Estimation of additional variance for Antarctic minke whales based on the abundance estimates from the revised OK method

Toshihide Kitakado[1] and Hiroshi Okamura[2]

[1] Tokyo University of Marine Science and Technology, Minato, Tokyo 108-8477, Japan

[2] National Research Institute of Far Seas Fisheries, Fisheries Research Agency, Kanagawa 236-8648, Japan

The corresponding author's email address: kitakado@kaiyodai.ac.jp

ABSTRACT

Approaches for estimating the additional variance attributed to inter-annual changes in whale distribution are outlined. Two different options, fixed and random effects models, can be assumed for annual rates of increase in abundance. Also, two models with and without adjustment of expectation in log-linear model for the abundance estimates are described although they essentially draw difference only in the prediction of abundance. The approaches are then applied to the actual IDCR/SOWER abundance estimates derived by the revised OK method (Okamura and Kitakado 2009). Two sets of abundance estimates using different use of duplicate identification, "Definite" and "Definite + Possible", were employed. The estimated values of the additional SD in log-scale of abundance under the fixed effect model without mean adjustment were 0.576 (SE= 0.170) for the abundance estimates using "Definite" duplicate ID and 0.583 (SE= 0.171) for the abundance estimates with "Definite + Possible". The standard errors of the best OK abundance estimates are inflated by taking into account the extent of the additional SD estimated from the fixed effect model (Okamura and Kitakado 2009).

INTRODUCTION

The issue of the additional variance arises from the fact that the estimated sampling variances for the abundance estimates do not account for yearly variation in abundance levels in areas due to inter-annual changes in distribution of whale populations. If the additional variance is ignored, uncertainty on abundance estimates tends to be underestimated.

To estimate the additional variance, some survey blocks, which we call the additional variance blocks (AVBs), must be defined beforehand as basic units for comparison of abundance estimates at least in two different year's surveys. Thanks to the two circumpolar series in the IDCR/SOWER surveys, the requirement of replication of surveys is satisfied. The six Management Areas themselves unfortunately does not work as a unit of the AVBs because any Management Areas were not

covered in a single year survey in CP3. Therefore, each Management Area must be divided into several blocks according to actual survey strata when the additional variance is estimated. The definition of AVBs was originally discussed in SC55 and finalized in the intersessional workshop held in the St.Andrews in 2009 (IWC 2009 and Figure 1). The estimation method was also finally specified in the intersessional workshop (IWC 2009). Here, the method is briefly explained with a remark and then applied to the actual IDCR/SOWER abundance estimates for the AVBs derived by the revised OK method (Okamura and Kitakado 2009).

METHODS

The agreed method

Let $\hat{N}_{b,y}$ be an abundance estimate in block b in year y, and $\tilde{N}_{b,y}$ be the true number of whales present in block b in year y. We suppose that the abundance estimate $\hat{N}_{b,y}$ is unbiased for $\tilde{N}_{b,y}$ and log-normally distributed as

$$\log \hat{N}_{b,y} = \log \tilde{N}_{b,y} + \varepsilon_{b,y}. \tag{1}$$

A vector $\boldsymbol{\varepsilon} = (\dots, \varepsilon_{b,y}, \dots)$, which expresses sampling errors in the abundance estimates, is assumed to have a multivariate normal distribution

$$\boldsymbol{\varepsilon} \sim N(-0.5diag(\tilde{V}), \tilde{V}), \tag{2}$$

where $\tilde{V}$ is a matrix transformed from an estimated variance-covariance matrix $V = (V_{b_i,y_i;b_j,y_j})$ as

$$\tilde{V}_{b_i,y_i;b_j,y_j} = \log\left(1 + \frac{V_{b_i,y_i;b_j,y_j}}{N_{b_i,y_i} N_{b_j,y_j}}\right) \tag{3}$$

so that $E[\hat{N}_{b,y}] = \tilde{N}_{b,y}$ and $Cov[\hat{N}_{b_i,y_i}, \hat{N}_{b_j,y_j}] = V_{b_i,y_i;b_j,y_j}$.

We also assume that the true abundance $\tilde{N}_{b,y}$ is expressed as a log-linear random-effect model with an annual rate of increase and a block effect:

$$\log \tilde{N}_{b,y} = \log N_{b,y} + \gamma_{b,y}, \tag{4}$$

$$\log N_{b,y} = \phi_{A(b)}(y - 2000) + \log \mu_b, \tag{5}$$

where, $\phi_{A(b)}$ is a growth rate in Management Area A to which block b belongs, and μ_b is a block-effect. The terms $\gamma_{b,y}$s are for expressing random effects due to the inter-annual changes in whale distribution and assumed to be independently and identically distributed as a normal distribution,

$$\gamma_{b,y} \sim N(-0.5\log(1+\tau^2), \log(1+\tau^2)), \tag{6}$$

so that $E[\tilde{N}_{b,y}] = N_{b,y}$ and $Var[\tilde{N}_{b,y}] = \tau^2 N_{b,y}^2$, where τ^2 is the additional CV.

The growth rates, $\phi_A(A =$ I,II,...,VI), can be assumed fixed effects. However, information on the growth rate drawn by the data in a single Management Area must be limited. Therefore we also explore a random-effect model as

$$\phi_{MA} \sim N(\phi_m, \sigma^2). \tag{7}$$

The reason behind the consideration of the random effects is "borrowing strength" of information on the annual rate of increase from other Management Areas.

The likelihood function for $\boldsymbol{\nu} = (\log \mu_1, \ldots, \log \mu_B)$, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_A)$ and τ is given by

$$L(\boldsymbol{\nu}, \boldsymbol{\phi}, \tau) = \frac{|D|^{1/2}}{(2\pi)^{p/2}} \, e^{-\frac{1}{2}\boldsymbol{y}^T D^{-1} \boldsymbol{y}}, \tag{8}$$

where $\boldsymbol{y} = \log \hat{\boldsymbol{N}} - \log \boldsymbol{N} + \frac{1}{2} diag(\tilde{V}) + \frac{1}{2} log(1+\tau^2)\mathbf{1}$, $D = \tilde{V} + \log(1+\tau^2)I$ and p is the number of data. The additional CV, τ, is estimated through an integrated likelihood function,

$$L(\tau) = \int L(\boldsymbol{\nu}, \boldsymbol{\phi}, \tau) d\boldsymbol{\nu} d\boldsymbol{\phi} \quad \text{(for fixed effect } \phi \text{ model)}, \tag{9}$$

or

$$L(\tau, \sigma) = \int L(\boldsymbol{\nu}, \boldsymbol{\phi}, \tau) f(\boldsymbol{\phi}|\phi_m, \sigma) d\boldsymbol{\nu} d\boldsymbol{\phi} d\phi_m \quad \text{(for fixed effect } \phi \text{ model)}, \tag{10}$$

where $f(\boldsymbol{\phi}|\phi_m, \sigma)$ is the probability density function of $\boldsymbol{\phi}$. A Laplace approximation of the integrated likelihood functions and the parameter estimation were conducted using ADMB-RE (Skaug and Fournier 2006).

The Management Area abundance and the circumpolar abundance can be estimated by summing up the predicted abundance given in the formula (5).

An alternative method

In the model above, the distribution of $\boldsymbol{\varepsilon}$ was given so that $\hat{N}_{b,y}$ was unbiased for $\tilde{N}_{b,y}$ and the standard deviation of $\hat{N}_{b,y}$ was exactly expressed as the product of its conditional CV and expectation. These requirements were also applied for γs. However, such adjustment for the distributions may produce larger values of predicted abundance estimates. The model with no adjustment of expectation can be given below:

$$\boldsymbol{\varepsilon} \sim N(0, \tilde{V}), \tag{11}$$

where $\tilde{V}$ is defined as

$$\tilde{V}_{b_i, y_i; b_j, y_j} = \frac{V_{b_i, y_i; b_j, y_j}}{N_{b_i, y_i} N_{b_j, y_j}}, \tag{12}$$

and

$$\gamma_{b,y} \sim N(0, \delta^2), \tag{13}$$

where δ is the additional standard deviation in log-scale.

Remark

The extent of additional variance estimated by the agreed model is almost same as that by the alternative model although the models have slightly different parameterizations. However, when the abundance is predicted using the models above, the difference is not negligible if the extent

of additional variance is large. The estimate of additional variance might be large for Antarctic minke whales, so the use of the alternative model must be safe for the prediction purpose.

RESULTS and DISCUSSION

Additional variance

We employed two sets of abundance estimates derived from the revised OK method (Okamura and Kitakado 2009) with different use of the duplicate identification, "Definite" and "Definite + Possible". The underlying abundance estimates used in this analysis were given in Table 1 ("Definite") and Table 2 ("Definite + Possible"). Table 3 showed the estimates of τ in the agreed model and δ in the alternative model for each of the abundance sets. The extent of the additional variance was high in each case, which means greater inter-annual variation of whale distribution.

In both the agreed and alternative models, the variance σ of the random effect for ϕ was estimated as 0, which means that all the parameter estimates $\phi_A(A = \text{I,II,...,VI})$ shrank toward the estimate of the mean parameter ϕ_m. This might be attributed to large uncertainty in the information of the annual rate of growth in each Management Area. However, the actual extent of change in the abundance estimate in surveyed area is expected to differ among the Management Areas, and therefore we use the estimate of additional variance in the fixed effect model.

The difference between the agreed and alternative models is essentially the treatment of mean and therefore it is a natural consequence that they gave almost same extents of the additional variance although the abundance tended to be estimated larger when the agreed model was applied. Taking the consistency of methodlogy into consideration, we used, in our best estimates (Okamura and Kitakado 2009), the estimate of additional variance given by the alternative method with fixed effect model.

Missing whales in sea-ice

In this analysis, we employed the abundance estimates in the surveys conducted in the open water north of the ice-edge. Howrver, whales may present in the pack ice. In fact, a survey in the Weddell Sea by a vessel and an autonomous underwater vehicle demonstrated that the krill density was fivefold greater than that of open water in an austral summer season (Brierley et al. 2002). Consequently, a large amount of krill would be present in pack ice in the Weddell Sea. This might be the case in other regions to a greater or lesser extent. This also implies that some extent of whales would also be in pack ice to feed on the krill there, and therefore the surveys may have remarkable chances to miss the whales in such unsurveyed.

The model outlined so far can be extended for simultaneous estimation of the numbers of whales in sea-ice. Let r be the proportion, which is a function of the unsurveyed area south of the ice-edge. Then, an extended model with ice-effects is given as follows:

$$\log \hat{N}_{b,y} = \log(1 - r(x_{b,y}; \alpha)) + \phi_{MA(b)}(y - 2000) + \log \mu_b + \gamma_{b,y} + \varepsilon_{b,y}.$$

The size of the unsurveyed area south of the ice-edge can be calculated by information on ice

concentration. If we could find better sea-ice indices and the function r, we might be able to explain reasons why the abundance estimates in the two survey periods were different.

ACKNOWLEDGEMENT

The authors thank the participants of the intersessional workshop in St.Andrews in 2009 for their helpful discussion.

REFERENCES

Brierley, A. et al. 2002. Antarctic Krill Under Sea Ice: Elevated Abundance in a Narrow Band Just South of Ice Edge Science 295: 1890 - 1892.

International Whaling Commission. 2009. Report of SOWER abundance workshop held in St.Andrews.

Okamura, H. and Kitakado, T. 2009, Abundance estimates and some diagnosis of Antarctic minke whales from historical IDCR/SOWER survey data using the revised OK method, Paper SC/61/IA6. XXX pp.

Skaug, H.J. and Fournier, D.A. 2006. Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. Comp. Statist. Data Anal., 51: 699-709.

Table 1. Abundance for AVBs estimated by the revised OK method using “Definite” duplicate ID (upper: estimate, bottom:CV).

		CPII						CPIII											
Area	Sector	1986	1987	1988	1989	1990	1991	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
I	W					31,166										9,912			
						0.365										0.377			
	M					70,860			33,497										
						0.283			0.496										
	E					23,221			64,977						8,260				
						0.324			0.345						0.316				
II	W		38,141										60,955						
			0.270										0.440						
	E		147,086									83,870							
			0.204									0.440							
III	W			113,588				34,785											
				0.326				0.207											
	E			12,335						41,098									
				0.434						0.269									
IV	W				26,482					13,760									
					0.440					0.297									
	E				58,407									39,445					
					0.227									0.466					
V	W	137,349															20,169		
		0.283															0.241		
	M	157,164																43,448	
		0.351																0.199	
	E	326,088																	159,771
		0.191																	0.146
VI	W						54,401				55,459								
							0.592				0.230								
	E						70,272									28,231			
							0.401									0.296			

Table 2. Abundance for AVBs estimated by the revised OK method using “Definite + Possible” duplicate ID (upper: estimate, bottom:CV).

		CPII						CPIII											
Area	Sector	1986	1987	1988	1989	1990	1991	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
I	W					30,702										9,838			
						0.361										0.855			
	M					68,593			30,022										
						0.286			0.415										
	E					22,504			55,383						6,990				
						0.330			0.348						0.284				
II	W		37,945										56,846						
			0.275										0.432						
	E		146,208									78,287							
			0.203									0.432							
III	W			109,759				31,649											
				0.319				0.197											
	E			11,676						37,965									
				0.436						0.264									
IV	W				24,665					12,517									
					0.432					0.283									
	E				54,059									36,990					
					0.216									0.458					
V	W	128,522															20,094		
		0.272															0.243		
	M	139,763																43,038	
		0.346																0.200	
	E	294,297																	157,429
		0.182																	0.145
VI	W						52,855				54,474								
							0.585				0.227								
	E						69,478									27,435			
							0.414									0.280			

Table 3. Summary of parameter estimates on the additional variance

Data set	Adjustment of expectation	Parameter	Fixed effect			Random effect		
			Estimate	SE	CV	Estimate	SE	CV
Definite	Yes	τ	0.630	0.215	0.341	0.538	0.157	0.292
	No	δ	0.576	0.170	0.295	0.498	0.131	0.263
Definite + Possible	Yes	τ	0.637	0.218	0.342	0.549	0.159	0.290
	No	δ	0.583	0.171	0.293	0.506	0.132	0.261

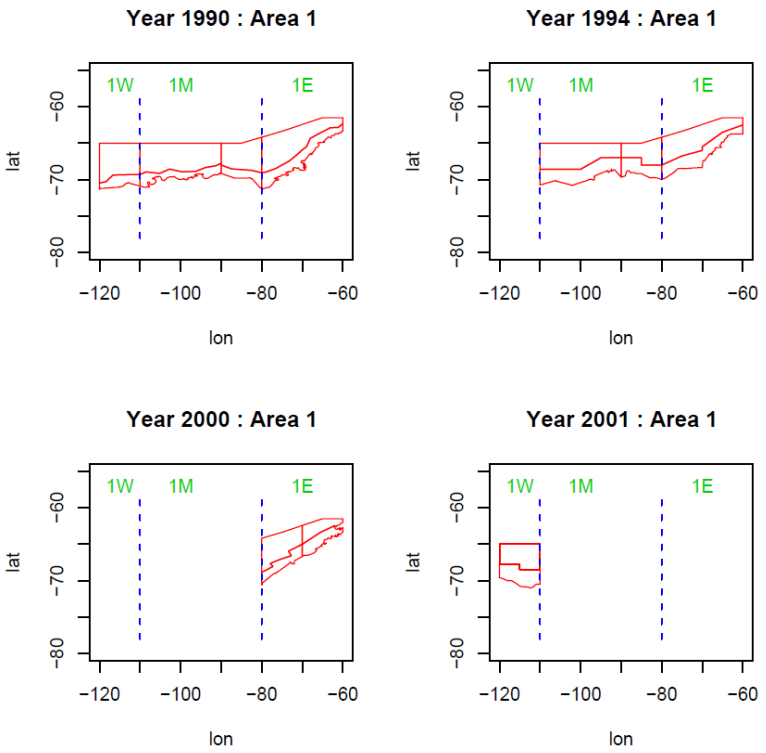

Figure 1(a) AVB map for Area I

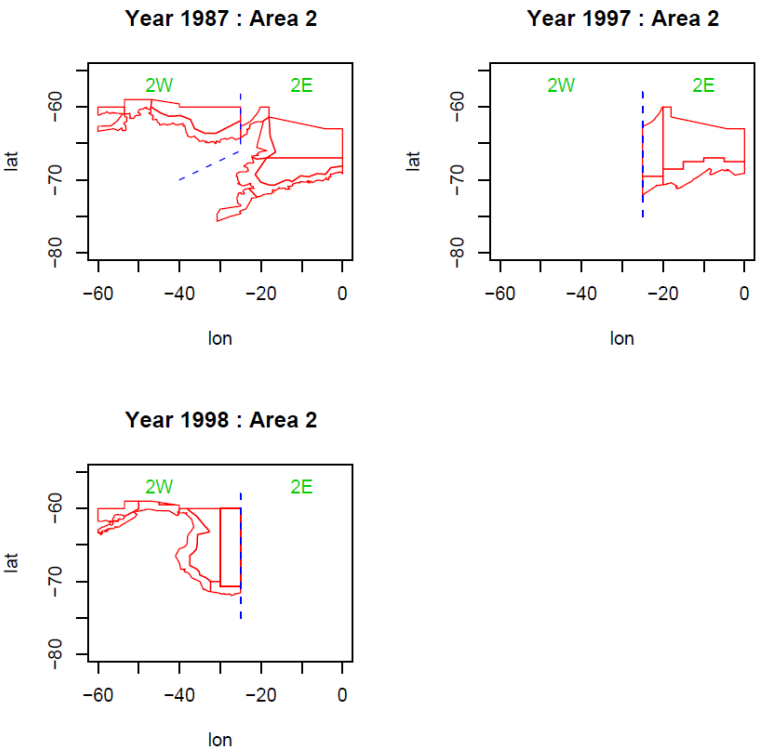

Figure 1(b) AVB map for Area II

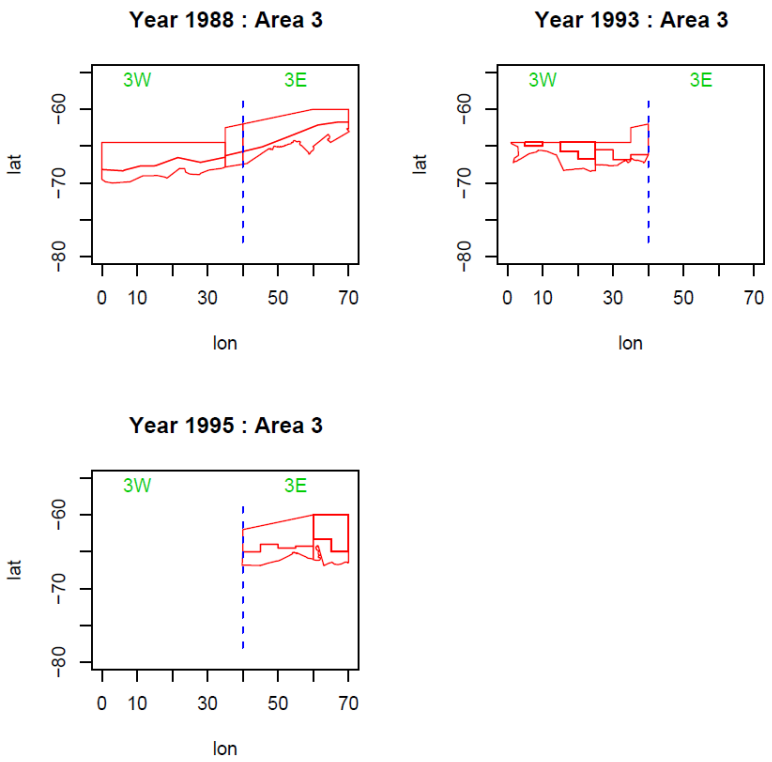

Figure 1(c) AVB map for Area III

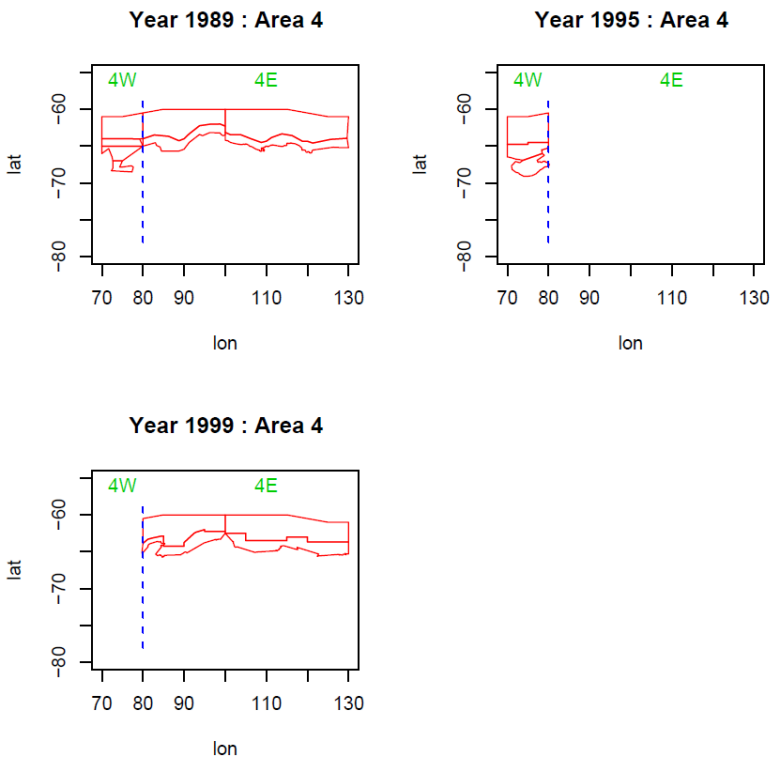

Figure 1(d) AVB map for Area IV

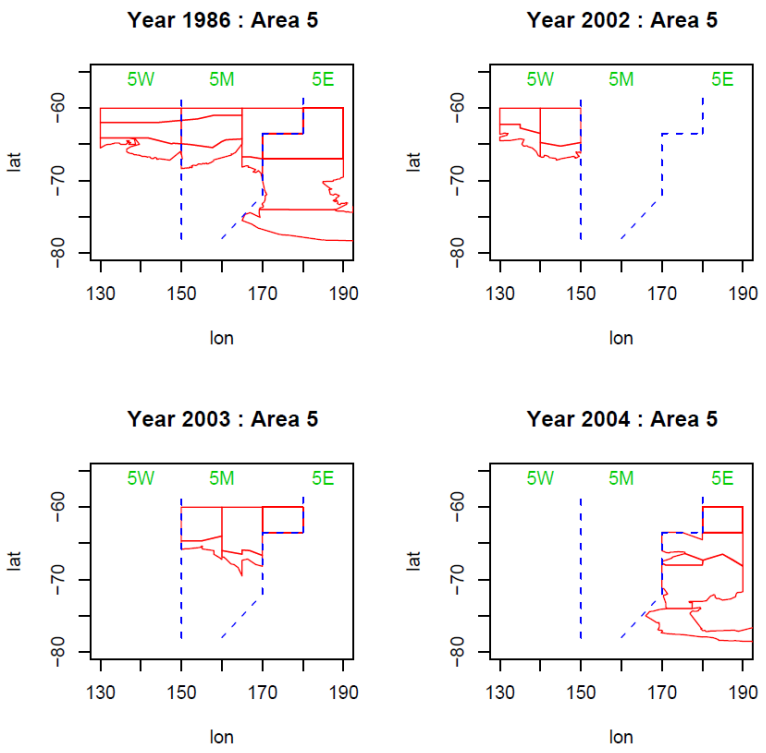

Figure 1(e) AVB map for Area V

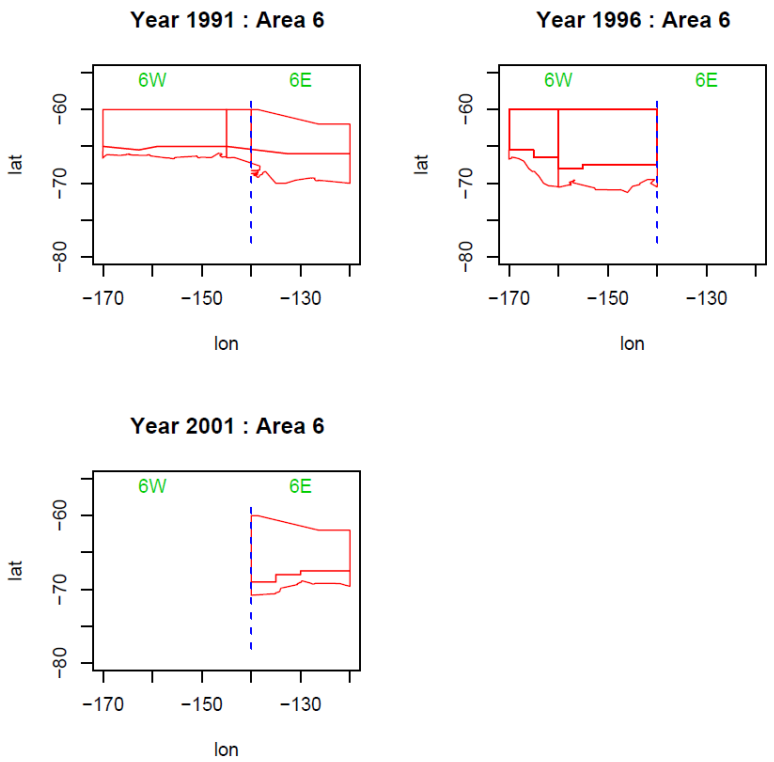

Figure 1(f) AVB map for Area VI