

# GIS Applications in Agriculture

*Series Editor*

**Francis J. Pierce**

*Washington State University, Prosser*

GIS Applications in Agriculture, *edited by Francis J. Pierce and David E. Clay*

GIS Applications in Agriculture, Volume Two: Nutrient Management for Energy Efficiency, *edited by David E. Clay and John F. Shanahan*

GIS Applications in Agriculture, Volume Three: Invasive Species, *edited by Sharon A. Clay*

# GIS Applications in Agriculture

*Volume Three: Invasive Species*

**Edited by**

**Sharon A. Clay**

**GIS APPLICATIONS IN AGRICULTURE SERIES**



**CRC Press**

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the  
Taylor & Francis Group, an **informa** business

---

# 15 Adapting Geostatistics to Analyze Spatial and Temporal Trends in Weed Populations

*Nathalie Colbach and Frank F. Forcella*

## CONTENTS

15.1 Executive Summary.....	320
15.2 Introduction .....	320
15.3 Analysis Steps.....	321
15.4 Data Collection .....	322
15.4.1 Experimental Field.....	322
15.4.2 Sampling Grid.....	322
15.5 Exploratory Data Analysis.....	323
15.5.1 Objective.....	323
15.5.2 Method.....	324
15.5.3 Results.....	325
15.6 Data Transformation.....	330
15.6.1 Objective.....	330
15.6.2 Method.....	330
15.6.3 Results.....	330
15.7 Detrending Data.....	330
15.7.1 Objective.....	330
15.7.2 Median Polishing.....	330
15.7.3 Estimating Trend with Linear Regressions.....	331
15.8 Empirical Semivariograms.....	334
15.8.1 Objective.....	334
15.8.2 Method.....	335
15.8.3 Results.....	335
15.9 Semivariogram Model Fitting .....	335
15.9.1 Objective.....	335
15.9.2 Method.....	340
15.9.3 Results.....	340
15.10 Analysis of Variogram Parameters .....	346
15.10.1 Objective.....	346

15.10.2	Method.....	346
15.10.3	Results .....	354
15.11	Kriging .....	354
15.11.1	Objective.....	354
15.11.2	Method.....	356
15.11.3	Results .....	356
15.12	Cross-Semivariograms and Cokriging .....	356
15.12.1	Objective.....	356
15.12.2	Cross-Semivariograms.....	360
15.12.3	Cokriging .....	360
15.13	Error Analysis .....	363
15.13.1	Prediction of Weed Means.....	363
15.13.2	Prediction of Weed Locations.....	363
15.14	Summary: Using Geostatistical Information for Decision Making .....	369
	Glossary .....	369
	References.....	370

## 15.1 EXECUTIVE SUMMARY

Geostatistics were originally developed for the mining industry to estimate the location, abundance, and quality of ore over large areas from soil samples to optimize future mining efforts. These methods have been adapted for many different situations. In this chapter, geostatistics are used to examine the weed distribution inside a single production field, variations of distribution over time, skewed data distributions, and correlations with species traits. A geostatistical study starts with selecting a sampling plan that “catches” the spatial relationships among the variables of interest. Exploratory data analysis then examines data distributions and checks whether the prerequisites for a geostatistical analysis are fulfilled. If necessary, data are transformed and detrended to meet these prerequisites. Then, empirical semivariograms are calculated and used to (1) explain small-scale spatial trends (e.g., weed patch shapes and progress in time as a function of species dispersal and germination traits), (2) determine the variances for unsampled distances to allow prediction of values in unsampled points and maps to be plotted, using kriging, and (3) reduce estimation errors at unsampled points. Cross-semivariograms and cokriging describe covariation of variables in space, and these relationships are used to estimate a sparsely sampled primary variable with the help of an extensively sampled secondary variable. Here, these methods were adapted to predict weed maps with past observations and variograms. Last, error analysis evaluates how close predicted weed maps are to observations and the risk of spraying insufficiently or unnecessarily when basing herbicide spraying in precision agriculture on weed maps predicted with past observations using cokriging.

## 15.2 INTRODUCTION

Geostatistics, originally developed for the mining industry, have been adapted to natural resource management, climatology, and ecology. The application of geostatistics to ecology and related disciplines is hindered by extremely skewed data resulting from aggregation in vegetation. For instance, weeds occur in patches<sup>1-9</sup> because they

tend to cluster where conditions, such as nutrient and soil moisture availability, are favorable for growth, persistence of propagule banks, and limited short distance dispersal of seed and propagules.<sup>4,10</sup>

In addition, the relevant working scale of a geostatistical application in an agricultural application can be quite different from the original large-scale mining applications. For many weed species, their abundance in a given field can mostly be explained by the past field history,<sup>11</sup> without any relationship to neighboring fields. Moreover, weed densities and locations, unlike ore deposits, vary over time. Indeed, fields are mostly infested by annual species,<sup>11</sup> which emerge, grow, reproduce, and die inside a single crop season. The success of individual species depends on weather as well as the crop sown by the farmer and on the choice of control techniques (e.g., herbicide and tillage) that interfere with the weed life cycle.<sup>12,13</sup> Last, agronomists and ecologists want not only to predict the future of weed populations but also to understand and explain these dynamics. The objective of this chapter is to demonstrate how geostatistics can be used to (1) describe small-scale spatial trends in weed populations and create weed contour maps, (2) relate the observed spatial variability to the weeds biological characteristics, (3) describe the development of the spatial trends over time and to predict the location of future weed patches, and (4) evaluate the usefulness of actual and predicted weed maps for site-specific herbicide spraying.

## 15.3 ANALYSIS STEPS

A geostatistical study starts with data collection, choosing an adequate sampling plan to “catch” the spatial relationships to be studied (Section 15.4). Exploratory data analysis then looks at data distributions and checks whether the prerequisites (i.e., normal or at least unskewed distribution, independence of mean and variance) for a geostatistical analysis are fulfilled (Section 15.5). If necessary, data are transformed to meet these prerequisites (Section 15.6). The next step checks for trends, i.e., large-scale spatial relationships, in data and tries to remove these with additional data transformations (Section 15.7). Only then are calculated empirical semivariograms to describe small-scale spatial trends, i.e., the variance of a variable between locations as a function of the distance between these points (Section 15.8). Fitting models to these data (Section 15.9) makes possible the estimation of variogram parameters, which can then be correlated to other data to explain small-scale spatial trends (Section 15.10). Semivariogram models are also used to estimate variance for unsampled distances and, thus, to plot maps with kriging (Section 15.11). Cross-semivariograms and cokriging (Section 15.12) describe the variance between two variables as a function of the distance between the sampling points and then use these relationships to estimate a sparsely sampled primary variable with the help of an extensively sampled secondary variable. Last, error analysis compares measured values with predicted values (Section 15.13).

Several software programs were used in the present work. Simple data analyses were carried out with Excel or SAS.<sup>14</sup> All the SAS program samples shown here were written for a UNIX environment (see an SAS manual,<sup>14</sup> for specific SAS script instructions). SAS was always run in the same directory where the data file was located; data files were obtained by saving the relevant excel sheets as text files (*prn* files, using blanks as separators). In the boxes that contain SAS code, procedure

names, options, etc. are shown in black, variable names, etc. chosen by the user are in **bold italics** (in SAS code, these will be blue) and comments are in *italics* (in SAS code, these will be green). SAS also was used for calculating empirical semi-variograms, fitting variogram fittings, and parameter analyses. GSLIB<sup>15</sup> (DOS version) was used for calculating empirical semivariograms and cross-semivariograms, kriging, and cokriging and maps. GSLIB is a free program that can be downloaded in several versions including DOS and a newer Windows (WinGslib version 1.5.6) version from the Web site <http://www.gslib.com/>. (Note: Program routines given in boxes are for GSLIB DOS version; however, these same parameters are asked for and used as inputs into the appropriate boxes for the WinGslib version.)

## 15.4 DATA COLLECTION

### 15.4.1 EXPERIMENTAL FIELD

The details can be found in Colbach et al.<sup>16</sup> with only the main points repeated here. A field survey of weed seedling populations was conducted from 1993 to 1997 at the Swan Lake Research Farm, Stevens Co., Minnesota. The field was 54 m wide (east–west) and 244 m long (north–south). Each year in mid-May, the field was planted with soybean, using a no-till planter. Rows were spaced 0.76 m apart and oriented north to south axis of the field. Weeds were treated with postemergence herbicides and interrow cultivation. An additional application of glyphosate was applied to control perennial weeds, either before planting or after harvest.

### 15.4.2 SAMPLING GRID

The sampling grid must be adapted to capture the spatial relationships that are to be analyzed. The minimum sampling distance must be carefully chosen, with this distance considerably lower than the distance at which spatial relationships are expected. Usually, regular sampling grids are preferred to avoid subconscious choices of the assessor (e.g., a preference for large and numerous plants) when placing quadrats in random sampling plans. In this study, weed seedlings were identified and counted in permanent 0.1 m<sup>2</sup> quadrats that covered both the crop-row and the interrow areas. Assessments were carried out once a year (except in 1995 when the data were not collected) prior to postemergence herbicide applications. Weed density by species was obtained at the same locations each year. Beginning at the field margins, samples were collected at 410 locations located on regular sampling grid with 10 rows and 41 columns (Figure 15.1).

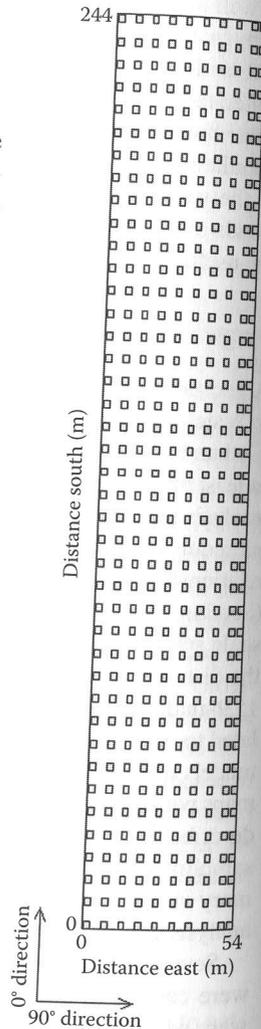


FIGURE 15.1 Sampling grid at the Swan Lake experiment station.

TABLE 15.1

Extract of Data File Found on CD in Chapter 15 Folder Labeled *Chapter\_15\_data.xls* for the Worksheet *Data1993*

xlocation	ylocation	aggre	amare	ascsy	cheal	cirar	setvi	sinar
6.1	0	50	10	0	30	0	50	0
6.1	6.1	10	0	0	0	0	20	0
6.1	12.2	0	10	0	0	0	0	0
6.1	18.3	0	10	0	20	0	0	0
6.1	24.4	0	0	0	0	0	20	0
6.1	30.5	0	0	0	0	0	10	0
6.1	36.6	0	10	0	0	0	60	0
6.1	42.7	0	0	0	0	0	10	0
6.1	48.8	0	0	0	0	0	20	0
6.1	54.9	0	0	0	0	0	20	0
[...]								

Note: The file contains several other worksheets that have weed density data by species and  $x$  and  $y$  locations from the Swan Lake Research Farm, Stevens Co., Minnesota.

Distances between grid points were 6.1 m in both the  $x$  and  $y$  directions, except between the last two sample rows that were separated by 3.05 m. Seven plant species were identified and analyzed in detail. These included redroot pigweed (*Amaranthus retroflexus* L.), common milkweed (*Asclepias syriaca* L.), wild mustard [*Brassica kaber* (DC.) L.C. Wheeler (= *Sinapis arvensis* L.)], common lambsquarters (*Chenopodium album* L.), Canada thistle (*Cirsium arvense* (L.) Scop.), quackgrass [*Elytrigia repens* (L.) Nevski], and green foxtail [*Setaria viridis* (L.) Beauv.].

In this study,  $x$  and  $y$  always indicate spatial coordinates (m), and  $z$ , the dependent variable, describes weed density (plants/m<sup>2</sup>). Table 15.1 shows an extract of the *Chapter\_15\_data.xls* data file for the *Data1993* worksheet (found on the CD in Chapter 15 file) containing the weed densities counted on the sampling grid. Figure 15.2 shows an example of densities counted for *S. viridis* in 1993 with the sampling grid shown in Figure 15.1.

## 15.5 EXPLORATORY DATA ANALYSIS

### 15.5.1 OBJECTIVE

The conditions for geostatistical analysis are that the values be normally distributed and independent. Quite often, this is not the case, especially when looking at vegetation data such as weeds. Weed densities tend to be skewed, with a large number of sampling points having zero or very small values (Figure 15.3). One of the first data analysis steps is to determine the frequency distribution and, if necessary, transform the data to better fit a normal distribution. This step involves calculating the mean, variance, skewness, kurtosis, and a test for normality. In addition, mean and variances

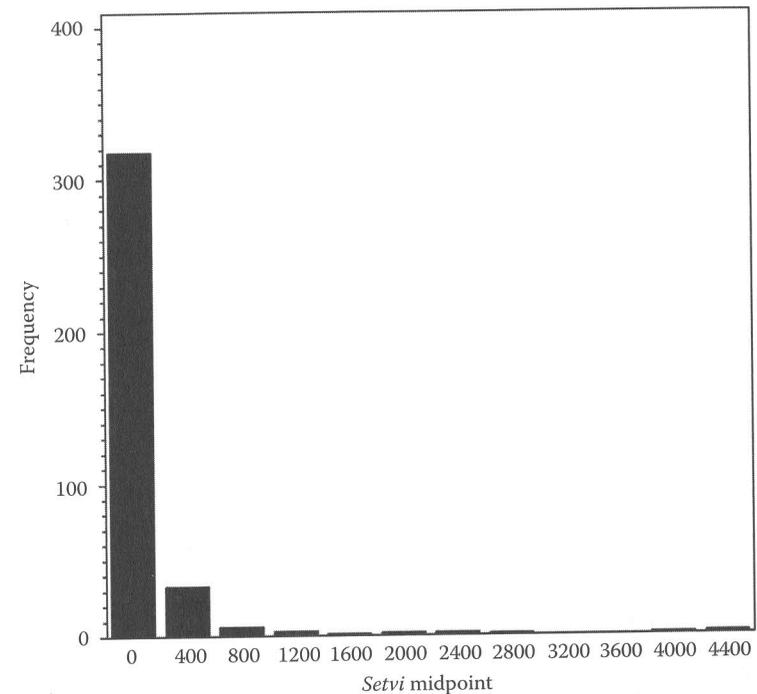


**FIGURE 15.2** Weed densities counted in the sampling quadrats at the Swan Lake experimental field. Example of *S. viridis* distributions in 1993 (a) and 1997 (b). Graph obtained with **PROC GMAP** of SAS using data from *Data1993* and *Data1997* worksheets for *Chapter\_15\_data.xls* (see Box 15.3).

were calculated along rows and columns, and a linear regression was used to check for a possible correlation between  $\log_e(\text{variance})$  and  $\log_e(\text{mean})$ <sup>18</sup> (Box 15.2).

### 15.5.2 METHOD

There are a range of different software applications that achieve these calculations and tests. In Excel, the statistical functions, **mean**, **standard error**, and **skew**, can be found in the drop-down menu Insert/Functions. In the Excel file available on the



**FIGURE 15.3** Distribution of *S. viridis* densities counted on the Swan Lake experimental field in 1993. Graph produced with **PROC GCHART** of SAS.

CD for Chapter 15 with the title *Chapter\_15\_data.xls* data, some of these variables have been calculated and placed on top of each data sheet. In SAS, these variables can all be calculated with the **PROC UNIVARIATE** function (Box 15.1). (*Note:* remember that the *.xls* files were converted to *.prn* files for upload into SAS, and, if using other file formats, use appropriate infile or datafile variables; see Section 15.3.) Box 15.1 shows part of the output produced by this function for *SETVI*. To examine the correlation among data points in both the *x* and *y* directions, the mean–variance correlation can be tested in SAS (see Box 15.2). First, the data for mean and variance are power transformed in each direction (Box and Cox test<sup>17</sup>), using the natural log [i.e.,  $\log_{\text{mean}} = \log_e(\text{mean})$  or  $\log_{\text{var}} = \log_e(\text{variance})$ ], and then the results are regressed using model  $\log_{\text{var}} = \text{constant} + \text{slope} \times \log_{\text{mean}} + \text{error}$  (Box 15.3).

### 15.5.3 RESULTS

Basic statistics (Table 15.2) indicate a wide variation in mean plant densities, ranging from 0 for *Asclepias syriaca* to over 150 plants/m<sup>2</sup> for *S. viridis*, and high values for standard deviations of the means. In the present example, none of the density distributions were normal, and all were highly skewed. In addition, variance and mean were always significantly correlated (*P* values <0.001) (see parameter estimate in output [Box 15.2] for dependent variable *logvar*).

## BOX 15.1

SAS program (top) for calculating descriptive statistics for weed species and extract of output for *SETVI* variable (bottom) of the *Data1993* worksheet of *Chapter\_15\_data.xls*, descriptive statistics for the other weeds are shown in Table 15.2.

```
data table1;
infile 'Data1993.prn' firstobs = 6;
input xlocation ylocation agrre amare ascsy cheal cirar setvi sinar;
proc univariate normal;
var agrre amare ascsy cheal cirar setvi sinar;
run; quit;
```

The UNIVARIATE Procedure  
Variable: setvi

		Moments	
N		369	369
Mean	151.273713	Sum Weights	55820
Std Deviation	490.041317	Sum Observations	240140.493
Skewness	6.43733769	Variance	47.2596874
Uncorrected SS	96815800	Kurtosis	88371701.4
Coeff Variation	323.943472	Corrected SS	25.510532
[..]		Std Error Mean	

## Tests for Normality

Test	--Statistic--		-----p Value-----	
Shapiro-Wilk	W	0.304304	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.378776	Pr > D	<0.0100
Cramer-von Mises	W-Sq	17.66014	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	86.49038	Pr > A-Sq	<0.0050

[..]

## BOX 15.2

SAS program (top) and extract of output (bottom) for analyzing mean-variance dependency of the *SETVI* variable. This test consists in regressing  $\log_e(\text{variance})$  against  $\log_e(\text{mean})$ ; if the slope is significant (i.e., if  $\text{Pr} > |t|$  is lower than .05), the dependent variable  $y$  should be transformed as  $y^{\text{slope}/2}$  (Box and Cox test<sup>17</sup>).

```
data table1;
infile 'Data1993.prn' firstobs = 6;
input xlocation ylocation agrre amare ascsy cheal
      cirar setvi sinar;

y = setvi;
proc sort; by xlocation;
proc means noprint;
var y; by xlocation;
output out=tablex mean = mean var = variance;

data table1; set table1;
proc sort; by ylocation;
proc means noprint;
var setvi; by ylocation;
output out=tabley mean = mean var = variance;

data table2; set tablex tabley;
logmean = log(mean);
logvar = log(variance);
proc reg;
model logvar = logmean;
run; quit;
```

The REG Procedure  
Model: MODEL1  
Dependent Variable: logvar

[..]

## Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	1.66556	0.61979	2.69	0.0312
logmean	1	1.78683	0.14511	12.31	<.0001

## BOX 15.3

SAS program for PROC GMAP used to produce the output of Figure 15.2.

```
*---distance between sampling points---;
%let length = 6.1;

*---reading data file with observed weed densities---;
data table1;
infile 'Data1993.prn' firstobs = 6;
input xlocation ylocation agrre amare ascsy cheal
      cirar setvi sinar;
x = xlocation;
y = ylocation;
plot = _N_; *creating a plot name from line number _N_;

*---create coordinate table for proc gmap---;
*for each (x,y) location, four coordinates (x+length/2,
y+length/2), (x+length/2,y-length/2), (x-length/2,
y-length/2) and (x-length/2,y+length/2) are created
where length = sampling distance;
data table3; set table1;
x = x+&length/2;
y = y+&length/2;
keep x y plot;
data table4; set table1;
x = x+&length/2;
y = y-&length/2;
keep x y plot;
data table5; set table1;
x = x-&length/2;
y = y-&length/2;
keep x y plot;
data table6; set table1;
x = x-&length/2;
y = y+&length/2;
keep x y plot;
data table3; set table3 table4 table5 table6;
keep x y plot;
proc sort; by plot;

*---draw map---;
goptions reset=all cback=white colors=(white grayee
graydd graycc graybb grayaa gray77 gray66 gray56
gray44 black); *color options for background and map;
```

## BOX 15.3 (continued)

```
pattern value=msolid; *each basic spatial unit is to
be filled;
proc gmap data=table1 map=table3;
*table1 contains the weed densities, table3 the
coordinates;
id plot; *plot is the identification of the basic
spatial unit;
choro setvi/coutline=same cempty=white ctext=black
midpoints= 0 to 150 by 15;
*setvi is the variable to represent on the map, the
color outline of each plot is the same as the color
used for filling, empty plots are drawn with white
lines, the legend is written in black and densities
are represented from 0 to 150 by 15;
run; quit;
```

TABLE 15.2

Exploratory Data Analysis for Weed Species of the *Data1993* Worksheet of *Chapter\_15\_data.xls* Calculated Using PROC UNIVARIATE and PROC GLM of SAS (for SAS Code, Refer to Boxes 15.1 and 15.2)

Weed Species	Bayer Code	Mean (Plants/m <sup>2</sup> )	Standard Deviation (Plants/m <sup>2</sup> )	Skewness <sup>a</sup>	Test for Normality <sup>b</sup>	Mean-Variance Correlation <sup>c</sup>
<i>Elytrigia repens</i>	aggre	4.6	42.9	15.86	<0.0001	<0.0001
<i>Amaranthus retroflexus</i>	amare	10.5	19.8	2.88	<0.0001	0.0011
<i>Asclepias syriaca</i>	ascsy	0.0	0.0			
<i>Chenopodium album</i>	cheal	2.7	10.2	5.23	<0.0001	<0.0001
<i>Cirsium arvense</i>	cirar	4.6	13.0	3.95	<0.0001	<0.0001
<i>Setaria viridis</i>	setvi	151.2	490.0	6.44	<0.0001	<0.0001
<i>Brassica kaber</i>	sinar	0.8	3.0	3.93	<0.0001	0.0010

<sup>a</sup> The more different the skewness value (calculated with PROC UNIVARIATE of SAS, see Box 15.1) is from zero, the more skewed (asymmetrical) the data distribution.

<sup>b</sup> Probability of error when stating that the distribution is not normal (Shapiro-Wilk test, see Box 15.1).

<sup>c</sup> P statistic for correlation between log<sub>e</sub>(variance) and log<sub>e</sub>(mean) calculated using PROC REG of SAS (see example in Box 15.2).

## 15.6 DATA TRANSFORMATION

### 15.6.1 OBJECTIVE

The above analysis indicated that the data were highly skewed, variances and means were correlated, and the dependent variable  $z$  had to be transformed before analysis. The objective of this step was to identify a suitable transformation that decorrelates the variance and mean values and converts the data to a normal, or less skewed, distribution. One approach frequently used to obtain a normal distribution is to remove outliers. Unfortunately, in the case of weed densities, skewed distributions are not just due to a single point (Figure 15.3). It is therefore necessary to find variable transformations that do not remove large portions of the observed data.

### 15.6.2 METHOD

Box and Cox<sup>17</sup> proposed a transformation to decorrelate variances and means. If the linear regression  $\log_e(\text{variance}) = a + b \log_e(\text{mean})$  is significant, then, a suggested transformation for the dependent variable  $z$  is  $z^{1-b/2}$ . Another interesting transformation that may be tried is  $\log_e(z + k)$  transformation where  $k$  is a constant. The constant,  $k$ , is needed because of the many zero values. In addition, a suitable chosen value for  $k$  can achieve normality or decrease skewness.<sup>18</sup> Other transformations tested on the Swan Lake data included  $z^k$ ,  $e^{kz}$ , and  $\log_e(\log_e(z + k))$ .

### 15.6.3 RESULTS

The best results (i.e., lowest skewness and decorrelation of mean–variance) were obtained using the  $\log_e(z + 1)$  transformation, even though none of the transformed variables showed a normal data distribution (Table 15.3).

## 15.7 DETRENDING DATA

### 15.7.1 OBJECTIVE

Detrending datasets aims to remove large-scale spatial trends. In the present case study, soil depth and texture could, for instance, vary in the experimental field and lead to a large-scale variation in weed densities.

### 15.7.2 MEDIAN POLISHING

Even with transformation, a mean–variance correlation was still significant in the dataset. To reduce this correlation, one should look for a large-scale spatial trend in the data and apply a detrending transformation such as median polishing:

$$z' = z - \text{row median} - \text{column median} + \text{overall median} \quad (15.1)$$

This transformation can only be used for gridded data and only works for additive trends. It is not adequate if the trend comprises an interaction between row and

TABLE 15.3

Variable Transformation Tested for the *S. viridis* (SETVI) Densities Observed at Swan Lake in 1993

Variable Transformation	Mean (Plants/m <sup>2</sup> )	Standard Deviation (Plants/m <sup>2</sup> )	Skewness <sup>a</sup>	Test for Normality <sup>b</sup>	Mean–Variance Correlation <sup>c</sup>
None	151.2	490.0	6.44	<0.0001	<0.0001
$y^{1-1.78/2}$	1.10	0.742	-0.627	<0.0001	0.2170
$\log_e(y + 0.1)$	2.22	3.14	-0.441	<0.0001	0.3432
<b><math>\log_e(y + 1)</math></b>	<b>2.91</b>	<b>2.22</b>	<b>0.053</b>	<b>&lt;0.0001</b>	<b>0.7779</b>
$\log_e(y + 10)$	3.75	1.40	0.92	<0.0001	0.0258
$\exp(y)$	10 <sup>258</sup>				<0.0001
$\log_e(\log_e(y + 1) + 1)$	1.13	0.759	-0.629	<0.0001	0.2146

Notes: Statistics were calculated with PROC UNIVARIATE and PROC GLM of SAS. The best transformation is indicated in bold.

- <sup>a</sup> The more different the skewness value (calculated with PROC UNIVARIATE of SAS, Box 15.1) is from zero, the more skewed (asymmetrical) the data distribution.
- <sup>b</sup> Probability of error when stating that the distribution is not normal (Shapiro–Wilk test with PROC UNIVARIATE of SAS, see Box 15.1).
- <sup>c</sup>  $P$  for correlation between  $\log_e(\text{variance})$  and  $\log_e(\text{mean})$  calculated with PROC REG of SAS (see Box 15.2).

column variables. Other transformations  $z' = f(z, \text{row}, \text{column})$  should then be tested. Moreover, median polishing does not work well if the mean plant density is low, for example, 1 plant/m<sup>2</sup>. In that case, row, column, and overall median values are nil, and Equation 15.1 will not modify variable  $z$ . If this occurs, the correlation of mean and variance is then not due to a large-scale trend, but simply to the extreme patchiness of the weed population.

With the 1993 Swan Lake data, median polishing was only possible for the more frequent species such as AMARE. In this instance, the transformed variable  $z'$  presented independent means and variances ( $P < 0.05$ ), and the density distribution was not skewed (0.16 vs. 0.67 for log-transformed data). Despite this apparent improvement of data distribution, median polishing was not satisfactory because it introduced artifacts into the data, i.e., weed patches are now present where there were none before, etc. This was probably because the seedling densities were so low.

### 15.7.3 ESTIMATING TREND WITH LINEAR REGRESSIONS

We tried other detrending transformations by estimating the large-scale trend by fitting a linear regression to the  $\log_e(z + 1)$  data (Box 15.4):

$$\log_e(z+1) = \text{constant} + \alpha \cdot \text{row} + \beta \cdot \text{column} + \gamma \cdot \text{row}^2 + \delta \cdot \text{column}^2 + \zeta \cdot \text{row} \cdot \text{column} + \text{error} \quad (15.2)$$

## BOX 15.4

SAS program (top) and output (bottom) for determining a large-scale trend in densities using *Setaria viridis* as an example. Trends for other species can be explored by changing the input variable (*SETVI*).

```

data table1;
infile 'Data1993.prn' firstobs = 6;
input xlocation ylocation agrre amare ascsy cheal cirar setvi sinar;
logsetvi=log (setvi+1);
proc glm;
model logsetvi = xlocation ylocation * xlocation * ylocation * ylocation * xlocation * ylocation * ylocation
/solution;
run; quit;

```

Dependent Variable: logsetvi

Source	DF	Squares	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	285.126316		57.025263	13.50	<.0001
Error	363	1533.399489		4.224241		
Corrected Total	368	1818.525805				
R-Square		Coeff Var	Root MSE	setvi Mean		
0.156790		70.43131	2.055296	2.918157		

Source	DF	Type I SS	Mean Square	F Value	Pr > F
xlocation	1	48.1963485	48.1963485	11.41	0.0008
ylocation	1	14.7703157	14.7703157	3.50	0.0623
xlocation*xlocation	1	83.9781075	83.9781075	19.88	<.0001
ylocation*ylocation	1	1.0143972	1.0143972	0.24	0.6244
xlocation*ylocation	1	137.1671471	137.1671471	32.47	<.0001
Source	DF	Type III SS	Mean Square	F Value	Pr > F
xlocation	1	34.6634823	34.6634823	8.21	0.0044
ylocation	1	12.5222972	12.5222972	2.96	0.0860
xlocation*xlocation	1	83.9781075	83.9781075	19.88	<.0001
ylocation*ylocation	1	1.0143972	1.0143972	0.24	0.6244
xlocation*ylocation	1	137.1671471	137.1671471	32.47	<.0001

Standard

Parameter	Estimate	t Value	Pr >  t
Intercept	3.612001877	5.78	<.0001
xlocation	-0.101280327	-2.86	0.0044
ylocation	0.011190925	1.72	0.0860
xlocation*xlocation	0.002458564	4.46	<.0001
ylocation*ylocation	0.000011263	0.49	0.6244
xlocation*ylocation	-0.0000554058	-5.70	<.0001

The final model using only those explanatory variables with  $\alpha \leq 0.05$  then constitutes the large-scale trend that should be retracted from the  $z$  variable. In the case of the *S. viridis* data from 1993, the final regression model (15.2) was

$$\begin{aligned} \text{trend} = & 88.06 - 19.48 \cdot x \text{ location} + 2.34 \cdot y \text{ location} \\ & + 0.64 \cdot x \text{ location} \cdot x \text{ location} - 0.10 \cdot x \text{ location} \cdot y \text{ location} \end{aligned}$$

The detrended variable then becomes

$$z' = \log_e(z+1) - \text{trend} \quad (15.3)$$

The  $R^2$  of this final model was very low (0.156) (Box 15.4), indicating that there was no large-scale trend. In addition, the resulting detrended  $z'$  variable presented worse distribution characteristics (skewness =  $-0.097$ ,  $P$  for mean–variance correlation = 0.0002) than the untrended  $\log_e(z+1)$  variable. Consequently, weed densities were only transformed with  $\log_e(z+1)$  for the subsequent analysis. As the distance over which the variograms were calculated was limited and as ordinary kriging (which does not assume a stable mean over the whole field) was used, the resulting bias would be negligible.

## 15.8 EMPIRICAL SEMIVARIOGRAMS

### 15.8.1 OBJECTIVE

Semivariograms describe small-scale spatial trends, i.e., the variance between locations as a function of the distance between these points. Semivariograms were developed for each species and year, using the following equation:

$$\gamma_h = \frac{1}{2N_h} \sum (z_{i+h} - z_i)^2 \quad (15.4)$$

where

- $\gamma_h$  is the empirical semivariance for the distance  $h$
- $N_h$  is the number of points separated by the distance  $h$
- $z_i$  is the weed density at location  $i$

This statistic is then plotted for each separation distance  $h$  (termed an empirical semivariogram) and characterizes the spatial variability of weed densities as a function of distance among locations. Separate empirical semivariograms were established in four directions: along the rows (i.e.,  $0^\circ$  or north-south direction), across the rows (i.e.,  $90^\circ$  or east-west direction), as well as along the diagonals (i.e.,  $45^\circ$  and  $135^\circ$ ) in order to check for anisotropy. For each of these four directions, only points located at an angle of  $\pm\alpha^\circ$  relative to the nominal direction (i.e.,  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , or  $135^\circ$ ) were used for the semivariogram. In the present study, the lowest possible angle was used to optimize the discrimination of the analyzed directions.

The maximum distance between points used for the semivariogram can be limited. Here, distance was limited to 50 m for both directions, as the greatest possible distance for both  $x$  and  $y$  axes (the field was only 54 m wide).

### 15.8.2 METHOD

In the original paper by Colbach et al.,<sup>16</sup> the semivariograms were calculated with **GSLIB** (DOS version) using the **GAMV** function, which is convenient for irregular data as the sampling grid in this research was not entirely regular. To run **GAMV**, parameterize the **GAMV.par** file located in **GSlib/DOSEXEC/GAMV** (see, e.g., Box 15.5) and double click on **GAMV.EXE** located in **GSlib/DOSEXEC/BIN** and enter the parameter file name (i.e., **GAMV.par** with the relevant path if applicable). The output will be located in the directory specified in the **GAMV.par** file. Alternatively, **gamv** function in the drop-down menu of **Variogram** of **WinGslib** could be used (Figure 15.4). On the **Files** page, **Input** and **Output** Files are defined, as well as the **number of lags**, **unit lag separation distance**, and **lags tolerance**. On the **Variograms** page, data for the **azimuth** (angles and associated information) and **Tail/Head** data are selected for the variograms. After data entry, click the **triangular “play”** button to run the program. The **“page”** button can then be clicked to see the output. Here, we also used **PROC VARIOGRAM** of SAS to calculate variograms, followed by **PROC GPLOT** to draw the graph (Box 15.6).

### 15.8.3 RESULTS

Figure 15.5 shows an example of the empirical semivariograms calculated along rows ( $0^\circ$  direction), columns ( $90^\circ$ ), and the two diagonals ( $45^\circ$  and  $135^\circ$ ) for SETVI in 2003. For both directions, variance was high at the lowest sampling distance (i.e., 6.1 m); this variance at low distances close to zero is called *nugget*. Variance increased more or less with increasing distance between analyzed data points. In the case of the  $135^\circ$  direction, variance reaches a threshold value (called *sill*) at approximately 25 m while in the  $0^\circ$  direction, variance continued to increase up to 40 m. These distances are called *ranges*, and a variation in range with direction points to geometric anisotropy. Final variance was much lower in the  $0^\circ$  direction than in any other direction, pointing to a marked zonal anisotropy.

The conclusions that can be drawn from a graphical analysis of empirical semivariograms are though limited. Thus, the next step fits models to the empirical semivariogram.

## 15.9 SEMIVARIOGRAM MODEL FITTING

### 15.9.1 OBJECTIVE

Empirical semivariograms allow description of spatial correlations. Fitting models to these data makes possible the estimation of (1) variance for unsampled distances, which is necessary to plot maps, and (2) variogram parameters, which can then be correlated to other data.

## BOX 15.5

Data file extract from *setvi199394.dat* (top) and parameter file *GAMV.par* (bottom) for running *GAMV.EXE* of *GSLIB* for calculating the 1993 semivariogram for *S. viridis* and the cross-semivariograms of 1993 with 1994 densities. The *setvi199394.dat* sheet was created manually by writing the six first lines and then by copying the relevant columns from the *Data1993.xls* and *Data1994.xls* datasheets. Both years of data were computed for four directions (0°, 45°, 90°, and 135°) and an angle tolerance (*atol*) of 0.1° (e.g., only data points within 45° ± 0.1° are considered for the 45° direction); 8 distance classes (**number of lags option**), **distance** by 6.1 m and **lag separation distance** option with a tolerance of 3.05 m were considered, thus using pairs of data points distanced up to 8·6.1 = 48.8 m in the 0° and 90° directions.

```
Setvi densities for 1993 and 1994
4
Xlocation
Ylocation
Setvi1993
Setvi1994
6.1 0 3.931825633 3.713572067
6.1 6.1 3.044522438 0
6.1 12.2 0 2.397895273
6.1 18.3 0 0
6.1 24.4 3.044522438 0
6.1 30.5 2.397895273 0
[...]
```

```
Parameters for GAMV
*****
START OF PARAMETERS:
c:/swanlake/setvi199394.dat
1 2 0
2 3 4
-98 1.0e21
c:/swanlake/setvi1993.sem
8
6.1
3.05
4
0.0 0.1 0.1 0.0 0.10 0.10
45.0 0.1 0.1 0.0 0.10 0.10
90.0 0.1 0.1 0.0 0.10 0.10
135.0 0.1 0.1 0.0 0.10 0.10
0
2
1 1 1 1
1 2 2 2
type 1 = traditional semivariogram
2 = traditional cross semivariogram
3 = covariance
4 = correlogram
5 = general relative semivariogram
6 = pairwise relative semivariogram
7 = semivariogram of logarithms
8 = semimadogram
9 = indicator semivariogram - continuous
10 = indicator semivariogram - categorical

\file with data
\columns for X, Y, Z coordinates (here, no z)
\number of variables,column numbers
\trimming limits
\file for variogram output
\number of lags
\lag separation distance
\lag tolerance
\number of directions
\azm,atol,bandh,dip,dtol,bandv
\azm,atol,bandh,dip,dtol,bandv
\azm,atol,bandh,dip,dtol,bandv
\azm,atol,bandh,dip,dtol,bandv
\standardize sills? (0=no, 1=yes)
\number of variograms
\tail=1st var., head=1st var., variogram type
\tail=1st var., head=2nd var., variogram type
```

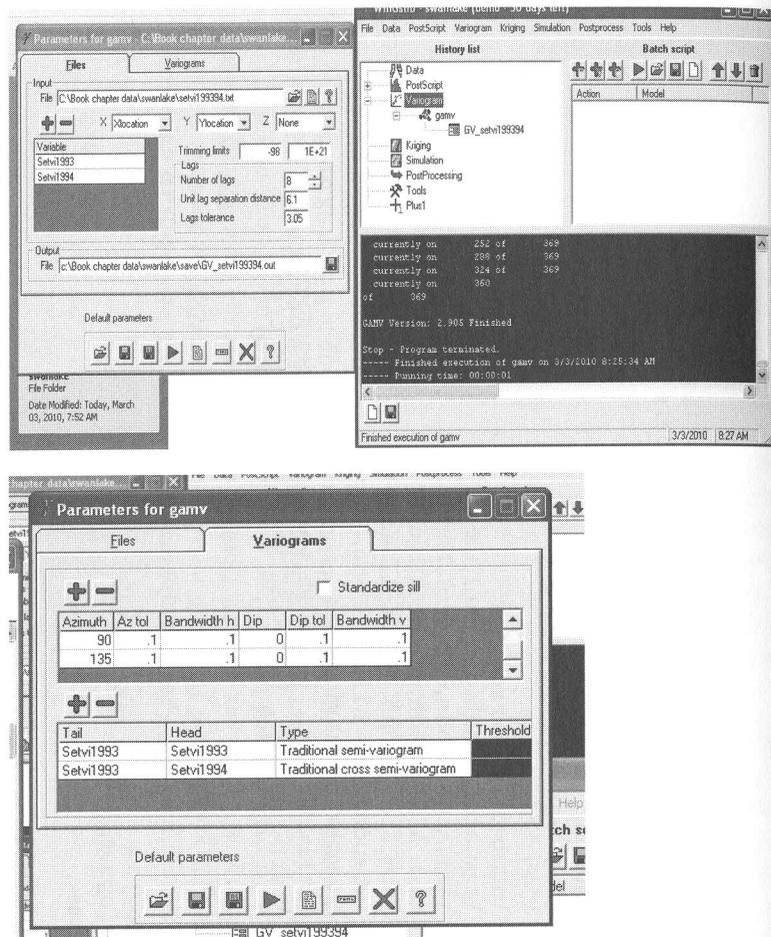


FIGURE 15.4 Using WinGslib to process Variogram information using **gamv** function for parameters given in Box 15.5 (DOS code for **GSLIB**).

Various contrasting models can be fitted to empirical variograms, for example, linear, power, spherical, exponential, or Gaussian models. The latter can be problematic as its derivative at distance = 0 is zero. In the present case, the linear and power models were inadequate because they do not allow for a sill value (i.e., variance reaching a threshold for large distances). Spherical (vs. exponential) models present the advantage of having a “real” range (i.e., the distance when variance reaches its sill). Here, a nested spherical model was used, i.e., a sum of two spherical models and a nugget value, with ranges depending on the directions.

$$\text{Model 1: } \begin{cases} \text{if } h < a_1 & \gamma_1(h) = c_1 \cdot \left[ 1.5 \cdot \frac{h}{a_1} - 0.5 \cdot \left( \frac{h}{a_1} \right)^3 \right] \\ \text{if } h \geq a_1 & \gamma_1(h) = c_1 \end{cases}$$

### BOX 15.6

SAS program for calculating (**PROC VARIOGRAM**) and drawing (**PROC GPLOT**) an empirical semivariogram. Output is shown in Figure 15.5. The information used for computing the semivariograms is described in Box 15.5.

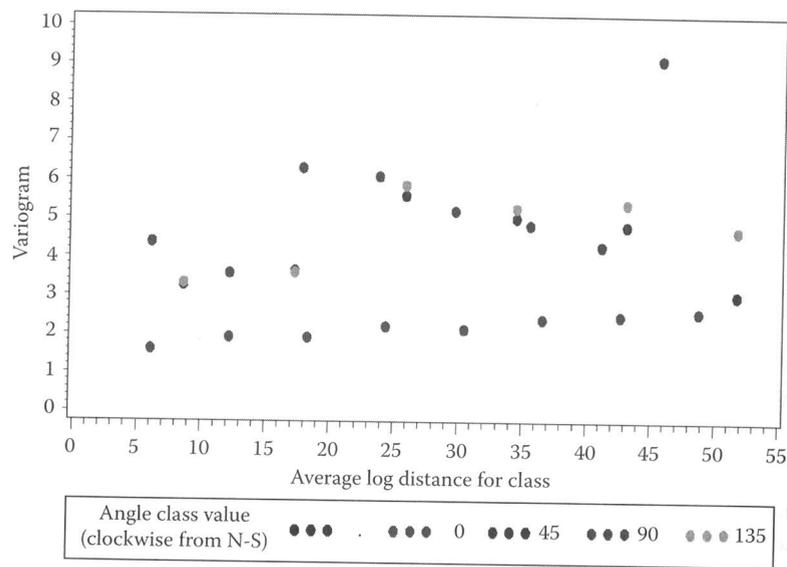
```
data table1;
infile 'Data1993.prn' firstobs = 6;
input xlocation ylocation agrre amare ascsy cheal cirar
      setvi sinar;
*---data transformation---;
logsetvi=log(setvi+1);
*---calculating empirical semivariogram---;
proc variogram outVar=tabVariogram;
compute angletolerance=0.1 lagdistance=6.1 ndirec-
tions=4 maxlags=8;
coordinates xcoord=xlocation ycoord=ylocation;
directions 0 (0.1) 45(0.1) 90 (0.1) 135(0.1);
var logsetvi;
*---drawing graph of empirical semivariogram---;
goptions reset=all cback=white colors=(black red blue
green orange);*
symbol1 v=dot;
axis1 order=(0 to 10 by 1);
axis2 order=(0 to 55 by 5);
proc gplot data=tabVariogram;
plot variog*distance=angle /vaxis=axis1 haxis = axis2;
run; quit;
* Color graph is on the CD accompanying this book.
```

$$\text{Model 2: } \begin{cases} \text{if } h < a_2 & \gamma_2(h) = c_2 \cdot \left[ 1.5 \cdot \frac{h}{a_2} - 0.5 \cdot \left( \frac{h}{a_2} \right)^3 \right] \\ \text{if } h \geq a_2 & \gamma_2(h) = c_2 \end{cases}$$

$$\text{Total model: } \gamma(h) = c_0 + \gamma_1(h) + \gamma_2(h) \quad (15.5)$$

where

- $c_0$  is the nugget (representing small-scale variation that cannot be described with the present sampling scheme)
- $c_1$  and  $c_2$  are the contributions of the first and second spatial structures to the total variance (sill)
- $a_1$  and  $a_2$  are the ranges (with different values for the  $0^\circ$  and  $90^\circ$  directions)



**FIGURE 15.5** Example of empirical semivariogram calculated with Equation 15.4 for *S. viridis* in 1993 with **PROC VARIOGRAM** of SAS (Box 15.6) (black circles appear in the legend because of empty lines in the SAS data table, which are considered as a fifth direction“.” by **PROC GPLOT**).

This model was fit to the empirical semivariogram for each species and year using an iterative least-squares procedure. Points with fewer than 50 pairs were excluded because they were considered unreliable.<sup>19–21</sup> Values for the ranges ( $a_1$ ,  $a_2$ ), contributions ( $c_1$ ,  $c_2$ ), and nugget ( $c_0$ ) were estimated using weighted least squares based on number of pairs,  $N_h$ .

Equation 15.5 has the advantage of covering a large range of possible situations. The most complicated situation occurs when all parameters are significantly different. A nested model is necessary when sills differ between directions (i.e., zonal anisotropy). If, however, the sills are identical and only the ranges differ, the analyzed variable presents a geometric anisotropy. Equation 15.5 can then be reduced to a single model with one sill, irrespective of direction. Only the ranges then vary with the direction.

### 15.9.2 METHOD

**PROC NLIN** of SAS was used to fit Equation 15.5 and estimate its parameters (Box 15.7).

### 15.9.3 RESULTS

Figure 15.6 shows an example of fitting a nested spherical model to an empirical semivariogram. For direction 0, the final variance (i.e., sill) is lower than for

### BOX 15.7

**SAS** program extract (top, must follow program shown in Box 15.6) and output extract (bottom, see also Figure 15.6) for fitting a variogram model to the empirical semivariogram. Explanations are given as comments in program (comment:). Model fitting is often difficult and requires bounding the parameters, using likely values seen of graphs. In the present example, bounds were used for *nugget*, *c1* and *a1\_0*.

```

data tabVariogram; set tabVariogram;
*to eliminate distances with less than 50 pairs;
if count >= 50;
proc nlin eformat method=gauss;
*initial parameter values to set off iteration (determined on graphs);
parms  nugget= 1.4
      c1 = 1.3
      c2 = 3.2
      a1_0 = 50
      a2_90= 17;
*two parameters are fixed as they are difficult to estimate because of insufficient sampling at distances close to zero and at larger distances;
a1_90= 0;
a2_0 = 400000;
h = distance; *renaming distance;
pi = 3.141592;
angle = angle*(pi/180); *transforming from degree to radian;

```

(continued)

## BOX 15.7 (continued)

```

*first model;
range1 = ( a1_90**2 * (sin(angle))**2 + a1_0**2 * (cos(angle))**2)**0.5;
if h < range1
then semi1 = c1 * ( 1.5 * (h/range1) - 0.5 * (h/range1)**3 );
else semi1 = c1;
*second model;
range2 = (a2_90**2 * (sin(angle))**2 + a2_0**2 * (cos(angle))**2 )**0.5;
if h < range2
then semi2 = c2 * ( 1.5 * (h/range2) - 0.5 * (h/range2)**3 );
else semi2 = c2;
*fixing limits on parameter values;
*logical limits; *Limits determined on graphs in case of bad fits;
bounds nugget >= 0; bounds nugget <= 1.4 ;
bounds c1 >= 0; bounds c1 <= 1.4 ;
bounds c2 >= 0;
bounds a1_0 >= 0; bounds a1_0 >= 45;
bounds a2_90 >= 0;
*sum of partial models;

```

```

model variog = nugget + semi1 + semi2;
*weighing least squares by the inverse of the number of pairs;
_weight_ = 1/count;
*preparing a data table with the predicted values;
output out=newp=pred;
*graph of predicted and observed variance vs. distance;;
data new; set new;
if angle = 0 or angle = 90; *to simplify graph;
goptions reset=all cback=white colors=(black );
symbol1 v=dot i=none c=red ;
symbol2 v=dot i=none c=green;
symbol3 v=none i=join c=red ;
symbol4 v=none i=join c=green;
axis1 order=(0 to 7 by 1);
axis2 order=(0 to 55 by 5);
proc gplot;
plot variog *distance=angle/vaxis=axis1 haxis = axis2;
plot2 pred*distance = angle/vaxis=axis1 haxis = axis2;
run; quit;

```

(continued)

## BOX 15.7 (continued)

## The NLIN Procedure

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Regression	2	2.5877	1.2939	-9.21	
Residual	25	0.3707	0.0148		
Uncorrected Total	27	2.9584			
Corrected Total	26	0.2342			

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	Label
nugget	1.4000	0	1.4000 1.4000	
c1	1.4000	0	1.4000 1.4000	
c2	2.1777	0.6746	0.7883 3.5671	
a1_0	45.0000	0	45.0000 45.0000	
a2_90	20.1176	26.3363	-34.1227 74.3579	
Bound4	0.1406	0.0368	0.0649 0.2163	
Bound6	0.1495	0.0407	0.0657 0.2333	
Bound1	0.000624	0.000368	0.00013 0.00138	

c1 <= 1.4  
nugget <= 1.4  
45 <= a1\_0

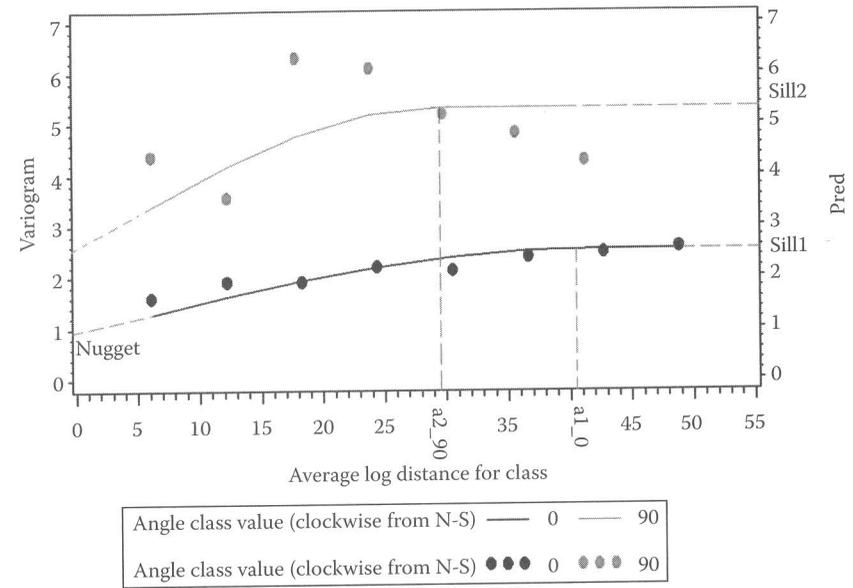


FIGURE 15.6 Example of fitting a nested spherical model (—) using Equation 15.5 to an empirical semivariogram for directions 0° and 90° for *S. viridis* in 1993 with **PROG NLIN** (Box 15.7). Lines and text were added manually to the **SAS** output graph.

direction 90 (i.e., sill2), pointing to lower variability in weed densities along vs. across crop rows. The nugget (i.e., the variance at zero distance) did not depend on the direction and is considered as “white noise,” i.e., small-scale variability that cannot be described with the present sampling scheme. The contribution  $c_1$  of the first model is the difference between sill1 and the nugget; it was reached after distance  $a1_0 = 40$  m in the 0° direction, whereas it was present at all distances in the 90° direction as  $a1_90$  was nil. In any other direction, the  $c_1$  contribution depended on the value of  $range1$  varying with the angle (see Box 15.7). The contribution  $c_2$  of the second model is the difference between sill2 and the nugget; it was reached after distance  $a2_90 = 29$  m in the 90° direction, whereas its effect was negligible in the 0° as the  $a2_0$  was infinite. In any other direction, the  $c_2$  contribution depended on the value of  $range2$  varying again with the angle (see Box 15.7). Model fitting was not always easy as variance did not increase smoothly (Figure 15.5). In addition, with nonlinear regressions, the result can vary considerably with the initial values used for the parameters or the method (DUD vs. GAUSS), and sometimes, it is necessary to manually limit the possible ranges for parameters (with the bounds option) to achieve not only convergence during **NLIN** iteration but also visually satisfactory results, i.e., fitted lines vs. observations.

In the field, this directional effect (anisotropy) was evident as elliptical weed patches that were longest in the direction of the crop rows (Figure 15.2). The most likely reason for the difference in ranges (geometric anisotropy) is that weed seeds and other propagules are moved in the direction of crop rows by agricultural implements,

such as tillage and harvesting equipment. Other factors, such as water and gravity, also may play a role in creating this distribution. A possible explanation for the difference in total variation (zonal anisotropy) would be the variation in performance of field implements (e.g., planter, cultivator, and combine). The differing speeds, depth adjustments, etc. that occur during north–south passes of these tools across the field may have contributed to an east–west heterogeneity in weed growth conditions.

## 15.10 ANALYSIS OF VARIOGRAM PARAMETERS

### 15.10.1 OBJECTIVE

The analysis of variogram parameters is pertinent when several variables are investigated in a site to examine the differences in weed densities and species. The simplest way is to explain variogram parameters as a function of a year and a weed species effect:

$$\text{Variogram parameter} = \text{constant} + \text{year effect} + \text{species effect} + \text{error} \quad (15.6)$$

This linear model was tested separately for each of the five parameters, i.e., the nugget, and the two contributions. The two ranges were analyzed together, adding a direction effect to model Equation 15.6.

The most interesting way is to look for relationships between the different variogram parameters and a series of explanatory variables that discriminate the datasets. Here, two types of discriminating variables are pertinent: environmental variables that explain the year effect and life traits that explain the differences among weed species. We concentrated on species characteristics by looking at the effect of pre- and postharvest seed production on spatial variability. The model used was as follows:

$$\begin{aligned} \text{Variogram parameter} = & \text{constant} \\ & + \text{year effect} \\ & + \beta \times \text{numbers of seeds dispersed before harvest} \\ & + c' \times \text{numbers of seeds dispersed during harvest} \\ & + \text{error} \end{aligned} \quad (15.7)$$

where  $\beta$  and  $\chi$  are the parameters associated to the covariates “seeds dispersed before harvest” and “seeds dispersed after harvest,” respectively. Seed production data were not collected during the field trials but were adapted from the means of the two values reported by Forcella et al.<sup>22</sup> for seed production by annual species. Other models looked at the effects of plant densities or germination behavior. In all cases, only explanative variables significant at  $P = 0.01$  were kept in the final model.

### 15.10.2 METHOD

Analyses were carried out with **PROC GLM** of SAS. Programs are given in Box 15.8 (Equation 15.6) and Box 15.9 (Equation 15.7).

### BOX 15.8

SAS program (top) and output extract (bottom) for analyses of variance of semivariogram parameters.

```

*---reading data file---;
data table1;
infile 'variogramParameters.prn' firstobs=2;
input year species$ nugget c1 c2 a1_0 a2_90;
*---aggregating 2 ranges into a single range variable---;
data table2; set table1;
range = a1_0;
direction = 0;
keep year species direction range;
data table3; set table1;
range = a2_90;
direction = 90;
keep year species direction range;
data table2; set table2 table3;
*---analysis of variance---;
proc glm data = table1;
class year species;

```

(continued)

## BOX 15.8 (continued)

```

model nugget c1 c2 = year species;
means year/lsd lines;
means species/lsd lines;

proc glm data = table2;
class year species direction;
model range = direction year species;
means year/lsd lines;
means species/lsd lines;
means direction/lsd lines;

run; quit;

```

Dependent Variable: range

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	5445.966816	544.596682	6.33	<.0001
Error	33	2840.190050	86.066365		
Corrected Total	43	8286.156866			
R-Square	Coeff Var	Root MSE	range Mean		
0.657237	57.68332	9.277196	16.08298		

[...]

Source	DF	Type III SS	Mean Square	F Value	Pr > F
direction	1	2119.878510	2119.878510	24.63	<.0001
year	3	231.823975	77.274658	0.90	0.4526
species	6	2873.895250	478.982542	5.57	0.0004

[...]

t Tests (LSD) for range

NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	33
Error Mean Square	86.06637
Critical Value of t	2.03452
Least Significant Difference	12.008
Harmonic Mean of Cell Sizes	4.941176

NOTE: Cell sizes are not equal.

Means with the same letter are not significantly different.

t Grouping	Mean	N	species
A	30.125	8	setvi
B	25.550	2	sinar

(continued)

## BOX 15.8 (continued)

B	A	C	20.188	4	agre
B	D	C	17.574	8	cirar
	D	C	11.411	8	amare
	D		8.099	8	cheal
	D		6.355	6	ascsy

t Tests (LSD) for range

NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	33
Error Mean Square	86.06637
Critical Value of t	2.03452
Least Significant Difference	5.6909

Means with the same letter are not significantly different.

t Grouping	Mean	N	direction
A	23.024	22	0
B	9.142	22	90

## BOX 15.9

SAS program (top) and output extract (bottom) for analyzing semivariogram parameters as a function of species traits.

```

data table1;
infile 'weednest.mod' firstobs=2;
input year variable$ stage$ species$ transfo r2 nugget c1 c2 almax a2min ;
*transformation to obtain relative variations and improve species comparison;
sill = nugget + c1 + c2;
nugget = nugget/sill;
c1 = c1/sill;
c2 = c2/sill;
*definition of seed production per plant;
if species = 'setvi' then seedprod = 43.2;
if species = 'amare' then seedprod = 61.2;
if species = 'cheal' then seedprod = 105.2;
if species = 'sinar' then seedprod = 117.8;
*proportion of seed shed after to harvest;
if species = 'setvi' then rec = 0.27;
if species = 'amare' then rec = 0.26;
if species = 'cheal' then rec = 0.40;
if species = 'sinar' then rec = 0.16;

```

(continued)

## BOX 15.9 (continued)

```

*seed amount shed before and after to harvest;
seedpost = rec*seedprod;
seedpre = (1-rec)*seedprod;
data table2; set table1;
range = almax;
direction = 0;
keep year species direction range seedpost seedpre;
data table3; set table1;
range = a2min;
direction = 90;
keep year species direction range seedpost seedpre;
proc glm data = table1;
class year species;
model nugget c1 c2 = year seedpost seedpre/solution;
data table2; set table2 table3;
proc glm data = table2;
class year species direction;
model range = direction year seedpost seedpre/solution;
run; quit;

```

Dependent Variable: range

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	2608.511848	434.751975	3.77	0.0121
Error	19	2189.420768	115.232672		
Corrected Total	25	4797.932616			
R-Square	Coeff Var	Root MSE	range Mean		
0.543674	62.26256	10.73465	17.24094		

[...]

Source	DF	Type III SS	Mean Square	F Value	Pr > F
direction	1	1037.526211	1037.526211	9.00	0.0074
Year	3	360.956840	120.318947	1.04	0.3958
seedpost	1	771.477725	771.477725	6.69	0.0181
seedpre	1	0.091953	0.091953	0.00	0.9778

Parameter	Estimate	Standard Error	t Value	Pr >  t
Intercept	20.17335864	B	7.40120899	0.0134
direction 0	12.63405782	B	4.21047543	0.0074
direction 90	0.00000000	B	.	.
Year 93	8.91374000	B	6.19765203	0.1666
Year 94	2.71061308	B	6.16775976	0.6653
Year 96	-1.07820041	B	6.19765203	0.8637
Year 97	0.00000000	B	.	.
seedpost	-0.52907114		-2.59	0.0181
seedpre	0.00423835		0.03	0.9778

### 15.10.3 RESULTS

The output extract in Box 15.8 indicates that the range was not significantly influenced by the survey year ( $\text{Pr} > F = 0.4525$ ). The most important effect (i.e., highest  $F$  value) was due to the direction, with significantly higher ranges in the  $0^\circ$  direction for all species. Ranges varied significantly among species, from 6m for ASCSY to 30m for SETVI.

This species effect was then further explained by covariables describing seed dispersal and germination behavior. The results of linear model Equation 15.7 indicate that an increase in seed dispersal before harvest resulted in a decrease in spatial variability in the direction of the crop rows ( $c_1$ ) but increased unexplained variability (nugget) (Table 15.4). Indeed, the sampling grid used in this work was not fine enough to model the effect on patch shape of preharvest seed dispersal, i.e., seeds falling close to their source plants. The simultaneous decrease in the magnitude of spatial variability between directions ( $c_2$ ) suggests that preharvest seed dispersal is spatially uniform in all directions. This is not surprising in fields planted with small-stature crops, like *Glycine max*, where weed seed heads are usually positioned above the crop canopy. Even if seeds were moved slightly further in one direction by wind gusts or animals, these movements probably would have resulted in dispersal more random than that of combines and would have contributed little to changes in directional spatial structure.

An increase in weed seed dispersal during harvest resulted in an increase in spatial variability in the direction of crop rows but had no effect on differences in spatial variability along the rows vs. across rows. Moreover, the range parameter tended to decrease as more seed is dispersed during harvest. This decrease is not easily explained but could be the result of an overlap with existing weed patches and greater dispersal distance as influenced by the combine harvester. Weed densities in the patch intersection would not be correlated to densities in the centers of existing patches.

## 15.11 KRIGING

### 15.11.1 OBJECTIVE

The next step in the analysis process is kriging to estimate seedling densities (or other variables) on unsampled locations and plot maps. Kriging is an interpolation technique that estimates the value of an attribute,  $z$ , at unsampled locations in the field based on available data at neighboring locations as well as semivariogram model parameters. Basically, there are two types of kriging, ordinary and simple kriging.

Simple kriging assumes a stable mean for the whole field that must be incorporated into the kriging analysis. This method is not convenient in the present case as the analyzed weed variables were often skewed, with mean–variance dependence, even after variable transformation.

Ordinary kriging, on the other hand, uses a variable mean. This kriging method is more adequate for analyzing weed densities due to the dataset values skewness and mean–variance correlation.

**TABLE 15.4**  
Effect of Direction, Year, and Seed Rain Timing on the Parameters of the Semivariogram Models

Variogram Parameter	Direction Effect		Year Effect		Seed Dispersal			
	$P(H=0)$	Regression Parameter	$P(H=0)$	Regression Parameter	Before Harvest		During Harvest	
Nugget	Not tested		0.4840		$P(H=0)$	Regression Parameter	$P(H=0)$	Regression Parameter
$c_1$	Not tested		0.0833		0.0015	0.0120	0.0019	-0.0192
$c_2$	Not tested		0.5293		0.0056	-0.00707	0.0001	0.0218
Range	0.0074	$0^\circ$	0.3958		0.0766	-0.00495	0.5464	-0.529*
		$90^\circ$			0.9778		0.0181	

\* Regression parameters were only shown for significant variables.

Notes: Results of linear model (Equation 15.7) with PROG GLM (see Box 15.9). Regression parameters were only shown for significant variables.

When estimating variable  $z$  at an unsampled location, kriging software programs use the measured  $z$  values of the neighboring locations located within a specified search radius. This radius should be smaller than the semivariogram range. Indeed, locations farther than the range are only randomly correlated, and this information, therefore, will not increase the quality of the estimation on the unsampled point.

### 15.11.2 METHOD

Various software programs exist for kriging and drawing maps, though with varying degrees of complexity. In the original work,<sup>16</sup> we used the **KB2D** and **KB3D** functions of **GSLIB** for bi-dimensional and three-dimensional kriging, respectively. These functions accept sums of any number of variogram models, including combinations of different models (e.g., a spherical + power model). The data resulting from kriging can then be processed with yet another function and/or software to draw maps, for instance, **PIXELPLT** of **GSLIB**, which produces a postscript file. Here, we used **PROC KRIGE2D** of **SAS** for kriging followed by **PROC GMAP** for drawing the map (Box 15.10).

### 15.11.3 RESULTS

Maps estimated with ordinary kriging for *S. viridis* plant densities in 1993 with the **SAS** program (Box 15.10) are presented in Figure 15.7. When compared to the original map showing the sampled densities (Figure 15.2), it clearly appears that the weed patches are located in the same parts of the field in both maps, but the kriged map has "smoothed" the raw map into a continuous and more detailed representation.

## 15.12 CROSS-SEMIVARIOGRAMS AND COKRIGING

### 15.12.1 OBJECTIVE

Cross-semivariograms are based on a similar principle as semivariograms, but instead of looking at correlations between locations for a given variable, they describe the variance between two variables as a function of the distance between the locations where the variables were measured. Quite often, the variables to be correlated are a biological (e.g., crop yield) and an environmental variable (e.g., soil nitrogen content). Cross-semivariogram models are then fitted and used for cokriging. While kriging only uses measurements of one variable to estimate this same variable, cokriging is usually used to estimate a sparsely sampled primary variable (e.g., soil nitrogen) with the help of an extensively sampled secondary variable (e.g., weed density). The idea is to estimate a variable that is expensive to sample with the help of an easy-to-measure variable. Cokriging requires the semivariograms for the primary and secondary variable, their cross-semivariogram, and their respective means.

### BOX 15.10

**SAS** program for kriging and drawing a density map for *S. viridis* in 1993; the output graph is shown in Figure 15.7. Kriging is ordinary, using only observations located in a radius of 20m for estimations (radius = 20); the observed variable is *setvi*. The variogram is a nested model of two spherical models (form option), using the parameter values determined with the program of Box 15.7. The nugget option indicates the nugget irrespective of direction. The scale option lists the contributions, and the range option lists the maximum ranges for the 90° and 0° directions (see angle option). The ratio option lists the ratio of minimum range/maximum range for the two directions, here 1/45 (Instead of 0/45 to avoid nil ranges that are not accepted by **KRIGE2D**) and 20.1/400000.

```
*---defining kriging step as macro-variable---;
%let length = 1 ;
*---reading data file---;
data table1;
infile 'Data1993.prn' firstobs = 6;
input xlocation ylocation agrre amare ascsy cheal
      cirar setvi sinar;
*---transforming weed densities---;
logsetvi=log(setvi+1);
*---kriging with nested anisotropic spherical model---;
proc krige2D data=table1 outest=table2;
coordinates xcoord=xlocation ycoord=ylocation;
grid X=0 to 54 by 1 Y=0 to 244 by &length; *grid of coordi-
nates for estimating densities;
predict radius=20 var=logsetvi ;
model form=(spherical spherical) nugget = 0.92
range=(40.9 400000)
ratio=(0.0244 0.0000732) scale=(1.60 2.7) angle=(0 0);
*---retransform weed densities---;
data table2; set table2;
estimate = exp(estimate)-1; *estimate is variable cre-
ated by proc krige2D;
plot = _N_ ; *creating a plot name from line number
_N_ ;
*---create coordinate table for proc gmap---;
*for each (x,y) location, four coordinates (x+length/2,
y+length/2),(x+length/2,y-length/2), (x-length/2,
y-length/2) and (x-length/2,y+length/2) are create where
length = kriging step.
```

(continued)

## BOX 15.10 (continued)

```

gxc and gxy are variables created by proc krige2D;
data table3; set table2;
x = gxc+&length/2;
y = gyc+&length/2;
keep x y plot;
data table4; set table2;
x = gxc+&length/2;
y = gyc-&length/2;
keep x y plot;
data table5; set table2;
x = gxc-&length/2;
y = gyc-&length/2;
keep x y plot;
data table6; set table2;
x = gxc-&length/2;
y = gyc+&length/2;
keep x y plot;
data table3; set table3 table4 table5 table6;
keep x y plot;
proc sort; by plot;
*---draw map---;
options reset=all cback=white colors=(white grayee
graydd graycc graybb grayaa gray77 gray66 gray56
gray44 black ); *color options for background and map;
pattern value=msolid; *each basic spatial unit is to
be filled;
proc gmap data=table2 map= table3; *table2 contains
the weed densities, table3 the coordinates;
id plot; *plot is the identification of
the basic spatial unit;
choro estimate/coutline=same cempty=white ctext=black
midpoints = 0 to 150 by 15;
*estimate is the variable to represent on the map, the
color outline of each plot is the same as the color
used for filling, empty plots are drawn with white
lines, the legend is written in black and densities
are represented from 0 to 150 by 15;
run; quit;

```



FIGURE 15.7 Maps of *S. viridis* seedling densities for 1993 at the Swan Lake Research Farm based on kriging with SAS program shown in Box 15.10.

In this case study, both methods were used somewhat differently. Cross-semivariograms were used to look for correlations between weed densities of a given species of years  $j$  and  $j + 1$ . The aim was to test whether cokriging could use the sampled data of year  $j$  to predict the weed distribution of year  $j + 1$ . Neither the semivariogram of year  $j + 1$  nor the cross-semivariogram for years  $j$  and  $j + 1$  could be known in advance. Therefore, we used the semivariogram and the mean density of year  $j$  also for year  $j + 1$  as well as the mean of past cross-semivariograms for all pairs of years  $i$  and  $i + 1$  with  $i \leq j$ .

### 15.12.2 CROSS-SEMIVARIOGRAMS

Empirical cross-semivariograms were calculated as follows:

$$\gamma_h = \frac{1}{2 \cdot N_h} \sum (z_{i+h,j} - z_{i,j})(z_{i+h,j+1} - z_{i,j+1}) \quad (15.8)$$

where  $z_{i,j}$  was the weed density of a given species at location  $i$  for year  $j$ . In the original work,<sup>16</sup> the empirical cross-semivariograms were calculated with the **GAMV** function of **GSLIB** (Box 15.5). The same type of variogram models as for semi-variograms can then be fitted to the empirical cross-semivariogram, using the same methods (see Section 15.9).

### 15.12.3 COKRIGING

The semivariogram and mean of 1996 as well as the cross-semivariogram of 1993 and 1994 (Table 15.5) were then used to predict the weed map of the 1997 from the observations of 1996. Cokriging was carried out with the three-dimensional **COKB3D** function of **GSLIB** (DOS version). These calculations are beyond the scope of this chapter although cokriging parameters are given in Box 15.11. The resulting maps (shown in Figure 15.8) were drawn with **PIXELPLT** (Box 15.12) after back transforming the output file produced by **COKB3D** with  $\exp(z) - 1$ .

Cokriging with the 1996 observations and mean for predicting 1997 weed densities predicted patch location correctly though weed densities were grossly over-estimated. Using the actual density mean of 1997, rather than the 1996 mean, considerably increased the prediction quality (Figure 15.8 vs. Figure 15.2). This is a feasible option, as estimating a mean density is considerably faster and cheaper than estimating patch locations.

**TABLE 15.5**  
Variogram Models Used for Predicting the 1997 Map of *S. viridis* from 1996 Observations Using Cokriging

Variogram Type	Year	Variogram Parameters						
		Nugget	Contribution		Range (m)			
			$c_1$	$c_2$	a1_90	a1_0	a2_90	a2_0
Semivariogram	1996	1.38	1.20	2.82	0	49.47	34.12	400,000
Cross-semivariogram	mean of 1993/1994 and 1996/1997	0.169	0.587	1.86	0.995	43.0	25.4	$5 \times 10^{18}$

### BOX 15.11

**COKB3D.PAR** file parameterized for cokriging 1997 *S. viridis* densities from 1996 observations (column 3 of **setvi1996.dat**) while disregarding 1997 observations (a dummy variable located in column 4 of **setvi1996.dat** was called instead and then eliminated), using the mean densities actually observed in 1996 (i.e.,  $4.7 = \log_e(109 + 1)$ ) and 1997 (i.e.,  $2.55 = \log_e(11.32 + 1)$ ), respectively, the 1996 semivariogram for both the main (semivariogram 1) and the secondary variable (2) and a mean cross-semivariogram (1 2). See Table 15.5 for variogram models.

Parameters for COKB3D

\*\*\*\*\*

START OF PARAMETERS:

```
setvi1996.dat
2          \file with data
1 2 0 4 3  \ number of variables primary+other
-98 1.0e21 \columns for X,Y,Z and variables dummy variable and setvi1996
0          \trimming limits: lower limit fixed to eliminate all values for
          \ dummy variable
          \co-located cokriging? (0=no, 1=yes)
3          \ file with gridded covariate
1          \ column for covariate = setvi 1996 densities
          \debugging level: 0,1,2,3
          \file for debugging output
          \file for output
          \number of x units,xmin,xsize
          \number of y units,ymin,ysize
          \nz,zmn,zsiz
          \x, y, and z block discretization
```

(continued)

## BOX 15.11 (continued)

```

1 12 8
20.0 20.0 20.0
20.0 20.0 20.0
0.0 0.0 0.0
1
2.55 4.7 0.0 0.0
1 1
2 1.38
1 1.20 0.0 0.0 0.0
49.47 0.00 1.0
1 2.82 0.0 0.0 0.0
400000 34.12 1.0
2 2
2 1.38
1 1.20 0.0 0.0 0.0
49.47 0.00 1.0
1 2.82 0.0 0.0 0.0
400000 34.12 1.0
1 2
2 1.6999E-01
1 5.8703E-01 0.0 0.0 0.0
4.3086E+01 9.954E-01 1
1 1.863E+00 0.0 0.0 0.0
5.49E+18 2.5417E+01 1
\min primary,max primary,max all sec
\maximum search radii: primary
\maximum search radii: all secondary
\angles for search ellipsoid
\kriging type (0=SK, 1=OK, 2=OK-trad)
\means for setvi1996 1997
\semivariogram for "i" and "j" (here 1996)
\ nb of structures, nugget effect
\ it,cc,ang1,ang2,ang3
\ a_hmax, a_hmin, a_vert
\ it,cc,ang1,ang2,ang3
\ a_hmax, a_hmin, a_vert
\ semivariogram for "i" and "j" (here 1996)
\ nb of structures, nugget effect
\ it,cc,ang1,ang2,ang3
\ a_hmax, a_hmin, a_vert
\ it,cc,ang1,ang2,ang3
\ a_hmax, a_hmin, a_vert
\ semivariogram for "i" and "j" (here mean cross-variogram)
\ nst, nugget effect
\ it,cc,ang1,ang2,ang3
\ a_hmax, a_hmin, a_vert
\ it,cc,ang1,ang2,ang3
\ a_hmax, a_hmin, a_vert

```

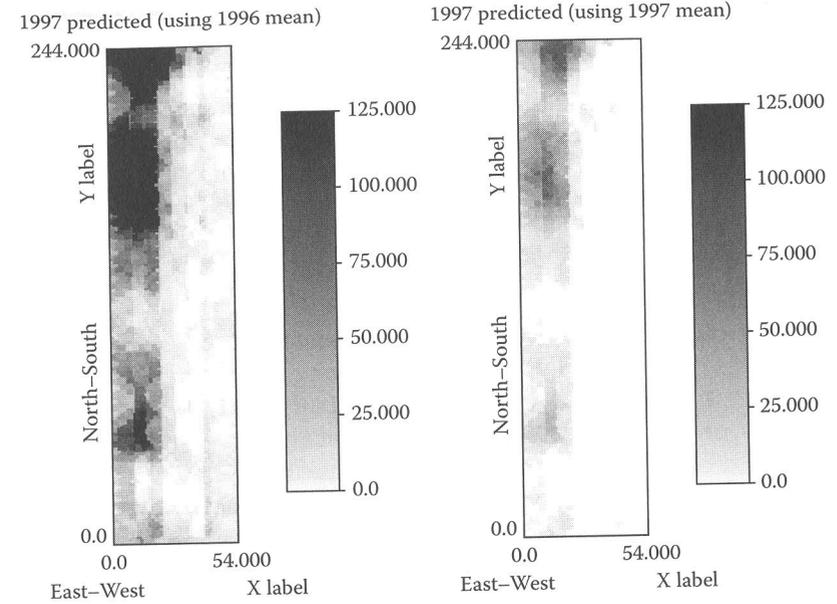


FIGURE 15.8 Maps of *S. viridis* seedling densities for 1997 based on cokriging from the weed densities sampled in 1996, using the 1996 mean density (a) or the 1997 mean density (b).

## 15.13 ERROR ANALYSIS

## 15.13.1 PREDICTION OF WEED MEANS

The mean *S. viridis* density observed in 1997 (i.e., 11.92) can be compared to the mean density calculated from the cokriged maps to determine the prediction error of the mean weed density. When cokriging with the 1996 mean, the predicted 1997 mean was 48.38, which is considerably higher than the actual mean density counted in 1997 (error = +306% = [(48.38 - 11.92)/11.92]\*100). This was due to the higher mean density in 1996 (i.e., 109). If the actual 1997 mean was used for cokriging, then the predicted 1997 was reduced to 5.11, with an error of -57% (=[(5.11 - 11.92)/11.92]\*100).

## 15.13.2 PREDICTION OF WEED LOCATIONS

Box 15.13 shows an SAS program for comparing kriged or cokriged weed densities to observed densities to calculate mean residual error and mean prediction error for the weed densities. Mean residual error is as follows:

$$\text{MRE} = \frac{\sum (z_i - \hat{z}_i)}{n} \quad (15.9)$$

where

$z_i$  are observed values (with mean  $\bar{z}_i$ )

$\hat{z}_i$  is the predicted values

$n$  is the number of values

## BOX 15.12

**PIXELPLT.PAR** file parameterized for drawing map from output of **COKB3D.EXE** with **PIXELPLT.EXE**. The content of set9697 was backtransformed with  $\exp(z) - 1$  before drawing the graph. Output is shown in Figure 15.8.

Parameters for PIXELPLT  
\*\*\*\*\*

## START OF PARAMETERS:

```
set9697.out      \file with gridded data
1               \ column number for variable
-1.0e21  1.0e21 \ data trimming limits
set9697.ps      \file with PostScript output
1             \realization number
54   0.5  1    \number of x units,xmin,xsize
244  0.5  1    \number of y units,ymin,ysize
1   0.0  1.0   \nz,zmn,zsiz
1           \slice orientation: 1=XY, 2=XZ, 3=YZ
1           \slice number
1997 predicted (Using real 1997 mean) \Title
East-West     \X label
North-South   \Y label
0             \0=arithmetic, 1=log scaling
0             \0=gray scale, 1=color scale
0             \0=continuous, 1=categorical
0.0  125.0  25.0 \continuous: min, max, increm.
7           \categorical: number of categories
1  9  Code_One   \category(), code(), name()
2  3  Code_Two
3  2  Code_Three
4  1  Code_Four
5  8  Code_Five
6  6  Code_Six
7  7  Code_Seven
```

Color Codes for Categorical Variable Plotting:  
1=red, 2=orange, 3=yellow, 4=light green, 5=green,  
6=light blue, 7=dark blue, 8=violet, 9=white,  
10=black, 11=purple, 12=brown, 13=pink, 14=inter-  
mediate green, 15=gray

## BOX 15.13

SAS program (top) and output (bottom) for calculating prediction errors and decision errors when using cokriged weed maps for estimating weed patches and making spraying decisions. **Data1997.prn** contains the observed weed densities; **set9697realMean.out** is an output file produced by **COKB3D** of **GSLIB**

```
***-reading data file with observations---;
```

```
data observation;
infile 'Data1997.prn' firstobs = 6;
input xlocation ylocation agrre amare ascsy cheal cirar setvi sinar;
x = floor(xlocation);
y = floor(ylocation);
obs = setvi;
proc sort; by y x;
***-reading cokriging output---;
data prediction;
infile 'set9697realMean.out' firstobs = 5;
input sim var;
*sim = exp(sim)-1; *only necessary if the content of the out file has not yet been back-
transformed manually;
***-creating coordinates for cokriging output---;
data coord;
do x = 0 to 54 by 1; output; end;
data coord; set coord;
```

(continued)

## BOX 15.13 (continued)

```

do y = 0 to 244 by 1; output; end;
proc sort; by y x;
*---merge all data tables---;
data prediction; merge coord prediction;
data table1; merge observation prediction; by y x;
if obsne ''; *to eliminate lines with missing values;
if simne '';
*---mean residual error---;
error = sim - obs;
e2 = error **2;
proc univariate;
var error ;
*---mean prediction error---;
proc means noprint;
var e2 ;
output out=new SUM = se2;
data new; set new;
racmsepe = (se2/_FREQ_)**0.5;
proc print; var racmsepe;
*---cases with inadequate spraying decision---;
data table1; set table1;

```

```

threshold = 27;*maximum acceptable weed density;
tminus= 0;*variable for summing cases without spraying though the threshold was exceeded;
tplus= 0;*variable for summing cases with spraying though the threshold was not exceeded;
if (obs > threshold and sim > threshold) or (obs <= threshold and sim <= threshold)
then do; *these are the cases with adequate spraying; end;
else do;
if obs > threshold and sim <= threshold
then tminus = 1;
else tplus = 1;
end;
Proc means; var tminus tplus;
run; quit;

```

## The UNIVARIATE Procedure

Variable: error

## Moments

N	400	Sum Weights	400
Mean	-5.6266984	Sum Observations	-2250.6794
Std Deviation	25.6789766	Variance	659.409839
Skewness	-2.6487863	Kurtosis	13.9112055
Uncorrected SS	275768.42	Corrected SS	263104.526
Coeff Variation	-456.37734	Std Error Mean	1.28394883

(continued)

## BOX 15.13 (continued)

Basic Statistical Measures

Location	Variability
Mean -5.62670	Std Deviation 25.67898
Median -0.33846	Variance 659.40984
Mode -0.81350	Range 265.57028
	Interquartile Range 12.76573

NOTE: The mode displayed is the smallest of 2 modes with a count of 2.

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----
Student's t	t -4.38234	Pr >  t  <.0001
Sign	M -32	Pr >=  M  0.0016
Signed Rank	S -7870	Pr >=  S  0.0006

Obs	racMSEP
1	26.2568

## The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
tminus	400	0.1125000	0.3163763	0	1.0000000
tplus	400	0.0500000	0.2182179	0	1.0000000

Mean prediction error is the root-square of the mean-squared error of prediction.<sup>23,24</sup>

$$rMSEP = \sqrt{\frac{\sum (z_i - \hat{z}_i)^2}{n}} \quad (15.10)$$

In the example where cokriging was carried out with the actual 1997 mean (Box 15.13), MRE = -5.62 plants/m<sup>2</sup> (significantly different from zero with Pr > |t| < .0001) and Skewness = -2.64, pointing to a tendency of underestimating densities. Mean error prediction was 25.9 plants/m<sup>2</sup>, which is approximately double the observed mean (i.e., 11.92).

If cokriging was done with the 1996 mean, MRE increased to 40.3 (significantly different from zero with Pr > |t| < .0001) and Skewness to 2.79, showing that weed densities were systematically overestimated. rMSEP was multiplied by nearly four (98.8 plants/m<sup>2</sup>).

## 15.14 SUMMARY: USING GEOSTATISTICAL INFORMATION FOR DECISION MAKING

The program of Box 15.13 also calculates the error frequency if herbicide spraying in precision agriculture was based on cokriged maps. In this example where the 1997 weed map was cokriged from 1996 observations and the 1997 means, 11% of the field was not sprayed though its weed density exceeded the maximum acceptable threshold of 27 plants/m<sup>2</sup>, whereas 4% was sprayed even though its density was sufficiently low; the remaining 85% was managed correctly, i.e., areas were treated when densities exceeded the threshold and left untreated when densities were low enough.

When the 1997 map was cokriged using the 1996 density, the decision error was different: 7% was erroneously left untreated, 30% was unnecessarily sprayed, and only 63% was managed correctly. The comparison between these two methods indicates that data used for making maps can significantly impact final outcomes.

## GLOSSARY

**Anisotropy:** It is present when spatial autocorrelation of a process changes with direction.

**Anisotropy (geometric):** It occurs when the range of the semivariogram changes with direction while the sill remains constant.

**Anisotropy (zonal):** It occurs when the sill of the semivariogram changes with direction while the range remains constant.

**Autocorrelation:**  $\rho_h = \frac{1-\gamma_h}{\sigma_j \cdot \sigma_{j+t}}$ , where  $\gamma_h$  is the autocorrelation and  $\gamma_h$  is the empirical semivariance for distance  $h$ ,  $\sigma_j$  and  $\sigma_{j+t}$  the standard deviations of variables  $z_j$  and  $z_{j+t}$ , respectively.

**Cokriging:** It estimates a sparsely sampled variable,  $z'$ , using the sampled data of this same variable and that of an extensively sampled variable,  $z$ , as well as the semivariograms of the variables  $z$  and  $z'$  and their cross-semivariogram.

**Contribution:** Difference between sill and nugget (if any).

**Cross-semivariogram (empirical):**  $\gamma_h = \frac{1}{2N_h} \sum (z_{i+h} - z_i)(z'_{i+h} - z'_i)$ , where  $\gamma_h$  is the empirical cross-semivariance for the distance  $h$ ,  $N_h$  the number of points separated by the distance  $h$ , and  $z_i$  and  $z'_i$  the data values of two variables measured at location  $i$ .

**Kriging:** Linear interpolation method that allows estimation of variable  $z_i$  at unsampled locations, using a weighted linear combination of available samples and a modeled semivariogram.

**Nugget:** It represents microscale variation that cannot be described with sampling plan used or measurement error.

**Range:** The distance (in any) at which data are no longer autocorrelated.

**Semivariance (empirical):**  $\gamma_h = \frac{1}{2 \cdot N_h} \sum (z_{i+h} - z_i)^2$ , where  $\gamma_h$  is the empirical semivariance for the distance  $h$ ,  $N_h$  the number of points separated by the distance  $h$ , and  $z_i$  a data value measured at location  $i$ . The semivariogram provides a description of how the data are related (correlated) with distance.

**Sill:** Value of semivariance  $\gamma_h$  for distance larger than range.

## REFERENCES

1. Bigwood, D.W. and Inouye, D.W., Spatial pattern analysis of seed banks: An improved method and optimized sampling, *Ecology*, 69, 497, 1988.
2. Marshall, E.J.P., Field-scale estimates of grass weed populations in arable land, *Weed Research*, 28, 191, 1988.
3. Van Groenendael, J.M., Patchy distribution of weeds and some implications for modeling population dynamics: A short literature review, *Weed Research*, 28, 437, 1988.
4. Thornton, P.K., Fawcett, R.H., Dent, J.B., and Perkins, T.J., Spatial weed distribution and economic thresholds for weed control, *Crop Protection*, 9, 337, 1990.
5. Nordmeyer, H. and Niemann, P., Möglichkeiten der gezielten Teilflächenbehandlung mit Herbiziden auf der Grundlage von unkräuterverteilung und Bodenvariabilität, *Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz*, 13, 539, 1992.
6. Wiles, L.J., Oliver, G.W., York, A.C., Gold, H.J., and Wilkerson, G.G., Spatial distribution of broadleaf weeds in North Carolina soybean (*Glycine max*) fields, *Weed Science*, 40, 554, 1992.
7. Halstead, S.J., Gross, K.L., and Renner, K.A., Geostatistical analysis of the weed seed bank, *Proceedings of the North Central Weed Science Society*, 45, 123, 1990.
8. Mortensen, D.A., Johnson, G.A., and Young, L.J., Weed distribution in agricultural fields. In: *Soil Specific Crop Management* (P.C. Robert, R.H. Rust, and W.E. Larson, eds), American Society of Agronomy, Madison, WI, p. 113, 1993.
9. Johnson, G.A., Mortensen, D.A., and Martin, A.R., A simulation of herbicide use based on weed spatial distribution, *Weed Research*, 35, 197, 1995.

10. Streibig, J.C., Gottschau, A., Dennis, B., Haas, H., and Polgaard, P., Soil properties affecting weed distribution, in *7th International Symposium on Weed Biology, Ecology, and Systematics*, Paris, vol. 7, p. 147, 1984.
11. Fried, G., Norton, L.R., and Reboud, X., Environmental and management factors determining weed species composition and diversity in France, *Agriculture, Ecosystems & Environment*, 128, 68, 2008.
12. Colbach, N., Chauvel, B., Gauvrit, C., and Munier-Jolain, N.M., Construction and evaluation of ALOMYSYS, modelling the effects of cropping systems on the blackgrass life-cycle. From seedling to seed production, *Ecological Modelling*, 201, 283, 2007.
13. Chikowo, R., Faloya, V., Petit, S., and Munier-Jolain, N., Integrated weed management systems allow reduced reliance on herbicides and long term weed control, *Agriculture, Ecosystems & Environment*, 132, 237, 2009.
14. SAS Institute Inc., *SAS/STAT User's Guide, Version 6*, SAS Institute Inc., Cary, NC, 1989.
15. Deutsch, C.V. and Journel, A.G., *GSLIB: Geostatistical Software Library and User's Guide*, Oxford University Press, New York, 369pp., 1998.
16. Colbach, N., Forcella, F., and Johnson, G.A., Temporal trends in spatial variability of weed populations in continuous no-till soybean, *Weed Science*, 48, 366, 2000.
17. Box, G.E.P., Hunter, W.G., and Hunter, J.S., *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, Wiley, New York, 1978.
18. Colbach, N., Duby, C., Cavalier, A., and Meynard, J.M., Influence of cropping systems on foot and root diseases of winter wheat: Fitting of a statistical model, *European Journal of Agronomy*, 6, 61, 1997.
19. Journel, A.G. and Huijbregts, C., *Mining Geostatistics*, Academic Press, New York, 1978.
20. Hamlett, J.M., Horton, R., and Cressie, N.A.C., Resistant and exploratory techniques for use in semivariogram analyses, *Journal of Soil Science Society of America*, 50, 868, 1986.
21. Cressie, N., *Statistics for Spatial Data*, John Wiley & Sons, New York, 1991.
22. Forcella, F., Peterson, D.H., and Barbour, J.C., Timing and measurement of weed seed shed in corn (*Zea mays*), *Weed Technology*, 10, 535, 1996.
23. Wallach, D. and Goffinet, B., Mean squared error of prediction in models for studying ecological and agronomic systems, *Biometrics*, 43, 561, 1987.
24. Wallach, D. and Goffinet, B., Mean squared error of prediction as a criterion for evaluating and comparing system models, *Ecological Modelling*, 44, 299, 1989.