DR. STAN KOTWICKI (Orcid ID : 0000-0002-6112-5021)

Article type : Original Article

Title:

The effect of random and density-dependent variation in sampling efficiency on variance of abundance estimates from fishery surveys.

Short running title:

The effect of q on variance estimates.

Authors:

Stan Kotwicki[1*], Kotaro Ono[2,3]

Affiliations and addresses:

[1]National Marine Fisheries Service, Alaska Fisheries Science Center, National Oceanic and Atmospheric Administration, 7600 Sand Point Way NE, Seattle, WA 98115, USA.

[2]Centre for Coastal Research (CCR), University of Agder, N-4604 Kristiansand, Norway.

[3]Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, P.O. Box 1066 Blindern. NO-0316 Oslo, Norway

Emails: stan.kotwicki@noaa.gov, kotaro.ono@ibv.uio.no.

*corresponding author:

stan.kotwicki@noaa.gov, tel: 206-526-6614, fax: 206 526-6723.

Abstract

Abundance indices (AIs) provide information on population abundance and trends over time, while AI variance (AIV) provides information on reliability or quality of the AI. AIV is an important output from surveys and is commonly used in formal assessments of survey quality, in survey comparison studies, and in stock assessments. However, uncertainty in AIV estimates is poorly understood and studies on the precision and bias in survey AIV estimates are lacking. Typically, AIV estimates are “design-based” and are derived from sampling theory under some aspect of randomized samples. Inference on population density in these cases can be confounded by unaccounted for process errors such as those due to variable sampling efficiency (*q*). Here, we simulated fish distribution and surveys to assess the effect of *q* and variance in *q* on design-based estimates of AIV. Simulation results show that the bias and precision of AIV depends on the mean *q* and variance in *q.* We conclude that to fully evaluate the reliability of AI, both observation error and variability in *q* must be accounted for when estimating AIV. A decrease in mean *q* and an increase in the variance in *q* results in increased bias and decreased precision in survey AIV estimates. These effects are likely small in surveys with mean $q \geq 1$. However, for surveys where $q \leq 0.5$, these effects can be large. Regardless of the survey type, AIV estimates can be improved with knowledge of *q* and variance in *q*.

Keywords:

additional variance, catchability, design-based estimate, fisheries-independent survey, gear efficiency, variance of variance.

Table of contents

Introduction

One of the main goals of scientific fishery-independent surveys (hereafter referred to as surveys) is to estimate either the absolute or relative index of abundance to provide information about status and trends of fish populations (Hilborn and Walters, 1992; Gunderson, 1993). Surveys belong to a group of methods that are based on the assumption that repeating the same sampling effort over time will lead to the observation of the same proportion of the population (Cochran, 1997; Schwarz and Seber, 1999). Surveys are also important because they provide indicators for establishing the ecological health of ecosystems (Nicholson and Jennings, 2004). Survey-derived abundance indices (AIs) and their variances (AIVs) are used in a wide variety of studies such as population dynamics (e.g. Sibly et al., 2005), ecological processes (Aydin and Mueter, 2007; Laman et al., 2015), experimental design (e.g. Overholtz et al., 2006), stock assessments (e.g. Ianelli et al., 2015), and ecological forecasting (e.g. Perry et al., 2005). The main role of the AIVs is to provide information about the reliability of AIs. For example, AIV have been used to compare sampling designs (e.g. Pennington and Volstad, 1991; Overholtz et al., 2006) to evaluate survey design (e.g. Cao et al., 2014), to compare different survey tools (Lingen et al., 1998; Doyle et. al., 2008), to assess improvement in survey methods (e.g. Smith and Gavaris, 1993, Smith and

Lundy, 2006; von Szalay et al., 2007), to weight different observations and data sources in integrated analysis and fishery stock assessments (e.g. Conn, 2010, Francis, 2011), or to achieve optimal effort allocation in stratified trawl surveys (e.g. Harbitz et al., 1998). However, the reliability (i.e. degree of both accuracy and precision) of AIV estimates have not been formally assessed. Similarly to AI, the AIV is just an estimate from the sample, and is itself a random variable. Reliability of the AIV estimate thus depends on the bias and precision of the AIV estimate. However, to date, studies on the reliability of AIV in fishery surveys are generally lacking, except for a few studies on the accuracy of confidence intervals (CIs) of AIs (e.g. Cadigan, 2011; Schnute and Haigh, 2003; Hoyle and Cameron, 2003) which indicated that inaccuracy in CI estimations are usually associated with small sample sizes or heavily skewed survey catch data. Additionally, Cadigan (2011) showed that CI accuracy can depend on estimation method and choice of catch model used.

The most common approach to obtain estimates of AI and AIV from fishery-independent surveys is by using design-based methods (e.g. Smith, 1990; Folmer and Pennington, 2000; Petitgas, 2001). Survey sampling designs range from simple to complex (e.g. Cochran, 1977; Lumley, 2004), and can have correspondingly simple and complex formulas for estimating AI and AIV (e.g. Strand, 2017). A common attribute of design-based estimators is the assumption that sampling efficiency (*q*, often also referred to as survey gear catchability) is constant across time and space (Chen et al., 2004). Sampling efficiency is defined here as the ratio of the survey estimate of abundance to the true abundance (e.g. Godø, 1994, Chen et al., 2004) at each sampling location. The assumption of constant *q* has been shown to be violated in numerous studies using survey gears (e.g. Aglen et al., 1999; Somerton et al., 2007; Kotwicki et al., 2005), which is known to cause variance estimates derived from samples alone to represent only a part of the total variance of the estimate of abundance (e.g. Maunder and Punt, 2004; Hjelvik et al., 2002; Cadigan, 2011). However, despite this knowledge, the effect of a violation of the assumption of constant *q* on the design-based AIV estimates is poorly understood and often ignored across numerous studies and in fishery management applications.

Taking into consideration the wide use of the AIV estimates in fisheries science and management, it is important to understand the implications of using unreliable AIV estimates and factors affecting uncertainty in AIV estimates. Therefore, the main goal of this study was to evaluate the effect of varying *q* on precision and bias of design-based survey estimates of AIV because, in addition to the well-known effects of sample size and spatial distribution on AIV estimates, the *q* value likely influences the uncertainty in AIV estimates (e.g. Pennington and Godø, 1995; Maunder and Punt, 2004). Variation in *q* has been observed to have a random component (e.g. Munro and Somerton, 2001; 2002), be depth-dependent (Benoít and Swain, 2003), environmentally driven (Somerton et al., 2013), or density-dependent (Kotwicki et al., 2014). The issue of environmentally driven q has been dealt with in the past using AI standardization methods (Maunder and Punt, 2004). On the other hand, the effects of random variation in q on AIV estimates have not been studied to date and studies of the effects of density-dependent q are very limited (e.g. Kotwicki et al., 2014). Density-dependent q deserves special consideration because it can result in hyperstability of AI (i.e. AI detects changes in the population abundance that are smaller than actual changes in the population abundance; Hilborn and Walters, 1992), which in turn can result in underestimated AIV (Kotwicki et al., 2014) and give a false perception of high reliability of the hyperstable index of abundance.

In the first part of this study, we attempt to answer the following questions: i. Does the variance in q propagate to the observed design-based AIV estimates from surveys?, ii. What is the expected relative bias in design-based AIV estimates due to variation in q?, iii. What is the impact of variation in q on precision in design-based AIV estimates?, and iv. What is the effect of density-dependent q on bias and precision of design-based AIV estimates? To provide answers to these questions, we simulated realistic spatial distributions of walleye pollock (*Gadus chalcogrammus,* Gadidae*;* hereafter referred to as pollock) observed in the eastern Bering Sea (EBS) during bottom surveys. Given a spatial map of simulated pollock densities (assumed as the "true" species distribution – the quotes are used here and thereafter because our inference is based on simulated data assumed to be "true"), we sampled this map to mimic surveys. We used this framework to vary the mean q and the amount of process error around q at each station and examined how this influenced the bias and precision of AI and AIV estimates.

In the second part of this study, we provide a review of the implications of our findings on the use of AIV estimates in fisheries science and management. Our study shows a possible strategy to estimate the accuracy and precision of AIV in presence of variable q and potential effects of variability in q on AIV estimates. This knowledge should help survey and stock assessment scientists and other fisheries researchers to make better informed decisions in applications of AIV estimates in fisheries science and management.

Methods

Bottom trawl survey data

The Alaska Fisheries Science Center has been conducting annual surveys in the eastern Bering Sea between June and July since 1982 over a fixed set of approximately 376 stations (using stratified systematic survey design) and has used the same standard trawl (83–112 eastern otter trawl) during all years. Surveys start in the south-eastern corner of the survey area and proceeded westward. Tow duration is approximately 30 min at 1.54 $m \cdot s^{-1}$ (3 knots) (see Stauffer [2004] for detail about survey protocol). The catch per unit effort (CPUE) is estimated using the area-swept method (e.g. Alverson and Pereyra, 1969) which determines the area-swept by multiplying the distance fished, as indicated by bottom contact sensor (Somerton and Weinberg, 2001), by the average distance between wing tips measured using acoustic spread sensors (see Weinberg and Kotwicki, 2008 for details).

General outline of the simulation approach

To examine the effect of variable sampling efficiency on the design-based estimates of AIV and AI, we first generated a map of simulated pollock distribution. This map was created by fitting spatio-temporal models to real EBS pollock data from 2005 to 2014 then drawing values from the distribution of predicted values. Once the simulated map was created, we then mimicked survey sampling procedure to generate simulated survey data and to finally estimate AI and AIV.

Simulating pollock distributions

Step 1: fitting a spatio-temporal model to the EBS pollock data

A spatio-temporal model that accounts for both environmental covariates and spatio-temporal dependency in catch was fit to the EBS pollock data from 2005 to 2014. The model followed the approach of Ward et al. (2015) or Ono et al. (2016) in which the analysis combined two models: one that tracked pollock occurrence and the other which tracked the density (in CPUE units) for tows where pollock were observed (see Supplementary Material). Fishing depth, surface, and bottom temperature, as well as sediment size, were included as covariates. Sediment size was estimated and interpolated at each station from historical data from grabs and dredges (Smith and McConnaughey, 1999). Sediment data were expressed in units of “phi” (negative $\log_2$ of the diameter in mm), where higher values correspond to smaller particle sizes (Wentworth, 1922). All covariates were modeled up to their quadratic terms in their original scale except for the depth variable that was log transformed first (in order to model a right skewed effect on the response variable). This choice was based on plots of raw CPUE data against each covariate (Zuur et al., 2010). Spatial and temporal dependency were also included through the use of Mátern covariance function and a first-order autoregressive process (AR1), respectively. Pollock occurrence was modeled using a binomial distribution with logit link and pollock CPUE was modeled using a Gaussian distribution with log link. All models were implemented using the R package *R-INLA* (Lindgren et al., 2011; Lindgren and Rue, 2015; Martins et al., 2013)

Step 2: Generating map of simulated pollock distribution

Annual maps of pollock distribution within the survey region were generated by calculating the predicted pollock density (product of predicted pollock occurrence and CPUE conditional on presence) on a nominal 1 km-grid (Fig. 1). These predictions were based on a single but random Markov chain Monte Carlo (MCMC) draw from the joint posterior distribution of the parameters (instead of the parameters' mean value) in order to account for uncertainty in parameter estimates and to create a patchier species distribution that is more reflective of "true" distribution as compared to the mean MCMC prediction. All environmental covariate values at the grid locations were kriged (ordinary kriging) based on semi-variogram model that best fitted the observed data for mapping purposes (see Supplementary Material). This was done through the function *autofitVariogram* in the R package *automap* (Hiemstra et al., 2009). All data were first converted into an Albers projection in order to preserve distances.

Survey simulations

One thousand surveys were simulated over modeled pollock distributions (we refer to it as the "true" distribution for the rest of the study) for each combination of year (n=10), average sampling efficiency ($\overline{q}$; n=12), and variance in sampling efficiency $V(q)$ (n=9; Table 1). This resulted in 1080 combinations which meant 1,080,000 simulated data sets of annual survey. We considered values of $\overline{q} > 1$ to address possibility of fish herding into the path of the trawl by the trawl doors or bridles (Somerton et al., 2007). To eliminate the effect of sample size on comparisons of AIV and AI estimates across varying values of *q*, each simulated survey data set consisted of a constant 376 sampled survey stations assuming simple random (SR) station allocation. The assumption of SR allocation was chosen to simplify interpretation of results even though the pollock data that were used for the simulation comes from a stratified survey. In survey simulations, we followed the often-postulated survey catch process that $u_{s,i} = q_{s,i}A_i$, where $u_{s,i}$ is the catch per unit effort at station *i*, during survey *s and* A_i *is* the "true" fish density at the station *i* (e.g. Schnute, 1994; Chen et al., 2004). For each simulated annual survey, $u_{s,i}$ was calculated using the A_i value from the simulated pollock distribution and randomly drawing $q_{s,i}$ from a gamma distribution with mean $\overline{q}$ and variance *V(q)*. The index of abundance for each annual survey was then estimated by obtaining the mean CPUE ($\overline{u_s}$) from all 376 survey stations.

Additionally, 1000 surveys were simulated for 5 values of density-dependent efficiency parameter (*a*) for all combinations of year and random $V(q)$ (Table 1), but assuming a $\overline{q} = 1$, resulting in an additional 450,000 annual survey simulations. Density-dependent sampling efficiency was modelled using the formula from Kotwicki et al. (2013):

$$u_{s,i} = \left(\frac{1}{q_{s,i}A_i} + \frac{1}{a}\right)^{-1},$$

where A_i is the true density at location *i* and q_i is the sampling efficiency drawn from a gamma distribution with mean $\overline{q}$ and variance *V(q)*. The parameter *a* represents density dependence of *q*. When fish densities are much lower than *a*, the term 1/*a* becomes negligible. With increased fish density, *a* becomes more influential, resulting in reduced sampling efficiency. For example, at fish density equal to the value of *a*, sampling efficiency will be approximately half of the efficiency at the lowest densities (Kotwicki at al., 2013).

Sample variance was estimated for each simulated annual survey using the following formula:

$$\sigma_S^2 = \frac{\sum_i (u_{s,i} - \overline{u_s})^2}{n},$$

where σ_S is the sample standard deviation of survey *s*, *n* in the number of samples in the survey (i.e. 376), $u_{s,i}$ is the CPUE of the *i*th sample from survey *s*, and $\overline{u_s}$ is the mean over the 376 survey sample.

As discussed in the Introduction, sample variance may not be an adequate approximation of the true variance of AI because of the effect of variable *q*. True variance of AI from real surveys is impossible to obtain because it would require conducting multiple surveys over the same population of fish. However, in the simulation framework, we were able to conduct multiple surveys over the same distribution to obtain "true" variance estimates under different scenarios. We defined "true" variance around the annual index of abundance (given a scenario) as:

$$\sigma_T^2 = \frac{\sum_s (\overline{u_s} - \overline{u_T})^2}{N},$$

where σ_T is a "true" standard deviation of the survey index of abundance, *N* is the number of simulated surveys (i.e. 1000), and $\overline{u_T}$ is the "true" mean fish density estimated from simulated density maps, assuming constant *q*=1. $\overline{u_T}$ is the same for all survey simulations within the same year.

Impact of variable sampling efficiency on abundance indices

The effect of variation in survey sampling efficiency on the estimate of AI ($\overline{u_s}$) was shown by producing box-and-whisker plots of AI relative errors (i.e. ($\overline{u_s}$ - $\overline{u_T}$)/ $\overline{u_T}$) based on the 1000 survey simulations for each value of simulated $\bar{q}$ and $V(q)$ presented in multi-panel plots. Box-and-whisker plots were also applied to illustrate the effects of $V(q)$ on survey coefficient of variation (CV) derived from the formula $\sigma_s^2/\overline{u_s}$. Other variables presented on box-and-whisker multi-panel plots included relative errors in σ_s estimates (i.e. $(\sigma_s - \sigma_T)/\sigma_T$, where and $\sigma_T$is the mean over the 1000 simulated survey σ_s), and CVs of σ_s (i.e. standard deviation(σ_s) / mean(σ_s). Finally, Box-and-whisker plots arranged in multi-panel plots were used to illustrate the relationship between survey AI and density-dependent efficiency and between survey CVs and density-dependent efficiency.

Results

Results indicate that random variability in *q* can bias an AIV estimate and reduce its precision. The degree of this effect depends on both the mean and variance of *q*. We present the result from one representative year because the results were similar across years. Results are presented for each year separately in the Supplementary Materials.

Random variation in sampling efficiency

The relationship between AI relative errors, $\bar{q}$, and $V(q)$ indicates that distribution of $\overline{u_s}$ may be highly skewed to low values indicating that $\overline{u_s}$ will likely to be biased low (i.e. unbiased mean, but biased median) in the presence of low but variable q (Fig. 2). For example, for $\bar{q}$ = 0.05 and $V(q)$ = 0.2 the median relative error was equal to -0.5. This bias increased with the decrease of $\bar{q}$ and increase in $V(q)$. However, for $\bar{q}$ > 0.4, median relative errors were close to 0 indicating unbiased estimates of AI.

The relationship between survey CV (i.e. $\sigma_s^2/\overline{u_s}$), $\bar{q}$ and $V(q)$ shows that survey CV depends on underlying spatial fish distribution and both $\bar{q}$ and $V(q)$. In Figure 3, for example, in case of a near constant *q* (top left), the survey CV of 0.2 can be attributed to the spatial distribution of the fish. In other panels in Figure 3, it is apparent that survey CV increases with a decrease of $\bar{q}$ and increase of $V(q)$. This result indicates that design-based estimates of σ_s capture some variability in the q (i.e. σ_s is not independent of $V(q)$). Survey CVs were high (> 0.3) for surveys which had low but highly variable q.

The relationship between relative error in σ_s (i.e. $(\sigma_s - \sigma_T)/\sigma_T$), q, and $V(q)$ shows that estimates of σ_s are biased low in surveys where $\bar{q}$ is low and $V(q)$ is high (Fig. 4). This result indicates that although estimates of σ_s capture some variability in q as it was shown on Figure 3, the "true" variance of the AI is still underestimated by σ_s, especially in surveys where $\bar{q}$ is low and $V(q)$ is high. For example, design-based estimates of σ_s from surveys with $\bar{q}$ = 0.1 and $V(q)$ = 0.2 were underestimated by approximately 40%, but when $\bar{q}$ > 0.6, median bias in $\sigma_s$was close to zero.

The relationship between CV of σ_s, $\bar{q}$ and $V(q)$ shows that the precision of σ_s estimates increases with an increase of $\bar{q}$ and decrease of $V(q)$(Fig. 5). Moreover, results also show that there are 2 components of σ_s variation. The first component is inherently associated with the "true" fish distribution (represented by blue line on Fig. 5) and second component is associated with $\bar{q}$ and $V(q)$. The first component varied between 0.1 and 0.2 between years (see supplementary materials). The influence of variable q on the estimate of σ_s was stronger than on the relative error in $\overline{u_s}$ indicating that survey estimates of σ_s are more sensitive to the variation in q than the $\overline{u_s}$ estimates (Fig. 2 vs. 3). For example, for surveys with $\bar{q} = 0.5$ and $V(q) = 0.4$, CV of σ_s was double of what it would've been under constant q , but $\overline{u_s}$ was close to unbiased.

Density-dependent sampling efficiency

The relationship between $\overline{u_s}$, CV, density-dependent effects, and $V(q)$ when $\bar{q} = 1$, shows a strong negative bias in the survey AI (Fig. 6) and survey CV (Fig. 7). Density-dependent effects result in similar bias across all values of $V(q)$. The effect of $V(q)$ was limited to increased uncertainty in the CV estimates and was much smaller that the density-dependent effects.

Discussion

General conclusions from the simulation study

The findings of our study indicate that information about sampling efficiency is necessary to assess the reliability of AIV and AI from fishery surveys because variation in *q* can affect both of these estimates. Although the results of lower precision in AI in response to the V(*q*) seem evident (Fig. 2), the results of effect of the V(*q*) on AIV are not. To the contrary, despite a clear evidence that the variability in *q* is a common feature of majority of surveys (e.g. Aglen et al, 1999; Somerton et al., 2007; Kotwicki et al., 2005) the effects of V(q) on AIV estimates are very poorly understood. This situation leads, as discussed in the Introduction, to wide usage of AIV estimates in fisheries science and management with or without limited consideration for their bias and precision. Our results provide answers to the 4 questions posed in the Introduction. First, it is apparent that the V(q) only partially propagates to the observed AIV (i.e. ${\sigma_s}^2$) estimates from surveys (Figs. 3 and 4) indicating that, in the presence of variable *q,* it is impossible to obtain unbiased AIV estimate from samples alone. Second, expected relative bias in ${\sigma_s}^2$ estimates due to V(q) can be approximately estimated, as presented here, using knowledge of the V(q) and simulated fish density distributions (Fig. 4). Third, the precision in AIV estimates depends on both underlying species' spatial distribution and V(q) (Figure S7, Fig. 5), and unfortunately, it cannot be estimated from the observations; however, it can be approximated using simulations. Fourth, density-dependent q has a strong effect on bias and precision of both AIV and AI, and it needs to be corrected for to obtain reliable estimates of AIV and AI. The correction for density-dependent q can be derived from auxiliary information about q (e.g. using different types of surveys conducted simultaneously; Kotwicki et al., 2014)

Our findings also show that when the assumption of a constant q is met, design-based estimates of AIV and AI are unbiased regardless of the value of q. However, after reviewing many studies of q (e.g. Somerton et al., 2013; Fraser et al., 2007; Kotwicki and Weinberg, 2005; Kotwicki et al., 2005; Munro and Somerton, 2001), and many others cited within this manuscript), we conclude that in reality this condition is likely never met and $V(q)$ is always > 0, which leads to potential bias in design-based estimates of AIV and AI. The degree of these problems depends on the $\bar{q}$ and its variability. It appears that for surveys with $\bar{q} \geq 1$, variation in q has only a small effect on the bias and precision of design-based estimates. However, for surveys with low q ($\bar{q} < 0.5$), this effect can be substantial and could result in imprecise and biased design-based estimates of AIV and AI. Additionally, our results also indicate that for small $\bar{q}$, V(q) can cause the error distribution in AI to be skewed low (Fig. 2). Because indices of abundance are usually used as time series, this may lead to negative bias in predictions of the majority of the estimates within time series.

Implications for using AIV estimates in fisheries science and management

Prevalence of random variation in q.

To assess the potential impact of $\overline{q}$ and V(q) on results from specific surveys, it is necessary to know their values. Although estimating these values is believed to be very complex and difficult (e.g. Somerton et al., 2007), numerous studies have done so. Values of $\overline{q}$ have been shown to vary widely among different survey gear types and species. For example, Fraser et al. (2007) and Kotwicki et al. (2005) estimated $\overline{q}$ for a large number of demersal species at different length categories, which ranged from 0.01 to close to 1. Low values of $\overline{q}$ are often observed for certain sizes of animals for which survey gear has poor selectivity. For example, bottom trawls have been found to have low values of $\overline{q}$ for small sizes of demersal fauna in the Arctic, while beam trawls have low $\overline{q}$ for larger animals as indicated by selectivity ratio (Kotwicki et al. 2017). Estimates of $\overline{q}$ > 1 imply that the survey catches considerably more fish than would be predicted to be in the path of the tow. For example: catchability > 1 was estimated for haddock in the North Sea and Ling in New Zealand (Harley and Myers 2001) and for species of flatfish from Alaska (Somerton et al., 2007). These high catchability values can be explained by fish being herded into the path of the trawl from the area beyond the net's wing tips (Dickson, 1993). Another example when $\overline{q}$ > 1 is seen with fish density estimates from combined acoustic and bottom trawl surveys (Kotwicki et al., 2018), where part of the water column is enumerated by acoustic gear and the other part by bottom trawl. It was shown that the combined estimate from these two techniques can have q > 1 due to the existence of an overlap zone sampled by both techniques. On the other hand, midwater trawl surveys for krill have been shown to have a very small q between 0.003 and 0.06 (Kasatkina, 1991).

Although existing literature rarely reports on the value of V(q), data presented in these studies show that V(q) can also vary widely between survey gears and species. Some authors report standard deviations of the $\overline{q}$ estimates. For example, Somerton et al. (2007) reported mean flatfish herding efficiency to be about 0.2 with a standard deviation of the mean ranging between 0.04 and 0.16. Assuming a sample size of 50, this would correspond to a V(q) in the range of 0.01 to 1.3. Kotwicki et al. (2015) estimated q for pollock acoustic and bottom trawl surveys. For bottom trawl, $\overline{q}$ = 0.9 with V(q) = 0.2 and for acoustic survey, $\overline{q}$ = 0.3 with V(q) = 0.4. Variances for midwater trawl q for krill, estimated from the standard deviations of the $\overline{q}$ (ranging from 0.01 – 0.30) assuming sample size of 50, would correspond to a V(q) ranging from 0.005 to 4.5 (Kasatkina, 1991). Mackinson et al. (2005) estimated $\overline{q}$ of the dredge used in the sandeel surveys to be in the range of 0.02 – 0.1 with V(q) ranging from 0.004 to 0.21. Because of small values of $\overline{q}$ in combination with relatively large V(q) in their experiment, we conclude that design-based variance estimates for the sandeel dredge survey are likely to be grossly underestimated (Fig. 5). Given the importance of V(q) on AIV and AI estimates, we advocate that future studies of survey gear efficiency should report V(q) alongside the $\bar{q}$.

Prevalence of density-dependent q

In the case of density-dependent q, estimates of AIV and AI may be significantly biased (Fig. 6) and they may appear overly precise (Fig. 7). This is troublesome because density-dependent q has been observed in a number of fishery surveys. Godø et al. (1999) found that q for Atlantic cod (*Gadus morhua*) and haddock (*Melanogrammus aeglefinus*) increased with fish density, while others found q decreased with fish density for capelin (*Mallotus villosus*; O'Driscoll et al., 2002), Atlantic croakers (*Micropogonias undulatus*), white perch (*Morone americana*; Hoffman et al., 2009), walleye pollock (*Gadus chalcogrammus*; Kotwicki et al., 2014), and Atlantic cod (Ono et al., 2018). Kotwicki et al. (2014) showed that density-dependent q can lead to a hyperstable index of abundance, which in turn can cause bias in stock assessment outcomes. Stock assessments based on a hyperstable index may fail to track population changes, which could potentially lead to overfishing (e.g. Hutchings, 1996; Walters and Maguire, 1996; Erisman et al., 2011). Moreover, the AIV of a hyperstable and density-dependent AI is likely to be underestimated (Kotwicki et al., 2014) leading to overweighting the influence of this index relative to other sources of information within the stock assessment model. Our results indicate that density-dependent q deserves special attention as it can lead to a large underestimation of AIV. Effects of density-dependent q are much larger compared to the effects of random variability in q, and if they exist, they need to be corrected for to avoid biases in stock assessments. The procedure to correct AI estimates for density-dependence has been proposed in Kotwicki et al. (2014), where they used near-bottom acoustic backscatter to derive a correction function for bottom trawl survey density estimates. They also reported that the AIV corrected for density-dependence was on average 55% higher than the AIV of not corrected index supporting results presented in this study. The causes of the density-dependent q are poorly understood and warrant further investigation. However, limited observations by O'Driscol et al. (2002), Hoffman et al. (2009), and Kotwicki et al. (2014) indicate that q may be affected by gear avoidance behavior or trawl saturation. Gear saturation has been also known to affect the CPUE of gear such as pots or traps (Bacheler et al., 2013), longlines (Rodgveller et al., 2011), and gillnets (Li et al., 2011), indicating that density-dependent effects should be considered when obtaining AIs from surveys using these gears. The prevalence of density-dependent effects on q in fishery surveys worldwide is unknown, but given the results of existing studies, we conclude that it may be common and should be considered in future for all fishery surveys and stock assessments.

Sampling design considerations for AIV estimation

Survey sampling design requires a finite population of unique and identifiable sampling units (e.g. location swept by trawl gear) for which a number (or weight) of animals is observed and a sampling plan is used that assigns a known probability of selection to the sampling units (Smith, 1990). It is preferable that the sampling units are independent and identically distributed, but it is not required as sampling designs that account for spatial correlation in animal distributions also exist (Petitgas, 2001; Thorson et al., 2015). However, in reality, the conditions required to obtain unbiased and precise estimates of AIV (i.e. constant q) are rarely met and it is impossible to assess reliability of these estimates based on samples alone. This is because sample value depends on two random variables: true fish abundance and q. Without knowledge of q and $V(q)$, it is impossible to reconcile the effect of these random variables on sample statistics.

Our results indicate that the reliability of AI derived from a fish survey cannot be assessed by the design-based variance estimate alone; but it also needs to take into account variability in q. This is not currently a common practice and analyses of data from fishery-independent surveys are usually confined to the estimation of AI and associated design-based AIV (e.g. Cochran, 1977; Smith, 1990; NPFMC, 2017). These estimates are commonly used in fishery stock assessment models, where AIs are used to predict population trends (e.g. Collie and Sissenwine, 1983; Gavaris, 1988), while AIVs are assumed to measure the quality of each AI relative to other AIs and are used as weights in the models (e.g. Ianelli et al., 2014). Improvements to survey methodology or design-based methods to obtain abundance estimates are often judged exclusively by their ability to reduce design-based AIV estimates (e.g. Smith and Gavaris, 1993; von Szalay, 2003; Pennington et al., 2002). We consider these approaches risky because they ignore the possible effects of the q and V(q) on the reliability of the AIV estimates.

The problem of underestimating AIV in presence of variable q has been previously acknowledged. For example, Pennington and Godø (1995) estimated that true AIV was approximately twice as large as the AIV estimate based on samples alone. A similar increase in AIV was shown when accounting for diurnal variation in the q for few species caught in the bottom trawl surveys in the Barents Sea (Hjellvik et al., 2002). In fisheries stock assessment models, several approaches have been used to account for this unknown level of additional AIV. One frequently used method is to model AIV $= \sigma_s^2 + \sigma_{add}^2$, where σ_s^2 is design-based sampling variance and σ_{add}^2 is an additional variance, which arises from variable survey catchability (Punt and Butterworth, 2003; Maunder and Punt, 2004). Alternatively, AIV and AI can be estimated using a statistical model-based approach by modelling catch process (e.g. Cadigan, 2011; Chen et al., 2004) or spatio-temporal modelling (e.g. Shelton et al., 2014; Thorson et al. 2015), which have become the standards used in certain parts of the world (e.g., the US West Coast). Both of these approaches have been shown to produce more precise estimates of population abundance and an improvement in variance estimation (Cadigan, 2011, Thorson et al., 2015). However, it remains unknown whether the variability in q affects these types of model-based estimates of AIV in a similar manner as the design-based approach because similar to design-based estimation, the majority of the model-based estimations use CPUE data exclusively and ignore existing knowledge about q.

Implications

In this study, we showed that the bias and precision of AIV can be approximated through simulations when the $\bar{q}$ and variation in q are known. Many studies have been performed to estimate q of a survey gear and predictions of $\bar{q}$ from these studies have been often useful (Somerton et al. 1999). However, the majority of them did not make use of estimates of V(q). In cases where the posterior distribution of q can be obtained, it can be used to propagate V(q) into estimates of AIV and AI (e.g. Kotwicki et al., 2014). Such studies can be performed on existing data from past experiments used to predict $\bar{q}$. We advocate that such studies should be undertaken especially for species for which $\bar{q}$ is estimated to be low and V(q) is high as these are likely to affect AIV and AI the most.

In fisheries stock assessments, estimates of AIV are often used as a measure of precision of the survey abundance estimates (e.g. Ianelli et al., 2014), assuming that the AI accuracy (or bias) remains constant over time. Potential sources of bias in AIs can result from multiple processes affecting q such as gear type, fish behavior, and sampling methodology. When AI is biased but q is constant, survey AI estimates can still be useful as they track temporal trends in abundance (Pennington and Strømme, 1998). AIs can be particularly informative if they cover long periods of time (Rose et al., 2000). However, when q changes over time, it can cause additional variation in the observed abundance trends, which is not fully reflected in the AIV estimate derived from samples (Kotwicki et al., 2014). In such cases, AIV depends not only on the underlying population structure and distribution, but also on the variability in q, which must be accounted for to fully evaluate the reliability of an abundance estimate (e.g., Pennington and Godø, 1995; Punt and Butterworth, 2003; Maunder and Punt, 2004).

Modern fisheries stock assessment models use multiple sources of information within the framework of integrated analysis (Fournier et al., 1998; Maunder and Punt, 2013). Data weighting is an important part of integrated analysis (Breen et al., 2003; Hulson et al., 2008; Francis, 2011) because the outcomes and the uncertainty of outcomes drawn from the application of integrated models can be strongly influenced by the choice of weightings (Deriso et al., 2007). Weightings in stock assessment models are represented by variance estimates for the corresponding data; however, the estimation of these weightings can be challenging (Francis, 2011; Maunder and Punt, 2013). These challenges often arise from incorrect variance estimates or from model misspecifications (Maunder and Piner, 2015). The most common approach to weighting the AI from surveys is to use design-based AIV estimate (i.e. observation error) as a measure of the uncertainty (e.g. Method and Wetzel, 2013; Maunder and Piner, 2015). Our results indicate that in the presence of variable q, this approach may result in overweighting of the AI influence in stock assessment models. To compensate for this, it is best to estimate AIV by accounting for V(q) as shown in this study. However, it needs to be done with care as survey inputs such as indices of recruitment, population age structure, and spatial distribution can also depend on V(q). Therefore, including a corrected AIV estimate for AI alone can cause overweighting of other inputs, some for which variance estimates have not been corrected. It was noted previously that down-weighting AI from surveys because of concerns about catchability may have the undesired effect of down-weighting perhaps the most valuable data in a stock assessment model (the abundance trend) in lieu of overfitting compositional data (Francis, 2011).

The information on population abundance and trends in abundance over time provided by AI is also important for studies other than stock assessments. In these studies, AI is used usually as a relative or absolute measure of population size and AIV is typically used in the form of CV to assess quality of the AI estimate. For example: Overholtz et al., (2006) used AI CVs to compare sampling designs between 3 acoustic surveys. Von Szalay et al. (2007) used survey AIV estimates to determine if the information from additional acoustic sampling can improve bottom trawl survey AI estimates. Survey stratification (e.g. Smith and Gavaris, 1993) and sample allocations (e.g. Smith and Lundy, 2006; Harbitz et al., 1998; Smith et al., 2011) are also usually based on AIV estimates without consideration for V(q). Coefficients of variation derived from AIV are often used to compare reliability of AI estimates derived from surveys using different tools or sampling designs (Pennington and Volstad, 1991). For example, Lingen et al. (1998) used CV to compare different survey tools for assessing abundance and distribution of fish eggs. Doyle et. al. (2008) used CVs to compare four types of sampling gear used to collect shovelnose sturgeon. Other examples are numerous (e.g. Dressel, 2005; Dissanayake and Stefansson, 2010) as using CV estimates from observations to assess quality of surveys appears to be a standard among survey scientists. Unfortunately, these assessments are usually performed without regard for accuracy or precision of AIV estimates. We argue that, as with all other statistics, accuracy and precision of AIV needs to be taken into account when CVs or other precision measures are used for comparisons between survey gears and modeling methods. To illustrate the danger of using CV estimates derived from observations alone as a means to choose appropriate survey method, we refer to Figure 7 where it is apparent that for surveys with strong density-dependent *q,* CV will likely be lower in comparison to the CV from surveys that may have higher variability in *q* but are not density-dependent. In such a case, using the CVs will lead to an incorrect choice of survey method with density-dependent *q*. Similarly, an incorrect choice may also occur when two surveys are not density-dependent but the difference between CVs is just a matter of chance.

Our results indicate that AIV estimates can be highly imprecise (Fig. 5) and should not be used *ad hoc* without scrutiny. Relationship between the CV of AI standard deviation and the *V(q)* presented in Figure 5 shows that there are clearly 2 components of the variance in the AIV. The first component is inherent to sampling design and underlying fish distribution and does not depend on variations in *q*. The second component contributes additional variance to AIV. This additional variance increases with a decrease in *q* and increase in V(*q*). Because *q* is likely variable in all fisheries surveys, it is safe to conclude that AIV derived from samples alone is always an underestimate of the total AIV. The degree of this underestimation depends on the underlying distribution of *q*. In addition, the precision of the AIV estimate can be very low resulting in high uncertainty when AIV estimates are used to compare survey and modelling methods. Ignoring this uncertainty can lead to the choice of a less reliable method just because the single AIV estimate for that method, by chance, happened to be lower when compared to more reliable methods. In stock assessment models, weights assigned to each AI may be spurious because the underlying uncertainty in the AIV is unknown. This can also lead to overestimating the reliability of survey abundance estimates in process studies (e.g. survey comparisons, ecosystem modeling, spatial dynamics studies) and fishery management decisions. In process studies, survey CVs are often used as a measure of the reliability of abundance estimates (e.g. Jakobsen et al., 1997; Francis, 1984; Overholtz et al., 2006), which in the presence of multiple sources of information, can lead to the wrong choice of data source used in research. Similarly, management decisions are often based on the most important piece of information available. Surveys with lowest sample CVs are often perceived as the best (e.g. Pennington and Volstad,. 1991); however, if CV is biased low due to low or variable *q*, it can lead to an erroneous choice of "best" survey. For example, a survey that does not sample a significant part of a contagious population can result in low AIs with very low AIVs, despite the true population being much larger.

Merit of AI and AIV estimates is highly dependent on bias and precision of these two statistics. The precision of an AI should be reflected in the estimate of the AIV given that AIV is adequately precise. However, in fishery science there is a general lack of studies on the precision of AIV estimates. This is surprising given the use of AIV estimates across wide range of studies in fisheries. Our study represents a first look at the effects of the underestimation of AIV uncertainty; however, more studies are needed to understand the impact of this underestimation. The variance of variance problems have been undertaken in statistics (e.g. Cho et al., 2005) as well as in applied studies other than fisheries (e.g. Dieters et al., 1995), and the common use of AIV in fishery science warrants a closer look into the uncertainty of these estimates.

Future directions and challenges

This study concentrated on the effects of the variation in *q* defined as sampling efficiency or gear catchability. However, there are other sources of uncertainty that could also affect reliability of the AIV estimates. These include the other two components of survey catchability: vertical and spatial availability (Edwards, 1968; Harley and Myers, 2001). The effects of variation in vertical availability (e.g. Kotwicki et al., 2015) in AIV would be similar to the effects of sampling efficiency presented in this study. This is because vertical availability can be considered to be part of spatial variation in *q* when fish density at a tow site is based on the whole water column and not just the water volume swept by the trawl. The effect of the variation in spatial availability on the AIV have not been considered in our study. However, it is important to acknowledge that they should not be ignored for many surveyed species. For example, in the Bering Sea there is evidence that some groundfish change their spatial distribution in response to environmental variability, density-dependent effects, and ontogeny (Nichol 1998; Kotwicki & Lauth 2013; Thorson et al. 2016). If these changes occur within the surveyed area, the derived AI can provide accurate information about trends in population abundance over time. However, when spatial distribution changes occur across survey boundaries and the entire population is not available to the survey, the trend in population abundance estimates is confounded by changes in the proportion of the stocks enumerated. Similarly to variation in sampling efficiency, ignoring variation in survey spatial availability can lead to underestimates of AIV estimates for affected species as well as biased trends in abundance indices (Thorson et al. 2013).

We are not aware of any studies that attempt to account for variation in spatial availability in the estimation of AIV. This task will likely be more challenging than accounting for sampling efficiency or vertical availability. Studies on variation in spatial distribution of species outside the surveyed areas are generally lacking because of the absence of data from unsurveyed areas. In some instances, data from the area outside the survey exists. This includes data from surveys performed by neighboring states and countries, or data from research studies conducted outside survey areas. The major difficulties with using the data from areas outside standard surveys come from differences in methodology used for collecting samples. To reconcile these differences, challenging and costly intercalibration studies are needed (e.g. Smith, 2002). The importance of such intercalibration has recently been increasing due to climate change and associated shifts in the distribution of species, which often occur across survey boundaries (e.g. Kotwicki and Lauth, 2013; Thorson et al. 2016). More studies are needed to develop methods that can determine the impacts of these movements on AIV and AI estimates.

Our approach to obtain "true" variance estimates using simulation has some shortcomings which could be addressed in future studies. First, the simulated spatial distributions, assumed to be "true", were done based on the existing survey data which could have been affected by *q*. Therefore, it is possible that the "true" variance estimates were biased because the true fish spatial distribution is unknown. However, we believe that these potential biases will usually be smaller than biases caused by variable *q* and regardless of the ability to predict true distribution incorporating *q* should produce more reliable AIV estimate. Future studies are needed to develop methods to simulate and test realistic species spatial distributions. Second, we have concentrated on describing the effects of variable *q* on AIV estimates using a simple random survey approach. This approach was chosen because we believe that the stated goals of the paper are best presented on that simple example. However, in reality, the majority of surveys are stratified (e.g. Staufer, 2004); therefore, in practical applications, it is necessary to perform simulations using correct survey design. Such studies are presently underway at the Alaska Fisheries Science Center for surveys conducted in the Aleutian Islands, Gulf of Alaska, and Bering Sea. Moreover, data from some surveys are used to obtain model-based (e.g. Shelton et al., 2014; Thorson et al. 2015) or model-assisted estimates (e.g. Chen et al., 2004, Cadigan, 2011). The AIV estimates from these modelling techniques are believed to be more reliable than ones from the design-based estimators; however, the effect of the random and density-dependent variation in *q* on these estimates needs further investigation.

Acknowledgments

We thank Dana Hanselman, Jim Ianelli, Jeff Napp, Michael Pennington, Paul Spencer, James Lee, Christine Baier, and 3 anonymous reviewers for reviews and discussions that greatly improved the quality of this paper. K.O. was funded by the Research Council of Norway through the Skagcore project (255675).

The data that support the findings of this study are available from the corresponding author upon reasonable request.

The findings and conclusions in the paper are those of the authors and do not necessarily represent the views of the National Marine Fisheries Service.

REFERENCES

Aglen, A., Engås, A., Huse, I., Michalsen, K., & Stensholt, B. K. (1999). How vertical fish distribution may affect survey results. *ICES Journal of Marine Science: Journal du Conseil*, 56, 345–360. doi:10.1006/jmsc.1999.0449

Alverson, D.L., & Pereyra, W.T. (1969). Demersal fish explorations in the northeastern Pacific Ocean -- An evaluation of exploratory fishing methods and analytical approaches to stock size and yield forecasts. *Journal of the Fisheries Research Board of Canada*, 26, 1985-2001. doi:10.1139/f69-188

Aydin, K., & Mueter, F. (2007). The Bering Sea—a dynamic food web perspective. *Deep-Sea Research Part II: Topical Studies in Oceanography,* 54, 2501-2525. doi:10.1016/j.dsr2.2007.08.022

NPFMC. (2017). *Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions for 2015.* North Pacific Fishery Management Council, Anchorage, AK.

Benoít, H.P., & Swain, D.P. (2003). Accounting for length-and depth-dependent diel variation in catchability of fish and invertebrates in an annual bottom-trawl survey. *ICES Journal of Marine Science: Journal du Conseil*, 60, 1298-1317. doi:10.1016/S1054-3139(03)00124-3

Breen, P.A., Hilborn, R., Maunder, M.N. & Kim, S.W. (2003). Effects of alternative control rules on the conflict between a fishery and a threatened sea lion (*Phocarctos hookeri*). *Canadian Journal of Fisheries and Aquatic Sciences*, 60, 527-541. doi:10.1139/f03-046

Cadigan, N.G. (2011). Confidence intervals for trawlable abundance from stratified- random bottom trawl surveys. *Canadian Journal of Fisheries and Aquatic Sciences*, 68, 781-794. doi:10.1139/f2011-026

Cao, J., Chen, Y., Chang, J.H., & Chen, X. (2014). An evaluation of an inshore bottom trawl survey design for American lobster (*Homarus americanus*) using computer simulations. *Journal of Northwest Atlantic Fishery Science*, 46, 27-39. doi:10.2960/J.v46.m696

Chen, J., Thompson, M.E., & Wu, C. (2004). Estimation of fish abundance indices based on scientific research trawl surveys. *Biometrics*, 60, 116-123. doi:10.1111/j.0006-341X.2004.00162.x

Cho, E., Cho, M.J., & Eltinge, J., (2005). The variance of sample variance from a finite population. *International Journal of Pure and Applied Mathematics*, 21, 3345-3350.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). Wiley, New York.

Collie, J.S., & Sissenwine, M.P. (1983). Estimating population size from relative abundance data measured with error. *Canadian Journal of Fisheries and Aquatic Sciences*, 40, 1871-1879. doi:10.1139/f83-217

Conn, P. B. (2010). Hierarchical analysis of multiple noisy abundance indices. *Canadian Journal of Fisheries and Aquatic Sciences,* 67, 108–120. doi:10.1139/F09-175

Deriso, R.B., Maunder, M.N., & Skalski, J.R. (2007). Variance estimation in integrated assessment models and its importance for hypothesis testing. *Canadian Journal of Fisheries and Aquatic Sciences*, 64, 187-197. doi:10.1139/f06-178

Dickson, W. (1993). Estimation of the capture efficiency of trawl gear. I: development of a theoretical model. *Fisheries Research*, 16, 239–253. doi:10.1016/0165-7836(93)90096-P

Dieters, M.J., White, T.L., Littell, R.C., & Hodge, G.R. (1995). Application of approximate variances of variance components and their ratios in genetic tests. *Theoretical and Applied Genetics*, 91, 15-24. doi:10.1007/BF00220853

Dissanayake, D.C.T., & Stefansson, G., (2010). Abundance and distribution of commercial sea cucumber species in the coastal waters of Sri Lanka. *Aquatic Living Resources*, *23*, 303-313. doi:10.1051/alr/2010031

Doyle, W., Paukert, C., Starostka, A., & Hill, T. (2008). A comparison of four types of sampling gear used to collect shovelnose sturgeon in the Lower Missouri River. *Journal of Applied Ichthyology*, 24, 637-642. doi:10.1111/j.1439-0426.2008.01113.x

Dressel, S.C., (2005). Using poststratification to improve abundance estimates from multispecies surveys: a study of juvenile flatfishes. *Fishery Bulletin*, 103, 469-488.

Erisman, B. E., Allen, L. G., Claisse, J.T., Pondella,D. J., Miller, E. F., & Murray, J. H. (2011). The illusion of plenty: hyperstability masks collapses in two recreational fisheries that target fish spawning aggregations. *Canadian Journal of Fisheries and Aquatic Sciences,* 68, 1705–1716. doi:10.1139/f2011-090

Folmer, O., & Pennington, M. (2000). A statistical evaluation of the design and precision of the shrimp trawl survey off West Greenland. *Fisheries Research,* 49, 165-178. doi:10.1016/S0165-7836(00)00196-X

Fournier, D.A., Hampton, J., & Sibert, J.R. (1998). MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. *Canadian Journal of Fisheries and Aquatic Sciences,* 55, 2105-2116. doi:10.1139/f98-100

Francis, R. C. (2011). Data weighting in statistical fisheries stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences*, 68, 1124-1138. doi:10.1139/f2011-025

Francis, R.I.C.C., (1984). An adaptive strategy for stratified random trawl surveys. *New Zealand Journal of Marine and Freshwater Research*, 18, 59-71. doi:10.1080/00288330.1984.9516030

Fraser, H.M., Greenstreet, S.P., & Piet, G.J. (2007). Taking account of catchability in groundfish survey trawls: implications for estimating demersal fish biomass. *ICES Journal of Marine Science: Journal du Conseil.* 64, 1800-1819. doi:10.1093/icesjms/fsm145

Gavaris, S., (1988). An adaptive framework for the estimation of population size. *Canadian Atlantic Fisheries Scientific Advisory Committee. Research Document 88/29,* St. Andrews, New Brunswick.

Godø, O. R. (1994). Factors affecting the reliability of groundfish abundance estimates from bottom trawl surveys. In A. Fernö, and S. Olsen (Eds.), *Marine Fish Behaviour in Capture and Abundance Estimation* (pp. 166–199). Fishing New Books, Oxford.

Godø, O.R., Walsh, S.J., & Engås, A. (1999). Investigating density-dependent catchability in bottom-trawl surveys. *ICES Journal of Marine Science: Journal du Conseil.* 56, 292–298. doi:10.1006/jmsc.1999.0444

Gunderson, D.R. (1993). *Surveys of fisheries resources.* John Wiley & Sons.

Harbitz, A., Aschan, M., & Sunnanå, K., 1998. Optimal effort allocation in stratified, large area trawl surveys, with application to shrimp surveys in the Barents Sea. *Fisheries Research*, 37, 107-113. doi:10.1016/S0165-7836(98)00130-1

Harley, S.J., & Myers, R.A. (2001). Hierarchical Bayesian models of length-specific catchability of research trawl surveys. *Canadian Journal of Fisheries and Aquatic Sciences,* 58, 1569-1584. doi:10.1139/f01-097

Hiemstra, P. H., Pebesma, E. J., Twenhofel, C. J. W., & Heuvelink, G. B. M. (2009). Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network. *Computers and Geosciences,* 35, 1711–1721. doi:10.1016/j.cageo.2008.10.011

Hilborn, R., & Walters, C.J. (1992). *Quantitative fisheries stock assessment: Choice, dynamics and uncertainty.* Chapman and Hall. doi:10.1007/978-1-4615-3598-0

Hjellvik, V., Godø, O. R., & Tjostheim, D. (2002). Diurnal variation in bottom-trawl survey catches: does it pay to adjust? *Canadian Journal of Fisheries and Aquatic Sciences,* 59, 33-48. doi:10.1139/f01-193

Hoffman, J.C., Bonzek, C.F., & Latour, R.J. (2009). Estimation of bottom trawl catch efficiency for two demersal fishes, Atlantic croaker and white perch, in Chesapeake Bay. *Marine and Coastal Fisheries: Dynamics, Management, and Ecosystem Science,* 1, 255-269. doi:10.1577/C08-048.1

Hulson, P.J.F., Miller, S.E., Quinn, T.J., Marty, G.D., Moffitt, S.D., & Funk, F. (2008). Data conflicts in fishery models: incorporating hydroacoustic data into the Prince William Sound Pacific herring assessment model. *ICES Journal of Marine Science: Journal du Conseil,* 65, 25-43. doi:10.1093/icesjms/fsm162

Hutchings, J.A. (1996). Spatial and temporal variation in the density of northern cod and a review of hypotheses for the stock's collapse. *Canadian Journal of Fisheries and Aquatic Sciences,* 53, 943–962. doi:10.1139/f96-097

Ianelli, J. N., Honkalehto T., Barbeaux S., & Kotwicki S. (2014). Assessment of the walleye pollock stock in the Eastern Bering Sea. In, *Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions for 2014* (pp. 55−156). North Pacific Fishery Management Council, Anchorage, AK.

Ianelli, J.N., Honkalehto T., Barbeaux S., & Kotwicki, S. (2015). Assessment of the walleye pollock stock in the Eastern Bering Sea. In, *Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions for 2015* (pp. 53−152). North Pacific Fishery Management Council, Anchorage, AK.

Jakobsen, T., Korsbrekke, K., Mehl, S., & Nakken, O. (1997). Norwegian combined acoustic and bottom trawl surveys for demersal fish in the Barents Sea during winter. *ICES Conference and Meeting documents,* 1997N: 17.

Kasatkina, S.M. (1991). Midwater trawl catchability as an aspect of a quantitative assessment of krill biomass conducted using a trawl census survey. *Selected Scientific Papers,* 1991, 257-272.

Kotwicki S., Buckley T., Honkaletho T., & Walters G. (2005). Variation in the distribution of walleye pollock with temperature and implications for seasonal migration. *Fishery Bulletin,* 103, 574-587.

Kotwicki, S, Horne, J. K., Punt, A. E., & Ianelli J. N. (2015). Factors affecting the availability of walleye pollock to acoustic and bottom trawl survey gear and bottom trawl sampling. *ICES Journal of Marine Science: Journal du Conseil,* 72, 1425–1439. doi:10.1093/icesjms/fsv011

Kotwicki, S., De Robertis, A., Ianelli, J. N., Punt, A. E., & Horne, J. K. (2013). Combining bottom trawl and acoustic data to model acoustic dead zone correction and bottom trawl efficiency parameters for semi-pelagic species. *Canadian Journal of Fisheries and Aquatic Sciences,* 70, 208–219. doi:10.1139/cjfas-2012-0321

Kotwicki, S., Ianelli, J. N., & Punt, A. E. (2014). Correcting density-dependent effects in abundance estimates from bottom trawl surveys. *ICES Journal of Marine Science: Journal du Conseil,* 71, 1107–1116. doi:10.1093/icesjms/fst208

Kotwicki, S, Ressler, P.H., Ianelli J. N., Punt, A. E., & Horne, J. K. (2018). Combining data from bottom trawl and acoustic surveys to estimate an index of abundance for semipelagic species. *Canadian Journal of Fisheries and Aquatic Sciences,* 75, 60–71. doi:10.1139/cjfas-2016-0362

Kotwicki, S., Lauth, R.R., Williams, K., & Goodman, S. (2017). Selectivity ratio a useful tool for comparing size selectivity of multiple survey gears. *Fisheries Research*, 191, 76-86. doi:10.1016/j.fishres.2017.02.012

Laman, E.A., Kotwicki, S., & Rooper, C.N. (2015). Correlating environmental and biogenic factors with Pacific Ocean Perch (*Sebastes alutus*) abundance and distribution in the Aleutian Islands, Alaska. *Fishery Bulletin,* 113, 270-289. doi:10.7755/FB.113.3.4

Lindgren, F., & Rue, H. (2015). Bayesian Spatial and Spatio-temporal Modelling with R-INLA. *Journal of Statistical Software,* 63, 1–25. doi:10.18637/jss.v063.i19

Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 73, 423–498. doi:10.1111/j.1467-9868.2011.00777.x

Lingen, C.D., Checkley, D., Barange, M., Hutchings, L., & Osgood, K. (1998). Assessing the abundance and distribution of eggs of sardine, Sardinops sagax, and round herring, Etrumeus whiteheadi, on the western Agulhas Bank, South Africa, using a continuous, underway fish egg sampler. *Fisheries Oceanography,* 7, 35-47. doi:10.1046/j.1365-2419.1998.00050.x

Lumley, T. (2004). Analysis of complex survey samples. *Journal of Statistical Software*, 9, 1-19. doi:10.18637/jss.v009.i08

Mackinson, S., Turner, K., Righton, D., & Metcalfe, J.D. (2005). Using acoustics to investigate changes in efficiency of a sandeel dredge. *Fisheries Research,* 71, 357-363. doi:10.1016/j.fishres.2004.09.002

Martins, T. G., Simpson, D., Lindgren, F., & Rue, H. (2013). Bayesian computing with INLA: new features. *Computational Statistics & Data Analysis,* 67, 68–83. doi:10.1016/j.csda.2013.04.014

Maunder, M.N., & Piner, K.R. (2015). Contemporary fisheries stock assessment: many issues still remain. *ICES Journal of Marine Science: Journal du Conseil*, 72, 7-18. doi:10.1093/icesjms/fsu015

Maunder, M.N., & Punt, A. E. (2004). Standardizing catch and effort data: a review of recent approaches. *Fisheries Research,* 70, 141–159. doi:10.1016/j.fishres.2004.08.002

Maunder, M.N., & Punt, A.E. (2013). A review of integrated analysis in fisheries stock assessment. *Fisheries Research*, 142, 61-74. doi:10.1016/j.fishres.2012.07.025

Methot, R.D., & Wetzel, C.R. (2013). Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research,* 142, 86-99. doi:10.1016/j.fishres.2012.10.012

Munro, P. T., & Somerton, D. A. (2001). Maximum likelihood and non-parametric methods for estimating trawl footrope selectivity. *ICES Journal of Marine Science: Journal du Conseil*, 58, 220-229. doi:10.1006/jmsc.2000.1004

Munro, P. T., & Somerton, D. A. (2002). Estimating net efficiency of survey trawl for flatfishes. *Fisheries Research*, 55, 267–279. doi:10.1016/S0165-7836(01)00280-6

Nichol, D.G. (1998). Annual and between-sex variability of yellowfin sole, *Pleuronectes aspera*, spring-summer distributions in the eastern Bering Sea. *Fishery Bulletin*, 96, 547-561.

Nicholson, M.D., & Jennings, S. (2004). Testing candidate indicators to support ecosystem-based management: the power of monitoring surveys to detect temporal trends in fish community metrics. *ICES Journal of Marine Science: Journal du Conseil*, 61, 35-42. doi:10.1016/j.icesjms.2003.09.004

O'Driscoll, R., Rose, G., & Andersen, J. (2002). Counting capelin: a comparison of acoustic density and trawl catchability. *ICES Journal of Marine Science: Journal du Conseil*, 59, 1062–1071. doi:10.1006/jmsc.2002.1262

Ono, K., Shelton, A.O., Ward, E.J., Thorson, J.T., Feist, B.E., & Hilborn, R. (2016). Space-time investigation of the effects of fishing on fish populations. *Ecological Applications,* 26, 392-406. doi:10.1890/14-1874

Ono, K., Kotwicki, S., Dingsør, G.E., & Johnsen, E. (2018). Multispecies acoustic dead-zone correction and bias ratio estimates between acoustic and bottom-trawl data. *ICES Journal of Marine Science: Journal du Conseil,* 75, 361-373. doi:10.1890/14-1874

Overholtz, W.J., Jech, J.M., Michaels, W.L., Jacobson, L.D., & Sullivan, P.J. (2006). Empirical comparisons of survey designs in acoustic surveys of Gulf of Maine-Georges Bank Atlantic herring. *Journal of Northwest Atlantic Fishery Science*, 36, 127-144. doi:10.2960/J.v36.m575

Pennington, M., & Volstad, J.H. (1991). Optimum size of sampling unit for estimating the density of marine populations. *Biometrics*, 1991, 717-723. doi:10.2307/2532157

Pennington, M., & Godø, O. R. (1995). Measuring the effect of changes in catchability on the variance of marine survey abundance indices. *Fisheries research*, 23, 301-310. doi:10.1016/0165-7836(94)00345-W

Pennington, M., & Strømme, T. (1998). Surveys as a research tool for managing dynamic stocks. *Fisheries Research*, 37, 97-106. doi:10.1016/S0165-7836(98)00129-5

Pennington, M., Burmeister, L.M., & Hjellvik, V. (2002). Assessing the precision of frequency distributions estimated from trawl-survey samples. *Fishery Bulletin*, 100, 74-80.

Perry, A.L., Low, P.J., Ellis, J.R., & Reynolds, J.D. (2005). Climate change and distribution shifts in marine fishes. *Science,* 308, 1912-1915. doi:10.1126/science.1111322

Petitgas, P. (2001). Geostatistics in fisheries survey design and stock assessment: models, variances and applications. *Fish and Fisheries*, 2, 231-249. doi:10.1046/j.1467-2960.2001.00047.x

Punt, A. E., & Butterworth, D. S. (2003). Specifications and clarifications regarding the ADAPT VPA assessment/projection computations carried out during the September 2000 ICCAT West Atlantic bluefin tuna stock assessment session. *ICCAT Collective Volume of Scientific Papers,* 55, 1041–1054.

Rose, G. A., Gauthier, S., & Lawson, G. L. (2000). Acoustic surveys in the full monte: simulating uncertainty. *Aquatic Living Resources*, 13, 367–372. doi:10.1016/S0990-7440(00)01074-3

Schwarz, C. J., & Seber, G. A. (1999). Estimating animal abundance: review III. *Statistical Science*, 1999, 427-456.

Shelton, A. O., Thorson, J. T., Ward, E. J., & Feist, B. E. (2014). Spatial semiparametric models improve estimates of species abundance and distribution. *Canadian Journal of Fisheries and Aquatic Sciences,* 71, 1655–1666. doi:10.1139/cjfas-2013-0508

Sibly, R.M., Barker, D., Denham, M.C., Hone, J., & Pagel, M. (2005). On the regulation of populations of mammals, birds, fish, and insects. *Science*, 309, 607-610. doi:10.1126/science.1110760

Smith, S.G., Ault, J.S., Bohnsack, J.A., Harper, D.E., Luo, J., & McClellan, D.B. (2011). Multispecies survey design for assessing reef-fish stocks, spatially explicit management performance, and ecosystem condition. *Fisheries Research*, 109, 25-41. doi:10.1016/j.fishres.2011.01.012

Smith, S.J., & Lundy, M.J., (2006). Improving the precision of design-based scallop drag surveys using adaptive allocation methods. *Canadian Journal of Fisheries and Aquatic Sciences,* 63, 1639-1646. doi:10.1139/f06-063

Smith, S.J., & Gavaris, S. (1993). Improving the precision of abundance estimates of eastern Scotian Shelf Atlantic cod from bottom trawl surveys. *North American Journal of fisheries management*, 13, 35-47. doi:10.1577/1548-8675(1993)013<0035:ITPOAE>2.3.CO;2

Smith, S.J. (1990). Use of statistical models for the estimation of abundance from groundfish trawl survey data. *Canadian Journal of Fisheries and Aquatic Sciences*, 47, 894-903. doi:10.1139/f90-103

Somerton, D., Ianelli, J., Walsh, S., Smith, S., Godø, O.R., & Ramm, D. (1999). Incorporating experimentally derived estimates of survey trawl efficiency into the stock assessment process: a discussion. *ICES Journal of Marine Science: Journal du Conseil*, 56, 299-302. doi:10.1006/jmsc.1999.0443

Somerton, D.A., & Weinberg, K.L. (2001). The effect of speed through the water on footrope contact of a survey trawl. *Fisheries Research,* 53, 17-24. doi:10.1016/S0165-7836(00)00272-1

Somerton, D.A., Munro, P.T., & Weinberg, K.L. (2007). Whole-gear efficiency of a benthic survey trawl for flatfish. *Fishery Bulletin*, 105, 278-291.

Somerton, D.A., Weinberg, K.L., & Goodman, S.E. (2013). Catchability of snow crab (*Chionoecetes opilio*) by the eastern Bering Sea bottom trawl survey estimate using a catch comparison experiment. *Canadian Journal of Fisheries and Aquatic Sciences,* 70, 1699–1708. doi:10.1139/cjfas-2013-0100

Stauffer, G. (2004). NOAA protocols for groundfish bottom trawl surveys of the Nation's fishery resources. *U.S. Department of Commerce, NOAA. Technical Memorandum,* NMFS-F/SPO-65, 205 p.

Thorson, J. T., Shelton, A. O., Ward, E. J., & Skaug, H. J. (2015). Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. *ICES Journal of Marine Science: Journal du Conseil*, 72, 1297-1310. doi:10.1093/icesjms/fsu243

von Szalay, P.G. (2003). The feasibility of reducing the variance of fish relative abundance estimates by integrating CPUE data from tow demersal trawl surveys in the Gulf of Alaska. *Alaska Fishery Research Bulletin,* 10, 1-13.

von Szalay, P.G., Somerton, D.A., & Kotwicki, S. (2007). Correlating trawl and acoustic data in the eastern Bering Sea: A first step toward improving biomass estimates of walleye pollock (*Theragra chalcogramma*) and Pacific cod (*Gadus macrocephalus*)? *Fisheries Research*, 86, 77-83. doi:10.1016/j.fishres.2007.05.005

Walters, C., & Maguire, J.J. (1996). Lessons for stock assessment from the northern cod collapse. *Reviews in Fish Biology and Fisheries,* 6, 125–138. doi:10.1007/BF00182340

Ward, E. J., Jannot, J. E., Lee, Y. W., Ono, K., Shelton, A. O., & Thorson, J. T. (2015). Using spatiotemporal species distribution models to identify temporally evolving hotspots of species co-occurrence. *Ecological Applications,* 25, 2198–2209. doi:10.1890/15-0051.1

Weinberg, KL, & Kotwicki, S. (2008). Factors influencing net width and sea floor contact of a survey bottom trawl. *Fisheries Research,* 93, 265-279. doi:10.1016/j.fishres.2008.05.011

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution,* 1, 3–14. doi:10.1111/j.2041-210X.2009.00001.x

Table

Table 1. Values of years, mean sampling efficiency, and variance in sampling efficiency over which surveys were simulated.

Years	2005 – 2014
Mean sampling efficiency ($\bar{q}$)	0.01, 0.05, 0.1, 0.2, 0.4, 0.6, 0.8, 1, 1.5, 2, 2.5, 3
Variance in sampling efficiency (V(q))	0.00001, 0.01, 0.1, 0.2, 0.4, 0.6, 0.8, 1, 2
Density-dependent efficiency parameter (a)	1, 100, 500, 2000, 50000

Figure captions

Figure 1. Simulated map of annual pollock density distribution in the EBS between 2005 and 2014. The map is based on a 1-km nominal grid and simulation runs were only performed within the EBS survey area. The dots in each map indicate the survey location.

Figure 2. Example (2010) of the relationship between relative error of sample derived abundance index (AI; (i.e. $\overline{u_s}$ - $\overline{u_T}$)/ $\overline{u_T}$) in relation to survey sampling efficiency ($\bar{q}$; x-axis) and variance in sampling efficiency ($V(q)$; panels). Grey line represents relative error equal to 0. Note that the abundance index is mean unbiased, and the effect of the sampling efficiency results in skewed index of abundance distribution.

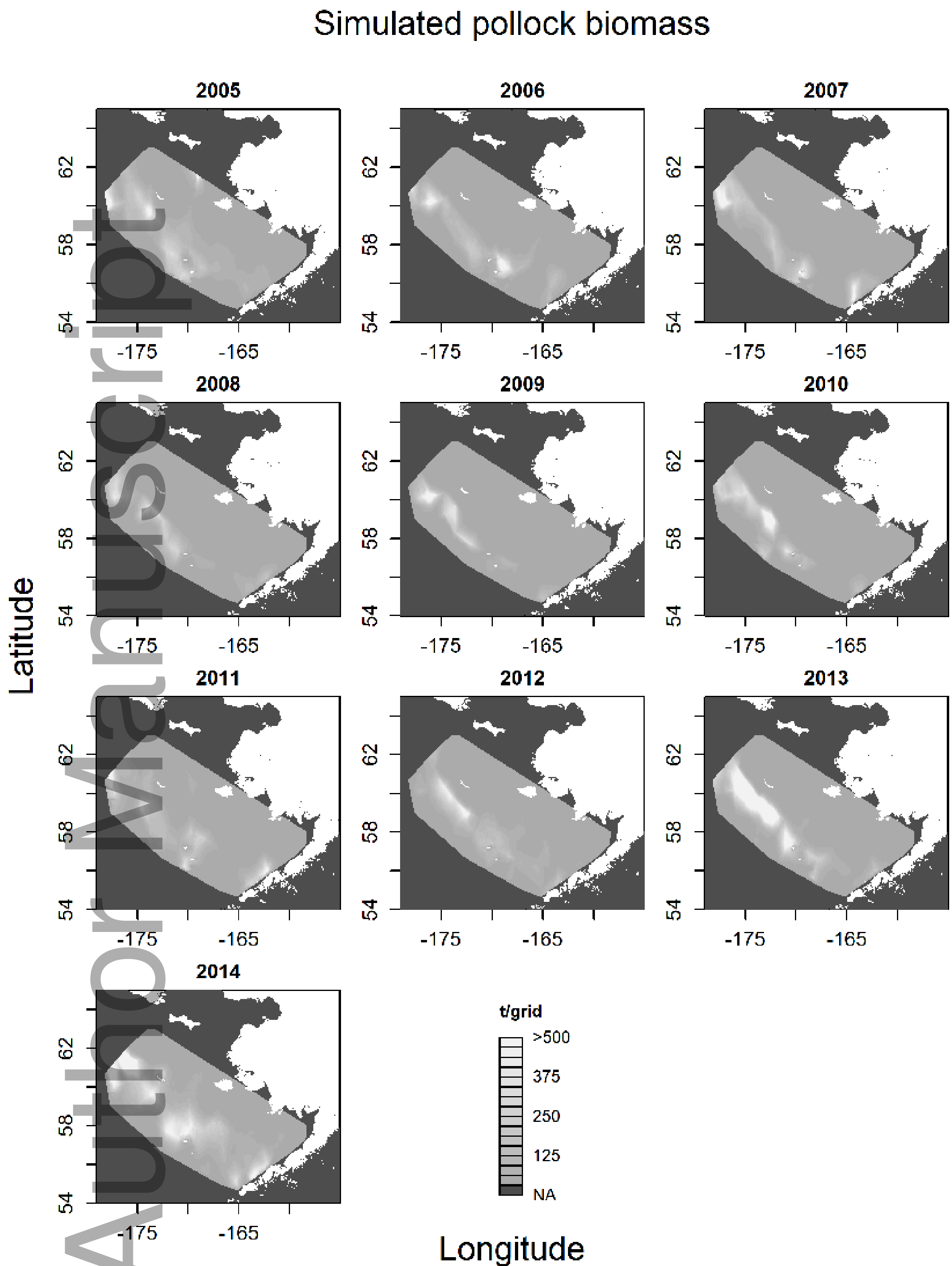
Figure 3. Example (2010) of the relationship between survey coefficient of variation (CV) of the abundance index ($\sigma_S^2/\overline{u_s}$) in relation to survey sampling efficiency ($\overline{q}$; x-axis) and variance in sampling efficiency ($V(q)$; panels).

Figure 4. Example (2010) of the relationship between relative error of sample standard deviation (SD) of abundance index (i.e. $(\sigma_s - \sigma_T)/\sigma_T$)) in relation to survey sampling efficiency ($\overline{q}$; x-axis) and variance in sampling efficiency ($V(q)$; panels). Grey line represents relative error equal to 0.

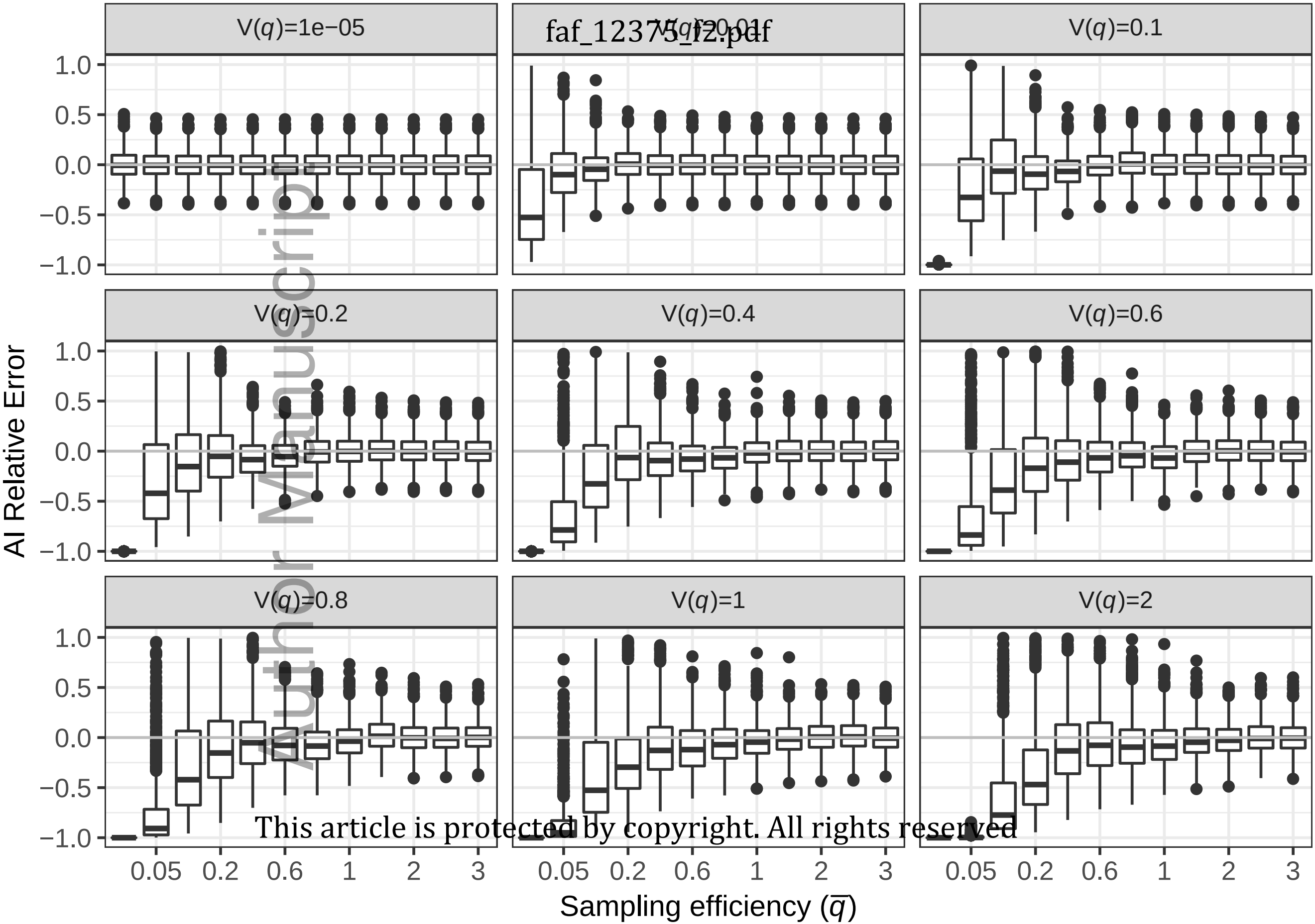
Figure 5. Example (2010) of the relationship between coefficient of variation (CV) of survey standard deviation (SD; i.e. standard deviation(σ_s) / mean(σ_s)) in relation to survey sampling efficiency ($\overline{q}$; x-axis) and variance in sampling efficiency ($V(q)$; panels). Black line represents expected of SD in case when sampling efficiency is constant.

Figure 6. Example (2010) of the relationship between survey abundance index (AI) in relation to density-dependent sampling efficiency (x-axis) and variance in sampling efficiency ($V(q)$; panels). Red line represents AI in case when sampling efficiency is not density-dependent.

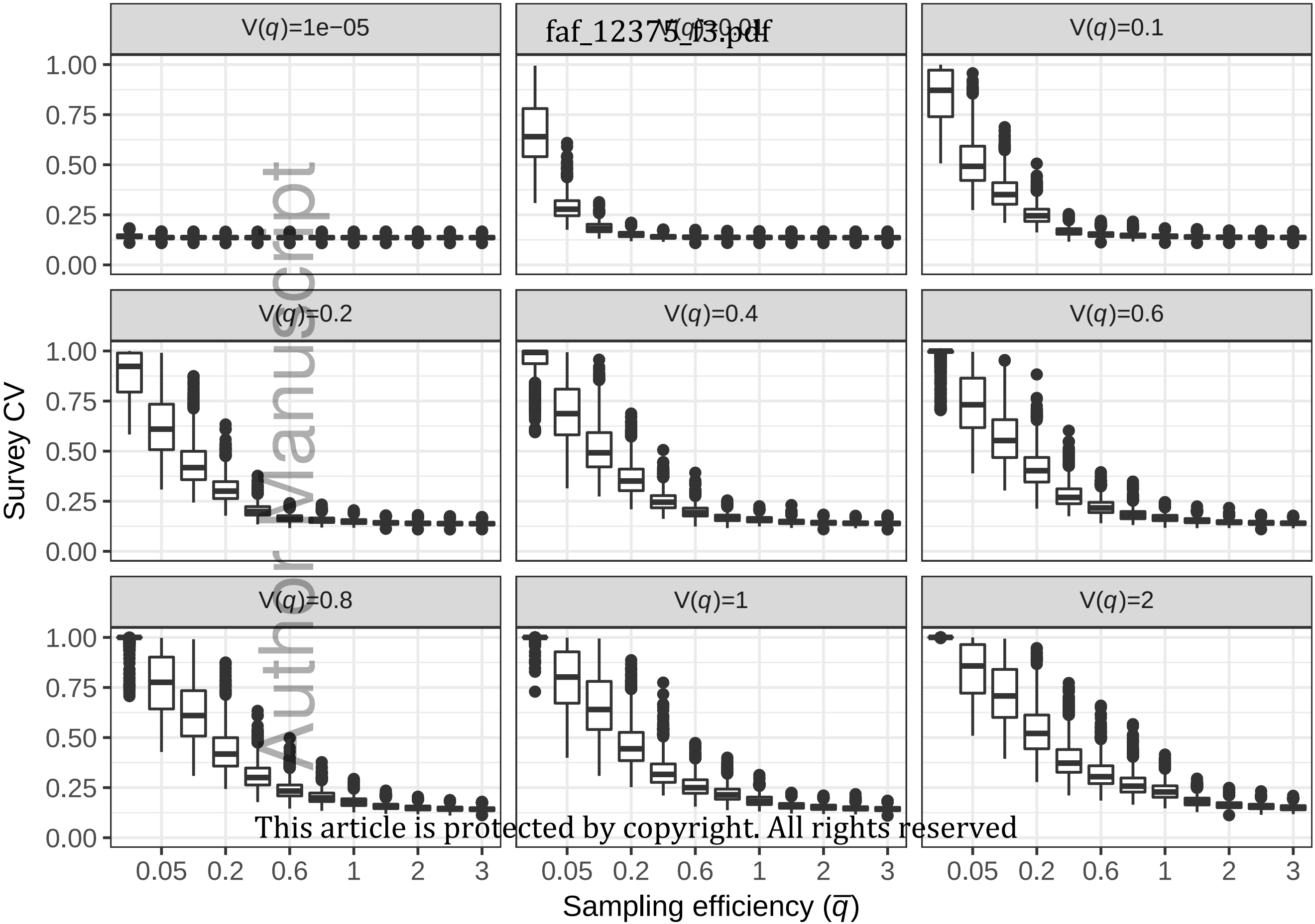
Figure 7. Example (2010) of the relationship between coefficient of variation of survey abundance index in relation to density- dependent sampling efficiency (x-axis) and variance in sampling efficiency ($V(q)$; panels). Red line represents expected CV in case when sampling efficiency is not density-dependent.

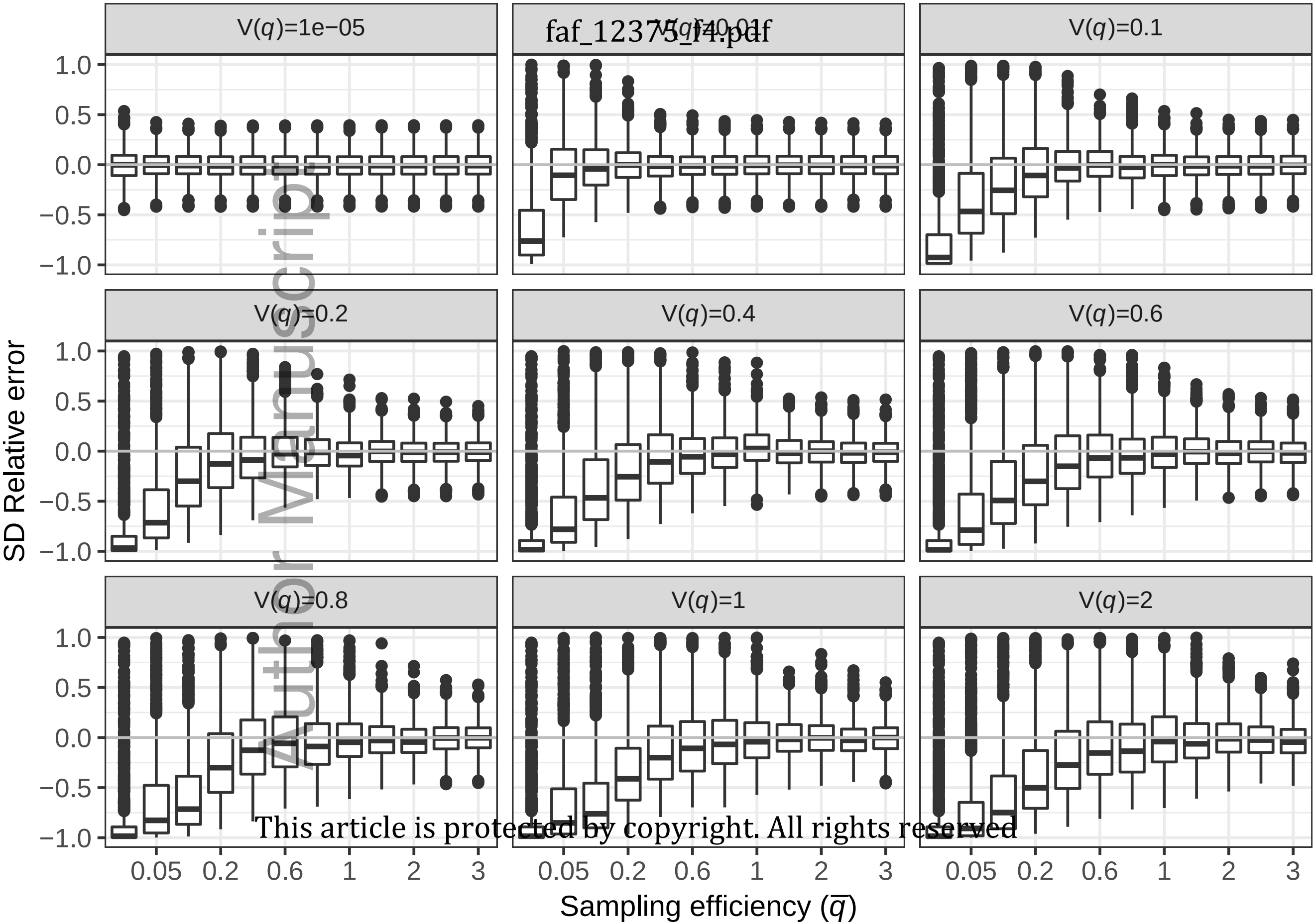


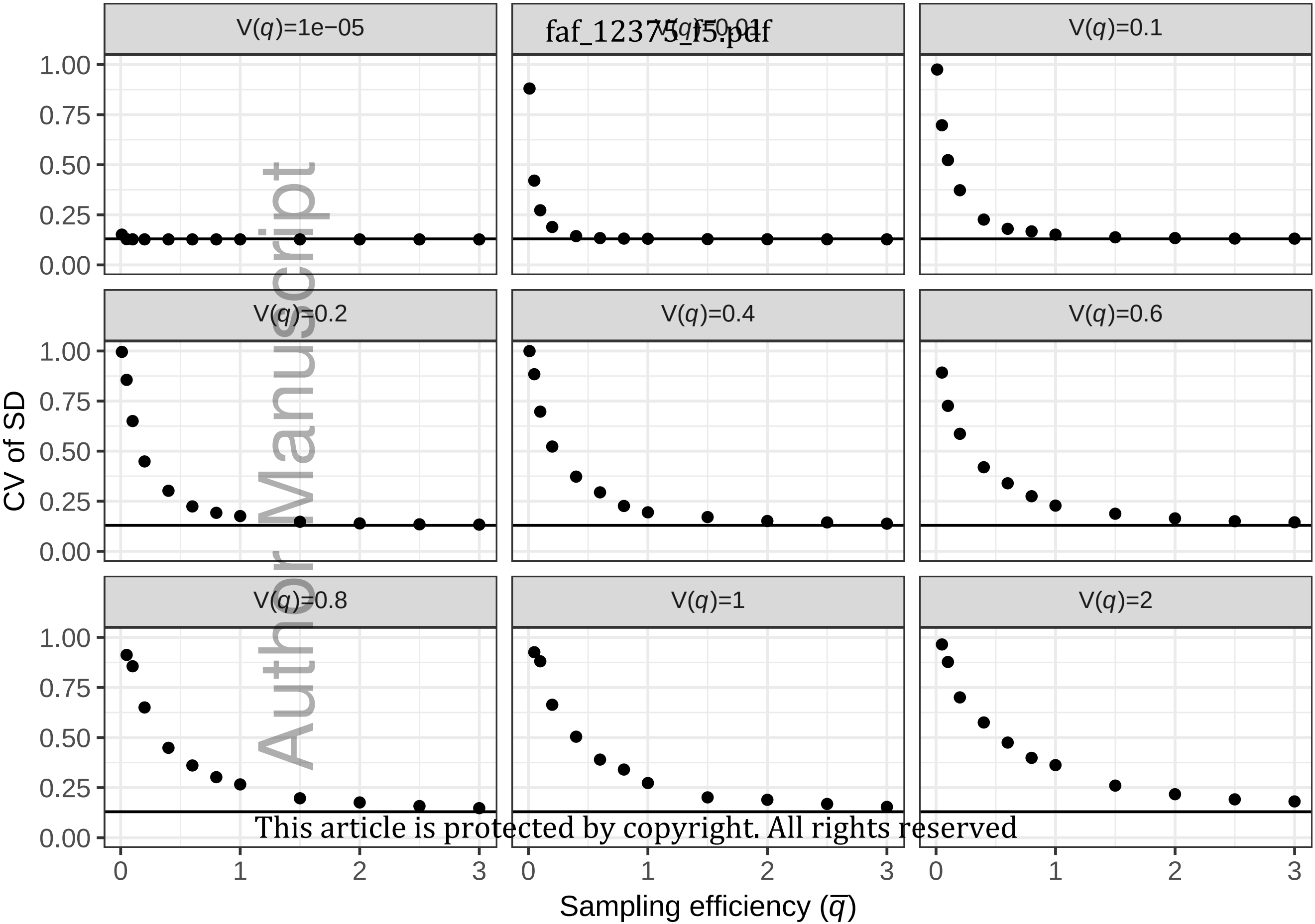
faf_12375_f1.png

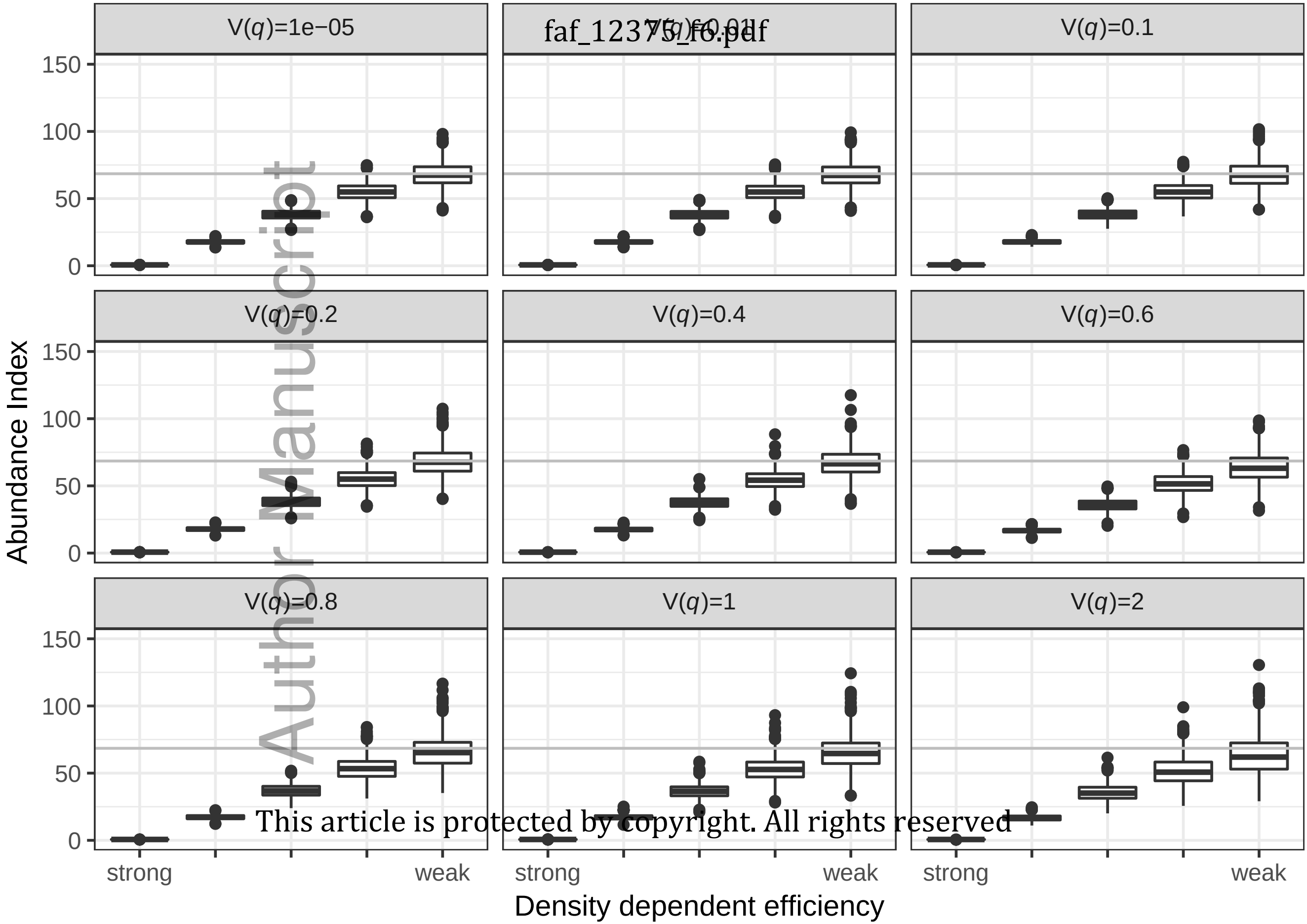

V(q)=1e−05
V(q)=0.01
V(q)=0.1
V(q)=0.2
V(q)=0.4
V(q)=0.6
V(q)=0.8
V(q)=1
V(q)=2
AI Relative Error
1.0
0.5
0.0
−0.5
−1.0
0.05
0.2
0.6
1
2
3
Sampling efficiency (q̄)

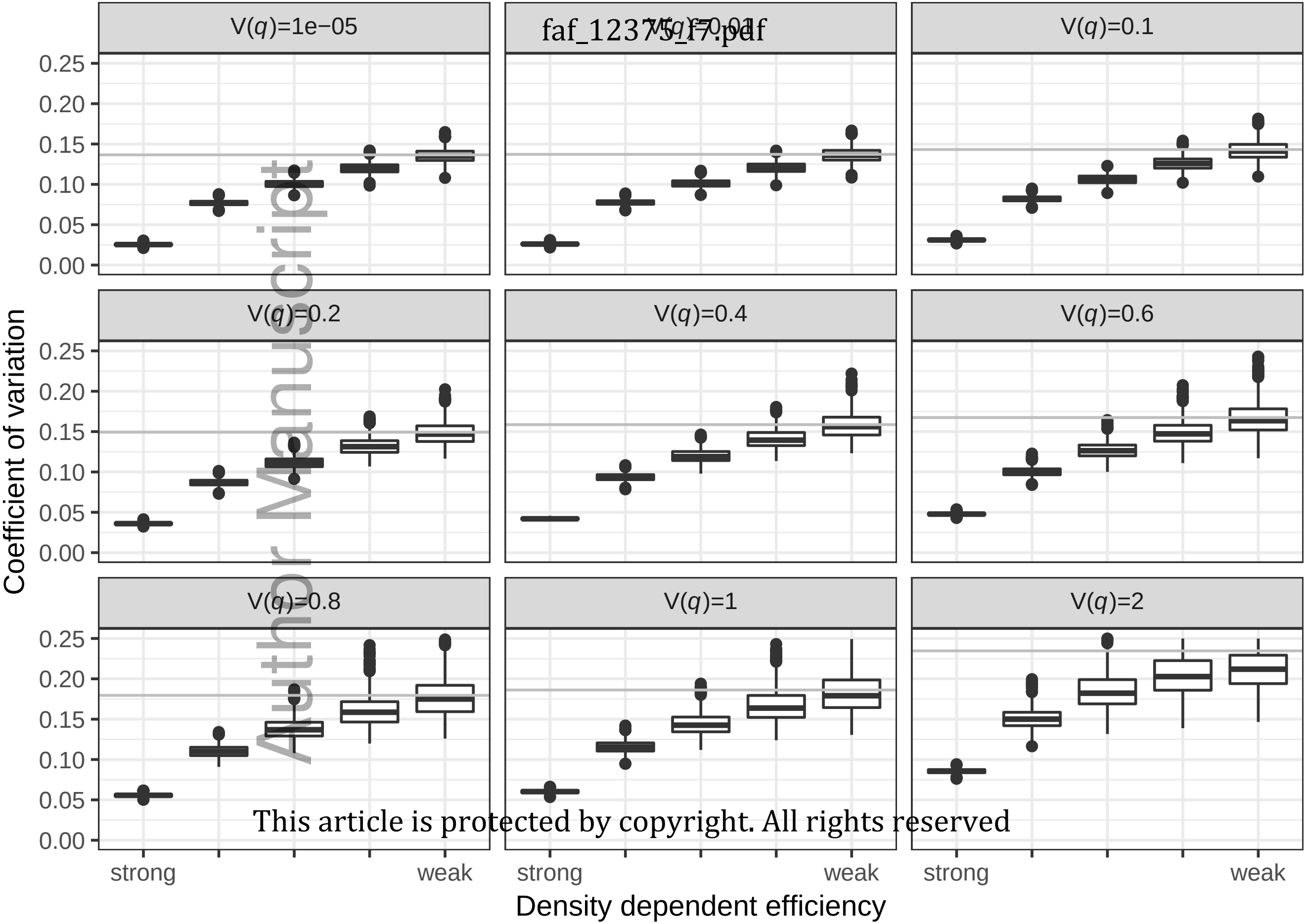
faf_12375_f3.pdf




V(q)=1e−05
V(q)=0.01
V(q)=0.1
V(q)=0.2
V(q)=0.4
V(q)=0.6
V(q)=0.8
V(q)=1
V(q)=2
SD Relative error
1.0
0.5
0.0
−0.5
−1.0
0.05
0.2
0.6
1
2
3
Sampling efficiency ($\overline{q}$)


V(q)=1e−05
V(q)=0.01
V(q)=0.1
V(q)=0.2
V(q)=0.4
V(q)=0.6
V(q)=0.8
V(q)=1
V(q)=2
CV of SD
Sampling efficiency ($\overline{q}$)
1.00
0.75
0.50
0.25
0.00
0
1
2
3


V(q)=1e−05
V(q)=0.01
V(q)=0.1
V(q)=0.2
V(q)=0.4
V(q)=0.6
V(q)=0.8
V(q)=1
V(q)=2
Abundance Index
0
50
100
150
strong
weak
Density dependent efficiency


V(q)=1e−05
V(q)=0.01
V(q)=0.1
V(q)=0.2
V(q)=0.4
V(q)=0.6
V(q)=0.8
V(q)=1
V(q)=2
0.25
0.20
0.15
0.10
0.05
0.00
Coefficient of variation
strong
weak
Density dependent efficiency