Predicting the probability that higher profits could be achieved by adopting PA

Conference or Workshop Item

Accepted Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).
Published version at: https://doi.org/10.3920/978-90-8686-916-9
To link to this article DOI: http://dx.doi.org/10.3920/978-90-8686-916-9_113

www.reading.ac.uk/centaur

Producing grain yield maps by merging combine harvester and remote sensing data

M. Karampoiki[1], A. Heiß[1], G.M. Sharipov[1], S. Mahmood[2], L.C. Todman[2], A.J. Murdoch[2], H.W. Griepentrog[1], D.S. Paraforos[1]
[1]University of Hohenheim, Institute of Agricultural Engineering, Technology in Crop Production, Garbenstr. 9, 70599 Stuttgart, Germany
[2]University of Reading, Agriculture, Policy and Development, Earley Gate, PO Box 237, Reading RG6 6EU, United Kingdom
maria.karampoiki@uni-hohenheim.de

Abstract

High-quality combine harvester yield data are very important to produce yield maps. However, errors that derive from the combine harvester may be corrected. A data source that could be used to improve the yield maps obtained from harvester data is satellite-based data (e.g. Sentinel 2 (S2) imageries) and unmanned aerial vehicles (UAV) based data that are commonly used to predict grain yield and also define fertilisation rates. In this study, yield data from a combine harvester, a multi-spectral camera mounted on a UAV and S2 images were collected to produce yield maps that offer more accurate representation. To calibrate and validate this method, biomass samples were acquired manually before harvesting while erroneous yield data were replaced by remote sensing data. A regression analysis between the ground-truth biomass samples and the filtered yield data had an R^2 value equal to 0.84, while the R^2 of the biomass with the yield data obtained by the proposed methodology was equal to 0.90.

Keywords: NDVI, Sentinel imagery, UAV data, yield data cleaning

Introduction

Accurate crop yield mapping is important as it provides information about the variability of yields at the field scale, mean yield values and helps for more productive and resilient agriculture. Reliable yield maps are also used to develop site-specific management strategies (Diker et al., 2004; Jin et al., 2017). However, yield maps are prone to error sources that mainly derive from combine harvester yield logging systems. Erroneous data need to be removed to ensure data quality (Arslan & Colvin, 2002; Blackmore & Moore, 1999). According to Lyle et al. (2013), those errors have been categorised into four groups: i) harvesting dynamics of the combine harvester; ii) continuous measurements of yield and moisture; iii) accuracy of the positioning system; and iv) harvester operator. During the process of filtering and removing these erroneous data, a high number of values are deleted making the creation of interpolated yield maps a challenging task.
Data related to remote sensing, such as unmanned aerial vehicles (UAV) data and Sentinel 2 (S2) data could be used to replace the missing values and improve the accuracy of the yield maps obtained. However, these data sources have some drawbacks due to weather conditions. Many studies have highlighted the role of S2 or UAV imagery to estimate the grain yield based on vegetation indices (Lambert et al., 2017). For instance, Hunt et al. (2019) focused on wheat yield mapping based on S2. The model was trained with the combine harvester data. On the other hand, Fu et al. (2020), estimated yield based on a multi-rotor UAV.

This study aims to develop a new methodology to produce grain yield maps combining three different sources of data (combine harvester data, S2 satellite and UAV based data). The novelty of the present work lies in the fact that the combination of these three aforementioned data sources is not commonly used in developing a yield map.

Materials and methods

To develop this methodology, five steps were followed: The first two steps were to filter and clean combine harvester yield data and filter remote sensing data. The next step was the identification of the time during the cultivation period with the best yield prediction from remote sensing data, correlating the yield with this prediction. The filtered data were replaced and validation was done with hand sampling measurements. These steps are being described more in detail in the following sections.

Experimental field
The field research was conducted in an experimental research farm at 'Ihinger Hof' of the University of Hohenheim (48°44'41.61"N, 8°55'26.42"E). The study area field (Riech Nord) of 11.5 ha was cultivated with winter wheat (*Triticum aestivum* L). Harvesting was carried out on July 29 2020 using a New Holland CRB 80 combine harvester (CNH Industrial N.V., London, UK). The yield logging system obtained 9448 measurements as can be seen in Figure 1, which were further analysed.

Remote sensing multispectral information
The remote sensing multispectral information used in this study was a combination of satellite and UAV obtained imagery. Earth observation imagery from S2 was obtained on March 27, April 26, May 6 and June 23, 2020. The data was downloaded from the European Space Agency (ESA) website[1]. From the 13 available bands, the necessary bands to calculate the normalised difference vegetation index (NDVI) (Eq. (1)) were the Band 4-Red (665 nm ± 30 nm), Band 5- Red edge (698 nm ± 713 nm), Band-8 near-infrared (NIR) (842 nm ± 115 nm) and Band 12 short-wave infrared (SWIR) (2100 nm ±

Figure 1. Satellite image with the experimental field and the yield logging positions from the combine harvester.

2280 nm). Both bands have a spatial resolution of 10 m. Using the software SNAP 8.0.0 (Sentinel Application Platform) and the plugin 'Sen2Cor' developed by ESA, an atmospheric correction was applied for the S2 images. Additionally, the normalized difference red edge index (NDRE) (Eq. (2)), the normalized difference water index (NDWI) (Eq. (3)) and the soil adjusted vegetation index (SAVI) (Eq. (4)), where the soil brightness correction factor (L) is equal to 0.5, calculated to compare to the NDVI values.

$$NDVI = \frac{NIR-Red}{NIR+Red} \tag{1}$$

$$NDRE = \frac{NIR-RE}{NIR+RE} \tag{2}$$

$$NDWI = \frac{NIR-SWIR}{NIR+SWIR} \tag{3}$$

$$SAVI = \frac{(NIR-Red)(1+L)}{NIR+Red+L} \tag{4}$$

UAV-based imagery was collected on March 31, April 15 and May 27, 2020, using a Matrice 100 UAV (DJI, Nanshan, Shenzhen, China Matrice 100), equipped with a compact Parrot Sequoia+ multispectral camera (Sensefly, Lausanne, Switzerland). The UAV with the camera flew 50 m above ground level. The camera provides a set of four bands: green (550 nm ± 40 nm); red (660 nm ± 40 nm); red edge (735 nm ± 10 nm); and near-infrared (790 nm ± 40 nm). The UAV-based data were analysed in Pix4Dmapper software (Lausanne, Switzerland) to produce NDVI, NDRE and SAVI maps. NDWI maps were not produced as the multispectral camera does not provide the SWIR band.

Ground-truth biomass measurements
In order to calibrate and validate the proposed methodology, ground-truth biomass samples were collected right before harvesting from 30 locations in the examined field. 65 per cent of the biomass samples were used to train the model and the rest 35 per cent to validate it. Each location corresponded to harvested crop plants that were within an area of 0.6 m × 0.8 m (Figure 2). The samples were collected in plastic bags and were dried at 60 °C until they reached a constant dry weight. The weight was converted to tonnes per hectare considering the sampling area.

Figure 2. Satellite image with highlighted locations of the acquired ground-truth biomass samples.

Yield data cleaning

The yield data that were acquired from the combine harvester were analysed using QGIS 3.10.12, a free and open-source geographic system. Initially, values that were outside the field boundaries or had the same sampling location were removed. The kriging interpolation method was applied to mask header cut width errors. Then, the field was separated into three zones to ease data post-processing After this, invalid values were removed (i.e. yield values that were under 0.5 t ha^{-1} or values that exceeded 15.3 t ha^{-1}. Additionally, parts, where partially-filled cutter bars were observed (single oriented trajectories), were deleted. Finally, values that were inconsistent with their adjacent points due to filling and emptying times and where the speed of the combine harvester changed were also removed.

Yield data replacement

To detect the date where the best yield prediction appears, correlations between NDVI, NDRE, NDWI, SAVI and biomass have been generated. In QGIS, NDVI, NDRE, NDWI and SAVI maps based on the S2 and UAV data were created from the period of March till June. The index values were extracted for the biomass samples using the extraction values to point tool. A scatter diagram was created to calculate R^2.

Regression Statistics

The analysis was based on the following criteria:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y} - y_i)^2} , ME = \frac{\sum_{i=1}^{n}(\hat{y} - y_i)}{n}$$

Where y_i: are the predicted values, $\hat{y}$: are the observed values and n: is the number of samples.

Results and discussion

Filtered yield data

The raw yield data obtained from the combine harvester, compared to the filtered yield data are presented in Figure 3. The percentage of the deleted values reached 48% resulting in 4535 valid loggings.

Vegetation indices and biomass correlations

As can be seen in Figure 4, the correlation coefficients between the biomass measurements and the vegetation indices increased from March till May. In June, it started to decrease. The indices which show higher correlations were NDVI and NDRE. NDVI and NDRE appeared the highest correlation on May 27 (R^2 equal to 0.85) and the lowest on March 27 (R^2 equal to 0.01 and 0.001 respectively). Regarding NDVI, on March 31, April 26, May 6, the R^2 was 0.02, 0.59, 0.67 and 0.74, respectively. On these dates, R^2 was 0.01, 0.61, 0.67, 0.79 for NDRE. In these cases, the highest correlation was achieved by the UAV-based data. The low correlation early in the season makes sense as in the first weeks of spring, the wheat crop plants are small and do not have high biomass to correlate with the biomass close to harvesting. On the other hand, late in the season,

and as the harvesting date approaches, due to the saturation effect, the multi-spectral information cannot identify with a high correlation, the variability in the biomass close to harvesting. This can explain the drop in the correlation at the end of June.
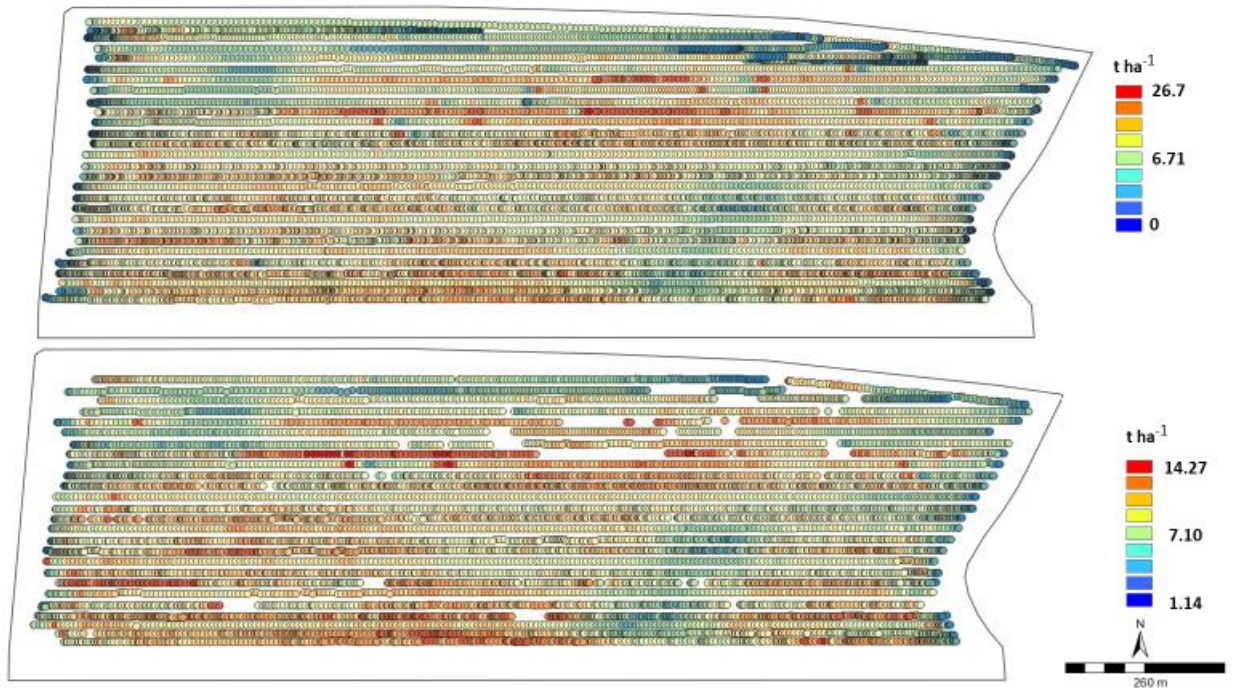

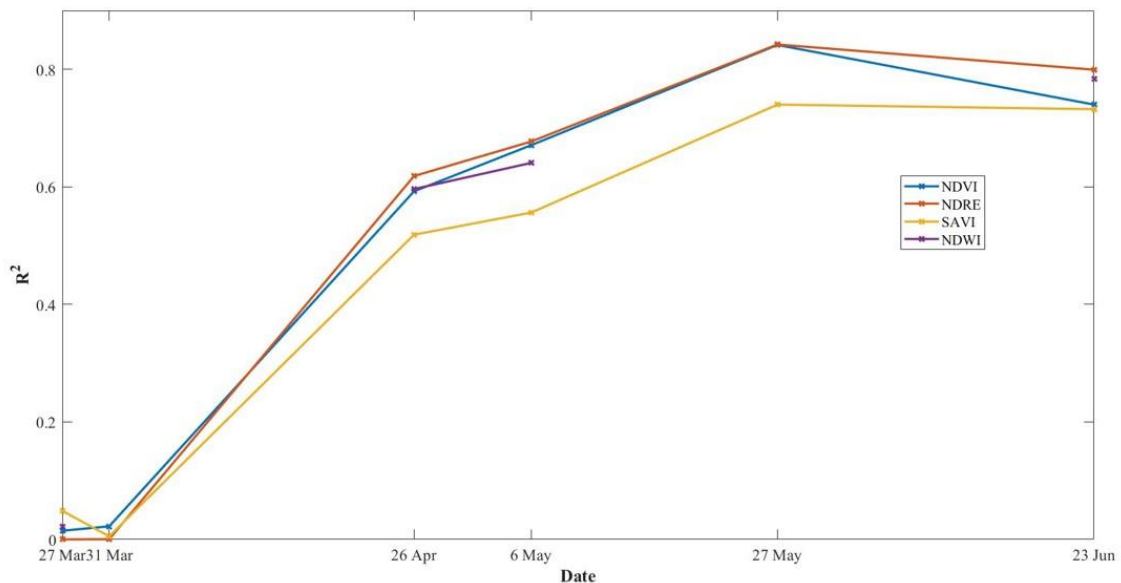Figure 3. (Top) combine yield logged data, and (bottom) filtered yield data.

Figure 4: Correlation between biomass and vegetation indices during the growing season.

Yield data replacement

Yield data replacement is based on the correlation between the NDVI on May 27, on June 23 and the ground-truth biomass measurements obtained, to compare the results between drone and S2. The biomass values were converted from kg m^{-2} to t ha^{-1}. Linear regression was used to identify the values where the filtered data could be replaced with the new

values. The linear regressions, which were used to replace the missing yield loggings in May and June can be seen in Figure 5. Even from this linear fitting, an error resulted as R^2 was equal to 0.85 and 0.82 respectively. However, it was under accepted limits since, in any other case, the filtered data had left missing information that was not helping the kriging spatial interpolation in creating a yield map.
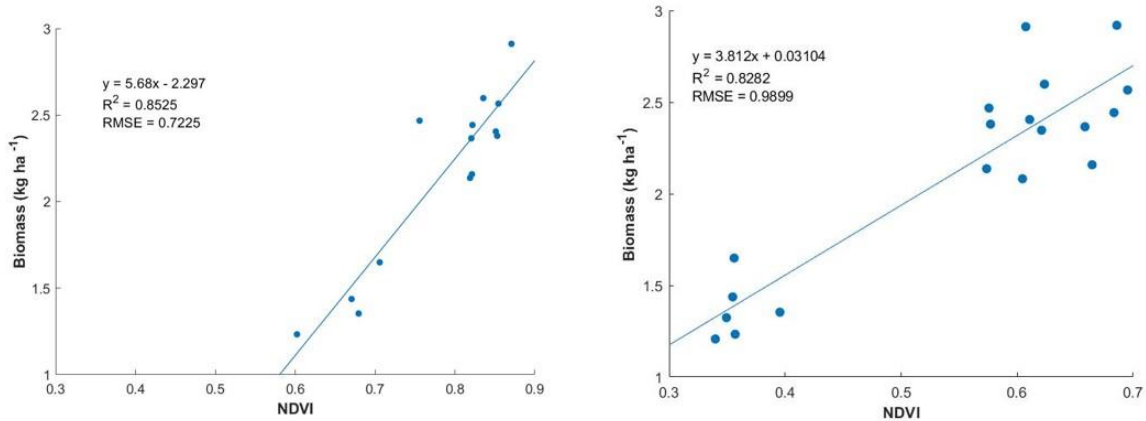

Figure 5. Linear regression analysis between the ground-truth biomass measurements and the NDVI on (left) May 27, 2020, and on (right) June 23, 2020, which was used for replacing the erroneous yield loggings.

Enhanced yield map

After calculating the new values, the join and relate tool was used in QGIS to identify the points based on their spatial location and replace the filtered yield values. The result of this method is presented in Figure 6. On top, the yield map based on the filtered data can be seen after performing ordinary kriging interpolation, while in the left and right the yield maps are produced by utilising the replaced values based on the proposed methodology and performing Kriging with an external drift (KED).
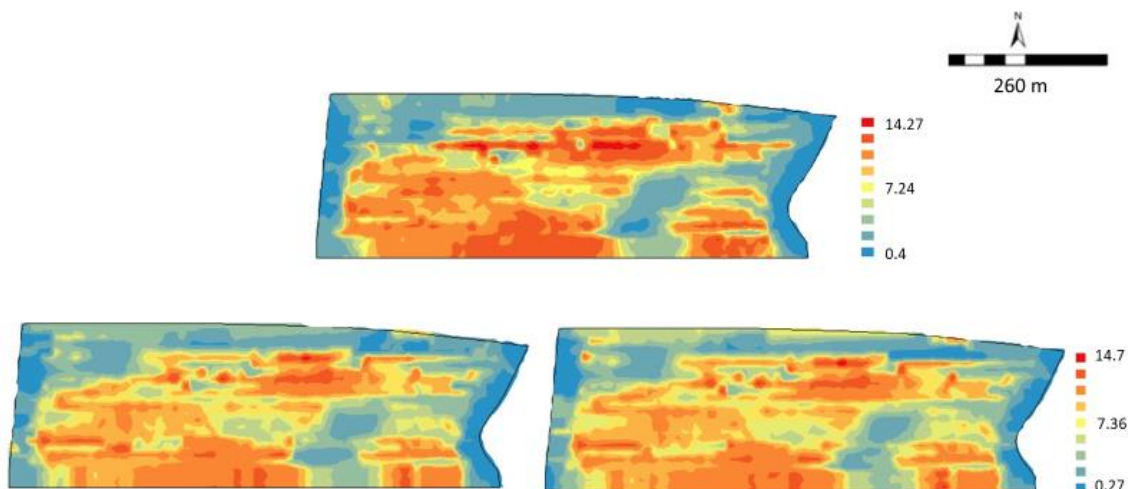

Figure 6. (Top) yield map based on the filtered data, (left) yield map based on replaced values in May, and (right) yield map based on replaced values in June.

After the creation of the yield maps, regression statistics were calculated to identify the improvement of the yield map. The yield values for the locations of the 35 per cent of ground-truth biomass measurements were extracted from the 3 maps in Figure 6. Table 1 shows the output summary of the regression statistics between the biomass values with new replaced and previous values respectively. The main result is that the value for R^2 increased from 0.84, for the map with the filtered data, to a value equal to 0.90 in May and decreased to 0.83 in June for the maps that were created by the proposed methodology. The RMSE decreased from 1.56 to 1.53 in May and increased to a value equal to 1.77 in June.

Table 1. Regression statistics of the three yield maps compared to the ground-truth biomass measurements.

Regression Statistics	Filtered yield data	Proposed methodology on May, 27	Proposed methodology on June, 23
R^2	0.84	0.90	0.83
RMSE	1.56	1.53	1.77
ME	0.63	0.61	0.82

Conclusion

Overall, this study explored the potential of using multi-source data consisting of combine harvester data and multispectral data from S2 and UAV to create grain yield maps by monitoring different vegetation indices and validating the results with 35% of the ground-truth biomass samples. The results showed that a strong correlation between the biomass samples and NDVI was in May and June. May 27 and June 23 were chosen as the best dates with the highest correlation. Based on these dates, a linear regression analysis was created to calculate the values from the given equations. The new values were calculated and the filtered data were replaced by them. New yield maps were created with the new values. Finally, regression statistics were calculated to compare the biomass values with the values from the combined yield logging data and the inserted values.

The presented methodology demonstrates the potential of combining data from three different sources and using KED method for kriging. As S2 images could have sometimes bad quality due to weather conditions, the UAV imagery could bridge this gap. Future work includes the value replacement based on the correlation of the obtained yield loggings, as ground-truth biomass measurements is a time-consuming task for the farmer.

Acknowledgements

technology-based products and services that will ultimately deliver a healthier and more sustainable lifestyle for all European citizens.

References

Arslan, S. and Colvin, T. 2002. Grain yield mapping: yield sensing, yield reconstruction, and errors. Precision Agriculture **3** 135-154.

Blackmore, B. S., and Moore, M. 1999. Remedial correction of yield map data. Precision Agriculture **1** 53 – 66.

Diker, K., Heermann, D.F., Brodahl, M.K., and Collins, F. 2004. Frequency analysis of yield for delineating yield response zones. Precision Agriculture **5** 435–444.

Fu, Z., Jiang, J., Gao, Y., Krienke, B., Wang, M., Zhong, K., et al. 2020. Wheat growth monitoring and yield estimation based on multi-rotor unmanned aerial vehicle. Remote Sensing **12** 508.

Hunt, M.L, Blackburn, G.A., Carrasco, L., Redhead, J.W. and Rowland C.S. 2019. High resolution wheat yield mapping using Sentinel-2, Remote Sensing of Environment **233** 111410.

Jin, Z., Azzari, G., and Lobell, D.B. 2017. Improving the accuracy of satellite-based high-resolution yield estimation: A test of multiple scalable approaches. Agricultural and Forest Meteorology **247** 207–220.

Lambert, M.J., Blaes, X., Traore, P.S., and Defourny, P. 2017. Estimate yield at parcel level from S2 time serie in sub-Saharan smallholder farming systems. In: 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), pp. 1-7.

Lyle, G., Bryan, B. and Ostendorf, B. 2013. Post-processing methods to eliminate erroneous grain yield measurements: review and directions for future development, Precision Agriculture **15** 377-402.