Multiscale Statistical Analysis of Weed Populations

T. Faechner and L. M. Hall, Alberta Agriculture and Rural Development, Edmonton, AB, Canada; and C. Deutsch, Univ. of Alberta, Edmonton, AB, Canada

Geostatistical Goals

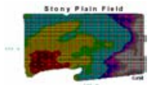
1. To increase our understanding of spatial and temporal variations of weed densities.
2. To develop predictive numerical models of weed distributions for risk qualified treatment.
3. To gain a better “biological” understanding through (geo)statistical analysis of multi-field and multi-year data.

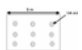
Project Objectives

1. To analyze weed population data for multiple species at a field near Edmonton, Alberta, Canada.
2. To investigate the spatial variability and dependence of weed species distribution on secondary data such as DEM and elevation.
3. To develop the methodology to assess uncertainty for decision making in the presence of sparse data for precision application of herbicides.

Stony Plain Field

The Stony Plain field is a 26 ha gently rolling field west of Edmonton, Alberta, Canada. This field was grid sampled using nine, ¼ m^2 grids at intervals of 35 meters, providing approximately one sample per 0.12 ha. The field was 585 m by 475 m. Sampling occurred 2 days prior to post-emergent herbicide application. The barley (*Hordeum vulgar L.*) was at the 3-4 leaf stage while weed stage varied from newly emerged seedling to 5-6 leaf.





DEM and Elevation Classification

Digital Elevation Map (DEM) units (MacMillan et. al. 1998) are used to describe and segregate landforms. The 15 DEM units are related to movement and accumulation of water in the landscape, and are significantly different in soil characteristics and crop yields. Landscape classes may be divided into slope classes with DEM units 1 to 3 defining upper slope areas, units 4-9 mid-slope areas and units 10-15 lower slope to depression areas. A DEM was prepared by Landwise Consulting for the above fields from aerial photos using methods described in MacMillan et. al. 1998.

Elevation was derived from aerial photos used to prepare the DEM.

Sampling

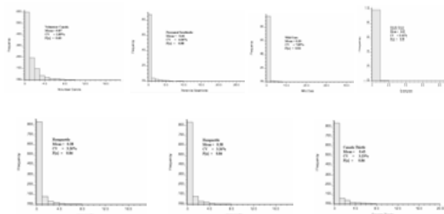
A nested grid sampling at 207 points spaced 35 m apart in a 26 ha field near Edmonton, Alberta was sampled for weed counts. At each point, 9 quadrants of ¼ m^2 were sampled for all weed species. If a weed species count exceeded 50 per ¼ m^2, and then a default value of 50 was recorded. The 1863 quadrants were geographically located for their spatial location. For this project, the following weed species were reported: wild oats (*Avena fatua* L.)(AVEFA), Canada thistle (*Cirsium arvense* (L.)Scop.)(CIRAR), hempnettle (*Galeopsis tetrahit* (L.)(GAETE), quackgrass (*Elytrigia repens* (L.)Nevski)(AGRRE), perennial sowthistle (*Sonchus arvensis* Beib.)(SONAR), volunteer canola (*Brassica* spp.)(BRSNN) and chickweed (*Stellaria media* (L.)(STEME).

Correlation Coefficients between Weed Species based on Indicator Probabilities

	CIRAR	STEME	GAETE	SONAR	AGRRE	BRSNN	AVEFA
CIRAR	0	0.326	0.134	0.076	0.078	0.190	0.103
STEME		0	0.323	0.341	0.399	0.224	0.377
GAETE			0	0.117	0.083	0.206	0.078
SONAR				0	0.067	0.185	0.076
AGRRE					0	0.202	0.035
BRSNN						0	0.206
AVEFA							0

Descriptive Statistics

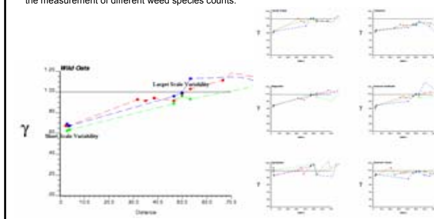
There is a high probability of no weeds of any species. The correlation coefficients were very low among the weed species implying a low association between weed species. Outliers and data sparsity have a significant effect on the correlation coefficients. With these limitations, we considered an indicator approach based on a presence/absence of weeds for this data set. Indicator means the data is coded as 1 or 0 depending a weed’s presence or absence, respectively at a location. The indicator correlation coefficients were much stronger compared to the data correlation coefficients.



Spatial Structure

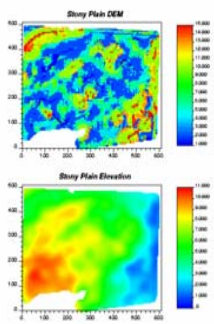
Variograms show a lack of spatial correlation, much less correlation than the shortest sampling distance of 5 m. The nugget effect was greater than 55% in the selected weeds analyzed. The sill was reached was reached in most variograms at 85 to 90 m. Omnidirectional, 45 and 135-degree variograms were modelled for directional trends and zonal anisotropy was observed in this data set.

There was a large nugget effect in the cross variograms between the different weed species which would suggest that the nugget effect in the traditional variograms is due mainly to micro-scale variation common to both weed species. This is under the assumption that errors associated with the measurement of different weed species counts.



Secondary Data

Elevation and DEM for this site appear smooth. Both these secondary data may be associated with weed distribution due to the accumulation and movement of water on a landscape. These variables could be used to help predict the distribution of weed species. The resolution of this data is each pixel represents a 5 by 5 m^2 area.

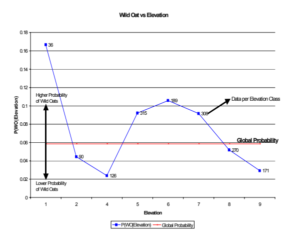


Conditional Probability

Conditional probabilities of wild oats given elevation were calculated to provide a method of calibrating the weed data to the secondary data. The six other weeds were also calibrated for elevation plus DEM. Elevation classes represent relative height in meters.

If the probability for a particular elevation is the global mean, then there is no new information from calibration; however, if there is a deviation from the global mean for a particular elevation class, then the elevation has additional information.

Wild oats appears to prefer the mid elevation range at this site (classes 5, 6 and 7) while avoiding the lower and upper elevation classes. This may mean that it enjoys a moderate microclimate, which has sufficient nutrients and water while it evades areas that are cooler and wetter as well as hotter and drier areas.



Merging Probabilities

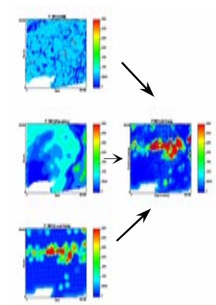
Three conditional probabilities for wild oats are available:

1. elevation
2. DEM
3. local data

These probabilities must be merged into a single posterior probability.

P{AVENFA|AllData}={PAVENFA|DEM)/P(AVENFA) *{P(AVENFA|Elevation)/P(AVENFA)}*{P(AVENFA| Local Data}

Using all this data increases the predictive power of the model.

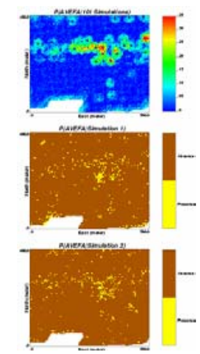


Simulations

Stochastic simulation of weed probabilities is a Monte-Carlo technique that allows the generation of realizations conditional to DEM, elevation and local data. A Gaussian random function model was used to describe the spatial variability of weed distribution with a zero mean and unit variance. One hundred and one realizations of this random function model were generated using truncated Gaussian simulation.

These realizations can be used to assess uncertainty in the likelihood of a weed's frequency distribution in the field. Additionally, this information can be used in risk qualified decision making for precision application of herbicides.

Realizations of wild oats at this site depict it as clustered in its distribution giving the appearance of “band” along the north half of the field. This clustered distribution makes it ideal for patch spraying.



Assumptions

Weed distribution is a parameter that changes from field to field depending on abiotic and biotic factors. For example, moisture and temperature may significantly influence weed density in other fields. Consequently, the appropriate variables must be investigated with each new growing region. The probability calibration to describe weed distribution would also need to be recalibrated for each new region.

Probabilistic models themselves have uncertainty. For example, the probabilities associated with elevation and DEM are uncertain; however, we would require yet another model to assess uncertainty in these model parameters. Although it would be difficult to quantify uncertainty in the model of uncertainty, future work consists of a sensitivity study on these important parameters.

There is significant short-scale variability in weed density due to environmental variability. This can be inferred from the nugget effect of 0.60 in the wild oat semivariogram. Thus, there are some features occurring at a scale smaller than the 5 m sample size. This may require a more intensive, consequently more expensive, sampling scheme to capture this shorter scale variability.

Conclusions

1. Weed density data from seven species was analyzed at a field near Edmonton, Alberta, Canada using (geo)statistical techniques. The variation within each species was low, as was their distribution. An indicator technique of quantifying the weed species spatial variability was utilized in the presence of this sparse data.
2. Spatial variability was evident for all weed species in this data set; however, the nugget effect was high due to variability at a scale of measurement smaller than 5 m. The variograms indicated anisotropy or directional trends for Canada thistle, wild oats, chickweed, hempnettle, quackgrass, volunteer canola, or perennial sowthistle.
3. Secondary data such as DEM and elevation was used to measure the dependence of weed species on these variables. Elevation and DEM are more continuous in their distribution and convenient if they could be used to predict weed distribution in fields. These conditional probabilities were used to calibrate the presence of weed species in the field.
4. Wild oats had a clustered distribution in its spatial variability. This clustered distribution may make it ideal to do patch spraying for this weed at this site. Wild oats can spread up to 1.5 m by natural means, which may partially explain its clustered distribution. The other weeds counted at this site had nugget effects ranging from 0.55 for Canada thistle to 0.87 for quackgrass.
5. Truncated Gaussian simulation is used to generate realizations that could be used to assess uncertainty for decision making in the presence of sparse data for precision application of herbicides. For wild oats, the realizations illustrate a clustered distribution pattern at this site implying that patch spraying compared to blanket spraying of a herbicide would be a more feasible means of weed control.

Future

Sample spacing which optimizes time and labor costs compared to gains in accuracy needs further study. This project had a 5 m grid sample spacing; however, this may be too expensive to implement on a field scale. Local calibration of ideal sample spacing could be validated in several fields. A small data set could be evaluated with various data spacing by leaving some data out. Then the cost of sampling could be calculated for each grid and compared to the cost of misclassification.

Temporal variation of weed species needs to be investigated using the sampling protocol established for this site. Environmental factors influence weed distribution, however the extent of weed stability from year to year requires analysis.

Spatial variability of weed species over large geographic distances or between different fields will be examined. This will allow numerical and (geo)statistical quantification of variability at different scales. The information can be used for modeling locally varying mean values together with locally varying means of variability.

UNIVERSITY OF ALBERTA

Alberta AGRICULTURE, FOOD AND RURAL DEVELOPMENT

CS-Jan. 2000