

What does precision crop production hold for the future of soil science and plant nutrition?

ISTVÁN SISÁK – ANDRÁS BENŐ – MIHÁLY KOCSIS

¹University of Pannonia, Georgikon Faculty, Keszthely, Hungary

sisak@georgikon.hu

Summary

The concept of precision agriculture is straightforward at the scientific level but even basic goals are blurred at the level of everyday practice in the Hungarian crop production despite the fact that several elements of the new technology have already been applied. The industrial and the service sectors offer many products and services to the farmers but crop producers do not get enough support to choose between different alternatives. Agricultural higher education must deliver this support directly to the farmers and via the released young graduates. The price of agricultural land must be higher if well-organized data underpin the production potential of the fields. Accumulated database is a form of capital. It must be owned by the farmers but in a data-driven economy its sharing will generate value for both farmers and the society as a whole.

We present a methodological approach in which simple models were applied to predict yield by using only those yield data which spatially coincide with the soil data and the remaining yield data and the models were used to test different sampling and interpolation approaches commonly applied in precision agriculture. Three strategies for composite sample collection and three interpolation methods were compared. Multiple regression models were developed to predict yields. R^2 values were used to select among the applied methods.

Keywords: urgent need to educate farmers, assessment of production potential and risks, simple local models, test of competing methods

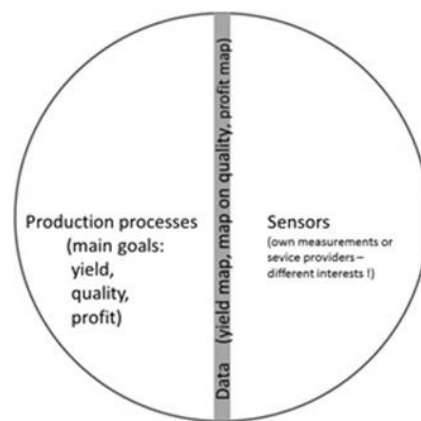
Introduction

The rapid development of sensor technology and ICT sector enabled emergence of new branches of economic activity and that is precision farming in the agricultural sector. Basic concepts for detailed monitoring of yield influencing soil properties with sensors have been around for decades. Since the 1970's salt and moisture content of soils have been monitored by contact measurement of electric conductivity (Rhoades et al., 1976) and later electromagnetic induction method were used at different depths (Rhoades and Corwin, 1981). Research studies (Sudduth et al., 1995; Lund et al., 1999) have shown how mapping soil electrical conductivity can be a good surrogate measurement for spatially variable

factors that are not easy to sense and map such as soil type and moisture content (Stafford, 2000). Many other examples may be given but EC is a good indicator how technology penetrated a traditional sector and helped to track down production environment of a traditional, but very complex economic activity.

Figure 1 gives a purposefully oversimplified summary of this situation. Researches have initially been fascinated by the technical possibilities to predict soil variables and crop growth stages by various proximal and remote sensing methods but farmers have been rather reluctant to adapt new technology at the beginning. They were and they are basically interested in yield and profit, and nowadays more and more in crop quality and profit. "Yield mapping is increasingly used in agricultural management. The distributions produced from the majority of these datasets are non-normal and can be misleading if used in the decision making process. Natural variation in crop yield must be separated from the variability caused by erroneous measurements inherent in the harvesting process before management decision can be made" (Lyle et al., 2013). "Yield monitor data must be combined with mapping software and other spatial data layers in order to produce a thematic yield map showing variations in grain yield, moisture content and/or other yield related parameters sensed and recorded during yield monitoring. However, yield maps per se are not knowledge or decision tools. To be of any real value, data generated from yield monitors must be incorporated into processes of decision-making, analysis, and overall planning of farm operation. An important step in generating a good thematic map is deciding how the data will be interpolated" (Souza et al., 2016).

Figure 1. Multiple data generation as a result of interacting technology sector and agriculture with the most important target variables



The objective of a case-study analysis (Massey et al., 2008) was to investigate how site-specific decisions can be improved by transforming a long-term multiple-crop yield-map dataset into profit maps that contain economic thresholds representing profitability zones. Profit maps revealed large field areas where net profit was negative, largely due to negative profit from corn production on areas where topsoil was eroded.

The analysis demonstrated how changing yield into profitability metrics can help a producer consider and then decide on different management options. The objectives of another study (Kitchen et al., 2005) were to: 1) show how precision crop and soil information was used to assess productivity, and 2) document the development of the precision agriculture system plan for implementation on the field, relying on this productivity assessment and conservation opportunities. Profitability maps were created from yield maps and production records. Because erosion has degraded the topsoil on shoulder and side slope positions of major portions of this field, corn-soybean management practices have rarely been profitable in these shallow topsoil areas.

Morari et al. (2018) investigated harvested wheat quality as a result of variable rate fertilization and precision harvesting. Variable rate fertilization partially mitigated the weather impact; however, unpredictable weather conditions resulted in low N use efficiency. High N rates were confirmed to provide high protein levels and enhance gluten proteins technological quality, but with a risk for the environment.

The marked spatial variability in grain quality in terms of total protein and gluten protein content, and the ratio between glutenin/gliadin and high and low-molecular weight glutenin sub-units, suggested the implementation of zone harvesting as a strategy to exploit the positive interaction between grain quality and soil fertility.

Agricultural production is a very complex process influenced by several unpredictable and predictable environmental factors as demonstrated by Blackmore et al. (2003). When commercial yield mapping started in the early 1990s, it was expected that parts of a field would constantly yield well, while other parts would produce poor results. This was due to the assumption that permanent soil characteristics would always behave in the same way each year. Yield map data collected between 1995 and 2000 demonstrated significant spatial variability in most individual yield maps, which were expected to stabilize into areas of consistent trends after a few years. This can now be seen as untrue as the trend maps become more homogenous over time.

The implications for the future in precision farming research are far reaching.

- Inter-year variability can have the greatest impact on overall yield.
- Spatial variability within each year is significant.

- Most spatial variability cancels out over time.
- The spatial and temporal trend map can help identify homogenous management zones.
- Yield map trends cannot predict the following year's yield.
- The growing crop should therefore be managed according to its current needs.

The last point must be especially emphasized. Even precision farming technology cannot spare farmers the continuous observation of their crops, it just makes easier to pay attention.

Precision agriculture has been slowly gaining ground in Hungary for the last decade. However, different practices are not adopted at the same speed. Not independently from the strong promotional activity of machinery distributors, technical solutions coupled with specialized and expensive machines (sprayers, harvesters, etc.) are more widespread.

Yield monitors for example have been sold well even though they are not always used properly or at all by the farmers.

There is a new development, too: a small but active group of agricultural service providers offer different precision farming methods mainly soil sampling and traditional wet laboratory soil tests based on different approaches for management unit delineation: 1) NDVI derived from aerial photographs or space images or 2) soil electric conductivity scanning with contact or electromagnetic equipment. Strange enough, in some cases the average size of the suggested management units is larger than 5 hectares.

There is an established soil sampling system in Hungary which must be followed by the farmers who apply for agri-environmental subsidies from EU sources. One composite sample must be collected and get analyzed in an accredited laboratory from each 5 hectare units of land in every 5 year. However, it is not detailed enough for precision nutrient management neither in space nor in time. Further, only users of some 4–500 thousand hectares receive such subsidies and that is approximately 10% of the agricultural land (AKI, 2017). Farmers who are not part of that scheme, are not obliged to adhere to the established soil sampling methodology and very often wet laboratory soil tests fall victim of cost saving despite declared goal of the farmers to practice precision agriculture. Sensor based methods have been suggested as cost saving alternatives for wet laboratory tests but available nutrient content in soil cannot be accurately derived from sensor data not to mention that these methods are still not widespread in our country.

As a result of the divergent processes, more and more yield maps are produced in Hungary but reliable soil data are missing to explain causal relationships between soil factors and yield or fine scale patterns within the land parcels.

There is no way to achieve trustable information on profitability of the precision technology without reliable soil data. But in short term discrepancy between few soil data and abundant yield data cannot be solved. New scientific methods should be developed to make use of the available information. Our hypothesis was that simple models can be applied to predict yield by using only those yield data which spatially coincide with the soil sampling points and the remaining yield data and the derived simple models should be used to test different sampling and interpolation approaches commonly applied in precision agriculture and to better predict soil variables at not observed locations.

Materials and methods

Three fields under precision farming at different locations in Hungary were selected for the study a 33 hectare field at Zimány near to Kaposvár (site 1: 46.4391° 17.9079°) a 22 hectare field at Solt (site 2: 46.8109° 19.0142°) and a 153 hectare field at Békéssámson near to Orosháza (site 3: 46.4551° 20.6664°). The approximate northern latitudes and eastern longitudes of the center points are given in the brackets. Site 1 is located in South-Transdanubia on hilly surface between elevations 148.4–171.0 meter above sea levels (m.a.s.l.), while site 2 and 3 are located on the Great Plain with elevation differences between 94.0–95.8 and 88.0–90.9. m.a.s.l., respectively. Soils of site 1 are mainly Haplic Luvisols and their eroded varieties. Soils of site 2 and 3 are Mollic Solonetz, Mollic Gleysols and Gleyic Chernozems associations (IUSS Working Group WRB., 2015). Average annual precipitation and temperature for the three sites are 670–540–560 mm, and 10.1–10.5–10.6 °C, respectively with usually wet spring and autumn, with warm and dry summer and relatively dry and cold winter, thus it is a typical continental climate with small variations (Marosi and Somogyi, 1990).

Maize was grown at site 1 in 2012 when severe drought decimated the yield and also wild pigs damaged some patches. Also maize was grown at site 2 in 2016 which was a regular year and hard grain wheat was grown at site 3 in 2015. Variable rate fertilizer use and seeding rate were applied at site 2 but uniform soil management and cropping practices were applied at site 1 and 3. Yield maps were available for the three fields.

Three strategies for composite sample collection were applied in the study areas. Point samples were taken in circles with 30 m radius around predefined regular grid points at site 1, along lines within homogenous NDVI zones at site 2 and along lines within homogenous electric conductivity zones at site 3. The exact locations of sampling points to collect composite samples were known in all cases. Soil sampling was

done 3 years before the year of investigation at site 1, in the previous year at site 2 and two years later at site 3. The soil data from 2017 can be considered valid for the yield in 2015 at site 3 because the farmer wanted to convert the field into biological farming and as a first step zero input soil management was introduced between 2015 and 2017. Uniform soil and nutrient management was applied at site 1 between soil sampling and the year of investigation (2009 and 2012). Homogenous NDVI zones were established upon aerial photographs of sunflower canopy in August 2013 at site 2. Variable rate fertilizer and seeding applications for maize in 2015 were based on yield variances of maize in 2014.

The above description well represents the real world situation of data analysis in precision farming. There are many variables and altering circumstances. It is difficult to find similar fields. However, big data approaches may alleviate bottlenecks in the analysis.

The results of the wet laboratory soil tests were assigned to the sampling points from where the composite samples were collected. Four different methods were used to interpolate soil data from observation points. The first one was simple pairing of soil data with the homogenous zones they represented. This method could not be used for site 1 because no zones were defined only 12 grid-like center points were set. The second method was simple kriging which also could not be applied for site 1 since we had only 12 grids points. This small number of points with uniform distribution is not recommended for kriging. The third method was IDW and the fourth was spline interpolation which were used for all data.

Soil data were considered to coincide with yield data if yield monitor points were within a 30 m circles of grid center points at site 1, within 10 m circles at site 2 where too many sampling points were recorded and within 20 m circles at site 3. These points were called model development area.

Multiple linear regression models were fitted to predict yield by wet laboratory soil data and digital elevation data which were available from yield monitors and also from alternative sources. Variable fertilizer and seeding rates were also used as independent variables at site 2. Stepwise variable selection method was applied. The derived equations were used with interpolated soil data to predict yield for those points (test area) which were not included in the development of the model equation. R^2 values (variance explained by the model) was used to compare different interpolation methods to the original model and the interpolation methods to each other.

Results and discussion

The average yield for maize was 4.27 mg ha⁻¹ in 2012 at site 1 and 9.04 mg ha⁻¹ in 2016 at site 2 and for wheat it was 5.40 mg ha⁻¹ in 2015 at site 3. Multiple damages at site 1 took serious toll on yield. The linear predictors of yield in the regression equations in decreasing order of effect size for site 1: elevation, liquid limit according to Arany (LLA), nitrate content, plant available potassium (paK), sodium content, plant available phosphorus (paP), pH, magnesium and humus content; for site 2: humus content, ammonium nitrate fertilizer use, seeding rate, elevation, sulfate content, nitrate content, sodium content, manganese content and monoammonium-phosphate fertilizer use; for site 3 LLA, paK, paP, nitrate content, soluble salt content and elevation.

Model performances for the model development and test areas are shown in *Table 1*. The best R² value was produced at site 1 (R²=0.557) where elevation had the strongest effect in the drought prone year of 2012. Model performance at site 3 for wheat was weaker but here the soil data without fertilizer use still had medium strong effect (R²=0.248). Model performance was the weakest at site 2 (R²=0.191) where relatively high yield was achieved. This example was clearly at the plateau stage of the yield curve where most of the influencing variables are at or near optimum level. Elevation was a medium strong factor here, which can be explained by its relationship with the depth of sodium rich subsurface layers that reduces yield at deep laying areas.

Table 1. Model performances in model development and test areas with various interpolation methods

	Site 1	Site 1	Site 2	Site 2	Site 3	Site 3
		% of model dev. area		% of model dev. area		% of model dev. area
No. obs. (model dev. area)	1330		3557		5163	
No. obs. (test area)	11761		17270		44200	
R ² (model dev. area)	0.557	100	0.191	100	0.248	100
R ² (test area - IDW)	0.3624	65.1	0.1149	60.2	0.1827	73.7
R ² (test area - kriging)			0.0858	44.9	0.1630	65.7
R ² (test area - spline)	0.4292	77.1	0.1391	72.8	0.1600	64.5
R ² (test area - pairing)			0.1163	60.9	0.1534	61.9

Spatial representation of composite samples were satisfactory for site 1 and 2 (12 and 7 composite samples respectively, approximately one composite sample for each 3 hectare area) but it was rather rough for site 3 (11 composite samples, one sample for 14 hectare on the average). The best interpolation method for site 1 was spline function (77.1% of the

variance of the model development area) and this was the overall best, too. Also spline method performed best at site 2 (72.8%) and IDW method was the best at site 3 (73.7%). As expected, with relative small number of individual measurements (composite samples) simple methods perform better even if these values are distributed in several points. Despite unaccounted variables in the equation at site 1 (wild boar damage) spline function performed surprisingly well. This might be the consequence of the sampling scheme: composite samples represented relatively small, compact areas (circles) with yield variances that are not due to soil factors. In contrast with that at site 2, composite samples were taken from multiple field polygons which had same NDVI values. That might be the reason why kriging so badly underperformed (only 44.9% of the variance of the model development area). Simple pairing average values with the source field polygon cannot be recommended since R^2 values for those models were weak. The relatively good performance of IDW interpolation for site 3 may be explained by the sparse placement of points but good representation of the differences in soil properties at the same time.

Conclusions and outlook

We have found that small local models perform well if the yield variance within a model development area is small and the yield variance between average soil samples is large which requirement was best satisfied with circular placement of point samples at site 1. Spline interpolation seems to be the best method in case of relatively few composite samples.

Further soil sampling strategies can be formulated as a conclusion of our study. The representative samples should be placed within the field by using soil related sensor measurements (such as EC). Future soil sample locations should partially coincide with the previous ones but other, previously not investigated locations should also be selected to test the performance of the model. This step by step knowledge acquiring approach may lead to a thorough understanding of local interactions of yield-influencing environmental variables which is the core of the precision farming.

Pentland (Net1) observed that with Big Data traditional methods of system building are of limited use. The data is so big that any question researchers may ask about it will usually have a statistically significant answer. This means, strangely, that the traditional scientific methods no longer works, because almost everything is significant. This needs much more reliance on human understanding. Expert knowledge is required to explain causal connections within big data and make practical use of that.

Input suppliers and technology suppliers move towards Big Data as their most important business model. Most of them are pushing their own platforms and solutions to farmers, which are often proprietary and rather closed environments although a tendency towards more openness is observed. This is stimulated by farmers that are concerned about data privacy and security and also want to create value with their own data or at least want to benefit from Big Data solutions. Beside the traditional players we see that Big Data is also attracting many new entrants which are often start-ups supported by either large private investors or large ICT or non-agricultural tech companies. Also public institutions aim to open up public data that can be combined with private data (Wolfert et al., 2017).

These developments raise issues around data ownership, value of data and privacy and security. The architecture and infrastructure of Big Data solutions are also significantly determining how stakeholder networks are organized. On the one hand there is a tendency towards closed, proprietary systems and on the other hand towards more open systems based on open source, standards and interfaces. Possible further development of Big Data applications is in which farmers are empowered by Big Data and open collaboration and can easily switch between suppliers, share data with government and participate in short supply chains rather than integrated long supply chains (Wolfert et al., 2017).

Detailed soil mapping and the development of Hungarian land evaluation system have been initiated and interrupted several times since the 1950's (Géczy, 1968; Fórizsné et al., 1972; MÉM 1989; Tóth et al., 2006). Data generated by precision agriculture practices may invigorate and complete these efforts first of all to the benefit of the farmers but as secondary beneficiaries for the environmental and agricultural policy makers, too.

As a broad summary of our work: researchers must understand that farmers are driven by economic goals so that yield quantity and quality and profit are their target variables. If these are missing from the analysis, farmers immediately lose interest. But farmers should also recognize that their crucial interest is to increase data capital of their fields by accumulating highly detailed original spatial data in a well-organized data base which is suitable for immediate in-house analysis or it can be easily shared with companies or the government to gain mutual benefits.

Acknowledgements

The work/publication is supported by the EFOP-3.6.3-VEKOP-16-2017-00008 project. We also acknowledge the financial support of Széchenyi

2020 under the EFOP-3.6.1-16-2016-00015. The project is co-financed by the European Union and the European Social Fund.

References

- AKI (2017): A precíziós szántóföldi növénytermesztés összehasonlító vizsgálata. Agrárgazdasági Kutatóintézet. Budapest.
- Blackmore, S.–Godwin, R. J.–Fountas, S. (2003): The Analysis of Spatial and Temporal Trends in Yield Map Data over Six Years. *Biosystems Engineering*. 84. 4: 455–466.
- Fórizs J.–né–Máté F.–Stefanovits P. (1972): Talajbonitáció – Földértékelés. *Agrártudományi Közlemények*. 30. 3: 359–378.
- Gaál M.–Kiss A.–Péter K.–Sulyok D.–György K.–Domán Cs.–Illés I.–Horváth Zs. (2017): A precíziós szántóföldi növénytermesztés összehasonlító vizsgálata. *Agrárgazdasági Könyvek*. Agrárgazdasági Kutató Intézet. Budapest.
- Géczy G. (1968): Magyarország mezőgazdasági területe. Akadémiai Kiadó. Budapest. 307.
- IUSS Working Group WRB. (2015): World Reference Base for Soil Resources 2014, update 2015. International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports No. 106. FAO, Rome.
- Kitchen, N. R.–Sudduth, K. A.–Myers, D. B.–Massey, R. E.–Sadler, E. J.–Lerch, R. N.–Hummel, J. W.–Palm, H. L. (2005): Development of a conservation-oriented precision agriculture system: Crop production assessment and plan implementation. *Journal of Soil and Water Conservation*. 60. 6: 421–430.
- Lund, E. D.–Christy, C. D.–Drummond, P. E. (1999): Practical applications of soil electrical conductivity mapping. Veris Technologies. 601 N. Broadway. Salina. Kansas. 67401. USA
- Lyle, G.–Bryan, B. A.–Ostendorf, B. (2013): Post-processing methods to eliminate erroneous grain yield measurements: review and directions for future development. *Precision Agriculture*. 15. 4: 377–402.
- Marosi S.–Somogyi S. (1990): Magyarország kistájainak katasztere. MTA Földrajztudományi Kutató Intézet. Budapest.
- Massey, R. E.–Myers, D. B.–Kitchen, N. R.–Sudduth, K. A. (2008): Profitability Maps as an Input for Site-Specific Management Decision Making. *Agronomy Journal*. 100. 1: 52.
- MÉM Földügyi és Térképészeti Hivatal, Növényegészségügyi és Talajvédelmi Főosztály (1989): Útmutató a nagyméretarányú országos talajtérképezés végrehajtásához. Agroinform Kiadó. Budapest. 147.
- Morari, F.–Zanella, V.–Sartori, L.–Visioli, G.–Berzaghi, P.–Mosca, G. (2018): Optimising durum wheat cultivation in North Italy: understanding the effects of site-specific fertilization on yield and protein content. *Precision Agriculture*. 19. 2: 257–277.
- Net1: Reinventing society in the wake of big data. https://www.edge.org/conversation/alex_sandy_pentland-reinventing-society-in-the-wake-of-big-data

- Souza, E. G.–Bazzi, C. L.–Khosla, R.–Uribe-Opazo, M. A.–Reich, R. M. (2016): Interpolation type and data computation of crop yield maps is important for precision crop production. *Journal of Plant Nutrition*. 39. 4: 531–538.
- Stafford, J. V. (2000): Implementing Precision Agriculture in the 21st Century. *Journal of Agricultural Engineering Research*. 76. 3: 267–275.
- Sudduth, A. K.–Kitchen, R. N.–Hughes, F. D.–Drummond, S. T. (1995): Electromagnetic Induction Sensing as an Indicator of Productivity on claypan soils. In *Site-Specific Management for Agricultural Systems*, American Society of Agronomy. Madison. USA.
- Tóth T.–Németh T.–Horváth E.–László P.–Bidló A.–Dér F.–Fekete M.–Fábián T.–Gaál Z.–Heil B.–Hermann T.–Kovács G.–Makó A.–Máté F.–Mészáros K.–Pásztor L.–Patocskai Z.–Speiser F.–Szűcs I.–Tóth G.–Várallyay Gy.–Vass J.–Vinogradov Sz. (2006): Új földminősítési rendszer megalapozása Magyarországon. III. Magyar Földrajzi Konferencia Tudományos Közleményei. MTA Földrajztudományi Kutatóintézet, 2006. szeptember 6–7. (CD).
- Wolfert, S.–Ge, L.–Verdouw, C.–Bogaardt, M. J. (2017): Big data in smart farming—a review. *Agricultural Systems*. 153: 69–80.