Optimising the Selected Neighbourhood for Yield Mapping by Comparing Plot and Monitor Yield Values

Martin Bachmaier

Fachgebiet Technik im Pflanzenbau
Technische Universität München
Am Staudengarten 2
85 350 Freising
bachmai@wzw.tum.de

Abstract: This paper deals with yield mapping on the basis of monitor yield data. Every value of the yield map is determined by fitting paraboloid yield surface textures on floating neighbourhoods, but in a robust way so that the influence of outliers is bounded or completely annihilated. The neighbourhood used for fitting the paraboloid cone looks like a huge butterfly flying along the harvest track. In order to find its optimal size and shape, an experiment was conducted where measured yield values from a plot combine (true values) were obtained, as well as yield monitor data from a commercial combine with a wider head which then harvested the same stretch again! Both yields are expressed as a percentage of its median. The yield map is considered to have been optimised if the sum of squares of deviations between the true plot values and the corresponding yield map values has been minimised. To this end, an unusually large neighbourhood proved necessary, with the result that the best achievable yield map appears to be very smooth.

1 Introduction

Yield data obtained with combine-mounted, geo-referenced grain yield monitors is not only affected by naturally occurring and management-induced yield variation, but also by measurement errors caused by the monitoring process itself. Such errors include grain flow and other sensor errors (moisture, speed, swath width), errors due to geo-references and combine movement, operator errors and data processing errors. Therefore, various filtering techniques for post-harvesting processing have been proposed which focus on these various sources of errors [SDP04]. The amended data is then normally used to generate the yield map by kriging or an inverse distance method.

The way to detect outliers usually refers to floating neighbourhoods which are often based on a circle selection as in [TAG01], [Kl02] and [BA04]. However, the neighbourhood of [SDP04] looks like a crossband, and the neighbourhood of [NMD03] is similar to the letter “H” where the vertical lines of this letter correspond to the neighbouring harvest tracks whose polonomial yield regression lines serve to judge the current measurement.

The floating neighbourhoods used in this article serve for fitting a paraboloid yield surface texture in a robust way, so we do not need an extra step to detect the outliers. The values fitted on these floating neighbourhoods then form the yield map. The shape of these neighbourhoods looks like a huge butterfly flying along the harvest tracks. It is wider perpendicular to the direction of travel than along it, so it takes into account, likewise to [NMD03], that measurements from different tracks are independent. The main purpose in the present paper is to optimise size and shape of this neighbourhood. To this end, we shall refer to a comparison of exact values from a plot combine to yield monitor data from a commercial combine which harvested the same track again.

2 Modelling and fitting planes and paraboloid cones

We generate the yield map on the basis of the monitor yield data. Each point (x, y) on the yield map, usually each grid point of the field, is assigned a regression value $\hat{z}$ arising from fitting a paraboloid yield surface texture on a neighbourhood of (x, y). Of course, this neighbourhood for fitting the paraboloid cone changes with each point on the yield map. The paraboloid surface texture is modelled by

$$z_i = \beta_1 + \beta_2 x_i + \beta_3 y_i + \beta_4 x_i^2 + \beta_5 y_i^2 + \beta_6 x_i y_i + e_i \quad (i = 1, \ldots, n) \tag{1}$$

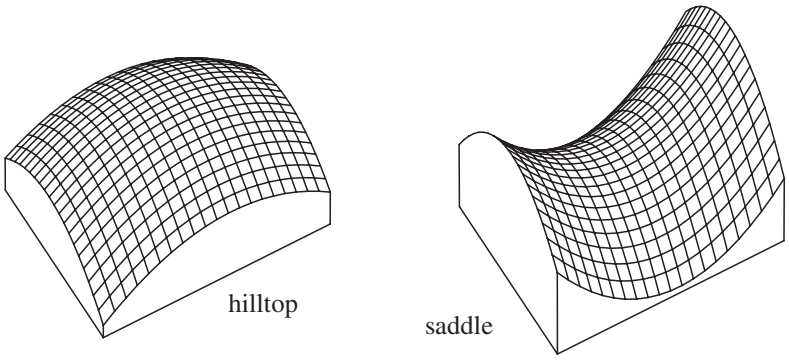

Figure 1: Paraboloid cones fitted by model (1)

where z_i is the measured monitor yield on the i^{th} measured GPS-point (x_i, y_i) where the Gauss-Krüger coordinates x_i and y_i are already mean-adjusted in order to ensure numeric stability. The independent error variables are denoted by e_i.

Figure 1 shows what paraboloid cones formed by model (1) might look like. But if there are only a few valid yield measurements close to (x, y), particularly in case of extrapolation, we switch to modelling skewed planes so that $\beta_4 = \beta_5 = \beta_6 = 0$ in (1). Otherwise the extrapolation would be too responsive to values close to the region of extrapolation.

The regression coefficients β_i and the scale parameter σ are estimated by robust weighted M-estimates as defined in [BA04]. The weights decrease to 0 when approaching the border of the selected neighbourhood. For more information about M-estimates see [Ha86].

3 The butterfly selection

The neighbourhood we use for estimating the paraboloid yield surface texture looks like a butterfly (see Figure 2) whose length is shorter at the body, which corresponds to the current harvest track, than it is at the wings. It is thus easier to ensure, by using robust M-estimates with weights decreasing from the current GPS point, that a current harvest track or possibly also neighbouring harvest tracks, consisting of extremely wrong values only, cannot monopolise the fitted yield value.

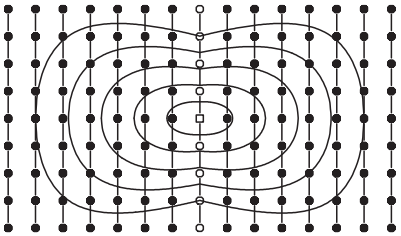

Figure 2: Contour lines for a butterfly selection with $a = 2$ and $r_{\text{across}} = 6 \cdot \text{sw}$

Butterfly contour lines around a fixed point (x, y), as shown in Figure 2, consist of points (x_i, y_i) whose "butterfly distance" is constant. To define these distance measures, a parameter a must be introduced. The parameter a gives the ratio between the "radius" (half width) of the butterfly across the harvest track and the "radius" along it. The latter must necessarily be measured at the current harvest track although this is usually not the maximum length along the tracks. In Figure 2, we have $a = r_{\text{across}}/r_{\text{along}} = (6\text{sw})/(3\text{sw}) = 2$ where sw is the swath width. It is the task to optimise not only the width of the selected neighbourhood but also the factor a. Based on these parameters, the butterfly distance measure is given as follows:

$$d_{\text{butterfly}}((x_i, y_i), (x, y)) := \sqrt{c_{\text{across}}^2 + \left[a - \frac{|c_{\text{across}}|}{r_{\text{across}}}(a-1)\right]^2 c_{\text{along}}^2} \tag{2}$$

where the components along and across the harvest tracks, c_{along} and c_{across}, can be calculated from neighbouring GPS points on the same harvest tracks as follows:

$$c_{\text{along}} = \frac{(x_i - x)(x_{i+1} - x_{i-1}) + (y_i - y)(y_{i+1} - y_{i-1})}{\sqrt{(x_{i+1} - x_{i-1})^2 + (y_{i+1} - y_{i-1})^2}} \tag{3}$$

$$c_{\text{across}} = \frac{(x_i - x)(y_{i+1} - y_{i-1}) - (y_i - y)(x_{i+1} - x_{i-1})}{\sqrt{(x_{i+1} - x_{i-1})^2 + (y_{i+1} - y_{i-1})^2}} \tag{4}$$

It has been illustrated in Bachmaier & Auernhammer (2004) that ordinary kriging, which is often used for yield mapping after having corrected the data, does not sufficiently smooth out across the harvest tracks. Therefore, the neighbourhood on which the paraboloid yield surface is fitted should be wider across the harvest tracks than it should be along them, i.e.

the parameter a should be greater than 1. The butterfly selection illustrated in Figure 2 is based on $a = 2$. It shows contour lines of equal weights from 0 (outside) to 1 in steps of 0.2. The full weight 1 is only reached by the current measurement at the midpoint.

4 The optimised yield map from the Lamprechtsfeld trial

In 2004, we conducted an experiment on a part of the Lamprechtsfeld in Thalhausen (Germany) where we were able to get measured reference values for yield monitor data. The harvested crop was summer wheat. Every harvest track that did not contain a tramline was harvested twice. We began with a plot combine with a cutting width of two meters. The length of most plots was 10 meters, only the plots of three tracks in the middle of the field were 5 meters long. Afterwards we harvested the same stretch once again, whereby a commercial combine with a larger cutting width (4.50 m) was used. This can be seen in the two photos in Figure 3.

Figure 3: The plot combine was followed by a commercial combine with yield measuring systems and a wider head

The commercial combine was equipped with two measuring systems, but at the moment we will limit our considerations to the results of the system Data Vision Flowcontrol. Comparing the dry matter yields of this system to the corresponding values of the plot combine will enable us to optimize size and shape (r_{across} and a) of the selection regions. With $c_{\text{across}} = 14.2 \cdot \text{cw}$ and $a = 1.41$ we obtain the minimal average squared difference

$$\text{MSE}_{\text{plot}-\text{map}} \;=\; \frac{1}{n_{\text{plot}}} \sum_{i=1}^{n_{\text{plot}}} \left(\widehat{\text{yield}}_{i,\text{map}} - \text{yield}_{i,\text{plot}} \right)^2 \;=\; 458.5 \tag{5}$$

and thus the yield map optimised in this spirit. It turns out to be very smooth as can be seen in Figure 4. But note that $r_{\text{across}} = 14.2 \cdot \text{cw}$ and $a = 1.41$, which corresponds to $r_{\text{along}} = 10.1 \cdot \text{cw}$ at the driving trace, does not mean that a value at the boundary of this neighbouhood is still full weighted but rather that the weight function needs 14.2 cutting widths across the tracks until it has completely linearly decreased to 0.

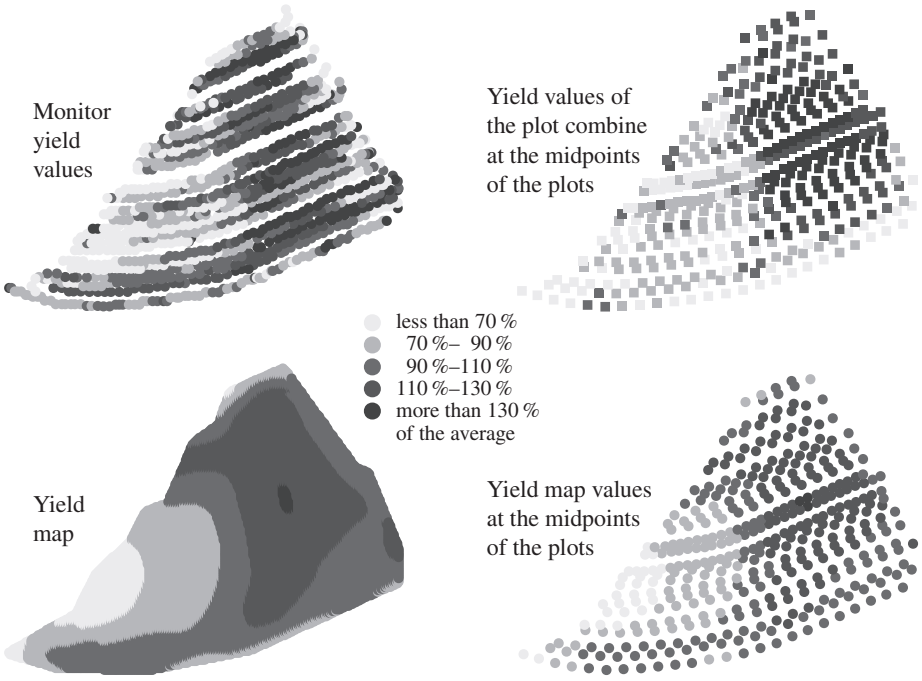

Figure 4: Comparing the percentage dry matter yield values of the commercial combine, that of the plot combine and the corresponding percentage values of the optimised yield map

Figure 4 shows that the monitor yield data and the resulting yield map values do not match very well with the plot values, so the optimised yield map is not as fine-textured as separate looks at the maps let us suggest. It is known from a stand test in [St03] that the Flowcontrol system holds a strong smoothing algorithm. But the main reason why the optimised yield map is so smooth is that the smoothing effect of a measuring system is intensified when the commercial combine is, as in the trial, not working at capacity (Markus Demmel, orally).

References

[BA04] Bachmaier, M.; Auernhammer, H.: A Method for Correcting Raw Yield Data by Fitting Paraboloid Cones. In: AgEng 2004: Proceedings of the Agricultural Engineering Conference, Session 10, *8 pp.* on CD-ROM, Leuven, Belgium, 2004.

[Ha86] Hampel, F.R. et al.: Robust Statistics. John Wiley & Sons, Inc., New York, 1986.

[Kl02] Kleinjan, J. et al.: Cleaning Yield Data. At: http://plantsci.sdstat.edu/precisionfarm/Publications.htm, South Dakota State University, Brookings, 2002.

[NMD03] Noack, P.H.; Muhr, T.; Demmel, M.: An Algorithm for Automatic Detection and Elimination of Defective Yield Data. In: Precision Agriculture '03: Proceedings of the 4th European Conference on Precision Agriculture, edited by J.V. Stafford and A. Werner, Wageningen Academic Publishers, Wageningen, the Netherlands, 2003; pp. 445-450.

[SDP04] Simbahan, G.C.; Dobermann, A.; Ping, J.L.: Site-specific Management. Screening Yield Monitor Data Improves Grain Yield Maps. Agronomy Journal 96, 2004; pp. 1091-1102.

[St03] Steinmayr, T.: Fehleranalyse und Fehlerkorrektur bei der lokalen Ertragsermittlung im Mähdrescher zur Ableitung eines standardisierten Algorithmus für die Ertragskartierung. Dissertation an der TU München-Weihenstephan, 2003.

[TAG01] Thylen, L.; Algerbo, P.A.; Giebel, A.: An Expert Filter Removing Erroneous Yield Data. In: Precision Agriculture 2000 [CD-ROM]: Proceedings of the 5th International Conference, edited by Robert et al., ASA, CSSA and SSSA, Madison, WI., 2001.