format images by one reader (*PFO*) for British Columbia rookeries in 1998. The solid lines represent the best fitting linear relationships, and the dashed line 1:1 agreement.

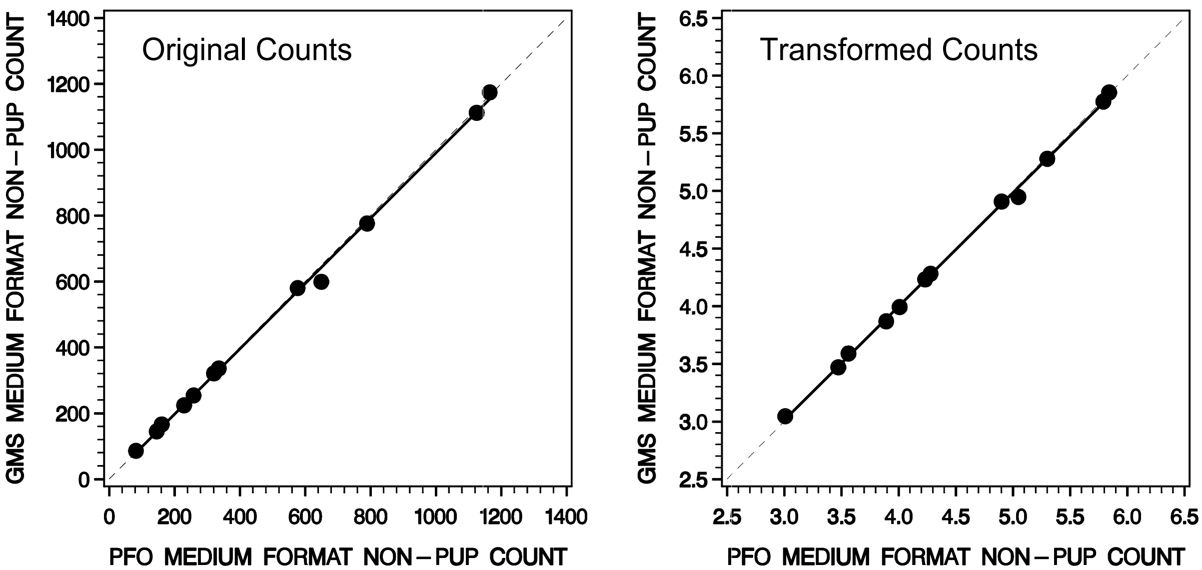

Figure 6. Comparison of non-pup counts from medium-format images made independently by two readers (*PO* and *GS*). The solid lines represent the best fitting linear relationships, and the dashed line 1:1 agreement.

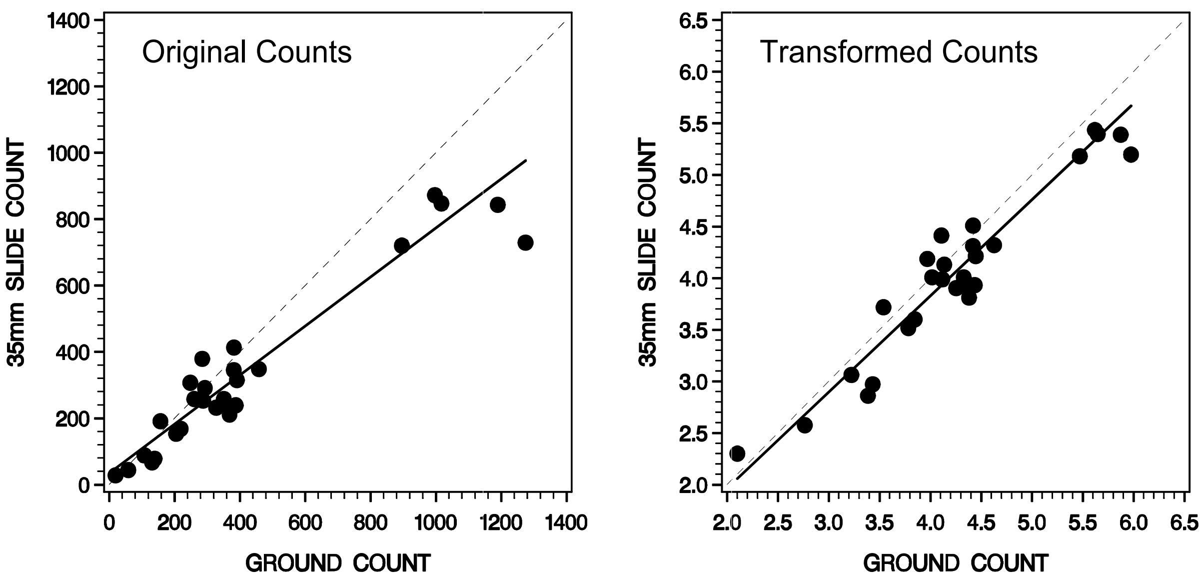

Figure 7. Comparison of 35mm slide counts and ground drive-counts for Forrester Island during 1994, 1995, 1997 and 1998. The solid lines represent best fitting linear relationships, and the dashed line 1:1 agreement.

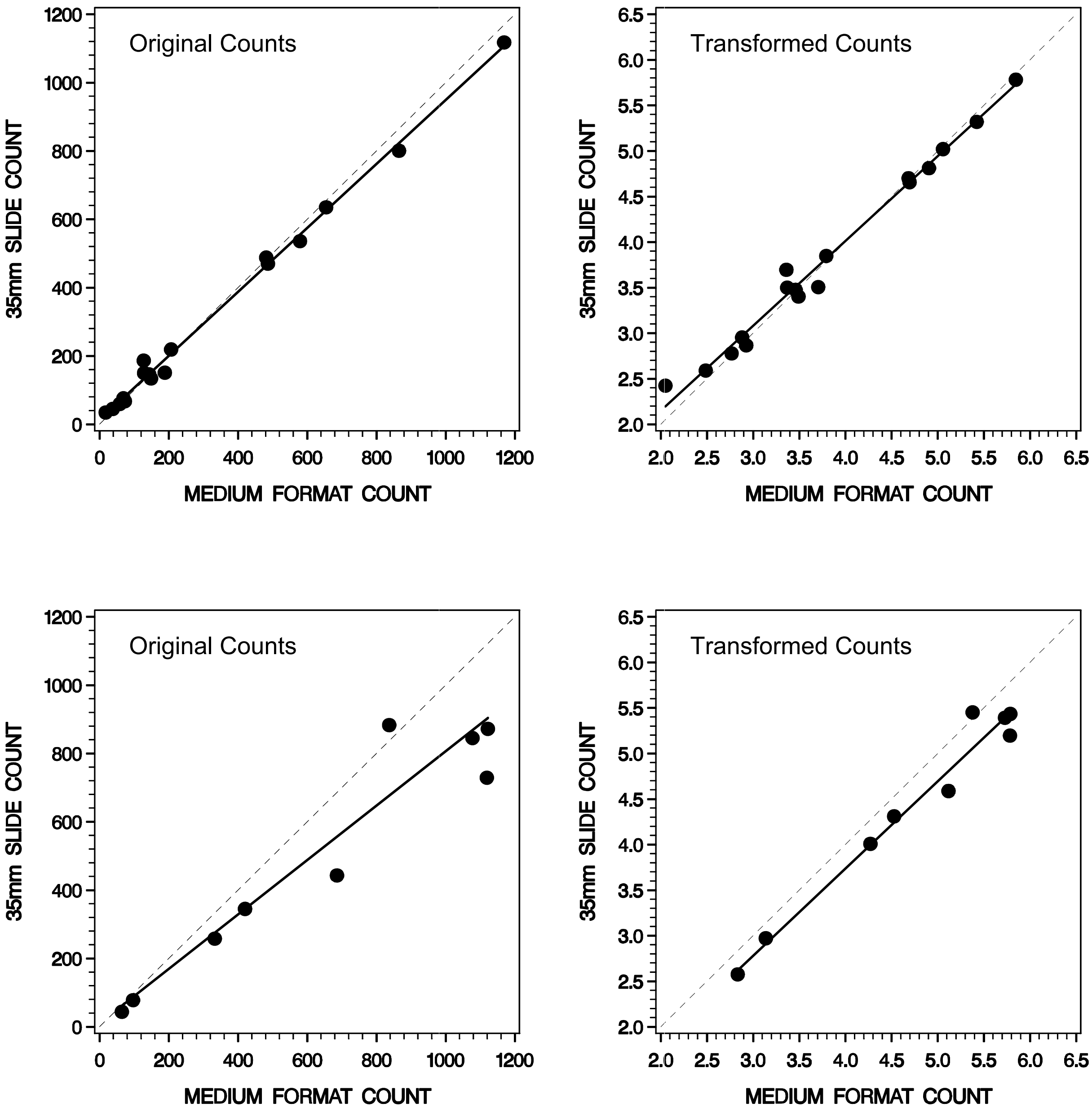

Figure 8. Comparison of 35mm counts and medium-format counts of pups for B.C rookeries in 1998 and 2002 (top panel), and Forrester Island in 1997, 1998 and 2002 (bottom panel). The solid lines represent the best fitting functional relationships, and the dashed line 1:1 agreement.

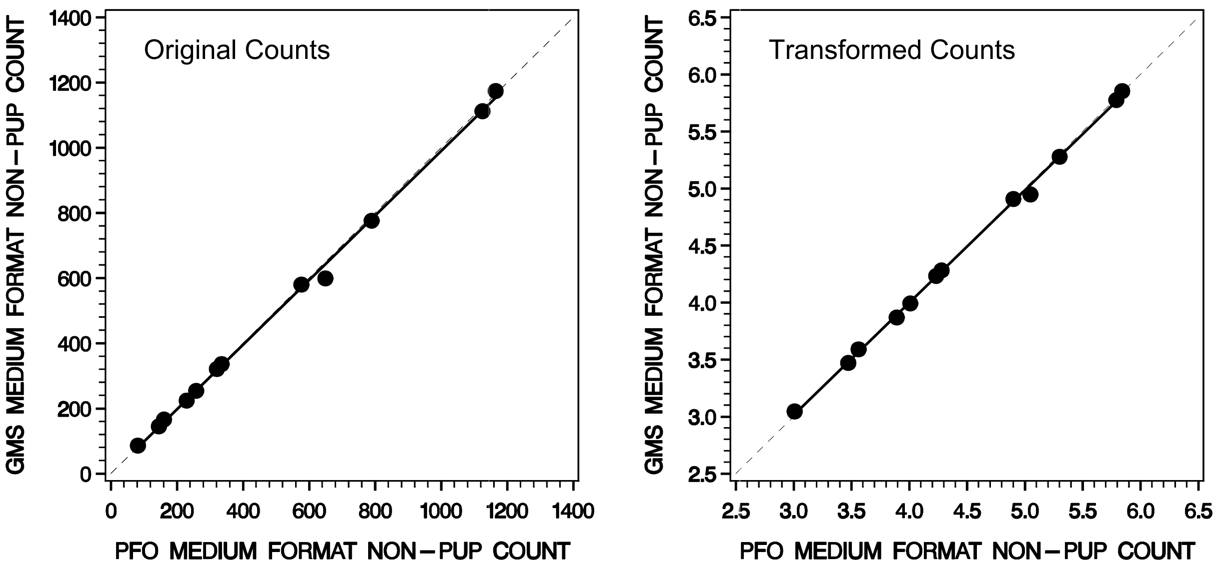

Figure 9. Comparison of 35mm counts and medium-format counts of non-pups for the Scott Islands in 2002. The solid lines represent the best fitting functional relationships, and the dashed line 1:1 agreement.

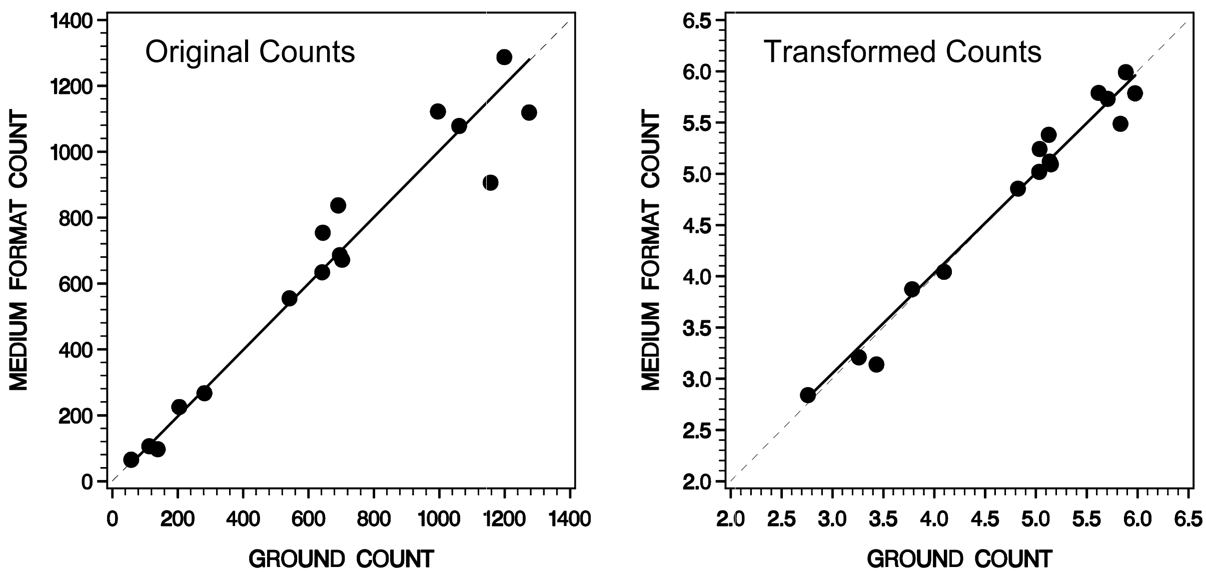

Figure 10. Comparison of ground drive-counts and medium-format counts for Alaskan rookeries in 1997 and 1998. Data are from Snyder et al. (2001). The solid line represent the best fitting linear relationship, and the dashed line 1:1 agreement.

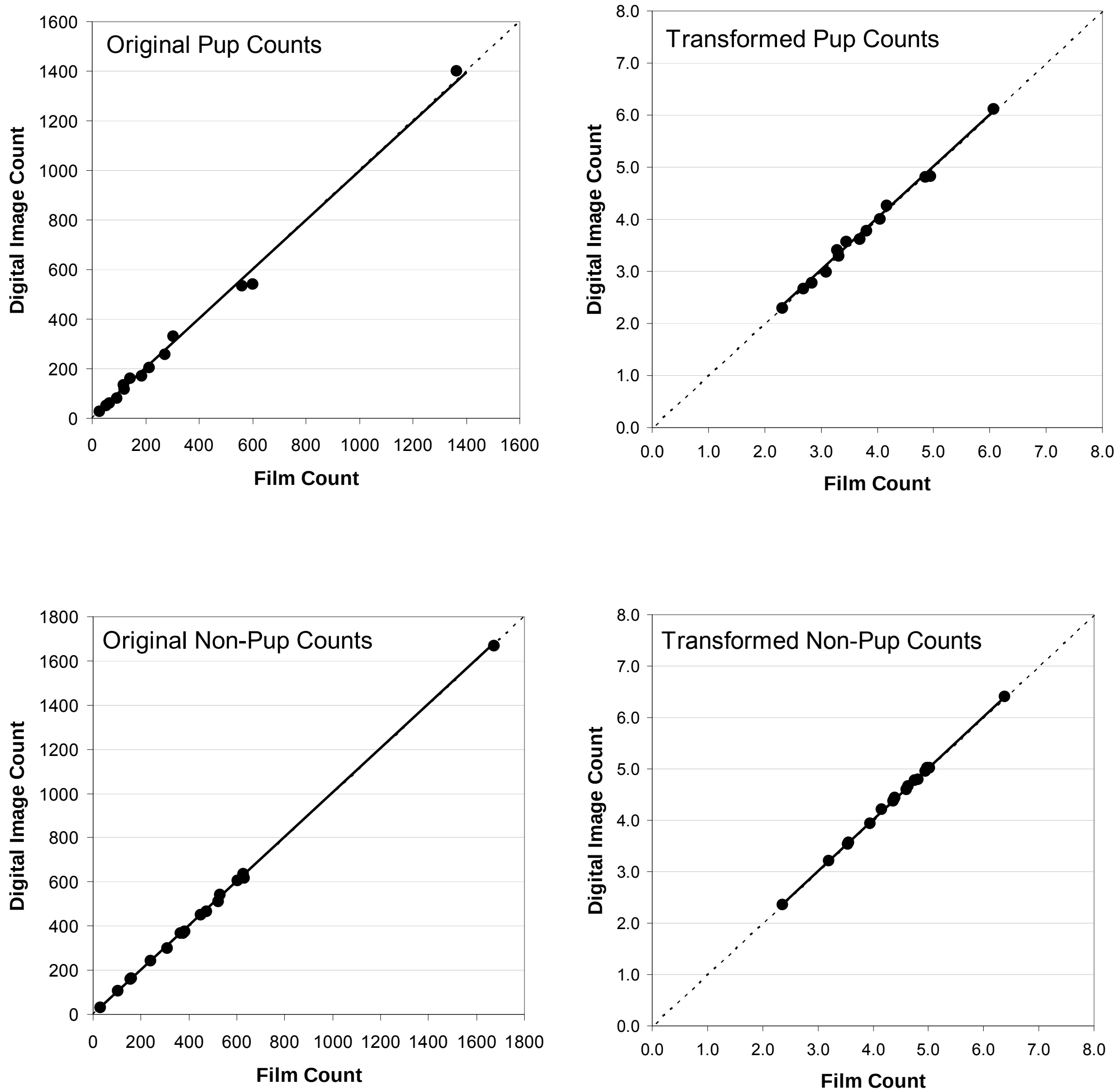

Figure 11. Comparison of pup (top panel) and non-pup (bottom panel) counts from oblique 35mm photographic slides with counts from 35mm oblique digital images for B.C. rookeries in 2006. The solid lines represent the best fitting linear relationships, and the dashed lines 1:1 agreement.