

On the adequacy of GIS-generated weed maps for Precision Farming

Matthias Backes¹, Lutz Plümer¹

¹Institute for Cartography and Geoinformation, University of Bonn, Meckenheimer Allee
172, D-53115 Bonn, Germany
backes@ikg.uni-bonn.de
pluemmer@ikg.uni-bonn.de

Abstract. Discrete sampling data is used in several environmental studies to create maps in order to support decision-making processes. The decision maps represent an increasing importance in modern Precision Farming. For the application of herbicides in a field, maps of weed distribution are necessary. The uncertainty of those, resulting from sparse sampling patterns is one major reason why farmers tend to be hesitant in applying the GIS-generated weed maps. In this paper, the lack of predictability and the problems of weed maps will be exemplified. For that purpose, approximately 2800 pictures of the surface of the ground that were taken on a maize field could be manually evaluated and compared to a well-established sampling grid for weeds using cross-correlation. The results prove low values of correlation between the produced weed maps and emphasise the uncertainty of these maps by a simulation experiment which will be described later on in this paper.

1 Introduction

In Precision Farming maps for decision support are produced from discrete sampling data. These maps are needed because of the increasing awareness that farmland is a particularly heterogeneous area. Meanwhile, in Precision Farming, fields are not treated as one unit but as an aggregation of site-specific subunits. This method can be helpful in both saving money and in the application of pesticides. Site-specific maps for the yield are produced in the same way as maps for fertiliser application. Measuring systems for these factors are well established and the use of the above described maps is undoubted in agricultural research. However, no effective site-specific use of herbicides in Precision Farming is presently available. Even though weeds seem to appear in patches, there is an enormous variance of dispersion within a field. There are serious doubts whether or not this variance can be displayed in a single application map, especially because the exact variance of the basic population has never been exactly calculated. This paper will focus on the uncertainties of GIS-generated weed maps. In recent studies several sampling strategies were used to count weeds in fields. Grids from 1 x 1 metres up to 36 x 40 metres can be found [4]. Several attempts to gather all weeds in subunits of fields have been carried out in the

past [1], [18]. In this study a regular sampling grid of 7.5 x 15 m (a) was compared with data collected from a camera based system [6], capturing images of the surface of the ground in a 2 x 3 m grid (b). On approximately 2800 single images weed abundance was counted manually. Visualisation of the risks in sampling and mapping weeds by geographic information systems was the objective of this paper.

2 Materials and Methods

The datasets were analysed by investigating the previously gathered data from a maize field near Bonn, Germany in summer 2001 [6], [7]. Data from grid (a) was captured manually in a quadrat of approximately 0.25 m² size. It was compared statistically with camera-based data (b), with a resolution of about 0.20 m².

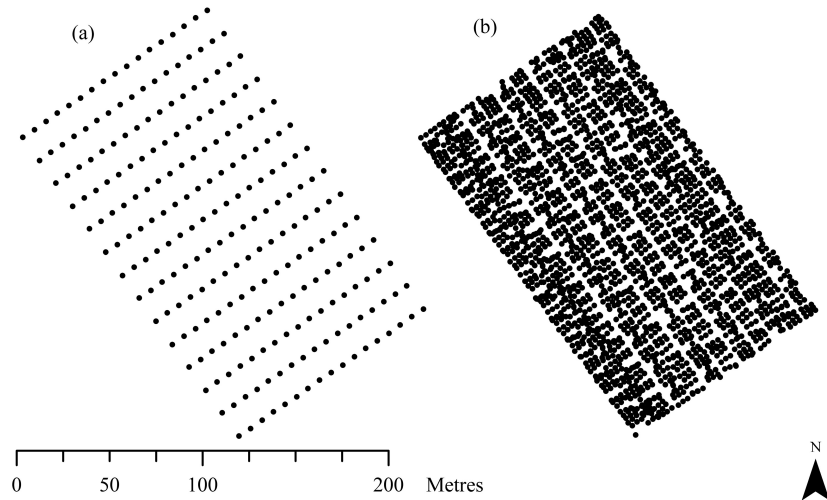


Fig. 1. Sampling grids (a) [7.5x15 metres] and (b) [2x3 metres] of the analysed field.

In order to compare the various datasets, they were normalised to a value of 1m² per unit. Data from grid (a) was transformed into a georeferenced point shapefile (a), with weed species and their added quantities as attributes. Image data from grid (b) had to be transformed into a point shapefile (b) of the same reference system (transverse mercator). Information about weeds counted on the images was attached to the shapefile as an attribute, distinguishing between up to 10 weed species. Both shapefiles were applied to generate weed distribution maps by 'Natural Neighbors' interpolation in ArcMap 8.2 [3], [15]. 'Natural Neighbors' interpolation was implemented in order to optimise the visualisation of weed distribution, because linear triangulation turned out to set the standard in current weed research studies [8]. Other approaches in weed mapping including kriging [1] and co-kriging [10] have not been focused in this study because the transferability of the above mentioned approaches on weed mapping is currently uncertain [5], [8], [14]. Shapefile (a) had a

total of 240 sampling points, which were counted manually in previous studies [6], [7]. Shapefile (b) had about 2800 which were also counted manually from automatically taken ground using standard image software (Fig. 1.). ArcView 3.3 was used to filter the point features in shapefile (b) that were next to shapefile (a), so that a direct comparison of both sampling methods could be realised. An ‘Avenue’ script called ‘Nearest Feature, with Distances and Bearings’ has been applied for that purpose (Fig. 2) [11].

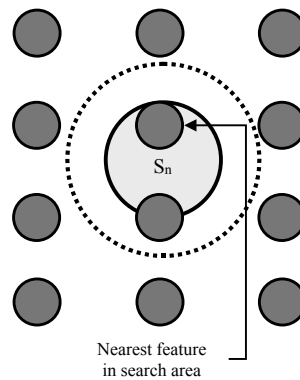


Fig. 2. Operating mode of the ‘Nearest Feature’ script. The nearest sampling point from the dense sampling grid (gray dots) was used to attach the information of weed abundance to the simulated shifted point shapefile (e.g. S_n) in a fixed search area. In this way of calculation, S_0 was produced by searching the nearest features of shapefile (b) to shapefile (a). Therefore, S_0 should represent approximately the same area as shapefile (a).

Subsequently, shapefile (b) was used to generate five further point shapefiles of 7.5 x 15 m size, starting at different positions in the above mentioned field to visualise areas that have not been sampled in grid (a). Every single shapefile was shifted by a distance of 1.5 m in using the move command in ArcMap (Fig. 3).

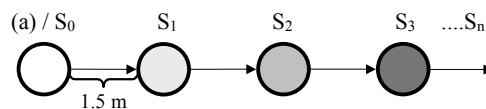


Fig. 3. Shifting realised by ArcMaps move command. The sampling points were shifted by 1.5 metres.

The results of this shifting (S_1 - S_5) were used to interpolate weed maps by ‘Natural Neighbor’ interpolation in ArcMap (Fig. 4, Fig. 5). In applying another ‘Avenue’ script called ‘Grid Plus’ [12] for ArcView, cross-correlation was calculated in order to enable a statistical comparison of the raster dataset. (Table 1, Table 2).

The cross-correlation in this script was calculated according to Equation (1).

$$r = \frac{Cov_{a,b}}{S_a S_b} \quad (1)$$

Remark 1. In the above presented equation, $Cov_{a,b}$ is the covariance of two compared raster datasets (a, b). S is the standard variation of the raster datasets.

3 Results

The analysis of both sampling strategies (a), (b), demonstrated that there are significant differences in the resulting weed maps (Table 1, Table 2). The sparse sampling design overestimates the occurrence of weeds compared with the dense sampling grid (Fig. 4).

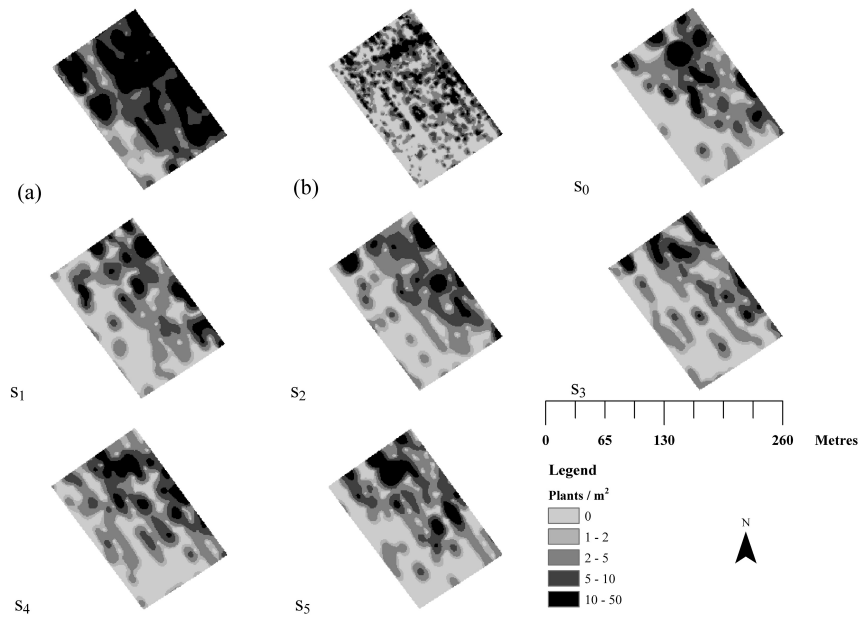


Fig. 4. Interpolated weed maps from sparse sampling (a), dense sampling (b) and five shifted starting points (shifted by 1.5 m) for sampling (S₁-S₅). S₀ is equal to (a) with weed quantity information from (b). The weed mapped is 'Common Lambsquarters' (*Chenopodium album* L.).

The dense sampling grid showed a much more differentiated picture of the weed distribution. Furthermore, a general problem has been visualised in the process of shifting the starting points in order to establish the weed distribution maps. The areas that have not been sampled in the survey produced clear evidence in the simulation

that there is an imminent risk of underestimating or even not realising the weeds or weed patches. Five significantly different maps have been produced from five different starting points of sampling (Fig. 3, Fig. 4, Fig. 5). Regarding the weed species analysed (*Chenopodium album* L., *Stellaria media* L.) in this paper and their biological features, there appears to be a close correlation between the credibility of the weed map produced from grid (a) and the distribution of the weeds in nature. Weeds occurring scattered all over the field appear to be largely overestimated by grid (a), whereas weeds that occur in a clustered way are mapped completely different or not at all (Table 1; Table 2). In this study six different weed species, including the category of all gramineous plants (e.g. *Echinochloa crus-galli* L. and *Avena fatua* L., not shown in a map) as one group were analysed. The exemplified species are representative for the other species which have not been mentioned in this paper.

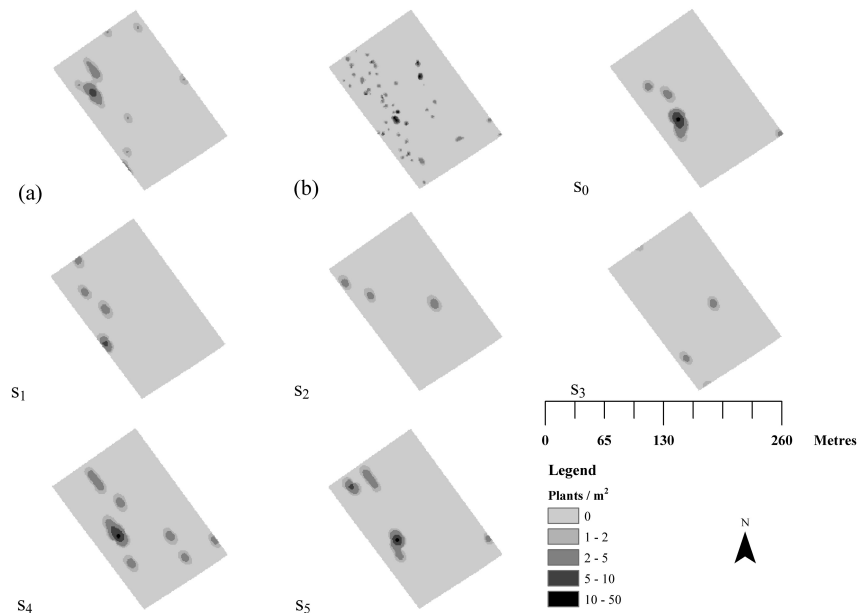


Fig. 5. Interpolated weed maps from sparse sampling (a), dense sampling (b) and five shifted starting points (shifted by 1.5 m) for sampling (S₁-S₅). S₀ is equal to (a) with weed quantity information from (b). The weed mapped is 'Common Chickweed' (*Stellaria media* L.).

The cross-correlation values that have been calculated for the different maps prove a weak relation between the maps which should ideally be close to a value of 1 (Table 1, Table 2) in case the weed maps were invariant to shifting. Concerning the 'clustered' weeds the relation between the datasets appears to be either random or neglectable. The 'scattered' weeds prove to have significantly higher levels of correlation.

As a further outcome of the analysis, the differences between the maps (a) and S₀ should be stressed. S₀ is supposed to be a simulation of (a) because the information of

shapefile S_0 contains the nearest features of (b) close to (a). The results of the study revealed a surprisingly weak correlation between the two datasets.

Table 1. Results of the cross-correlation (r) between different raster datasets. ‘Scattered’ weed e.g. *Chenopodium album* L. (CS = Camera sampling, $S_0 - S_5$ = Shifted datasets)

	CS (b)	7.5x15m (a)	S_0	S_1	S_2	S_3	S_4	S_5
CS (b)	1	0.22	0.55	0.39	0.44	0.50	0.52	0.53
7.5x15m (a)	0.22	1	0.30	0.39	0.28	0.23	0.29	0.22
S_0	0.55	0.30	1	0.48	0.39	0.51	0.46	0.54
S_1	0.39	0.39	0.48	1	0.49	0.42	0.36	0.28
S_2	0.44	0.28	0.39	0.49	1	0.61	0.39	0.34
S_3	0.50	0.23	0.51	0.42	0.61	1	0.52	0.47
S_4	0.52	0.29	0.46	0.36	0.39	0.52	1	0.64
S_5	0.53	0.22	0.54	0.28	0.34	0.47	0.64	1

Table 2. Results of the cross-correlation (r) between different raster datasets. ‘Clustered’ weed e.g. *Stellaria media* L. (CS = Camera sampling, $S_0 - S_5$ = Shifted datasets)

	CS (b)	7.5x15m (a)	S_0	S_1	S_2	S_3	S_4	S_5
CS (b)	1	0.11	0.38	0.13	0.07	0.06	0.37	0.40
7.5x15m (a)	0.11	1	0.15	0.16	0.36	0.01	0.02	0.03
S_0	0.38	0.15	1	0.06	0.02	-0.04	0.54	0.71
S_1	0.13	0.16	0.06	1	0.02	-0.02	0.064	-0.02
S_2	0.07	0.35	0.02	0.02	1	0.41	-0.05	-0.05
S_3	0.06	0.01	-0.04	-0.02	0.41	1	-0.02	-0.03
S_4	0.37	0.02	0.54	0.06	-0.05	-0.02	1	0.56
S_5	0.40	0.03	0.71	-0.02	-0.05	-0.03	0.56	1

4 Discussion

Although weed maps are occasionally used in decision-making processes in Precision Farming, the accuracy of these maps is still questionable. As mentioned above the scope of this study was not to compare different mapping techniques [2], or geostatistical methods [9] (e.g. kriging) for the assessment of weed spatial distribution but to visualise and objectify the evident risks in weed sampling and mapping by geographic information systems. Geostatistical approaches to assess weed spatial distribution and adequate sampling strategies have increased in recent years [1], [2], [13], [17]. Even though it is relatively clear that the sampling strategies applied in practice are pragmatically motivated, further information available to researchers needs to be implemented in order to find out to what extent they can rely on these maps for decision-making. The most important factors for the choice of the sampling strategy are the cost-benefit analysis and the time required for manual sampling. Alongside these facts it is still vital to remember that the resolution used for sampling weeds is also the resolution that is appropriate for application maps. The objective for GIS workers in Precision Farming is certainly not the production of aesthetically pleasing maps.

This case-study showed the consequences of a sparse sampling strategy for weeds. It is imperative to apply adjusted sampling strategies for the different biological distributions of weeds. Regarding the ‘scattered’ (s.a.) weeds, the production of application-maps, according to the information obtained from the ‘sparse’ samples (s.a.) is clearly realistic in case low thresholds for herbicide treatment have been chosen previously. Most parts of the field will be treated by herbicides for selected ‘scattered’ species of weeds (Fig. 4.) if the thresholds for herbicide application are low. The risk of missing weeds by herbicide application based upon these maps seems to be neglectable, even though the cost- and resource saving effect is also relatively low. Weeds occurring in a clustered way are more jeopardising factors in sparse sampling designs (Fig. 5, Table 2). Underestimating these weeds may have harmful consequences for the yield. The most prominent example for this is Catchweed (*Galium aparine* L.), which is already treated with herbicides in accordance to an economic threshold of 0.1 plants per m² [16]. Underestimating this particular weed can cause significant yield losses due to different influences on the crop.

A general problem that tends to occur when sampling weeds manually in a field becomes rather evident when considering the low correlation values between map (a) and the simulation of the same area (S_0) (Table 1, Table 2). In fact, there should be little or no differences between them, because S_0 is the result of attaching the information from shapefile (b) to a shapefile that is analogue to shapefile (a). Nevertheless, there are vast differences between them, one reason of which being the error-prone counting of weeds in a field, often carried out by different investigators. Another plausible explanation of the above described differences could be the divergence in the resolution of the sampling area (0.25 m² compared to 0.20 m²) [2].

However, even though the dense sampling strategy reveals big differences compared to the sparse strategy, it is vital to realise that a final assessment of this strategy cannot yet be carried out. The real variation of weeds is neither represented in the sparse, nor in the dense sampling strategies that have been investigated earlier in this study. The efforts in research required for a complete survey of weeds in a field would be enormous but also essential for a final assessment of all weed-sampling strategies and adequate mapping of weeds for herbicide application in a geographic information system.

Acknowledgements

We would like to thank Dr. Roland Gerhards, Dr. Markus Sökefeld and Dominik Dicke for providing us with the required sampling data.

References

1. Cousens, R.D., Brown, R.W., McBratney, A.B., Whelean, B., Moerkerk, M. Sampling strategy is important for producing weed maps: a case study using kriging, *Weed Science*, 50 : 542-546, 2002.
2. Dille, J. A., Milner, M., Groeteke, J.J., Mortensen, D.A., Willimas II, M.M. How good is your weed map? A comparison of spatial interpolators, *Weed Science* 51 (1) : 44-55, 2002.
3. Booth, B. Using ArcGIS 3D Analyst, *Environmental Systems Research Institute Inc. (ESRI)*, 2000.
4. Garibay, S. V., Richner, W., Stamp, P., Nakamoto, T., Yamagishi, J., Abivardi, C. and Edwards, P.J. Extent and Implications of Weed Spatial Variability in Arable Crop Fields, *Plant Production Science*, 4 (4) : 259-269, 2001.
5. Gerhards, R., Sökefeld, M., Knuf, D., Kühbauch, W. Mapping and geostatistical analysis of weed distribution in sugarbeet fields for site-specific weed management, *J. Agronomy and Crop Science* 176 : 259-266, 1996.
6. Gerhards, R., Sökefeld, M., Nabout, A., Therburg, R.-D., Kühbauch, W. Online weed control using digital image analysis, *Journal of Plant Diseases and Protection*, Special Issue XVIII : 421-427, 2002.
7. Gerhards, R., Sökefeld, M., Timmermann, C., Kühbauch, W. Site-specific weed control in maize, sugar beet, winter wheat and winter barley, *Precision Agriculture* 3 (1) : 25-35, 2002.
8. Gerhards, R., Wyse-Pester, D.Y., Mortensen, D.A. Characterizing spatial stability of weed populations using interpolated maps, *Weed Science*, 45 : 108-119, 1997.
9. Gregg, A. J., Mortensen, D. A., Gotway, C. A. Spatial and Temporal Analysis of Weed Seedling Populations Using Geostatistics, *Weed Science*, 44 : 704-710, 1996.
10. Heisel, T., Ersboll, A. K., Andreassen, C. Weed Mapping with Co-Kriging Using Soil Properties, *Precision Agriculture*, 1 : 39-52, 1999.
11. Jenness, J. Nearest Feature, with Distances and Bearings (v. 3.6c), 2002. *Online resource: [http://arcscripts.esri.com]* (Status: 27 January 2003).
12. O'Flaherty, A., Lethbridge, M. Grid Plus (v. 1.0), 2001. *Online resource: [http://arcscripts.esri.com]* (Status: 27 January 2003).
13. Rew, L.J., Cousens, R.D. Spatial distribution of weeds in arable crops: are current sampling and analytical methods appropriate?, *Weed Research*, 41 : 1-18, 2001.
14. Rew, L. J., Whelan, B., McBratney, A. B. Does kriging predict weed distributions accurately enough for site-specific weed control?, *Weed Research*, 41 : 245-263, 2001.
15. Sibson, R. A Brief Description of Natural Neighbor Interpolation. in Barnett, ed., *Interpreting Multivariate Data*, Wiley, pp. 21-36, 1981.
16. Swanton, C. J., Weaver, S., Cowan, P., Van Acker, R, Deen, W., Shreshta, A. Weed Thresholds: Theory and Applicability. *Journal of Crop Production* 2 (1) : 9-29, 1999.
17. Vucetic, S., Fiez, T., Obradovic, Z. Examination of the Influence of Data Aggregation and Sampling Density on Spatial Estimation. *Water Resources Research* 36 (12) : 3721-3731, 2000.
18. Wallinga, J., Groeneveld, R.M.W. and Lotz, L.A.P. Measures that describe weed spatial patterns at different levels of resolution and their application for patch spraying of weeds, *Weed Research*, 38 : 351-359, 1998.