



Interpolation of Data Measured by Field Harvesters: Deployment, Comparison and Verification

Tomáš Řezník, Lukáš Herman^(✉), Kateřina Trojanová,
Tomáš Pavelka, and Šimon Leitgeb

Department of Geography, Faculty of Science, Masaryk University, Kotlářská 2,
611 37 Brno, Czech Republic

tomas.reznik@sci.muni.cz, {herman.lu, ktrojanova,
pavelka.tomas, leitgeb}@mail.muni.cz

Abstract. Yield is one of the key indicators in agriculture. The most common practices provide only one yield value for a whole field according to the weight of the harvested crop. On the contrary, precision agriculture techniques discover spatial patterns within a field to minimise the environmental burden caused by agricultural activities. Field harvesters equipped with sensors provide more detailed and spatially localised values. The measurements from such sensors need to be filtered and interpolated for the purposes of follow-up analyses and interpretations. This study verified the differences between three methods of interpolation (Inverse Distance Weighted, Inverse Distance Squared and Ordinary Kriging) derived from field sensor measurements that were (1) obtained directly from the field harvester, (2) processed by global filters, and (3) processed by global and local filters. Statistical analyses evaluated the results of interpolations from three fully operational Czech fields. The revealed spatial patterns, as well as recommendations regarding the suitability of the interpolation methods used, are presented at the end of this paper.

Keywords: Data filtering · Field harvester · Interpolation · Inverse Distance Squared · Inverse Distance Weighted · Ordinary Kriging · Yield mapping

1 Introduction

The main goals of precision agriculture (or precision farming) generally include the minimisation of negative environmental impacts on the one hand, and the maximisation of economic profit on the other hand [3, 17]. Geospatial information is highly valuable for these purposes [24, 25], in particular when based on Semantic web principles [16]. A differentially corrected Global Navigation Satellite Systems (GNSS) equipped yield monitoring system on field harvesters enables collection of georeferenced yield data [10, 18]. These data can be processed within Geographic Information system (GIS) using several interpolation techniques in order to generate detailed yield maps [5, 19]. On their basis, farmers can more precisely determine where exactly to put which inputs and in what quantities, because data from field harvesters represent the most detailed source of yield information. Unfortunately, these measurements usually contain errors

which influence results of complex spatial analyses [7]. As suggested for example by [2], as well as [4], such errors might arise for the following reasons: the occurrence of unexpected events during the harvesting process leading to unusual behaviour on the part of the machine; the trajectory of the field harvester; and errors caused by the wrong calibration of the yield monitor. Therefore, the measurements need to be processed and filtered (different types of filters can be used; see Sect. 2.2 for more details).

Other aspects requiring further investigation include the influence of the individual interpolation methods on the quality of the resulting yield maps. The objective of this paper is to investigate the influence of three interpolation methods (Inverse Distance Weighted, Inverse Distance Squared and Ordinary Kriging) commonly used to generate of yield maps. To compare them, we used descriptive statistics, Mean Prediction Error, Root Mean Square Prediction Error and Map Algebra. Interpolation methods were applied on three fields from Rostěnice Farm (Czech Republic). See details in Sect. 2.1.

2 Materials and Methods

2.1 Study Site

Data measured by a cereal field harvester were used to analyse and evaluate the approaches of spatial filtering and interpolation. Data acquisition was conducted at the Rostěnice cooperative farm in the south-eastern part of the Czech Republic (Fig. 1). The farm, Rostěnice a.s. (N49.105 E16.882), manages over 8,300 ha of arable land in the South Moravian region of the Czech Republic (see Fig. 2).

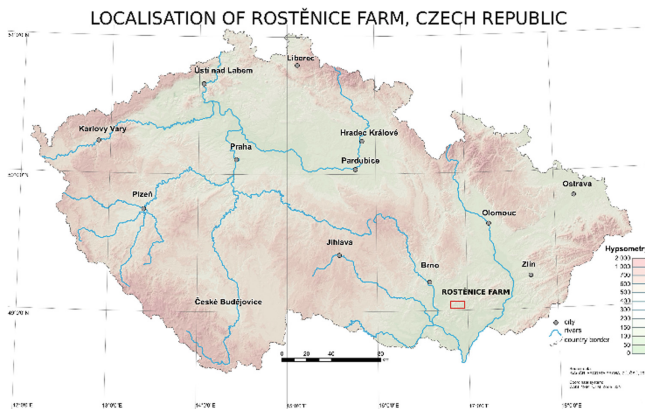


Fig. 1. Overview map of the Rostěnice farm.

The average annual temperature is 8.8 °C and the average annual rainfall is 544 mm. Within the managed land, the most prevalent soil types consist of Chernozem, Cambisol, haplic Luvisol, Fluvisol near bodies of water, and, occasionally, also Calcic

Leptosols. The fields are located mainly in sloping terrain. The main programme consists in plant production, where the main focus is on the cultivation of malting barley, maize for grain and biogas production, winter wheat, oilseed rape, and other crops and products such as soybean and lamb. The high spatial variability of soil conditions in the southern part of farm has led to the adoption of precision farming practices, such as the variable application of fertilisers (since 2006) and crop yield mapping by field harvesters (since 2010).

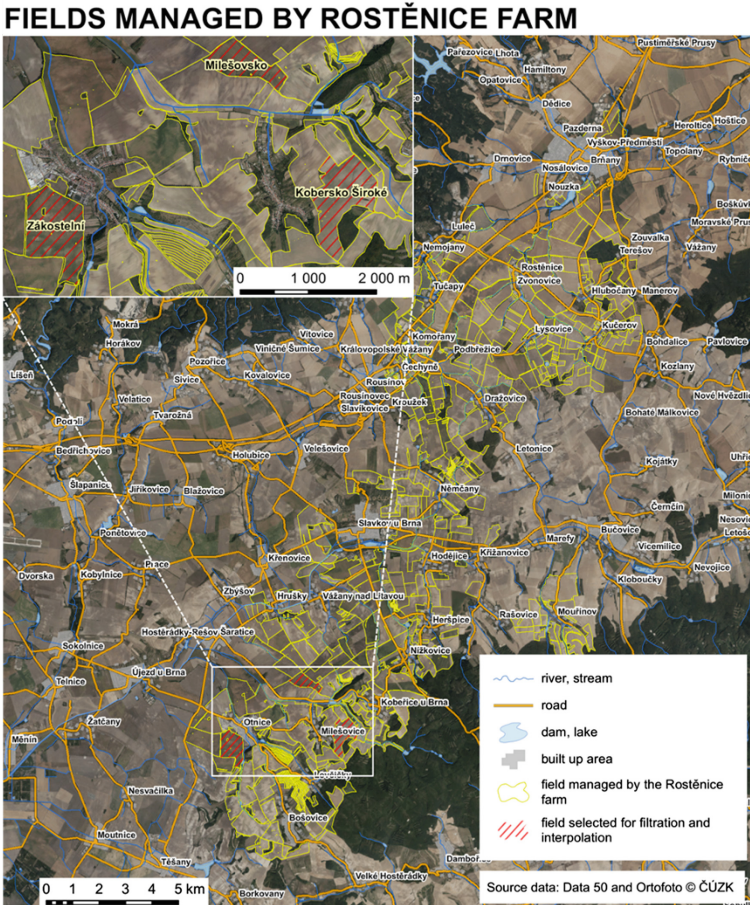


Fig. 2. Detailed map of the Rostěnice farm with the studied fields highlighted.

2.2 Sensor Measurement and Processing

Data were measured for three fields by a CASE IH AXIAL FLOW 9120 field harvester equipped with an AFS Pro 700 monitoring unit in 2017. The measurements were of GNSS-RTK (Real Time Kinematics) quality. Measurements were taken continuously

each second at an average speed of $1.55 \text{ m}\cdot\text{s}^{-1}$, recommended as optimal at the Ros-
těníce farm for cereal harvesting by the CASE IH AXIAL FLOW 9120 harvester. The
harvesting width was 9.15 m. The monitoring of yield was conducted as on-the-go
mapping by recording grain flow and moisture continuously over the whole plot area.
The crop type did not influence the spatial density of the performed measurements. The
measurements were stored directly in the field harvester and manually copied to a USB
flash drive after the end of operations on the pilot fields.

As mentioned above, data from field harvesters contain different types of errors.
These data errors corrupt the results, which means the datasets need to be processed
and filtered. Filtering of data from field harvesters was described, e.g., by [6, 12, 22].
We used the approach introduced [20]. This approach comprises two subsequent steps
– global filtering and local filtering. Global filtering removes non-reliable measure-
ments (data point values) within the whole dataset by means of a statistical analysis of
measurement values and related attributes. Local filtering then focuses on some parts of
the dataset in a higher detail (Table 1), and it is mostly based on the analysis of the
neighbourhood of data point values.

In the approach used [20], global filters detect incorrect outliers based on:

- the range of possible yield values,
- the speed of a field harvester,
- the direction of harvesting.

Local filtering brings the most accurate results regarding domain knowledge, e.g.
measurements, data processing and yield history, as well as knowledge of the data, of
the situation, and of the whole range of issues in general. Local filtering comprises a set
of subjective methods (points are excluded manually). In the approach used [20] local
filters identify potentially incorrect values when the following situations occur in
datasets:

- the crossings in harvester trajectory,
- the neighbouring rows in harvesting trajectory are too close to each other,
- the gaps in measurements within one row of trajectory.

Table 1. Absolute number of sensor measurements, relative percentage of data points after
global and local filtering.

Name	Measured data		Global filtering		Global and local filtering	
	Points	%	Points	%	Points	%
Zákostelní	28 658	100.00	24 509	85.52	23 877	83.32
Milešovsko	8 406	100.00	6 652	79.13	6 331	75.32
Kobersko Široké	18 462	100.00	13 498	73.11	12 817	69.42

2.3 Interpolation Methods

There are many spatial interpolation methods applied in various environmental-related disciplines and many diverse factors affect the performance of these methods [13]. From the wide range of interpolation methods mentioned by [13], only a few are used in precision agriculture to process field harvester data. [11] used punctual and block kriging. [21] applied IDW in addition to the aforementioned two methods. [23] compares Ordinary Kriging (OK), Inverse Distance Weighted (IDW) and Inverse Distance Squared (IDS). We decided to test these three interpolation methods, firstly because they are used in precision agriculture and, secondly, because they are the most commonly used interpolation methods in environmental sciences (see [13]).

IDW is a popular deterministic method for spatial analysis. IDW is multivariate interpolation, which means it is a function for more than one variable. The method is used to predict the unknown points in specific locations. The unknown values are calculated as a weighted average from the known available values [15]. Inverse Distance Squared (IDS) is very similar to IDW, but power of IDW is equal to 1 and power of IDS is equal to 2. IDW is referred to as linear interpolation, while equation of IDS is squared. IDW and IDS are deterministic methods [1, 23].

Table 2. Basic settings of the OK interpolation for the selected fields.

	Zákostelní			Milešovsko			Kobersko Široké		
	Measured data	Global filtering	Global and local filtering	Measured data	Global filtering	Global and local filtering	Measured data	Global filtering	Global and local filtering
Lag size	0.000354	0.000619	0.000875	0.001199	0.000033	0.001369	0.000067	0.001523	0.01493
Number of lags	12	12	12	12	12	12	12	12	12

Kriging – also known under the acronym BLUE (Best Linear Unbiased Estimator) – is a stochastic (geostatistical) interpolation method which uses geostationary estimation methods. The interpolated values are calculated by a Gaussian process and controlled by covariances. Local estimate is used for kriging, which means that the expected values of the variable are calculated from available data in a relatively small neighbouring area. There are several types of kriging, where ordinary kriging is the most frequently used one. Ordinary Kriging (OK) is spatial interpolation, where the error variance is minimised [8]. OK provides estimate values in points or in blocks for which a variogram is known. Data in the neighbourhood of the predicted value are used for the estimate [26]. An alternative to OK that is another variant of kriging and that can be used to process yield data is, for example, Simple Kriging (see [20]).

The three algorithms mentioned above represent both deterministic (IDW, IDS) and stochastic methods (OK). IDS and IDW methods can be considered simpler in terms of setting their input parameters. All three interpolations were computed in ArcGIS 10.6 software. The parameters were computed by means of Exploratory Spatial Data Analysis.

2.4 Verification Methods

The differences between observed and calculated values can be evaluated using simple descriptive statistics, Mean Prediction Error, Root Mean Square Prediction Error and Map Algebra. Within the methods of descriptive statistics, we investigated the mean, and especially the minimum and maximum, to determine whether the studied interpolation methods overestimate or underestimate the results compared to the original data.

Mean Prediction Error (MPE) is the average difference between the measured and the predicted values (Formula 1). The error values of the estimates should be impartial and their average should be zero [9].

$$MPE = \frac{\sum_{i=1}^n (\hat{Z}(s_i) - z(s_i))}{n} \quad (1)$$

Where:

$\hat{Z}(s_i)$: predicted value,

$z(s_i)$: measured value,

n : number of observations.

Root Mean Square Prediction Error (RMSE) is the standard deviation of the prediction errors. It is the square root of the average of squared differences between predicted and measured values (see Formula 2). Smaller value of the RMSE means the model is more suitable, because the calculated values are closer to the measured values [9].

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{Z}(s_i) - z(s_i))^2}{n}} \quad (2)$$

Where:

$\hat{Z}(s_i)$: predicted value,

$z(s_i)$: measured value,

n : number of observations.

To visualise and describe spatial patterns of the differences between interpolation algorithms, Map Algebra was used. Map Algebra and especially the relative difference method provide insights into variances between values in overlapping raster data [14]. The relative differences (d_v) were used as defined in Formula 3. Relative differences are often used as a quantitative indicator of quality assurance and quality control for repeated measurements/calculations where the results are expected to be similar or the same.

$$d_v = \frac{i_a - i_b}{i_a} \times 100 \quad (3)$$

Where:

- i_a : reference interpolated value,
- i_b : compared interpolated value.

All these verification methods were computed in ArcGIS 10.6 software. Verification methods were calculated for all three studied fields (Zákostelní, Milešovsko and Kobersko Široké – see Fig. 2) and interpolated surfaces from all three steps of data filtration (Measured data; Global filtering; Global and Local filtering).

3 Results

Interpolations were made for all datasets, with nine different interpolated surfaces for each field (see the example of Kobersko Široké field provided in Fig. 3). Relative yield values were used for more suitable comparison. The size of the pixels of the interpolated surfaces was 3.5×3.5 m. OK parameters are presented in Table 2. For filtered data, spatial extents were smaller due to the filtration process, which removes measurement points at the edges of fields. In order to achieve homogeneous (consistent) and comparable results, we decided not to use extrapolation methods because their precision in the respective areas would be debatable.

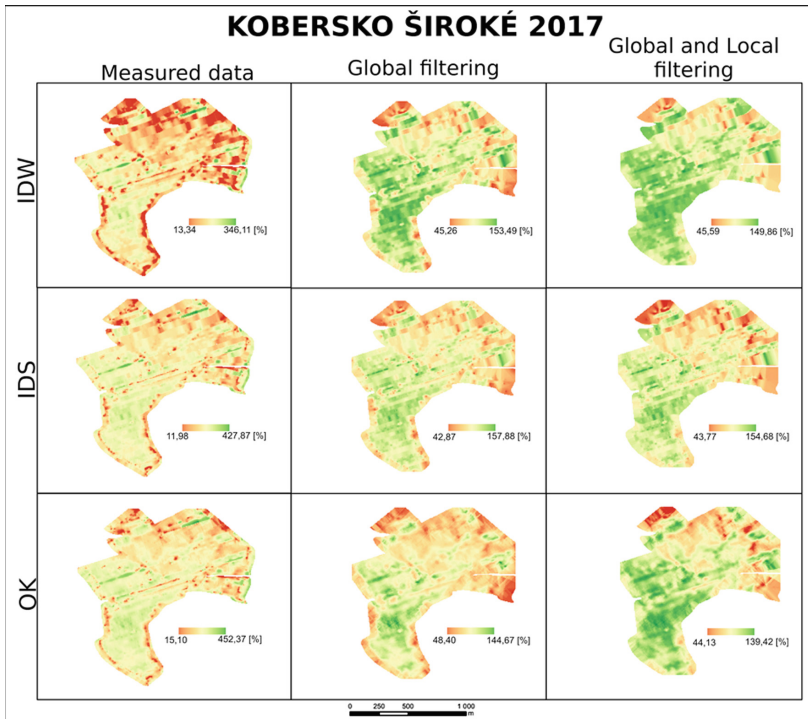


Fig. 3. Interpolations of measured, globally filtered, and both globally and locally filtered data for the Kobersko Široké field [relative yield values in %].

As the first step in comparison of the interpolation algorithms used, descriptive statistics were calculated. Figure 4 shows the mean values and the total range of values, both for the input data and the interpolated surfaces. At the same time, it is clear that especially global filtering removes extreme outliers and considerably reduces the overall range of values (Fig. 4 – compare the columns Measured data and Global filtering). A similar but much smaller effect in reducing the range to the measured data is also detectable in all three interpolation algorithms used (Fig. 4 – compare the columns Input data and the others).

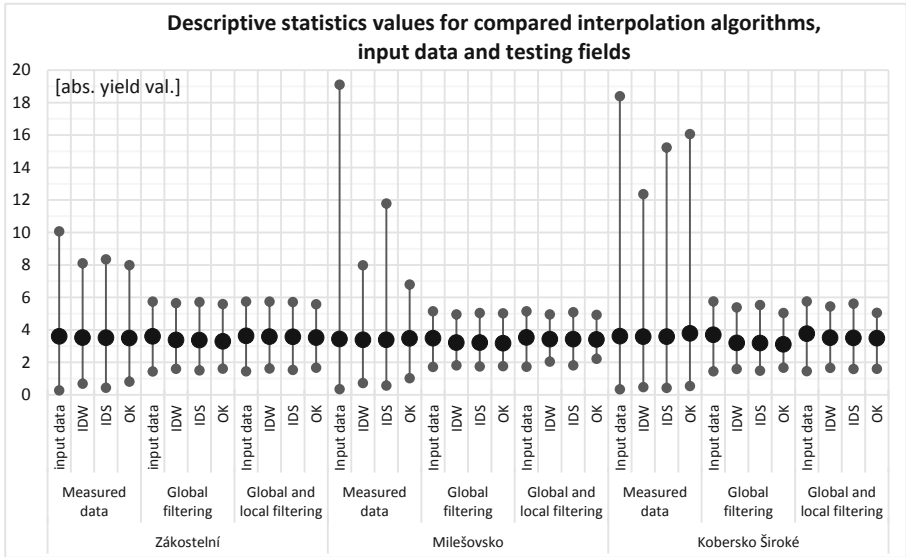


Fig. 4. Minimal, mean and maximum values for three compared interpolation algorithms, input data and three testing fields [absolute yield values].

The second step in comparison of the interpolation algorithms used was based on MPE and RMSE calculation. Results are presented in Fig. 5. Both MPE and RMSE express average model prediction error, but RMSE also expresses extreme errors.

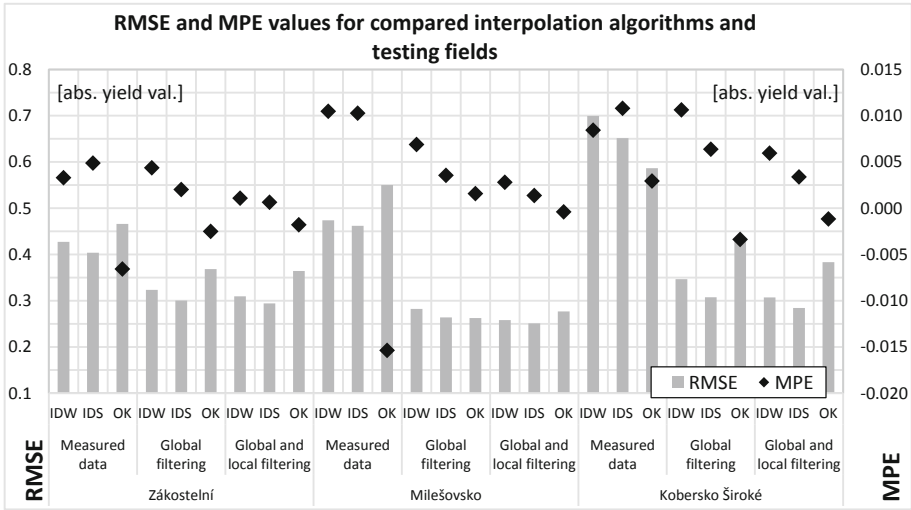


Fig. 5. RMSE and MPE values for three compared interpolation algorithms and three testing fields [absolute yield values].

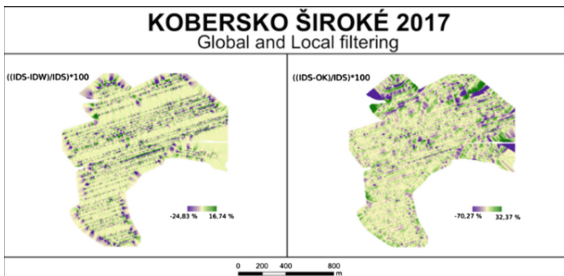


Fig. 6. Relative differences between IDS and IDW (left) and IDS and OK (right) for both globally and locally filtered data from Kobersko Široké field [in %].

Figure 6 depicts the relative differences between the compared interpolation methods. We chose the Kobersko Široké field for this visualization, because there are evident differences between interpolation algorithms (see Fig. 5). Figure 6 shows also the spatial pattern of the calculated relative differences. Relative differences occur in both pairs of compared algorithms at the field edges. However, the biggest relative differences exist between OK and IDS in some field passes; this is since IDS is an exact interpolator, which gives the most weight to near input points, while OK is rather a spatial estimator.

4 Discussion

When comparing individual interpolation algorithms (IDW, IDS and OK), we focus primarily on their applicability within the three above-mentioned steps of processing data from field harvesters (measured data, global filtering and local filtering). The analysis of differences between fields in which algorithms have been compared lies beyond the scope of this paper as these differences are influenced by multiple natural and artificial factors (i.e., field shape, topography, soil, water regime, fertilisation, harvesting strategy). A set of various methods have been used in the comparison: descriptive statistic, MPE, RMSE and Map Algebra. RMSE is the most commonly used method for comparing the individual interpolation methods. However, a combination of multiple methods provides a more comprehensive comparison, although it cannot be interpreted as unambiguously as the result of a single indicator. [21] also reach similar conclusions.

It seems that IDW and IDS are more suitable for interpolation of directly measured (unfiltered) data. For the data used in this paper, IDS is characterised by the fact that the average, minimum and maximum values in the interpolated surface are closer to the input data. IDS is also characterized by lower RMSE values. These differences can be partially explained by the number of neighbouring measurement points used in the interpolations. For example, IDS increases weight of the nearest measurement points and, therefore, it essentially reduces the number of measurement points used in the interpolation, thus better capturing the local variability in the data. Similar conclusions are also provided by [23].

When comparing interpolation algorithms for filtered data (for globally filtered and especially globally and locally filtered data), the differences between IDW, IDS and OK have been considerably smaller. Also, with regard the filtered data interpolations, the lowest RMSE values were achieved for IDS. Regarding MPE it was also closest to zero on all three fields. An interesting fact is that MPE values for OK were less than zero in all three fields, which shows that OK tends to underestimate the interpolated results. In general, ordinary kriging acted as a spatial estimator, rather than an exact interpolator [21].

Essential settings for IDW/IDS calculation are simpler than the settings necessary for OK interpolation. Basically, it is necessary to set the cell size of the resulting raster and the power value (1 for IDW; 2 for IDS). Thus, we can conclude that IDW and IDS are more suitable for less experienced users of the GIS technology such as farmers. OK is more difficult to set up, but these additional parameters have a positive effect on the accuracy of the interpolation, especially if there are large gaps without measurement points in the input datasets appearing, e.g., when yield data are filtered.

The obvious limit of this research certainly lies in the fact that we have tested only a limited number of interpolation algorithms (and their settings). However, these interpolation algorithms are also used by other authors for yield data processing [11, 21, 23] and, according to [13], IDW, IDS and OK are generally the most commonly used interpolation algorithms in the environmental sciences.

5 Conclusions and Future Work

This paper compared three interpolation techniques in terms of their usability in yield mapping: Inverse Distance Weighted, Inverse Distance Squared and Ordinary Kriging. The measurements from a field harvester equipped with a GNSS unit need to be filtered and interpolated for follow-up analyses. This study verified the differences between the three aforementioned methods of interpolation used on data that were derived from field sensor measurements. These measurements were (1) obtained directly from the field harvester, (2) processed by global filters, and (3) processed by both global and local filters. Statistical analyses evaluated the results of interpolations from three fields (Zákostelní, Milešovsko and Kobersko Široké) cultivated by a fully operational farm (Rostěnice, the Czech Republic).

So far, existing approaches evaluated positional accuracy only with respect to the Root Mean Square Prediction Error. The outcomes of this study have confirmed that a different interpolation method has to be chosen when taking into account: (1) solely the Root Mean Square Prediction Error, or (2) a combination of Mean Prediction Error, Root Mean Square Prediction Error, in combination with descriptive statistics and Map Algebra. In general, Inverse Distance Squared seems to be the most suitable interpolation method, especially when it comes to interpolating unfiltered data. Both Inverse Distance Squared and Inverse Distance Weighted methods are exact interpolators and it is relatively easier to define their input parameters. Ordinary Kriging appears to be relatively the least suitable, but its importance grows especially when a dataset contains missing/removed measurements. However, as in previous studies, no universally valuable advice could be given as to which interpolation method is the best. More general recommendations, but only for yield mapping domain, can be formulated after additional testing on multiple datasets from different fields with different crops etc.

The conducted study will also serve as a resource for further research that will attempt to compare the measured yield with that predicted based on yield productivity zones. Interpolated surfaces serve as the yield productivity zones for both measurement and prediction. The resulting interpolated surfaces are considerably influenced by the applied interpolation method. Interpolation algorithms and their settings therefore influence the evaluation of yield productivity predictions when confronted with the measured values.

Acknowledgments. This paper is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 818346 titled "Sino-EU Soil Observatory for Intelligent Land Use Management" (SIEUSOIL). Kateřina Trojanová, Tomáš Pavelka and Šimon Leitgeb were also supported by funding from Masaryk University under grant agreement No. MUNI/A/1576/2018. The authors would like to thank all persons from the Rostěnice Farm who participated in the study.

References

1. Almasi, A., Jalalian, A., Toomanian, N.: Using OK and IDW methods for prediction the spatial variability of a horizon depth and OM in Soils of Shahrekord, Iran. *J. Environ. Earth Sci.* **4**(15), 17–27 (2014)

2. Arslan, S., Colvin, T.S.: Grain yield mapping: yield sensing, yield reconstruction, and errors. *Precision Agric.* **3**(2), 135–154 (2002)
3. Auernhammer, H.: Precision farming – the environmental challenge. *Comput. Electron. Agric.* **30**(1–3), 31–43 (2001)
4. Blackmore, S., Moore, M.: Remedial correction of yield map data. *Precision Agric.* **1**(1), 53–66 (1999)
5. Charvat, K., et al.: Advanced visualisation of big data for agriculture as part of databio development. In: *IEEE International Geoscience and Remote Sensing Symposium*, pp. 415–418 (2018)
6. Gozdowski, D., Samborski, S., Dobers, E.S.: Evaluation of methods for the detection of spatial outliers in the yield data of winter wheat. *Colloquium Biometricum* **2010**(40), 41–51 (2010)
7. Hoskova-Mayerova, S, Talhofer, V., Hofmann, A., Kubicek, P.: Spatial database quality and the potential uncertainty sources. In: *Studies in Computational Intelligence*, pp. 127–142 (2013). https://doi.org/10.1007/978-3-642-32903-6_10
8. Huisman, O., By, R.A.: *Principles of Geographic Information Systems: An Introductory Textbook*, 4th edn. International Institute for Geo-Information Science and Earth Observation, Enschede (2009)
9. Krivoruchko, K.: *Spatial Statistical Data Analysis for GIS Users*, 1st edn. ESRI Press, Redlands (2011)
10. Kubicek, P., Kozel, J., Stampach, R., Lukas, V.: Prototyping the visualization of geographic and sensor data for agriculture. *Comput. Electron. Agric.* **97**(9), 83–91 (2013)
11. Lee, K.H., Chung, S.O., Choi, M.-C., Kim, Y.-J., Lee, J.-S., Kim, S.-K.: Post processing software for grain yield monitoring systems suitable to Korean full-feed combines. In: *Proceedings of the 13th International Conference on Precision Agriculture*, pp. 1–15 (2016)
12. Leroux, C., Jones, H., Clenet, A., Dreux, B., Becu, M., Tisseyre, B.: A general method to filter out defective spatial observations from yield mapping datasets. *Precision Agric.* **19**(5), 789–808 (2018)
13. Li, J., Heap, A.D.: A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors. *Ecol. Inform.* **6**(3–4), 228–241 (2010)
14. Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W.: *Geographic Information Science and Systems*, 4th edn. Wiley, Hoboken (2015)
15. Lu, G.Y., Wong, D.W.: An adaptive inverse-distance weighting spatial interpolation technique. *Comput. Geosci.* **34**(9), 1044–1055 (2008)
16. Palma, R., Reznik, T., Esbrí, M., Charvat, K., Mazurek, C.: An INSPIRE-based vocabulary for the publication of agricultural linked data. In: Tamma, V., Dragoni, M., Gonçalves, R., Ławrynowicz, A. (eds.) *OWLED 2015. LNCS*, vol. 9557, pp. 124–133. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-33245-1_13
17. Reznik, T., et al.: Open farm management information system supporting ecological and economical tasks. In: Hřebíček, J., Denzer, R., Schimak, G., Pitner, T. (eds.) *ISESS 2017. IAICT*, vol. 507, pp. 221–233. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-89935-0_19
18. Reznik, T., et al.: Monitoring of in-field variability for site specific crop management through open geospatial information. *ISPRS Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **XLI-B8**, 1023–1028 (2016)
19. Reznik, T., et al.: Disaster risk reduction in agriculture through geospatial (big) data processing. *ISPRS Int. J. Geo-Inf.* **6**(8), 1–11 (2017)

20. Reznik, T., Pavelka, T., Herman, L., Leitgeb, S., Lukas, V., Sirucek, P.: Deployment and verifications of the spatial filtering of data measured by field harvesters and methods of their interpolation: Czech cereal fields between 2014 and 2018. *Sensors* **19**(22), 1–25 (2019)
21. Robinson, T.P., Metternicht, G.: Comparing the performance of techniques to improve the quality of yield maps. *Agric. Syst.* **85**(1), 19–41 (2005)
22. Spekken, M., Anselmi, A.A., Molin, J.P.: A simple method for filtering spatial data. In: 9th European Conference on Precision Agriculture, pp. 259–266 (2013)
23. Souza, E.G., Bazzi, C.L., Khosla, R., Uribe-Opazo, M.A., Reich, R.M.: Interpolation type and data computation of crop yield maps is important for precision crop production. *J. Plant Nutr.* **39**(4), 531–538 (2016)
24. Stampach, R., Kubicek, P., Herman, L.: Dynamic visualization of sensor measurements: context based approach. *Quaestiones Geographicae* **34**(3), 117–128 (2015)
25. van Wart, J., Kersebaum, K.C., Peng, S., Milner, M., Cassman, K.G.: Estimating crop yield productivity zones at regional to national scales. *Field Crops Res.* **143**(1), 34–43 (2013)
26. Wackernagel, H.: Ordinary kriging. In: *Multivariate Geostatistics*, pp. 74–81 (1995)