

Avaliação dos estimadores de amostragem por distâncias usando processos de Poisson não homogêneos

Anabela Afonso

CIMA e Departamento de Matemática da Universidade de Évora - aafonso@uevora.pt

Russell Alpizar-Jara

CIMA e Departamento de Matemática da Universidade de Évora - alpizar@uevora.pt

Resumo: Neste artigo avaliamos, através de um estudo de simulação, o desempenho dos estimadores convencionais da amostragem por distâncias quando a distribuição espacial dos indivíduos na área em estudo segue um processo de Poisson não homogêneo, nas seguintes situações: 1) os indivíduos na área em estudo estão fixos, mas obtidos de uma população não homogênea, e em cada simulação colocaram-se transectos aleatórios; 2) os transectos estão fixos e os indivíduos são gerados aleatoriamente em cada simulação, a partir de processos de Poisson não-homogêneos.

Palavras-chave: amostragem por distâncias, enviesamento, função intensidade, processos de Poisson.

Abstract: In this paper we assess, through a simulation study, the performance of conventional distance sampling estimators when the spatial distribution of individuals in the study area follows a non-homogeneous Poisson process: 1) when individuals in the study area are fixed, but belong to a non-homogenous population (in each simulation different transects are randomly allocated); 2) when transects are fixed and individuals are randomly generated in each simulation by a non-homogenous Poisson process.

Keywords: distance sampling, bias, intensity function, Poisson processes.

1 Introdução

Na teoria convencional da amostragem por distâncias, consoante a abordagem considerada, assume-se que na área coberta a distância dos indivíduos ao transecto tem distribuição uniforme ou relaxa-se este pressuposto garantindo a colocação aleatória dos transectos na área em estudo. A abundância de indivíduos, N , na área de estudo, A , é estimada por

$$\hat{N} = \frac{n}{\hat{P}_c \hat{P}_a}. \quad (1)$$

sendo n o número de indivíduos detectados na área coberta a , P_c a probabilidade de cobertura e P_a a probabilidade de detecção que é estimada a partir função

de detecção $g(x) = P(\text{detectar} | \text{distância } x)$ ajustada às distâncias observadas (Buckland et al., 2004).

Recentemente foram publicados alguns trabalhos com críticas a esta técnica (Barry e Welsh, 2001; Melville e Welsh, 2001) e outros que concluem que os métodos da amostragem por distâncias têm um bom desempenho (Fewster e Buckland, 2004; Fewster et al., 2005). Afonso e Alpizar-Jara (2007a) realizaram um estudo de simulação, com populações não homogêneas fixas e função de detecção conhecida, e concluíram, para as populações consideradas, que o estimador convencional da amostragem por distâncias é não viesado mas com variância muito sensível à distribuição espacial dos indivíduos na área em estudo. Verificaram ainda que tanto a disposição sistemática dos transectos como o aumento do número de transectos permitem reduzir a variabilidade deste estimador. Neste trabalho, vamos generalizar estes resultados considerando que a distribuição espacial dos indivíduos é estocástica. Em Afonso e Alpizar-Jara (2007b) é apresentada uma versão resumo deste trabalho.

2 Simulação

Nas simulações realizadas consideraram-se populações com $N = 1000$ indivíduos distribuídos numa área de estudo quadrada $A = (0; 1) \times (0; 1)$.

De um modo geral, a localização dos indivíduos é determinada por diversos factores como sejam as características topográficas do terreno, tipo de *habitat*, relacionamento entre os indivíduos, entre muitos outros factores relacionados com o meio ambiente ou intrínsecos dos elementos da população. Deste modo, as localizações (x, y) dos indivíduos foram geradas a partir de processos de Poisson não homogêneos com diferentes funções intensidade, $D(x, y)$, as quais são apresentadas na figura 1. Também se considerou a distribuição homogênea (cenário 1), para efeitos comparativos.

No processo de detecção foi garantida a igual cobertura, não foram observados indivíduos a distâncias superiores a w , e considerou-se que a função de detecção era conhecida e pertencente à família *half-normal*, ou seja, $g(x) = \exp(-x^2/(2\theta^2))$, para $0 \leq x \leq w$ e $\theta > 0$. Assumiu-se que $\theta = 0,01$. Um dos critérios sugerido por Buckland et al. (2001) para a determinação da distância de truncatura, w , é que a função de detecção no ponto de truncatura seja $g(w) = 0,15$ ou $g(w) = 0,10$ para os transectos lineares ou pontuais, respectivamente. Deste modo, neste trabalho considerámos para os transectos lineares $w = 0,0195$ e para os transectos pontuais $w = 0,0215$.

No processo de simulação foram dispostos 1, 5 e 10 transectos lineares e 27 e 135 transectos pontuais, de forma aleatória ou sistemática com início aleatório. Na disposição aleatória dos transectos, verificou-se que, por exemplo, quando se consideraram 10 transectos lineares cerca de 17% da área coberta era amostrada por mais do que 1 transecto. Desta forma, optou-se por avaliar os estimadores convencionais usando duas formas de disposição aleatória de transectos: a) sem qualquer tipo de restrição; b) evitando a sobreposição das áreas cobertas pelos

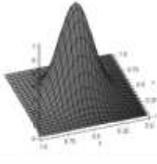
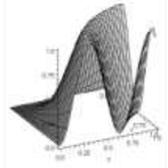
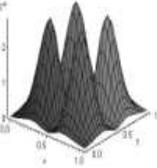
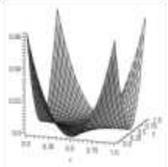
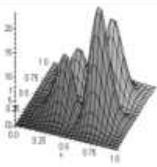
Cenário	$D(x, y)$	Função intensidade	Cenário	$D(x, y)$	Função intensidade
1	λ		4	$\frac{\frac{1}{2}\left[\left(\frac{x-0,5}{0,15}\right)^2 + \left(\frac{y-0,5}{0,15}\right)^2\right]}{2\pi \cdot 0,15^2}$	
2	$\sin^2(5xy)$		5	$(x-0,5)^2(y-0,5)^2 \times \frac{\frac{1}{2}\left[\left(\frac{x-0,5}{0,15}\right)^2 + \left(\frac{y-0,5}{0,15}\right)^2\right]}{2\pi \cdot 0,15^2}$	
3	$(x-0,5)^2(y-0,5)^2$		6	$(x^2-0,25)^2(y^2-0,3)^2 \times \frac{\frac{1}{2}\left[\left(\frac{x^2-0,25}{0,15}\right)^2 + \left(\frac{y^2-0,3}{0,15}\right)^2\right]}{2\pi \cdot 0,15^2}$	

Figura 1: Funções intensidade para cada cenário.

transectos. Os transectos lineares foram colocados horizontalmente, excepto no cenário 2 onde se optou pela colocação vertical devido ao gradiente de densidade apresentado pela distribuição espacial dos indivíduos na área em estudo. A colocação sistemática dos transectos pontuais consistiu na geração de uma grelha de pontos igualmente espaçados entre si em linha e em coluna. De modo a comparar os resultados obtidos com os transectos lineares com os dos transectos pontuais, tentou-se obter áreas cobertas de igual dimensão com ambas as técnicas de amostragem.

2.1 Abordagens

No processo de avaliação do desempenho dos estimadores convencionais da amostragem por distâncias foram consideradas duas abordagens:

- Abordagem 1: simularam-se populações fixas (figura 2), usando cada uma das funções intensidade apresentadas na figura 1. Seguidamente, em cada simulação dispuseram-se os transectos de forma aleatória ou sistemática;
- Abordagem 2: fixaram-se os transectos e, em cada simulação, geraram-se aleatoriamente as localizações dos $N = 1000$ indivíduos da população de

acordo com as funções intensidade do processo de Poisson não homogêneo apresentadas na figura 1.

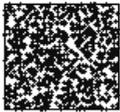
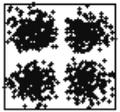
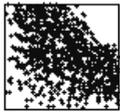
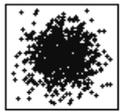
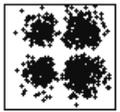
Cenário	População simulada	Cenário	População simulada	Cenário	População simulada
1		3		5	
2		4		6	

Figura 2: Populações simuladas na abordagem 1.

2.2 Resultados

Em cada uma das abordagens consideradas, e para cada cenário atrás descrito, realizaram-se 100 simulações de Monte Carlo. Em cada um dos cenários foi calculada a abundância estimada de indivíduos, \hat{N} , baseada nas n distâncias amostrais detectadas. Nas figuras 3 e 4 apresentam-se os resultados obtidos para os transectos lineares e pontuais, respectivamente, para a abundância estimada, \hat{N} , dimensão da amostra, \bar{n} , e respectivo coeficiente de variação, cv .

Relativamente aos transectos lineares, verifica-se que quando se usa a abordagem 1 o estimador é não enviesado em todos os cenários, mas a sua variância é muito sensível à distribuição espacial dos indivíduos na área em estudo, atingindo valores extremamente elevados especialmente com a disposição aleatória dos transectos. Esta variabilidade pode ser reduzida com o aumento do número de transectos e evitando a sobreposição das áreas cobertas pelos transectos aleatórios. No entanto, verifica-se um aumento mais significativo na precisão do estimador dos transectos lineares quando se colocam os transectos de forma sistemática, o que pode ser explicado pela melhor cobertura da área em estudo. No caso da abordagem 2, verifica-se que o estimador pode apresentar, em alguns cenários, um severo enviesamento quando se considera apenas um transecto linear ou 5 transectos aleatórios, o qual é originado pela má localização dos transectos. Se os transectos forem colocados de forma sistemática, o enviesamento é negligenciável. Ao nível da variabilidade, de um modo geral, esta é muito baixa quando se opta pela colocação sistemática dos transectos lineares.

Sempre que possível, deve-se realizar um estudo piloto na área em estudo que permita aprofundar o conhecimento sobre esta área e verificar a validação

Cenário	a	N.º TL	Abordagem 1			Abordagem 2			
			\hat{N}	\bar{n}	cv	\hat{N}	\bar{n}	cv	
1	0,039	1 A H	1036,9	24,7	19%	989,8	23,5	21%	
		5 A H	1012,1	120,4	9%	998,3	118,7	8%	
		5 Ab) H	1004,5	119,5	8%				
	0,195	5 S H	997,2	118,6	10%	1003,0	119,3	8%	
		10 A H	1007,1	239,5	6%	1004,7	239,0	6%	
		10 Ab) H	1004,9	239,0	6%				
0,390	10 S H	995,8	236,8	6%	1004,7	239,0	5%		
	2	0,039	1 A V	1013,7	24,1	36%	1139,9	27,1	18%
			5 A V	1014,3	120,6	16%	1005,0	119,5	9%
5 Ab) V			989,2	117,6	15%				
0,195		5 S V	1001,0	119,0	6%	944,9	112,4	9%	
		10 A V	1020,7	242,8	11%	1013,7	243,9	6%	
		10 Ab) V	1009,6	240,1	8%				
0,390	10 S V	997,0	237,1	4%	1001,9	238,3	5%		
	3	0,039	1 A H	1162,2	27,6	82%	419,2	10,0	32%
			5 A H	1084,1	128,9	34%	939,9	111,8	8%
5 Ab) H			942,8	112,1	32%				
0,195		5 S H	995,1	118,3	9%	989,6	117,7	9%	
		10 A H	1028,9	244,7	27%	987,9	235,0	5%	
		10 Ab) H	1004,1	238,8	19%				
0,390	10 S H	1006,4	239,4	6%	988,8	235,2	6%		
	4	0,039	1 A H	990,2	23,6	89%	1253,0	29,8	18%
			5 A H	991,1	117,9	42%	1048,2	124,7	9%
5 Ab) H			999,6	118,9	38%				
0,195		5 S H	1004,0	119,4	9%	977,7	116,3	9%	
		10 A H	1007,2	239,5	29%	1081,7	257,3	6%	
		10 Ab) H	968,3	230,3	20%				
0,390	10 S H	1002,3	238,4	7%	993,4	236,3	5%		
	5	0,039	1 A H	1075,1	25,6	67%	1918,1	45,6	14%
			5 A H	1048,4	124,7	30%	876,7	104,3	9%
5 Ab) H			969,8	115,3	27%				
0,195		5 S H	1008,8	120,0	8%	999,2	118,8	9%	
		10 A H	1033,3	245,8	22%	735,1	174,8	7%	
		10 Ab) H	997,4	237,2	15%				
0,390	10 S H	995,6	236,8	5%	993,5	236,3	6%		
	6	0,039	1 A H	898,1	21,4	92%	1721,4	40,9	16%
			5 A H	1020,0	121,3	44%	1103,4	131,2	8%
5 Ab) H			1030,4	122,5	34%				
0,195		5 S H	1006,9	119,7	7%	1004,4	119,4	8%	
		10 A H	991,2	235,8	31%	754,5	179,5	8%	
		10 Ab) H	1011,6	240,6	18%				
0,390	10 S H	987,9	235,0	5%	1003,4	238,6	6%		

Figura 3: Resultados obtidos para a amostragem por transectos lineares (TL), com as abordagens 1 (populações fixas) e 2 (transectos fixos), 100 simulações, $\theta = 0,01$, $w = 0,0195$ e os transectos dispostos de forma: A) aleatória, Ab) aleatória sem sobreposição, S) sistemática, H) horizontal e V) vertical.

Cenário	a	N.º TP	Abordagem 1			Abordagem 2		
			\hat{N}	\bar{n}	cv	\hat{N}	\bar{n}	cv
1	0,039	27 A	1003,1	15,3	24%	1011,6	15,5	26%
		27 Ab)	999,8	15,3	24%			
		27 S: grelha 3×9	985,4	15,1	23%	1025,3	15,7	22%
	0,195	135 A	1008,3	77,1	9%	1012,0	77,3	11%
		135 Ab)	998,9	76,3	10%			
		135 S: grelha 9×15	1017,5	77,8	11%	1004,4	76,8	10%
2	0,039	27 A	1065,3	16,3	28%	1005,7	15,4	24%
		27 Ab)	1041,7	15,9	30%			
		27 S: grelha 3×9	990,0	15,1	23%	890,5	13,6	24%
	0,195	135 A	1013,3	77,4	14%	991,6	75,8	11%
		135 Ab)	979,0	74,8	10%			
		135 S: grelha 9×15	1012,6	77,4	12%	1017,1	77,7	9%
3	0,039	27 A	1071,1	16,4	43%	1334,8	20,4	20%
		27 Ab)	1073,8	16,4	42%			
		27 S: grelha 3×9	978,2	15,0	26%	1047,6	16,0	26%
	0,195	135 A	1008,6	77,1	19%	1068,9	81,7	12%
		135 Ab)	1036,6	79,2	14%			
		135 S: grelha 9×15	1008,8	77,1	10%	985,4	75,3	12%
4	0,039	27 A	1092,1	16,7	34%	1167,3	17,8	22%
		27 Ab)	1104,5	16,9	31%			
		27 S: grelha 3×9	1009,0	15,4	26%	991,3	15,2	28%
	0,195	135 A	1027,3	78,5	16%	1026,4	78,4	11%
		135 Ab)	1000,7	76,5	13%			
		135 S: grelha 9×15	999,3	76,4	9%	998,2	76,3	11%
5	0,039	27 A	1009,0	15,4	35%	819,9	12,5	25%
		27 Ab)	1006,4	15,4	33%			
		27 S: grelha 3×9	984,1	15,0	31%	1163,4	17,8	23%
	0,195	135 A	988,0	75,5	15%	886,6	67,8	11%
		135 Ab)	999,2	76,4	13%			
		135 S: grelha 9×15	1010,4	77,2	13%	1009,5	77,1	11%
6	0,039	27 A	1067,2	16,3	41%	725,7	11,1	33%
		27 Ab)	1064,6	16,3	44%			
		27 S: grelha 3×9	988,0	15,1	44%	1489,9	22,8	21%
	0,195	135 A	1010,5	77,2	19%	984,8	75,3	12%
		135 Ab)	998,0	76,3	15%			
		135 S: grelha 9×15	1002,6	76,6	10%	969,3	74,1	11%

Figura 4: Resultados obtidos para a amostragem por transectos pontuais (TP), com as abordagens 1 (populações fixas) e 2 (transectos fixos), 100 simulações, $\theta = 0,01$, $w = 0,0215$ e os transectos dispostos de forma: A) aleatória, Ab) aleatória sem sobreposição e S) sistemática.

dos pressupostos exigidos pelos diferentes desenhos amostrais possíveis, e deste modo decidir qual o desenho mais apropriado. Neste caso particular, este estudo permitiria determinar mais eficazmente a localização dos transectos na área em estudo e assim minimizar o enviesamento.

No caso dos transectos pontuais, as conclusões são similares às dos transectos

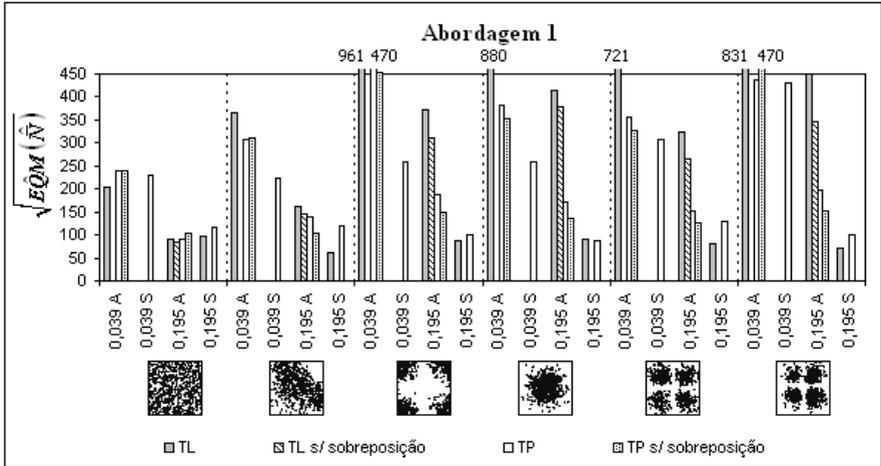


Figura 5: $\sqrt{EQM(\hat{N})}$ por área coberta, a , para a amostragem por transectos lineares (TL) e pontuais (TP), com a abordagem 1 (populações fixas) e os transectos dispostos de forma: A) aleatória e S) sistemática.

lineares. Na abordagem 1 verifica-se que o estimador é não enviesado, mas com coeficiente de variação elevado nos cenários não homogêneos. Aqui, uma redução significativa na variabilidade apenas é alcançada com o aumento do número de transectos. De salientar que não se verifica uma diferença significativa nos resultados obtidos com a colocação aleatória e aleatória sem sobreposição dos transectos pontuais. Na abordagem 2, em alguns cenários não homogêneos, o enviesamento do estimador é bastante elevado bem como a sua variabilidade quando se consideram 27 transectos pontuais. Com o aumento do número de transectos, o enviesamento é anulado e o coeficiente de variação é ligeiramente inferior.

Para melhor comparar os resultados obtidos com os transectos lineares e pontuais, por abordagem, calculou-se o erro quadrático médio da abundância média de indivíduos estimada, $\widehat{EQM}(\hat{N})$. Nas figuras 5 e 6 representam-se os valores obtidos para a raiz quadrada desta medida, para as abordagens 1 e 2, respectivamente, correspondendo o eixo das abcissas à área coberta, a , pelos transectos.

Como se pode observar nestas figuras, existem EQM estimados muito elevados que são originados pela elevada variância (abordagem 1) ou pelo grande enviesamento (abordagem 2) nas estimativas. No caso em que a área coberta é 0,039, que corresponde à área coberta por 1 transecto linear ou 27 transectos pontuais, verifica-se que o EQM estimado é menor nos transectos pontuais

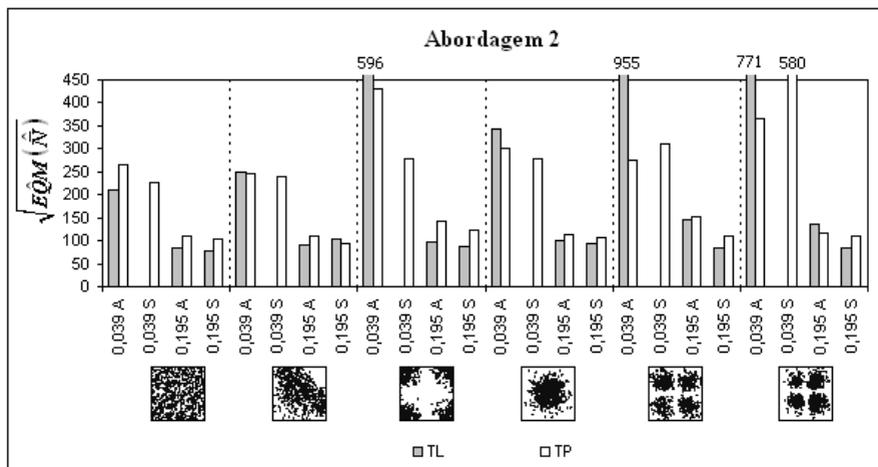


Figura 6: $\sqrt{EQM(\hat{N})}$ por área coberta, a , para a amostragem por transectos lineares (TL) e pontuais (TP), com a abordagem 2 (transectos fixos) e os transectos dispostos de forma: A) aleatória e S) sistemática.

do que nos transectos lineares, excepto no cenário 1. Esta situação pode ser explicada pelo facto de se obter uma melhor cobertura da área em estudo com os 27 transectos pontuais do que com 1 transecto linear. Quando a área coberta é 0,195, a qual é relativa à colocação de 5 transectos lineares ou 135 transectos pontuais, os resultados obtidos com os transectos pontuais apenas são melhores do que os dos transectos lineares na abordagem 1 e quando os transectos são colocados aleatoriamente, uma vez que tanto com a colocação sistemática como com a abordagem 2 verifica-se o inverso. Em ambas as abordagens, verifica-se que a disposição sistemática dos transectos diminui o EQM estimado, sendo esta redução bastante acentuada nos cenários não homogêneos.

3 Conclusão

Neste trabalho avaliou-se o desempenho dos estimadores convencionais da amostragem por distâncias dos transectos lineares e pontuais quando as populações são não homogêneas.

Face aos resultados obtidos nas simulações, verifica-se que quando se considera um número muito pequeno de transectos, dependendo da abordagem considerada, podem existir alguns problemas ao nível do enviesamento e da precisão nos estimadores convencionais. No entanto, desde que a amostra tenha uma dimensão razoável, estes estimadores têm um bom desempenho tanto ao

nível do não enviesamento como da precisão, mesmo em populações distribuídas de forma espacial não completamente aleatória. A colocação de linhas sistemáticas é preferível às linhas aleatórias uma vez que aumentam substancialmente a precisão do estimador da abundância. O mesmo não se pode afirmar sobre a colocação de pontos sistemáticos ao invés de pontos aleatórios, uma vez que não se consegue identificar um padrão no seu comportamento tendo em conta tanto do número de transectos pontuais como da distribuição espacial dos indivíduos.

Comparando os resultados obtidos pelos transectos lineares com os dos transectos pontuais, verifica-se que quando a área coberta é pequena o estimador de transectos pontuais dá melhores resultados. No entanto, à medida que aumenta a área coberta os resultados obtidos pelo estimador dos transectos pontuais são superados pelos resultados dos transectos lineares, quando os transectos são colocados de forma sistemática.

Agradecimentos

Os autores são membros do CIMA-UE, centro de investigação financiado no âmbito do FEDER pelo Programa de Financiamento Plurianual da FCT.

Referências

- [1] Afonso, A. e Alpizar-Jara, R. (2007a). *Estimação em amostragem por distâncias com distribuições espaciais não homogéneas*. Em *Estatística Ciência Interdisciplinar* (Ferrão, M.E., Nunes, C. e Braumann, C. A., eds), 201-210. Lisboa: Edições SPE.
- [2] Afonso, A. e Alpizar-Jara, R. (2007b). Assessing distance sampling under non-homogeneous Poisson processes. *Bulletin of the International Statistical Institute* (Proceedings of 56th Session). Lisboa: ISI 2007.
- [3] Barry, S.C. e Welsh, A.H. (2001). Distance sampling methodology. *Journal of the Royal Statistics Society, Series B*, 63, 31-53.
- [4] Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D. L. e Thomas, L. (2001). *Introduction to distance sampling. Estimating abundance of biological populations*. Oxford: Oxford University Press.
- [5] Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D. L. e Thomas, L. (2004). *Advanced distance sampling: estimating abundance of biological populations*. Oxford: Oxford University Press.
- [6] Fewster, R.M, Laake, J.L. e Buckland, S.T. (2005). Reader reaction: Line transect sampling in small and large regions. *Biometrics*, 61, 856-861.
- [7] Fewster, R.M e Buckland, S.T. (2004). *Assessment of distance sampling estimators*. Em *Advanced distance sampling: estimating abundance of biological populations* (Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D.L. e Thomas, L., eds.), 281-306. Oxford: Oxford University Press.
- [8] Melville, G.J. e Welsh, A.H. (2001). Line transect sampling in small regions. *Biometrics*, 57, 1130-1137.