Processing Satellite Imagery for Decision-Making in Precision Agriculture

Panagiotis Bouros
Institute of Computer Science
Johannes Gutenberg University Mainz, Germany
bouros@uni-mainz.de

Ira Assent
Department of Computer Science
Aarhus University, Denmark
ira@cs.au.dk

Jacob Høxbroe Jeppesen
Department of Engineering
Aarhus University, Denmark
jhj@eng.au.dk

Thomas Skjødeberg Toftegaard
Department of Engineering
Aarhus University, Denmark
tst@eng.au.dk

ABSTRACT

Empowered by geo-locating and sensor-based technologies, precision agriculture brings a data-intensive paradigm into farming. In this spirit, we investigate the role of outlier detection and visualization in decision-making for precision agriculture. We discuss two analysis tasks for visually monitoring fields that exhibit problematic crop growth compared to their neighbors, and for visualizing problematic areas inside a field. As a proof of concept, we analyze satellite imagery to case-study our tasks in the context of the Future Cropping project.

1 INTRODUCTION

The proliferation of the Global Positioning System (GPS) and of sensor-based technologies have given rise to a different paradigm of agriculture. *Precision agriculture* as this paradigm is known, strikes to improve economic returns and reduce the environmental impact [4, 7–9]. Harvesters mounted with sensors and GPS receivers allow for instant collection of position-aware crop data. Contemporary Geographic Information Systems (GIS) are able to store and analyze collected crop data, in many cases combined with other types of information from multiple sources such as satellites and meteorological stations. Decision-making plays a key role in the effort of exploiting this abundance of data. Traditionally, decision-making heavily relies on farmers' experience and empirical knowledge. In contrast, in precision agriculture, information is quantified and decision-making is supported by a systematic and data-intensive analytical procedure that can be visualized for novel insights and improved farming decisions.

Anomaly or *outlier detection* has been widely studied in data mining to support decision-making; the goal is to identify objects that significantly deviate from the expected pattern in a dataset [1]. There exist essentially three types of ourliers. *Point* outliers generally studied in multidimensional data management, are objects inconsistent with the rest of the dataset. Detection of this type mainly involves statistical and distribution-based methods. *Contextual* outliers are defined based on a set of *contextual* and *behavioral* attributes; these objects exhibit significantly different values on their behavioral attributes compared to dataset objects with similar values on the contextual attributes. The key for contextual outlier detection is to identify those objects with similar contextual attributes also known as the *neighborhood* of an object. Spatial outliers [12] are a type of contextual outliers where the context is captured by object geometries and their distance in space. Finally, a *collective* outlier is a group of clustered objects with low variance among them, which are together inconsistent to the rest of the dataset.

In precision agriculture, outlier detection is mainly studied for pre-processing and cleaning of the collected data [13]. The focus has been on detecting errors introduced by measurement failures, e.g., related to the accuracy and the calibration of sensor and positioning devices, or even to signal loss. In contrast, anomalies directly related to poor decision-making, e.g., crop establishment, fertilization or herbicide application, and to natural causes including climate, topography, and soil-landscape features, have received less attention.

In an attempt to fill this gap, we propose the tasks of *inter-field* (i.e., between fields) and *intra-field* (i.e., within a field) analysis. In the former, our goal is to identify fields, which exhibit a problematic behavior, e.g., unexpectedly low collected crop mass, compared to their neighboring fields. [1] The key challenges for this task are twofold; (i) how to define field neighborhood and (ii) how to quantify the importance/influence of each neighbor. Different from traditional spatial outlier detection that relies solely on arbitrary polygon distances, we consider domain-specific data properties such as soil/crop type and the variation of crop-growth. On the other hand, intra-field analysis is triggered in an effort to interpret the results of inter-field analysis. The goal is to identify and visualize the areas inside a particular problematic field that significantly deviate from the average or normal behavior of the field. Practically, these are the areas where the farmer may need to take action [4, 7], e.g., by increasing fertilization, changing crop type or building drains. Overall, the proposed tasks are of great value not only to farmers but also to contractors and consultants of agriculture.

Recent studies [5, 6] showed how precision agriculture may benefit from processing *satallite imagery*. Imagery datasets have become readily available by open access to NASA Landsat in 2008 [14] and to ESA Sentinel satellites. In line with this trend, we conduct a case study for our inter- and intra-fields tasks in the context of the Future Cropping project. Our study employs satellite imagery in two manners. First, we calculate a vegetation index to monitor the crop growth on the fields. Second, our intra-field task properly partitions a field to take full advantage of the detail and the resolution of the available imagery data.

[1] Note that besides problematic, we can also identify fields that outperform their neighbors; in other words, we are able to detect both *negative* and *positive* outliers.

Table 1: Notation

symbol	description
α	Attribute of interest
$f.\alpha$	Value for the attribute of interest on field f
$f.\widehat{\alpha}$	Predicted value for the attribute of interest on field f
N_f	Neighborhood of field f
w_i	Weight of neighbor field n_i
P_f	Partitioning of field f
$p.\alpha$	Value for the attribute of interest on partition p
$dist(f_1, f_2)$	Spatial distance between fields f_1 and f_2
MAX_DIST	Maximum allowed distance between two fields

2 ANALYSIS TASKS

Besides spatial geometries (e.g., polygons and/or point data), our outlier detection methodology uses for every field (or parts of it) a set of *numerical* attributes that store measurements such as the collected crop mass, and a set of *categorical* attributes which store properties of the field such as its soil and crop type. Our analysis focuses on a particular numerical *attribute of interest* denoted by α, which monitors the crop growth on a field; Table 1 summarizes the notation used throughout the rest of this paper.

2.1 Inter-field Analysis

We model inter-field analysis as a spatial outlier detection task [12] with attribute of interest α as the *behavioral* attribute of our analysis while the geometry (polygon), the categorical attributes and the rest of the numerical attributes are the *contextual* attributes used to define the neighborhood around each field.

Conceptually, the process of inter-field analysis involves three steps. First, the neighborhood of each field f is defined on the basis of the contextual attributes. Under this, selecting the proper distance measure *dist* between two fields is critical; *dist* takes into account the spatial proximity of the fields and potentially the values of other contextual attributes, e.g., the soil type. Let N_f be the set of neighbors for a field f. The second step is to predict the value for the attribute of interest α on f by aggregating the values on f's neighbor fields in N_f; typically, the *mean* value is used for this purpose. [2] In our study, we calculate $f.\widehat{\alpha}$ as the *weighted mean* value, which allows us to weight the importance and the contribution of each neighbor, considering both the spatial distance to field f and other factors as discussed in Section 3.2. Formally, we have:

$$f.\widehat{\alpha} = \frac{\Sigma_{i=1}^{|N_f|}\{w_i \cdot n_i.\alpha\}}{\Sigma_{i=1}^{|N_f|} w_i} \tag{1}$$

where w_i denotes the weight of neighbor field n_i. The third step is to report the outlying fields. The value of $|f.\widehat{\alpha} - f.\alpha|$ quantifies the deviation of α's predicted value from the actual value on a field f. In practice, there exist two approaches for determining outliers. We can either rank the fields in decreasing order of their $|f.\widehat{\alpha} - f.\alpha|$ deviation and output the top k, or return all fields with an extreme deviation value. Probabilistic and statistical models can be used to determine such extreme deviation values [1].

The above procedure allows us to identify fields whose predicted value on attribute of interest α significantly deviates from the actual value. In fact, we distinguish between two types of outliers based on the sign of the $f.\widehat{\alpha} - f.\alpha$ deviation. If $f.\widehat{\alpha} - f.\alpha > 0$ then field f behaves worse than expected based on its neighborhood; we call f a *negative* outlier in this case. Otherwise, f is a *positive* outlier that exceeds our expectations. Both cases are important for decision-making in precision agriculture as negative outliers may require immediate action by the farmers and positive outliers may provide important feedback to improve crop growth in other fields.

[2] As an alternative, median can be used to smooth the impact of extreme α values.

2.2 Intra-field Analysis

Conceptually, the process of intra-field analysis involves three steps. First, the extent of each field is split into a number of *partitions*; each partition represents an area of interest. In practice, such a partitioning may for instance be provided by the farmer or generated by a partitioning process. For instance, in our case study of Section 3, we used the image resolution to partition our satellite data. Similar to the entire field, we assume that each partition is associated with the same attribute of interest α. [3] We next define $|p_i.\alpha - f.\alpha|$ as the deviation of α's value on a particular partition p_i, from the value on the entire field f. Last, we can employ the two approaches discussed in the previous subsection for the inter-field analysis to report the outlying areas of a field, i.e., either by ranking the partitions by their $|p_i.\alpha - f.\alpha|$ deviation value, or by reporting the partitions with an extreme deviation value.

3 CASE-STUDY

We case-studied our outlier detection tasks for the inter- and the intra-field analysis in the context of the Future Cropping project, which collects and analyzes agricultural data in Denmark. [4] Our study involves two types of data; (i) the geometries (i.e., the polygons) of all 590,490 fields in Denmark and their crop type, and (ii) a 6.3GB vegetation map derived from ESA Sentinel-2 satellite imagery captured on May 8, 2016. [5] We conducted our case-study on QGIS. [6]

3.1 Processing Satellite Imagery

In the past, a number of vegetation indices [2, 3, 10, 11] has been proposed to monitor the crop growth on fields, primarily by estimating properties of vegetation such as the concentration of biomass or chlorophyll. Similar to previous studies [5, 6], we consider the *Normalized Difference Vegetation Index* ($NDVI$) for our case-study. ESA Sentinel-2 satellites provide 13-band multispectral images with a spatial resolution down to $10m \times 10m$. $NDVI$ values can be calculated combining two particular spectral bands:

$$NDVI = \frac{NIR - VIR}{NIR + VIR} \tag{2}$$

where NIR and VIR are the near-infrared and the visible red bands, respectively.

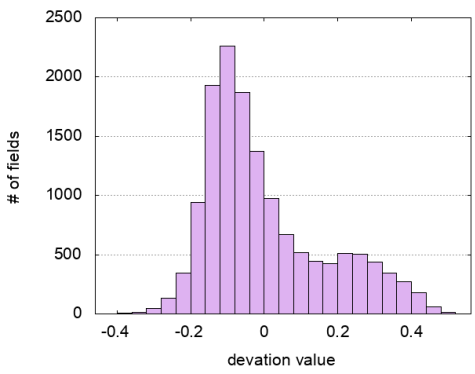
For our analysis, we use $NDVI$ as the attribute of interest α. Specifically, we first calculate the index value on every pixel of the satellite images using Formula 2. Then, for each field f, we compute *mean* μ_f and *standard devation* σ_f of its $NDVI$ values based on the overlapping pixels, and set $f.\alpha = \mu_f$. Figure 1b

[3] Essentially, the value of attribute α on both the entire field and a partition is computed by aggregation. For instance, harvesters store position-aware crop data for the collected mass on specific geographic locations inside a field; we can aggregate these values to compute the overall collected mass for the entire field or for particular areas.

[4] https://futurecropping.dk

[5] https://kortdata.fvm.dk/download/Index?page=Markblokke_Marker, https://scihub.copernicus.eu

[6] https://qgis.org



(a)

(b)

Figure 1: Inter-field analysis on winter rapeseed crop fields: (a) distribution of $f.\widehat{\alpha} - f.\alpha$ values with mean value $\mu_d = 0.0013$ and standard deviation $\sigma_d = 0.1645$, (b) example of a negative outlier (in red) with $f.\alpha = 0.176$ and $f.\widehat{\alpha} = 0.671$.

Table 2: Top-10 most important neighbors of f in Figure 1b.

field	α (*NDVI*)	$dist(f, n_i)$ (in meters)	σ_i	w_i
f	0.176	–	–	–
n_1	0.183	14	0.071	0.928
n_2	0.764	264	0.054	0.921
n_3	0.762	311	0.056	0.914
n_4	0.777	452	0.056	0.901
n_5	0.724	435	0.059	0.900
n_6	0.604	75	0.098	0.895
n_7	0.629	5	0.110	0.889
n_8	0.753	1,277	0.067	0.814
n_9	0.787	1,644	0.028	0.813
n_{10}	0.347	408	0.169	0.797

shows a snapshot of our *NDVI* map for the fields in Denmark; the greener a field (or parts of it) appears, the higher its $f.\alpha$ (i.e., *NDVI*) value is.

3.2 Inter-field Analysis

We now present our case-study for the inter-field analysis and exemplify its results. We apply Formula 1 on each field f in Denmark to predict its *NDVI* value, i.e., $f.\widehat{\alpha}$. For performance, previous works in outlier detection typically define neighborhood N_f as the k closest fields. However, these neighbors (or part of them) could be in fact very distant to field f, which will potentially affect the quality of the results. Instead, we narrow the extent of a neighborhood; hence, N_f contains all fields of the same crop type as f within a radius of $10km$, as suggested by our agricultural project partners. Formally, we define the distance of f to a neighbor field n_i as:

$$dist(f, n_i) = \begin{cases} ED(f, n_i) & \text{if fields } f, n_i \text{ have same crop type} \\ MAX\_DIST & \text{otherwise} \end{cases} \tag{3}$$

where $ED(f, n_i)$ is the Euclidean distance of fields f, n_i polygons and MAX_DIST is the maximum $10km$ distance.

To weight the importance of a neighbor $n_i \in N_f$, we take into account two factors; (1) its spatial distance $dist(f, n_i)$ to f and (2) the variation of crop growth inside the field. [7] Essentially, we prioritize neighbors located close to field f, with crops evenly grown on their entire extent. These are fields with a similar *NDVI* value (visually, with a similar shade of green) on the majority of their overlapping pixels; hence, low variance in crop growth for neighbor n_i can be captured by a low σ_i standard deviation of the *NDVI* values. Formally, weight w_i is defined as:

$$w_i = \left[1 - \frac{dist(f, n_i)}{MAX\_DIST}\right] \cdot [1 - \sigma_i] \tag{4}$$

With Formulas 1, 3 and 4, we compute the $f.\widehat{\alpha} - f.\alpha$ deviation of the predicted value from the actual $f.\alpha$ *NDVI* value, on every field f. As discussed in Section 2.1, one option for our inter-field analysis, would be to rank all fields by their deviation and recommend the ones with the k highest values as potential *negative* outliers and those with the k lowest deviation values as *positive* outliers. However, such recommendations have little practical merit for the end-user; for instance, even the highest $f.\widehat{\alpha} - f.\alpha$ value may in fact be too low to mark field f as problematic. Instead, we focus on identifying the fields that exhibit the most extreme deviation values.

Figure 1a reports on the distribution of the $f.\widehat{\alpha} - f.\alpha$ values for a particular type of crop called winter rapeseed. The mean value and the standard deviation for this distribution are $\mu_d = 0.0013$ and $\sigma_d = 0.1645$, respectively. We observe that over 99.5% of the fields have a $f.\widehat{\alpha} - f.\alpha$ value inside $[\mu_d - 3{\cdot}\sigma_d, \mu_d + 3{\cdot}\sigma_d]$. A similar conclusion was drawn for the rest of the crop types in our case-study. Based on this observation, we follow a typical statistical approach for identifying extreme $f.\widehat{\alpha} - f.\alpha$ values [12], namely those that deviate from mean μ_d at least three times the standard deviation σ_d, i.e., $\frac{|f.\widehat{\alpha} - f.\alpha - \mu_d|}{\sigma_d} \geq 3$ holds. [8] The nature of the outlier (positive or negative) is determined by the sign of $f.\widehat{\alpha} - f.\alpha$.

Figure 1b visually exemplifies our inter-field analysis for the winter rapeseed crop; the negative outlier field f is highlighted by a red border while a subset of its neighborhood is given in black. Further, Table 2 lists the characteristics of the 10 most important neighbors of f, sorted by their weight w_i (Formula 4); notice how the closest field to f, n_7, is not its most important neighbor as n_7 exhibits a higher variation of crop growth compared to fields n_1 to n_6. For field f, we have its actual *NDVI* value $f.\alpha = 0.176$ while using Formula 1, we predict that $f.\widehat{\alpha} = 0.671$. The extreme deviation of the predicted from the actual *NDVI* value is captured in Figure 1b where the majority of the shown neighbors are

[7] The importance of low-variation fields has been previously studied e.g., in [5].

[8] Note that μ_d, σ_d are statistics computed for the particular crop type of field f.

Figure 2: Intra-field analysis on a winter rapeseed crop field: (a) pixel grid partitioning and (b), (c) outlying partitions.

significantly greener than field f, indicating that f's actual $NDVI$ value should have been higher than 0.176.

3.3 Intra-field Analysis

We briefly discuss our case-study for the intra-field analysis. As mentioned in Section 2.2, the first step is to partition a field. For this purpose, we use the pixel grid of the satellite imagery in order to best exploit the $10m \times 10m$ resolution. Figure 2a illustrates the partitioning of a winter rapeseed crop field; each partition (grid cell) p_i corresponds to a pixel of the $NDVI$ map and its $p_i.\alpha$ is set to the $NDVI$ value of that pixel. For every partition p_i, we then compute the $|p_i.\alpha - f.\alpha|$ deviation of its $NDVI$ from the value of the entire field. Last, we employ the approach of Section 3.2 to find outliers. Figure 2b visualizes the results of our analysis; the $|p_i.\alpha - f.\alpha|$ value of the partitions drawn in the figure deviates from mean μ_d, at least 3 times the standard deviation σ_d. Observe how our analysis is able to report the partitions that cover the light colored problematic areas of the field. Additional problematic areas can be identified by reporting partitions with a $|p_i.\alpha - f.\alpha|$ value deviating 2 times, as shown in Figure 2c.

4 CONCLUSIONS

We studied the role of outlier detection in decision-making for precision agriculture and how such tasks can benefit from processing satellite imagery. For this purpose, we discussed the tasks of inter- and intra-field analysis, addressing the special challenges raised by the agricultural domain. As a proof of concept, we conducted a case-study supported by ESA satellite images within the Future Cropping project.

In the future, we plan to extend our study in multiple directions; (i) incorporate more types of data, e.g., collected from harvesters or rain distribution data, (ii) employ visualization tools to collect and evaluate farmers' feedback, (iii) enhance intra-field analysis for recommending explicit actions to farmers, (iv) further analyze satellite imagery to monitor the outlying fields or problematic areas through the course of time and (v) investigate techniques from Machine Learning for predictive analytics.

ACKNOWLEDGEMENTS

This work was funded by Innovation Fund Denmark as part of the Future Cropping project (J. nr. 5107-00002B).

REFERENCES

[1] Charu C. Aggarwal. 2013. *Outlier Analysis.* Springer. https://doi.org/10.1007/978-1-4614-6396-2

[2] Jan G.P.W. Clevers, Steven M. de Jong, Gerrit F. Epema, Freek van der Meer, Wim H. Bakker, Andrew K. Skidmore, and Elisabeth A. Addink. 2001. MERIS and the red-edge position. *International Journal of Applied Earth Observation and Geoinformation* 3, 4 (2001), 313 – 320.

[3] Anatoly A. Gitelson, Yoram J. Kaufman, and Mark N. Merzlyak. 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment* 58, 3 (1996), 289 – 298.

[4] Peter. Skøjdt, Poul Møller Hansen, and Rasmus Nyholm Jørgensen. 2003. *Sensor Based Nitrogen Fertilization Increasing Grain Protein Yield in Winter Wheat.* Technical Report Risø-R-1389. Risø National Laboratory, Roskilde, Denmark.

[5] Jacob Høxbroe Jeppesen, Rune Hylsberg Jacobsen, Rasmus Nyholm Jørgensen, Andes Halberg, and Thomas Skjødeberg Toftegaard. 2017. Identification of High-Variation Fields based on Open Satellite Imagery. *11th European Conference on Precision Agriculture.*

[6] Jacob Høxbroe Jeppesen, Rune Hylsberg Jacobsen, Rasmus Nyholm Jørgensen, and Thomas Skjødeberg Toftegaard. 2016. Towards Data-Driven Precision Agriculture using Open Data and Open Source Software. In *International Conference on Agricultural Engineering.*

[7] D. R. Kindred, A. E. Milne, R. Webster, B. P. Marchant, and R. Sylvester-Bradley. 2015. Exploring the spatial variation in the fertilizer-nitrogen requirement of wheat within fields. *The Journal of Agricultural Science* 153, 1 (2015), 25–41.

[8] M. Mamo, D. J. Mulla G. L. Malzer, D. R. Huggins, and J. Strock. 2003. Spatial and Temporal Variation in Economically Optimum Nitrogen Rate for Corn. *Agronomy Journal* 95, 4 (2003), 958–964.

[9] S. M. Pedersen, S. Fountas, B. S. Blackmore, M. Gylling, and J. L. Pedersen. 2004. Adoption and perspectives of precision farming in Denmark. *Acta Agriculturae Scandinavica, Section B – Soil & Plant Science* 54, 1 (2004), 2–8.

[10] T. B. Raper and J. J. Varco. 2015. Canopy-scale wavelength and vegetative index sensitivities to cotton growth parameters and nitrogen status. *Precision Agriculture* 16, 1 (01 Feb 2015), 62–76.

[11] J. W. Rouse, Jr, R. H. Haas, J. A. Schell, and D. W. Deering. 1974. Monitoring vegetation systems in the Great Plains with ERTS. In *3rd Earth Resources Technology Satellite-1 Symposium. Volume 1: Technical Presentations, section A.* 309–317.

[12] Shashi Shekhar, Chang-Tien Lu, and Pusheng Zhang. 2003. A Unified Approach to Detecting Spatial Outliers. *GeoInformatica* 7, 2 (2003), 139–166.

[13] Peter Chu Su. 2011. *Statistical Geocomputing: Spatial Outlier Detection in Precision Agriculture.* Master's thesis. UWSpace, Waterloo, Canada. http://hdl.handle.net/10012/6347.

[14] Michael A. Wulder, Jeffrey G. Masek, Warren B. Cohen, Thomas R. Loveland, and Curtis E. Woodcock. 2012. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment* 122, Supplement C (2012), 2–10. Landsat Legacy Special Issue.