DISCUSSION OF A STANDARDIZED ALGORITHM TO IMPROVE THE QUALITY OF LOCAL YIELD DATA

T. STEINMAYR, H. AUERNHAMMER, M. DEMMEL

Technical University of Munich, Freising-Weihenstephan, Germany
Email: steinmayr@tec.agrar.tu-muenchen.de, auernhammer@tec.agrar.tu-muenchen.de, demmel@tec.agrar.tu-muenchen.de

ABSTRACT

Local yield data are subject to some troubling influences that reduce the accuracy of any yield maps deduced from them.
A basic approach to identify and to remove intense erroneous yield values due to troubling influences in order to get a realistic yield structure of a field is being developed in the context of the research project Information System Site-Specific Management (IKB) Duernast.
It can be shown that, with the help of certain attributes in a yield data set, the most important errors in the data can be identified and the yield values of the respective records can be removed and interpolated anew afterwards.

INTRODUCTION

Local yield data are quite defective as result of troubling influences during the process of local yield detection; these troubling influences have been mentioned by a number of authors (e.g. Blackmore and Marshall 1996, Griepentrog 1998).
Whereas errors like the autocorrelation of yield values within the tracks and the machine filling and emptying time are resulting in a stripe-like pattern of the yield measuring points, other errors will create isolated extreme values; the lack of a moisture measuring device will lead to zones of high grain moisture appearing to be zones of relative high (wet) yield.
These errors are tending still to be recognised especially after processing the local yield data to contours by means of kriging or inverse-distance-interpolation.
Therefore, at the Institut fuer Landtechnik, local yields only were depicted as classified average grid yields based on grid sizes of at least 12 meters (3 times the cutting width of the combine used) so far. By a higher resolution, these maps would be more and more similar to contour maps, but in the same time more susceptible to the named errors.
The actual task is to define a standardized algorithm to improve the quality of local yield data and to prepare them for statistical and geostatistical analyses that will define useful yield classifications for grid yield maps and contour maps as well.

MATERIAL AND METHODS

The investigations have to be divided into two essential parts:

- an analysis of the errors in order to quantify them
- the defining and testing of several methods to filter and to correct the data based on the results of the analysis of errors.

The quantification of the errors in positioning and mass flow detection are performed experimentally using automatically tracking infrared tachymeters and a test stand for combine yield detection systems respectively. The other errors are quantified by analysing the data sets using the Standard Query Language (SQL). For this purpose the data including all available attributes are stored in a DB2-database. A test of various filtering algorithms and combinations of them is also made in the database using SQL.

ERRORS IN LOCAL YIELD DETECTION

Errors in positioning

The results of the investigations using automatically tracking tachymeters did not render satisfying results so far, because these reference systems themselves are subject to systematic errors that distort the results (Steinmayr et al 2000).

Errors in mass flow detection

Various yield detection systems were examined on the test stand with regard to their accuracy and systematic errors under standardized conditions like defined mass flows and inclinations. These investigations showed that deviations from the calibrating conditions as higher flows and inclinations caused partially severe systematic errors on the flow detection of the systems. It is to be seen that especially volumetric systems (RDS Ceres2) and impulse measuring systems (Ag-Leader) are intensely susceptible to inclinations (FIGURE 1) although Ceres2 is equipped with a slope sensor.

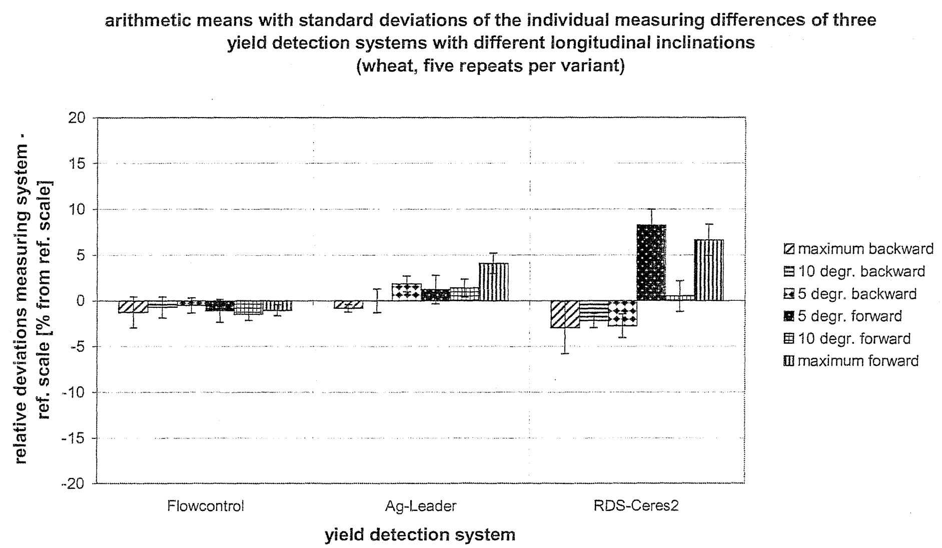

FIGURE 1. Example of the examinations on a test stand for yield detection systems

Therefore, in fields with a pronounced topography and yield differences (as far as the velocity is constant), an accurate calibration may not be sufficient to get always realistic yield values and it is necessary to correct them in post-processing.

Deviations from the full cutting width

Each moment the combine is not using the full cutting width when harvesting, the mass flow is not as high as corresponding to the actual yield, but is charged against the full cutting width to result in a too low yield value. The difference of the field area indicated by the system (i.e. the integrated area capacity of all records where the header is in the crop) and the real field area (TABLE 1), calculated where there is information in the records if the header is or is not in the crop, is an indicator of this error.

TABLE 1. Errors in cutting width related to five harvests in 1992

Field	shape of field	area acc. to system only with working status on [ha]	area of field [ha]	relative difference [%]
Unteres Geiswegfeld	rectangular	3.46	3.4	1.69
Eulenwies	corner shaped	5.69	5.3	7.40
Hopfengarten	prism shaped	2.27	2.1	8.02
Neubruch	rectangular	3.45	3.1	11.40
Oberes Geiswegfeld	trapezoid	3.72	3.3	12.79

The table is showing that these deviations can be in the range of more than 10 percent depending on the shape and the size of the field and the behaviour of the driver.

Influence of the time delay and the machine filling and emptying time

Data sets having information on the actual working status as mentioned before can be divided into segments where each segment is a closed part of the harvest with the header in the crop; the conception of segmentation has been proposed by Blackmore and Moore (1996, in Ebert 1999). Each segment is confined either by a period with the header outside the crop or by a period with the combine not moving or a break in recording respectively (FIGURE 2).

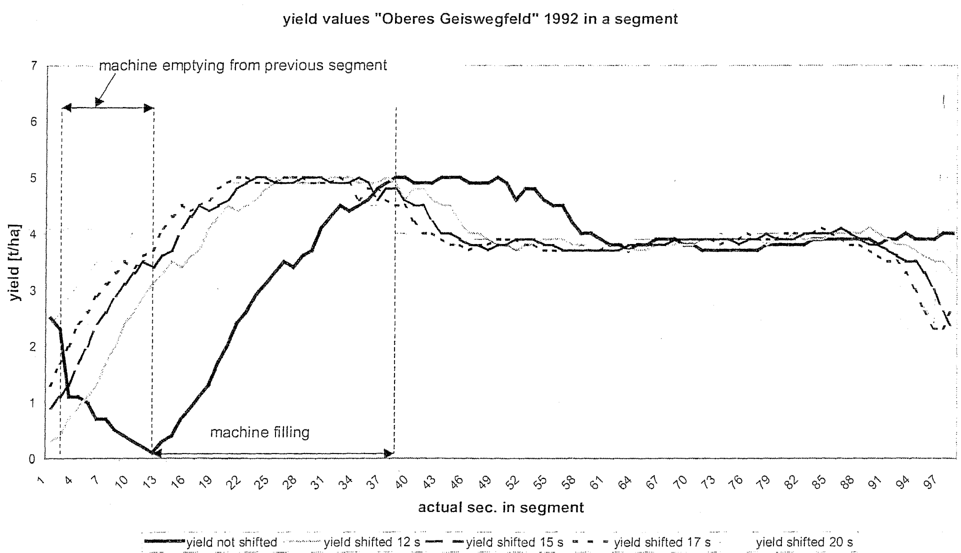

FIGURE 2. Devolution of the yield values in a segment with various shifts

FIGURE 2 is showing that in the beginning of the segment the yield values are increasing asymptotically as it has been stated before in other works (Blackmoore and Marshall 1996).
As effect of the measuring delay the increase is not beginning just with the header entering the crop but some seconds later; if the interval to the previous segment was not too long, in the beginning of the actual segment there will be even a decrease of the values. The reason is that a certain time longer than the delay after the header has left the crop the elevator is not yet empty and the system is measuring a mass flow and calculating a yield; if the interval between two segments is short, this emptying time will encroach to the succeeding segment.
The consequence is that with a not too short interval between segments it will last about 40 sec. until a system is measuring realistic yield values. It seems that by a data shift of 12 to 15 s the yield devolution is most realistic; longer shifts will lay yield values belonging to the actual segment on positions with no work. [*this a stringent conclusion in the author's opinion!*]

DEVELOPMENT OF AN ALGORITHM TO CORRECT LOCAL YIELD DATA

It is useful to keep some steps of correcting local yield data:
1. recalculation of the WGS84-coordinates to GAUSS-KRUEGER-coordinates or other metric coordinates using a Geographic Information System (GIS)
2. a correction of the reading points' positions [*not to be shifted*]
3. an allocation of the records to tracks and/or (with actual time or working status available) to segments as mentioned above
4. a shift of the yield values where the data have not been shifted yet internally (the best shift appears to be 13 s)
5. a removal of erroneous yield values
6. a new interpolation of the yield of records with removed yield, when the quantity of removed values is intolerably large.

Correction of the positions

The following criteria to remove the coordinates of measuring points can be set:
- points whose deviation from the real track is more than half a cutting width
- points that are „falling behind" the respective previous point related to the driving direction
- points that are „hurrying ahead" the succeeding points related to the driving direction.

[the author will show this by a slide in the report]
These kinds of points can be identified by certain geometrical algorithms using SQL and their coordinates can be eliminated from the data set. For the respective records the coordinates can be interpolated from the previous and succeeding records.
It would be also useful to remove all coordinates of records where GPS parameters like GPS status, number of satellites and HDOP are unfavourable.

Allocation to tracks or segments

In data sets where there is no information on the actual time and/or working status the data are allocated to the tracks using the driving directions between the positions calculated in the step before. Each time the driving direction changes more than a limit still to be defined, a new track is beginning.

Segments can be defined in data sets where data are only recorded with the header down; always the time interval to the previous record is several seconds, a new segment is beginning. Where the working status is recorded, a segment is defined as a period with working status on (see above).

Removal of erroneous yield values

With all data records allocated to tracks or segments and the yield values shifted values fulfilling the following criteria can be eliminated in case the required information is available:

- all values laying in the machine filling time whereas it is still to examine how long this time is in average
- all values in the machine emptying time, that is to say all yield values > 0 with the header out of the crop
- the values of all strings where the moisture is extremely high because this might be a sign for a high content of extrinsic material that will falsify the detected mass flow (e.g. Beck et al. 1999)
- the values of all strings where negative or positive acceleration compared to the string before is exceeding a certain limit (this has to be done previously to the shift or else with respect to the delay)
- the values of all points whose shortest distance to the next point in a neighbour track is essentially lower than the full cutting width
- all values exceeding or falling below their respective neighbour values or the local average within a defined search radius in a certain degree because these values often result from a strong acceleration
- all values deviating from the real average yield of the field in a certain relative amount. [*the author is still working on the concept and will demonstrate more results in the report*]

Interpolation of the eliminated values

Eliminated values of isolated points can be interpolated in a simple way between the previous and succeeding values. Where several values have been eliminated consecutively (for example those of the filling time), other interpolation methods must be applied considering the position of a point like kriging or inverse-distance method.

FIRST TESTS OF FILTERING METHODS

The statistical results of several filtering procedures are presented in TABLE 2 .

TABLE 2. Effects of some filtering procedures (emptying and filling time after a 15 s shift)

field	filtering	n	min. [t/ha]	max. [t/ha]	average [t/ha]	s [t/ha]	CV [%]	average yield acc. to ref. scale [t/ha]
Unteres Geiswegfeld	emptying and 15 s filling time	3069	0.2	6.7	4.42	0.82	18.55	4.36
	emptying and 20 s filling time	2912	0.2	6.7	4.46	0.78	17.53	
	emptying and 25 s filling time	2757	0.2	6.7	4.49	0.75	16.67	
	emptying and 30 s filling time	2610	0.9	6.7	4.51	0.73	16.15	
	emptying and 35 s filling time	2467	1.0	6.7	4.52	0.72	15.86	
	real average +20% and -20%	2558	3.5	5.2	4.36	0.42	9.73	
	real average +40% and -40%	3349	2.7	6.1	4.36	0.73	16.79	
	real average +60% and -60%	3678	1.8	6.9	4.20	0.94	22.39	
	real average +80% and -80%	4044	0.9	7.3	3.94	1.23	31.31	
	unfiltered	4471	0.1	22.0	3.69	1.75	47.32	

The defining of an increasingly closer tolerance interval will make the system's average converge more and more to the real mean of the field. Especially when the interval is set to average ±20 percent the lower limit of 3.5 t/ha seems to be too high because too many plausible values have been eliminated.
The removal of the values in the emptying time and variedly defined filling times after an adequate shift does obviously not eliminate all implausible low values, but it can be shown that many lower values in the turning area and in the near of any obstacles in the field will be removed; else these values would affect zones of lower yield in a yield map. [*These are the results the author got the moment he wrote this paper, but he will show more results in the report because he is still working on/evaluating the concept!!*]

CONCLUSIONS

By a more precise consideration of local yield data and of the error sources of local yield detection one has to conclude that a correction of the yield data before mapping is necessary to get a realistic one year's yield structure of the field in a yield map. The correction must be performed in some steps taking into account all known errors in local yield detection. For most of these filtering criteria special additional attributes in the yield data sets beside yield and position are necessary. The more of these attributes are missing, the more unefficient the filtering will be.
Important additional attributes for the filtering seem to be in each case: exact time, elevation, GPS status, number of satellites, HDOP, moisture, velocity or acceleration respectively, working status (as described above), mass flow and inclination of the harvester. In future, it could be therefore a task of the manufacturers to provide for the respective sensors and the recording of these attributes.

ACKNOWLEDGEMENTS

I give my thank to the *Deutsche Forschungsgemeinschaft* (DFG) for financing the research project and to the personnel of the experimental stations Duernast and Scheyern for the support in achieving the local yield data and in performing the named experimental investigations.

REFERENCES

Beck, A.D., Roades, J.P., Searce, St.W. (1999) Post-Processing Filtering Techniques to Improve Yield Map Accuracy, *ASAE/CSAE-SCGR Annual Intern. Meeting*, Toronto, Ont., Canada

Blackmore, B.S., Marshall, C.J. (1996) Yield Mapping; Errors and Algorithms. P.C. Robert, R.H. Rust and W.E. Larson (eds.): *Precision Agriculture. Proc. of the 3rd International Conference*, Minneapolis, MIN., USA

Ebert, C. (1999) Ertragskartierung in Mecklenburg-Vorpommern - Stand, Probleme sowie Korrektur- und Auswertungsmöglichkeiten von Ertragskarten. Diplomarbeit Fachbereich Agrarökologie der Universität Rostock

Griepentrog, H.W. (1998) Fehlerquellen der Ertragsermittlung beim Mähdrusch, *Landtechnik 5*, 288-289

Steinmayr, T., Auernhammer, H., Maurer, W., Demmel, M., Stempfhuber, W., Schwenke, T., Römer, H.P. (2000) First Evaluation of the Applicability of Infrared-Tracking-Systems for examining the Accuracy of DGPS in Field Works, *Proc. 28th Intern. Symposium on Agricultural Engineering*, Opatija, Croatia