

Incorporating Environmental data into Poisson Kriging Approaches for Mapping Patterns of Herbivore Species Abundance in Kruger National Park, South Africa

Kerry, R.¹, Goovaerts, P.², Smit, I.³ and Ingram, B. R.⁴

¹Department of Geography, Brigham Young University, Provo, Utah, USA.
ruth_kerry@byu.edu

²Biomedware Inc., Ann Arbor, Michigan, USA.
goovaerts.pierre@gmail.com

³South African National Parks, Skukuza, South Africa.
izaks@sanparks.org

⁴Department of Computer Science, University of Talca, Talca, Chile.
ingrambr@googlemail.com

1. Introduction

Kruger National Park, South Africa, provides 19,485 km² of protected habitats for the unique species of the African savanna, several of which are endangered. For the last forty years annual aerial surveys to monitor large herbivore populations have been conducted. These have been used to understand population trends and the environmental factors and management actions that influence herbivore density and distribution patterns. From 1980-1993, the whole park was surveyed annually, but this was costly and time consuming. In 1998, the park-wide census approach was replaced by a sampling strategy whereby the number of animals is recorded along 800 m wide East-West transects, spaced at intervals of 2.5-5.6 km (Kruger et al. 2008). However, such strip transects leave “gaps” in the data spatially. The park currently use the Distance method (Thomas et al. 2004) but several assumptions of the method are not met especially for rare species or species that tend to be clustered in space.

Geostatistical methods at first glance might seem ideal for populating the gaps in survey data and for estimating the total numbers of each animal in the park in a given year. However, the histogram of animal count data for the park is usually highly positively skewed, especially for the rarer species or those that tend to cluster spatially. The histograms tend to approach the Poisson distribution. This hampers the estimation of the variogram by the traditional method of moments. Kerry et al. (2010a) compared an Auto-Indicator kriging approach (Goovaerts, 2009) and Poisson kriging (Monestiez et al. 2006) as potential methods for populating the data gaps between transects and to create continuous surfaces of species abundance. It was thought that an auto-indicator approach could be used to efficiently compute and model variograms for numerous thresholds representing each count. However, the study showed that variograms for the rare high counts were pure nugget and so the number of thresholds had to be reduced. This meant that the number of large counts of each animal was underestimated. Also, the nature of the data meant that unless the data were preprocessed to migrate the data to a grid, there were no zero counts and this meant that low counts were over-estimated. In contrast to the Auto-Indicator approach, with and without pre-processed data, two Poisson approaches produced markedly smaller, and sometimes an order of magnitude smaller, mean absolute errors (MAEs) in cross-validation. An initial investigation showed that incorporating environmental data into a simple 0/1 Indicator approach reduced MAEs slightly. Here we illustrate a method of incorporating environmental data into the Poisson kriging approaches and compare the errors associated with this to the errors when no environmental data are included.

2. Methods

Poisson kriging of count data was performed using two types of denominator:

- (1) observational area (ratio = spatial density, Fig. 1a)
- (2) total number of animals in a given area (ratio = proportion, Fig. 1b).

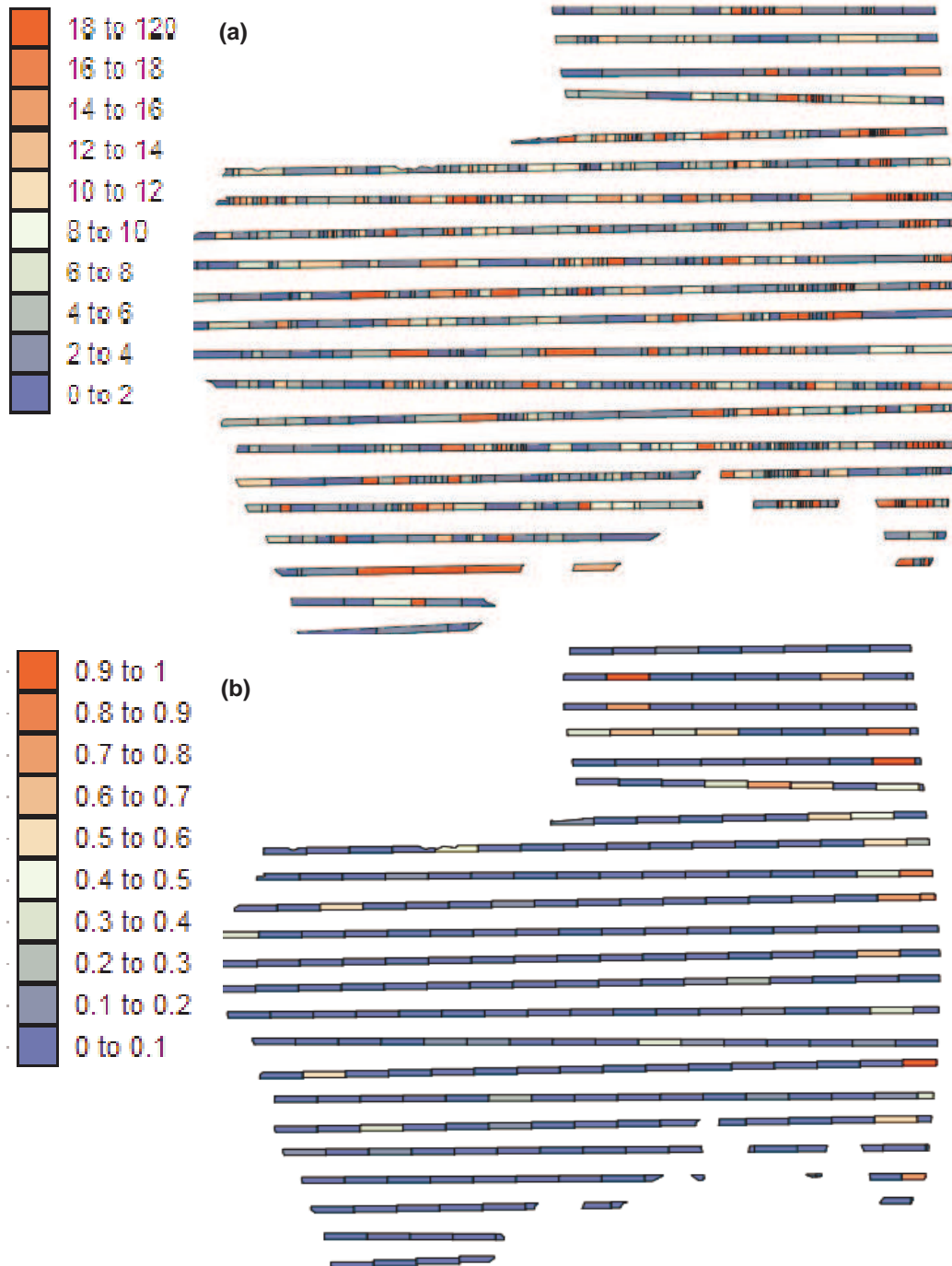


Figure 1. Calculation of (a) spatial density from 800 m wide transect data for Poisson approach (1) and (b) proportion of each animal from 5 km long blocks of the 800 m wide transect data (e.g. number of impala/total number of all animals in 5km by 800m block) for Poisson approach (2).

Both Poisson approaches result in sightings of rare animals in sparsely populated areas (i.e. small numbers) being down-weighted for variogram computation and kriging. However, Approach (2) is only suitable for accurately mapping the distribution of individual species in the park.

The following observational area/population-weighted estimator adjusts for the small number problem:

$$\hat{\gamma}_{Rv}(\mathbf{h}) = \frac{1}{2 \sum_{\alpha, \beta} \frac{n(v_\alpha)n(v_\beta)}{n(v_\alpha) + n(v_\beta)}} \sum_{\alpha, \beta}^{N(\mathbf{h})} \left\{ \frac{n(v_\alpha)n(v_\beta)}{n(v_\alpha) + n(v_\beta)} [z(v_\alpha) - z(v_\beta)]^2 - m^* \right\}, \quad (1)$$

where $N(\mathbf{h})$ is the number of pairs of areas (v_α, v_β) whose observational area/population-weighted centroids are separated by the vector \mathbf{h} and m^* is the observational area/population-weighted mean of the N area rates. The usual squared differences, $[z(v_\alpha) - z(v_\beta)]^2$, are weighted by a function of their respective observational area/population sizes, $n(v_\alpha)n(v_\beta)/[n(v_\alpha) + n(v_\beta)]$, which gives more importance to more reliable data pairs based on large observational areas/large total counts of animals (Monestiez et al. 2006, see also Kerry et al. 2010b).

The animal density/proportion and the associated kriging variance for a location X are estimated as:

$$\hat{r}(X) = \sum_{i=1}^K \lambda_i z(v_i), \text{ and} \quad (2)$$

$$\sigma^2(X) = \bar{C}_R(X, X) - \sum_{i=1}^K \lambda_i \bar{C}_R(v_i, X) - \mu(X), \quad (3)$$

The kriging weights (λ_i) and the Lagrange parameter $\mu(X)$ are computed by solving the ‘‘Poisson kriging’’ system:

$$\sum_{j=1}^K \lambda_j \left[\bar{C}_R(v_i, v_j) + \delta_{ij} \frac{m^*}{n(v_i)} \right] + \mu(X) = \bar{C}_R(v_i, X), \quad i = 1, \dots, K, \quad (4)$$

$$\sum_{j=1}^K \lambda_j = 1,$$

where $\delta_{ij}=1$ if $i=j$ and 0 otherwise. The covariances are estimated from the results of a deconvolution of the model fitted to variogram (1), see Goovaerts (2008). The ‘‘error variance’’ term, $m^*/n(v_i)$, leads to smaller weights for rates measured over smaller areas/populations.

Various environmental data (Figure 2) was incorporated into the mapping by kriging the residuals from a Poisson regression between environmental and animal data.

Leave-one-out (LOO) cross-validation was used to assess the relative performance of the different methods for estimating counts of all species, and of representatives of the key feeding groups grazers, browsers and mixed feeders such as giraffes, impala and zebra for the whole park. Rarer species or those species that tend to cluster in herds were also investigated.

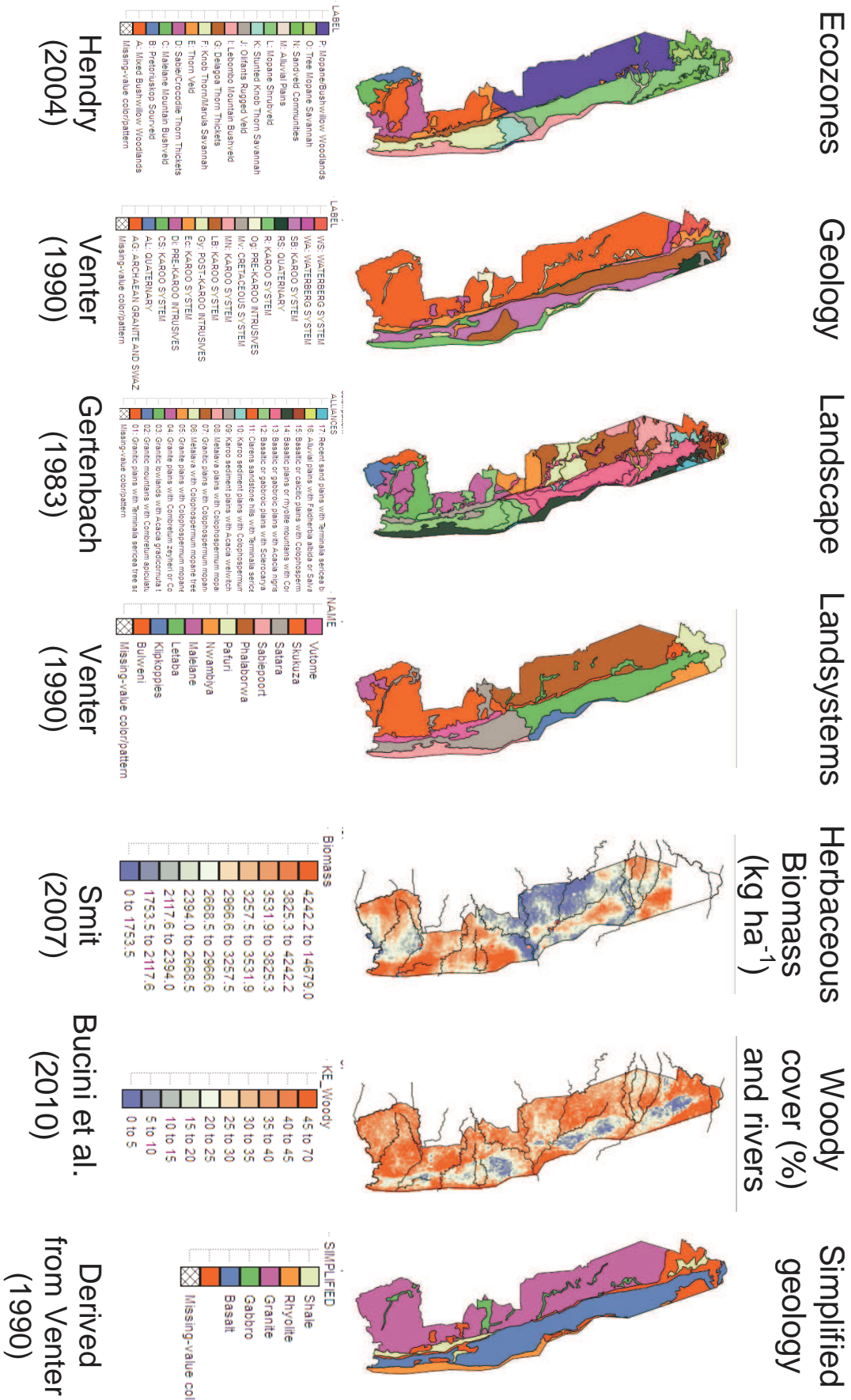


Figure 2. Environmental Data used in Poisson Regression

3. Results and Discussion

Table 1 shows that for estimating numbers of giraffe, impala and zebra, Poisson approach (1) yielded smaller errors. It also created patterns that are more sensible than those of Poisson approach 2 when compared to the observed counts (Figure 3). The MAEs indicate that approach (1) produces its best estimates when there are more animals i.e. looking at counts of all animals, or more abundant animals such as impala. Poisson approach (2), however, leads to best estimates for the rarest animals or those that tend to occur in isolated herds. The effects on MAEs of incorporating environmental data such as biomass, tree cover, geology and ecotypes into both Poisson approaches (results not shown here) will be discussed in the presentation.

Table 1. Mean Absolute Errors (MAEs) from Leave-One-Out Cross-validation for Poisson kriging using spatial density (approach 1) or proportion of animals (approach 2).

Data	MAE	
	Poisson approach (1)	Poisson approach (2)
All animals 1998	0.0528	*
All animals 2000	0.0401	*
All animals 2001	0.0463	*
All animals 2005	0.0448	*
<i>Key feeding groups</i>		
Giraffe 2000	0.1337	0.1481
Impala 2000	0.1668	1.3791
Zebra 2000	0.2630	0.3909
<i>Rarer species</i>		
Elephant 2000	0.4494	0.0996
Kudu 2000	0.6264	0.1142
Waterbuck 2000	0.7747	0.0516
Warthog 2000	0.8165	0.0460
Wildebeest 2000	0.3133	0.1290
White rhino 2000	0.5348	0.0850

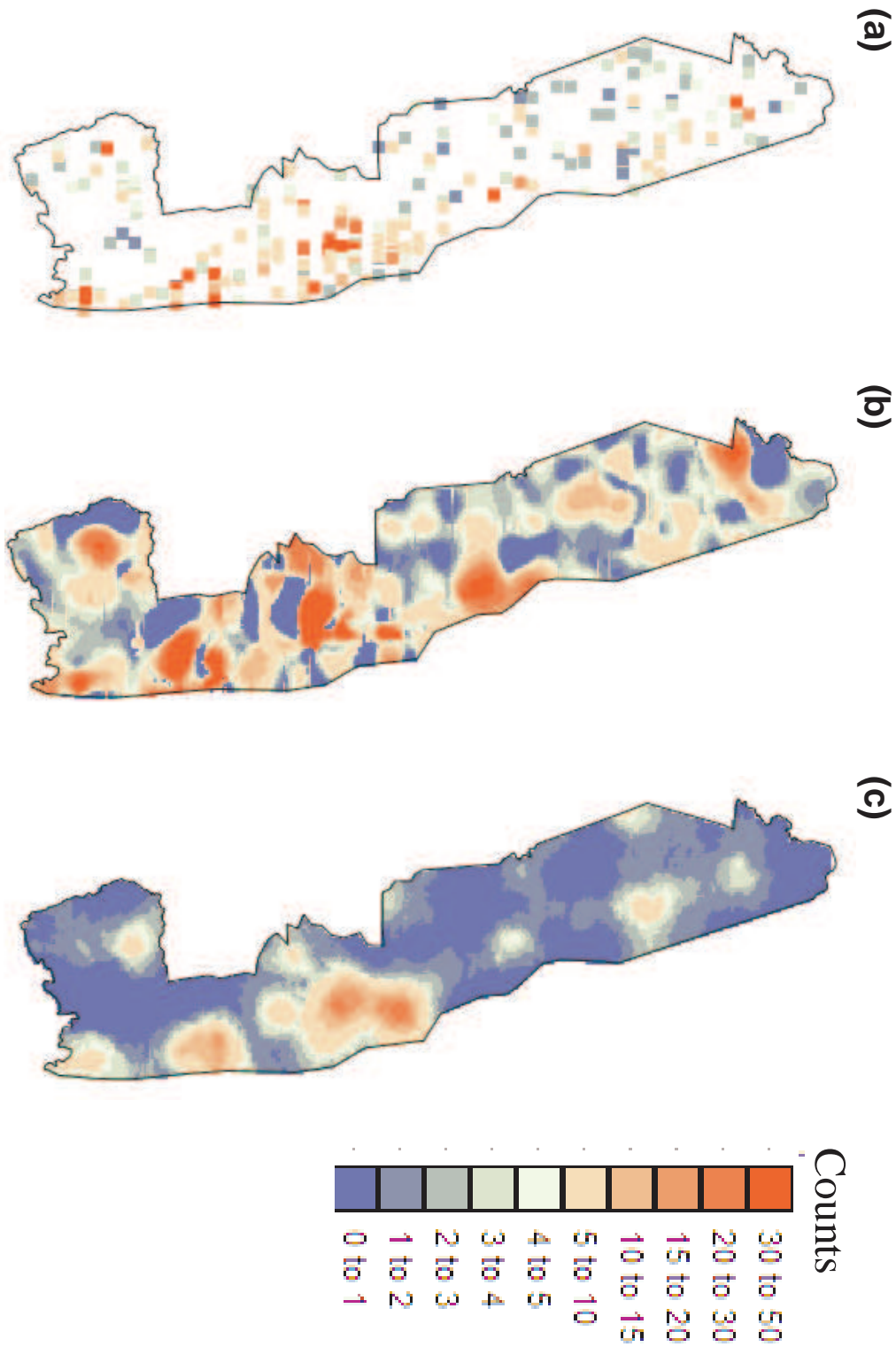


Figure 3. (a) Observed counts of zebras in 2000 and kriged maps of counts produced by (b) Poisson approach (1), and (c) Poisson approach (2).

5. References

- Bucini, G, Hanan, NP, Boone, RB, Smit, IPJ, Saatchi, S, Lefsky, MA & Asner, GP, 2010. Woody fractional cover in Kruger National Park, South Africa: remote-sensing-based maps and ecological insights. *Ecosystem function in savannas: measurement and modeling at landscape to global scales* (ed. Hill, M.J. & Hanan, N.P.), CRC/Taylor and Francis, pp 219-237.
- Gertenbach, WPD, 1983. Landscapes of the Kruger National Park. *Koedoe*, 26:9-121.
- Goovaerts P, 2009. AUTO-IK: a 2D indicator kriging program for the automated non-parametric modeling of local uncertainty in earth sciences. *Computers & Geosciences*, 35:1255-1270.
- Goovaerts, P, 2008. Kriging and semivariogram deconvolution in presence of irregular geographical units. *Mathematical Geosciences*, 40(1), 101-128.
- Hendry O. 2004. *Kruger Park Ecozone Map*. Jacana Media. 16p.
- Kerry R, Goovaerts P, Haining RP and Ceccato V 2010b. Geostatistical Analysis of Car Theft and Robbery in the Baltic States. *Geographical Analysis*. 42:53-77.
- Kerry R, Goovaerts P, Smit I and Ingram BR, 2010a. Comparing the Accuracy of Indicator and Poisson Kriging for Investigating Patterns of Herbivore Species Abundance in Kruger National Park, South Africa. *Proceedings of 9th Ninth International Symposium on Spatial Accuracy Assessment in Natural resources and Environmental Sciences, Leicester, United Kingdom, July 2010*.
- Kruger JM, Reilly, BK and Whyte IJ, 2008. Application of distance sampling to estimate population densities of large herbivores in Kruger National Park. *Wildlife Research*, 35:371-376.
- Monestiez P, Dubroca L, Bonnin E, Durbec JP and Guinet C, 2006. Geostatistical modelling of spatial distribution of *Balaenoptera physalus* in the Northwestern Mediterranean Sea from sparse count data and heterogeneous observation efforts. *Ecological Modelling*. 193:615-628.
- Smit IPJ, 2007. *Artificial surface-water provision in a semi-arid savanna: a spatio-temporal analysis of herbivore distribution patterns in relation to artificial waterholes under different habitat, rainfall and management scenarios in the Kruger National Park, South Africa*, Unpublished PhD thesis, University of Cambridge, Cambridge.
- Thomas L, Laake JL, Strindberg S, Marques FFC, Buckland ST, Borchers DL, Anderson, DR, Burnham KP, Hedley SL, Pollard JH and Bishop JRB 2004. *Distance 4.1. Release 2. (Research Unit for Wildlife Population Assessment, University of St Andrews: St Andrews, UK*.
- Venter FJ, 1990. *A classification of land for management planning in the Kruger National Park*. Unpublished PhD thesis, University of South Africa, Pretoria.