Distribution Mapping of Wildebeest in Serengetti Park, Tanzania

Geostatistical Mini Project Report

By

Table of Contents

Introduction.. 3
Materials and Method 4
Results... 6
Discussion.. 9
Conclusion 10
Reference 11

Table of Figures

Figure 1: Study Area.. 4
Figure 2: Sample data distribution.. 6
Figure 3: Variogram Model 7
Figure 4: Probability map 7
Figure 5: Standard deviation of Kriging 7
Figure 6: True Indicator Map/Plot... 8
Figure 7: Histogram of sample data... 9

Introduction

The assessment of wildlife populations and distributions is important for the analysis of interaction of certain species with environmental factors such as vegetation cover, drainage and human settlement (Khaemba 2001). The analysis of their interaction can be carried out with GIS tools and as a result explanation for observed wildlife distribution.

Wildlife park managers require accurate information on the abundance and distribution of the wildlife animals for effective management (K. Ottichilo, de Leeuw et al. 2001). The modeling of abundance and distribution requires accurate and up to date information, which is obtained efficiently and speedily for large parks through aerial surveys (Khaemba 2001). On the other hand this type of surveys can be expensive due to the utilization of small aircraft and equipment such as digital aerial photographs.

The wildebeest (large herbivore – grazer) is a keystone species in the both the Serengetti park, in Tanzania and Masai Mara, in Kenya, ecosystem because of its occurrence in large numbers and its annual migrations within and outside the ecosystem(K. Ottichilo, de Leeuw et al. 2001).

Indicator kriging has provided well for environmental protection and wildlife management(Webster and Oliver 2008). It involves the transformation of data; it converts a variable that has been measured on a continuous or discrete scale to several indicator variables, each taking the values 1 or 0, which denote the presence or absence respectively, at the sample sites and estimating their values elsewhere through simple or ordinary kriging of indicators (Webster and Oliver 2008).

Problem Statement

Wildlife abundance in tropical areas is commonly modelled using the Jolly II method, which interpolates sample characteristics like species density from data collected through aerial surveys to larger management regions. Spatial dependence within wildlife populations is ignored and abundance estimates obtained through this method often have wide confidence limits(Khaemba 2001).

Objective

To determine the distribution of wildebeest in the South of Serengetti Park Tanzania, using a presence probability distribution map.

Questions

1. Is the presence distribution of wildebeest spatially dependent?
2. What is the accuracy of the predicted verses the observed values?

Materials and Method

The study was carried out in the South East of Serengetti National Park in Tanzania as shown below:

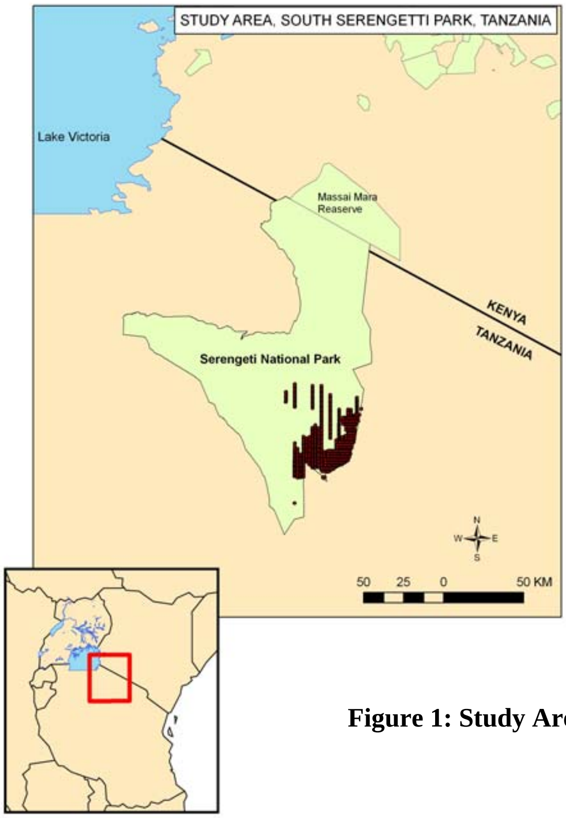

Figure 1: Study Area

The materials used for carrying out the analysis of the presence probability mapping of wildebeests was an excel sheet of 734 points of the counts and density of wildebeests from an aerial photographic survey of 1998 to 2000 at an extent of approximately 75 km North by 50 km East.

For illustrating the study area ArcGIS 9.3 software was used and for the analysis part R statistic software with geostat package was used.

Indicator Variables

Creation of a Presence (1) absence (0) data column in the sample sheet as indicator variables i.e. *ind <- (Owb$Count > 0) and Owb.Prob <- data.frame(Owb, ind=ind, IndNum = as.numeric(ind))*

Any wildebeest count that is above zero (0), it is assigned a value of one (1), which denotes presence and if zero (0), it is assigned a value of zero (0), which denotes absence of wildebeest.

Empirical Variogram

In order to model the spatial variation of the presence absence data, the bin width was retained at 1/15 of the maximum separation of paired points with a cutoff around half the extent of the study area (30 km) was used to create the empirical variogram i.e. *vi <- variogram(IndNum ~ 1, loc = Owb.Prob, cutoff = 30000)*

Fitting the Model Variogram

For fitting the model variogram onto the empirical variogram an exponential model was used with a partial sill of 0.08, nugget of 0.12 and a range of 25000 meters i.e.

Owb.Prob.vimf <- fit.variogram(vi, vgm(0.20, "Exp", 25000, 0.12))
plot(vi, fit.variogram(vi, vgm(0.08, "Exp", 25000, 0.12), fit.method=7…

The model variogram weighing method was used in order to improve the fitting to the empirical variogram i.e. fit method 7 (Nj/h^2j) where Nj = number of paired points and h^2j = lag distance.

The bounding grid for probability mapping

A grid platform was created for obtaining the probability map for the distribution of wildebeest in the study area, with a grid size of 100 by 100 meters. This was done using the expand grid command i.e.

grid <- expand.grid(UTM_E = seq(698186, 745220, by = 100), UTM_N = seq(9652412, 9726244, by = 100))

Interpolation by Indicative Kriging

The sample points were interpolated using Indicative Kriging to obtain a probability surface of the presence of wildebeest i.e.

Owb.Prob.Krig < krige(IndNum~1,Owb.Prob,grid,model=Owb.Prob.vimf)

In order to visually illustrate the probability surface a map was created, using the following command: *spplot(Owb.Prob.Krig,zcol="var1.pred"...)* . To illustrate the standard deviation of Kriging , the following command was used: *spplot(Owb.Prob.Krig,zcol="var1.var"...)*

Cross validation to establish the model accuracy

A cross validation was done in order to establish the accuracy of the probability map i.e.

Owb.Prob.CV <- data.frame(CrossV = (Owb$Count > 0))
Kr.CV <- krige.cv(CrossV ~ 1, loc = Owb.Prob.CV, model = Owb.Prob.vimf)

Visualization of the cross validation was then carried out using the following command:

plot(coordinates(Kr.CV), asp = 1, col = ifelse(Kr.CV$observed, "green", "red"),…

Results

Following the methodology shown above the main findings of the analysis were presented as follows:

1. The presence and absence distribution of Wildebeest in the study area is shown below:

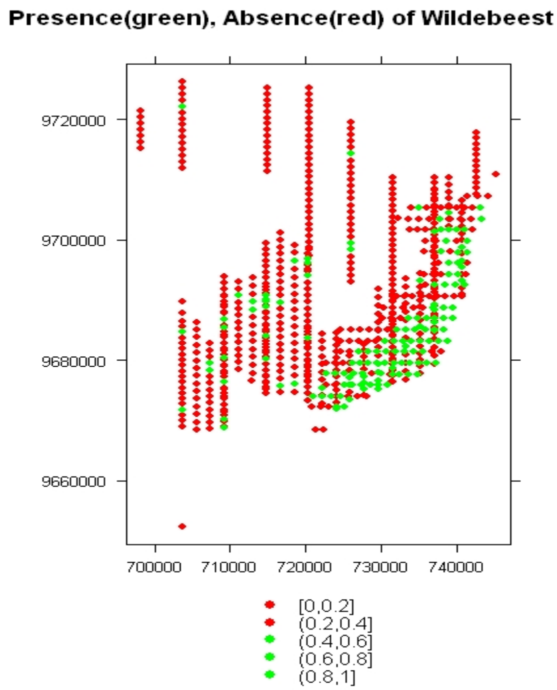

Figure 2: Sample data distribution

There is a generally a high distribution of absence data of wildebeest in the study area. The South East and the East part tend to have higher distributions of presence data while the rest of the area is dominated by absence data of wildebeest occurrence.

2. Fitting of the exponential variogram model onto the empirical variogram was carried out as shown below on figure 3.

The exponential variogram was preferred to other authorized models because of its lowest sum of square error compared to other (as shown in table 1) and it is a simple and commonly used model. The blue dots, in figure 3, is the average semivariance within a bin and the label represents the number of point pairs, for example, the average semivariance of the first bin is around 0.145m^2 and the number of point pairs are 2531.

Table 1

Variogram Model	Sum of Square Error
Exponential	$2.078 * 10^{-8}$
Circular	$2.082 * 10^{-8}$
Spherical	$2.082\ 8\ 10^{-8}$
Gaussian	$2.285 * 10^{-8}$

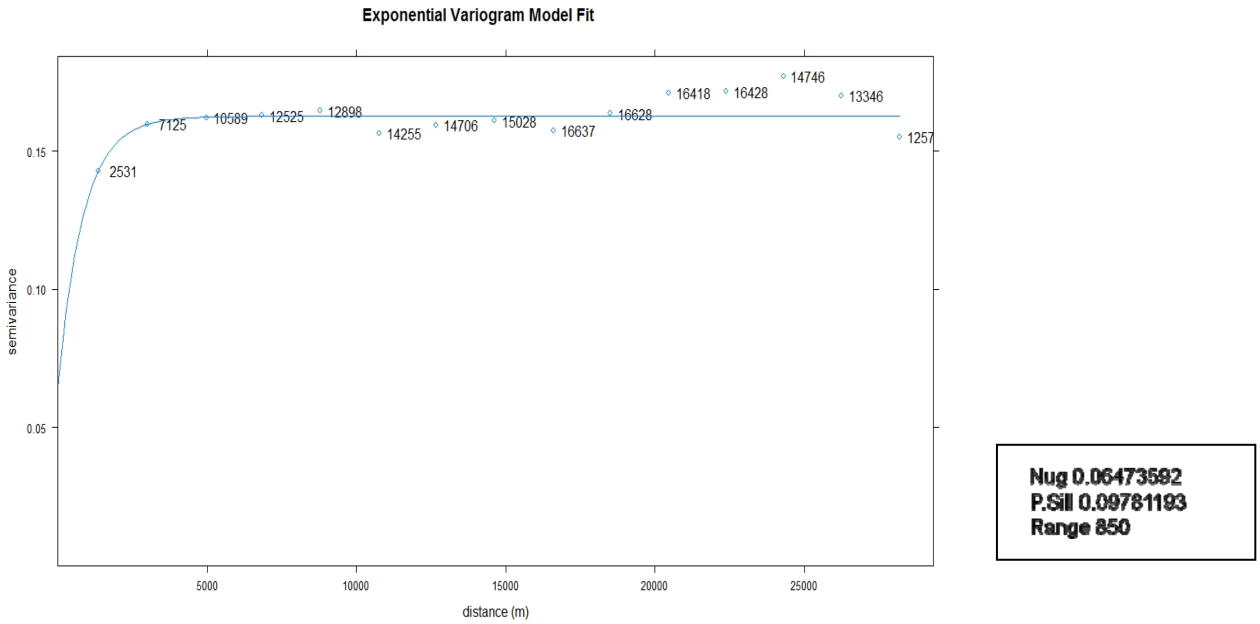

Figure 3: Variogram Model

3. Illustration of the presence probability of the occurrence of wildebeest and the standard deviation of the interpolation through indicative kriging is shown in figure 4 and 5.

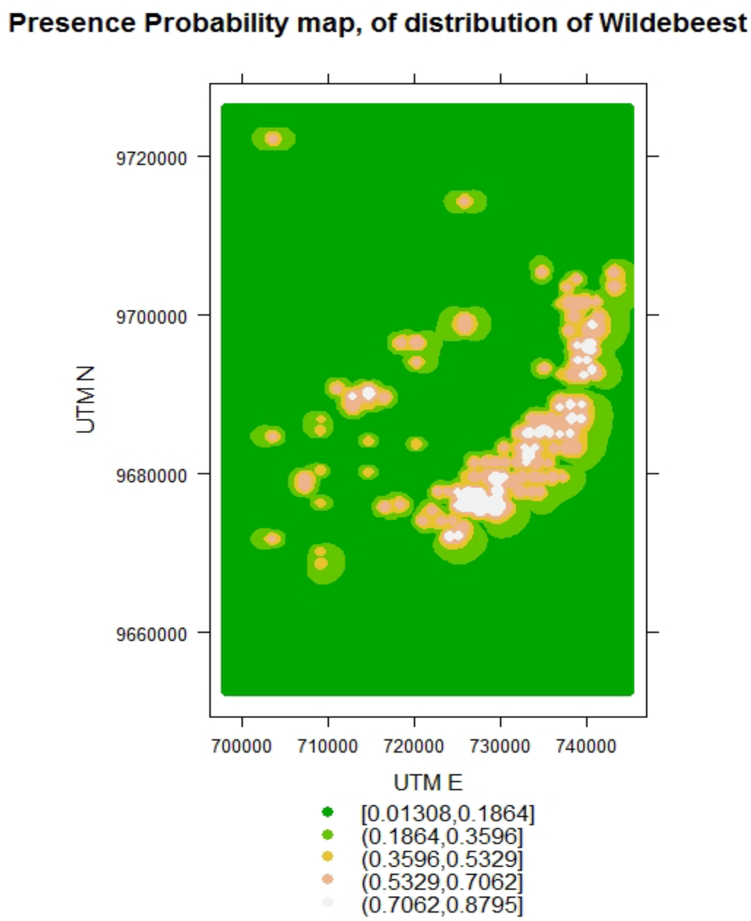

Figure 4: Probability map

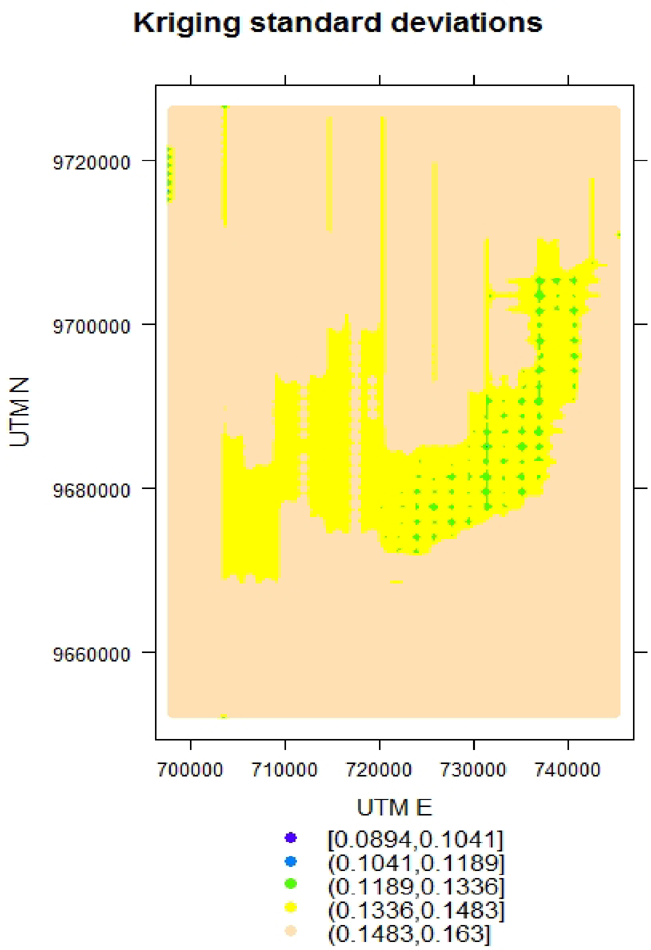

Figure 5: Standard deviation of Kriging map

According to figure 4, the probability of wildebeest occurrence is high at the mid South East and Eastern part and to a less extent at the mid west part of the study area, shown as white (70 to 87% chance) and brown (50 to 69% chance) patches. Most parts the occurrence of wildebeest in the study area is very low, shown as dark green color (13.1 to 18.6% chance) with some patches of low occurrence, shown as light green with 18.6 to 36 % chance.

The standard deviation of indicator kriging is generally high as shown in figure 5 but reduces towards the vicinity of the sampled data especially for the presence data.

4. The accuracy of the model i.e. the predicted presence probability for the distribution of wildebeest was done by cross validation, using the sample data that was used to create the model and a true indicator plot/map was created to visualize the accuracy, as shown in figure 6:

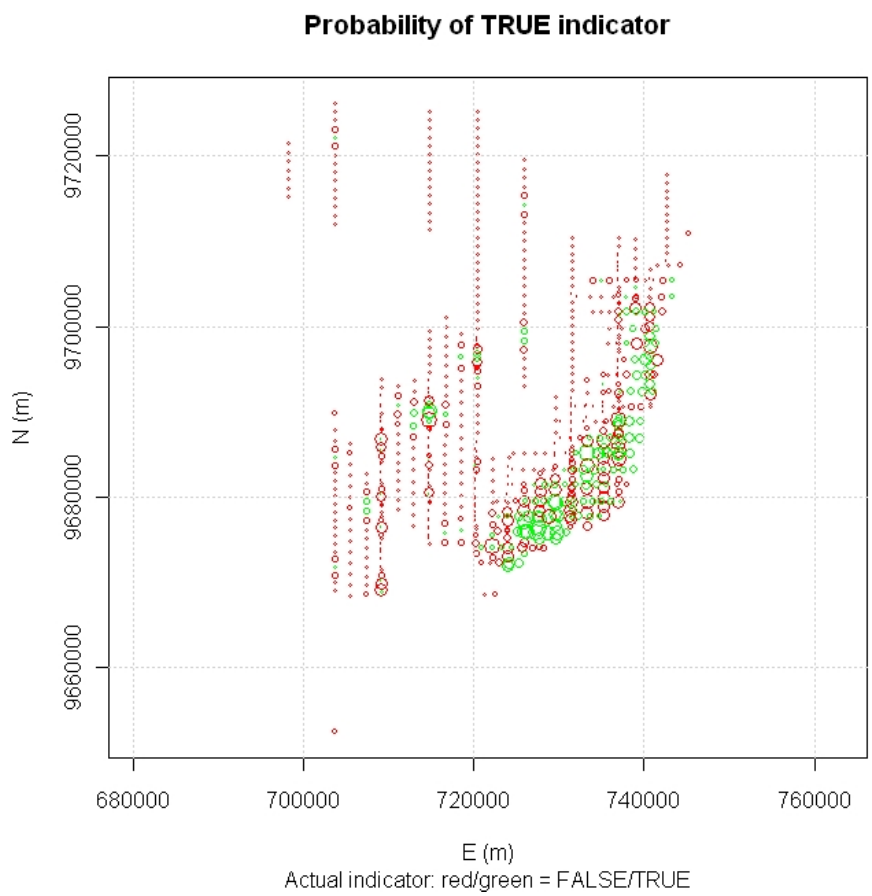

Figure 6: True Indicator Map/Plot

The green circles are where the predicted probability of presence is true while the red circles means the predicted probability of presence is false. The ultimate goal of the probability of true indicator, in order to obtain the very good model, is to have the red circles very small and the green circles big. According to *figure 6* the accuracy of the model is fair given that there are equal amount of green and red circles that are around the same size.

Discussion

The initial intention of the study was to predict the density distribution of wildebeest across the study area through Ordinary Kriging (OK) and further more using Regressional Kriging (RK) with vegetation type as an external variable to predict the distribution of the wildebeest and compare the different results of the OK and RK or in other words whether vegetation type affects the distribution of wildebeest in the study area.

OK and RK interpolate values between sampled measurements, such as wildebeest counts and density in the case study. However, it requires normally distributed data, which is commonly invalidated in ecological censuses(Walker, Balling Jr et al. 2008). The sample data of the study showed a typical skewed distribution of ecological (wildebeest) censuses as shown on *figure 7*, even after log transformation. Furthermore the study data had a lot of observation of zero (0) i.e. absence data, which could not produce numerical predictions of measurements.

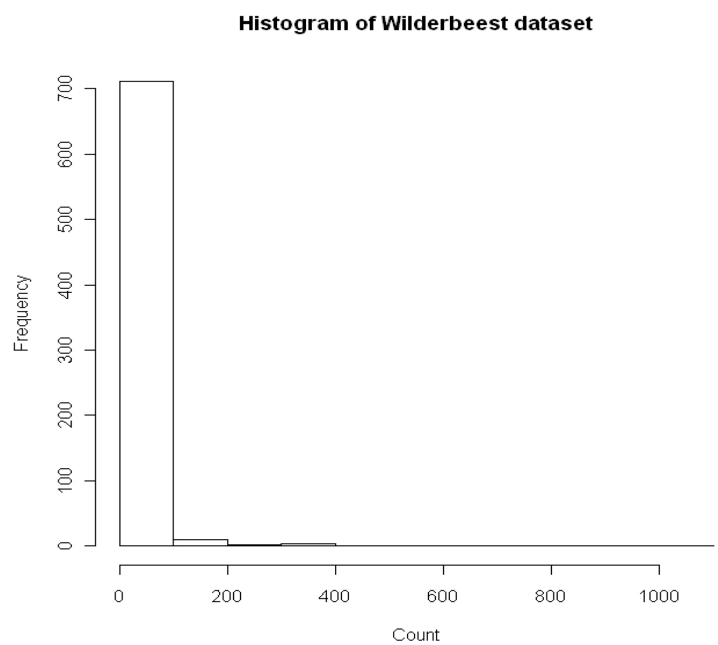

Figure 7: Histogram of sample data

On the other hand, indicator Kriging has the advantage of not requiring normally distributed data and require fewer statistical decisions, but it is not able to produce numerical predictions of measurements(Walker, Balling Jr et al. 2008). Given the skewness of the data, indicator Kriging would be a more appropriate method of interpolation but it limited the analysis to only mapping the presence probability of wildebeest in the study area.

Using Indictor Kriging (IK) the results demonstrate that the probability of finding wildebeest in the study area is higher in the East and South East part than the rest. This can be helpful information in guiding visitors who want to observe the famous wildebeest animal in the park but on the other hand provide little information for effective management of the wildebeests and other dependent species.

Some of the assumptions of the model and presence probability map and limitations of the study are discussed below.

Assumptions

- Other external covariates have no influence: Wildebeest are large herbivores that are savanna grazers that live in the wild. These external covariates (factors) that can influence the distribution of wildebeest are the vegetation type and predators such as lions and other big cats. Other factors include conversion of natural land into agriculture land (land fragmentation) and human settlement(K. Ottichilo, de Leeuw et al. 2001). For the case of the model such external covariates had no influence to the distribution of the wildebeest.
- The wildebeest don't move beyond the grid size: The grid size created for the study area was 100 by 100 meters. All individual grids had a presence probability value of wildebeest occurrence on condition that they don't move beyond the assigned grid; this means that the wildebeest are confined in an area of 100 by 100 meters.

Limitations

- Statistical abnormalities, for example: the lack of normal distribution in the sample data, instead it was highly skewed hence limiting measurement prediction; spatially independent data, values that have high count/density values would be close to very low count/density data or even absence(0) data.
- Data deficiency: Lack of environmental covariates e.g. vegetation type, land use, human settlement distribution or predator distribution.
- Only probability indicator of presence of wildebeest without multiple/categorical indicator levels e.g. low, medium or high distribution.
- Lack of accuracy quantification: The cross validation results only showed a visual impression of the model accuracy and not a numeric figure such as percentage to quantify the accuracy. The accuracy of the model could also be done by initially dividing some of the sample data for cross validating with predicted values.

Conclusion

Animal abundance and distribution in tropical ecosystem is usually modeled using Jolly II method which according to(Khaemba 2001), is a method that doesn't honor spatial dependence of the wildlife populations. According to the research study carried out the distribution of the wildebeest can be predicted using presence probability mapping by indicative kriging, which honors the spatial dependence of the wildebeest population. The predicted presence probability of wildebeest was found to be spatially dependent and the accuracy was fair but can be improved if the outlined assumptions are surmounted.

Reference

K. Ottichilo, W., J. de Leeuw, et al. (2001). "Population trends of resident wildebeest [Connochaetes taurinus hecki (Neumann)] and factors influencing them in the Masai Mara ecosystem, Kenya." Biological Conservation 97(3): 271-282.

Khaemba, W. M. (2001). "Spatial point pattern analysis of aerial survey data to assess clustering in wildlife distributions." International Journal of Applied Earth Observation and Geoinformation 3(2): 139-145.

Walker, J. S., R. C. Balling Jr, et al. (2008). "Birds of a feather: Interpolating distribution patterns of urban birds." Computers, Environment and Urban Systems 32(1): 19-28.

Webster, R. and M. A. Oliver (2008). "Geostatistics for environmental scientists 2nd ed.

."