disy
Kontraktadresse:
julian.bruns@disy.net

Präzisionsackerbau

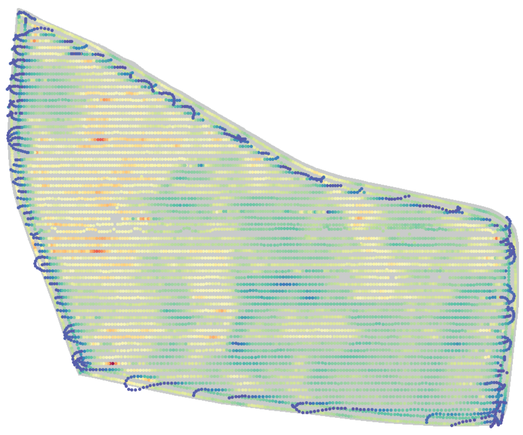
Geospatial Intelligence

Oliver Hennhöfer, Peter Ullrich, Julian Bruns — Disy Informationssysteme GmbH, Karlsruhe
Andreas Heiß — Universität Hohenheim, Institut für Agrartechnik

Zusammenfassung Die Bedeutung der Bereinigung landwirtschaftlicher Daten für den Einsatz in Forschung und Praxis soll durch Anwendung und Vergleich von Methoden zur Datenbereinigung hervorgehoben werden. Die Vor- und Nachteile für zwei Methoden zur Datenbereinigung werden dabei beispielhaft herausgearbeitet. Zuletzt wird ein Ausblick auf mögliche datenanalytische Methoden gegeben, mit denen sich die bereinigten Daten analysieren lassen und deren Ergebnisse als Entscheidungsgrundlage in der Landwirtschaft dienen können.

Datenbereinigung

Um die Validität der Ergebnisse aus der eigentlichen Datenanalyse zu gewährleisten, muss die Datengrundlage zuvor um fehlerhafte Messungen bereinigt werden. Dazu wurden zwei Algorithmen zur Identifikation von Ausreißern in räumlichen Daten auf Feldertragsdaten von Winterweizen angewendet.



Unbereinigte Ertragsdaten mit charakteristischen Fehlmessungen an den Feldrändern. Da das Feld anfangs ringsherum geerntet wird, kommt es beim Einbiegen in eine neue Fahrbahn im Feldinneren zu Mindermessungen, da bereits geerntete Fläche überfahren wird.

Methoden zur Datenbereinigung

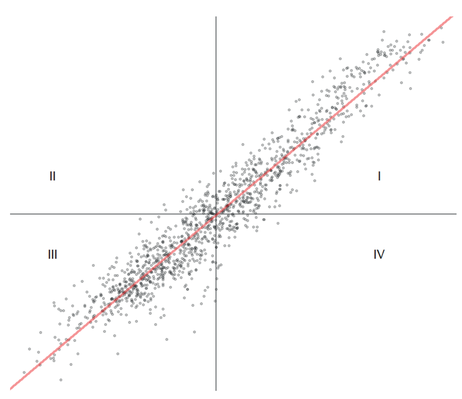
Die angewandten Methoden bestehen aus einem Algorithmus, der speziell zur Bereinigung landwirtschatflicher Feldertragsdaten konzipiert wurde (Vega, 2019) und einem Algorithmus für die allgemeine Erkennung räumlicher Ausreißer (Kou, 2006).

Der Algorithmus nach Vega bereinigt die Feldertragsdaten zuerst mithilfe klassischer statistischer Ausreißererkennung, zudem werden Punkte an der Feldgrenze entfernt. Anschließend wird das lokale Moran's I (Anselin, 1995) berechnet um räumliche Ausreißer zu identifizieren. Das resultierende räumliche Modell wird abschließend mithilfe von Methoden der Regressionsdiagnositik (Belsley et al., 1980) von übrigen Ausreißern bereinigt.

Formel für das lokale Moran's I:

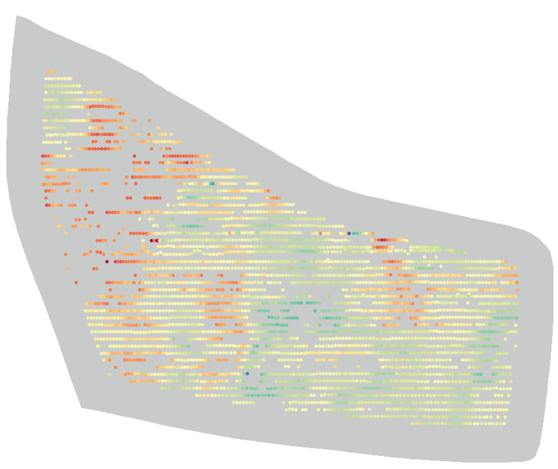
$$I_i = \frac{(y_i - \bar{y})}{s_i^2} \sum_{j=1}^{n} w_{ij}(y_j - \bar{y})$$

$$s_i^2 = \frac{\sum_{j=1}^{n}(y_j - \bar{y})^2}{n-1}$$

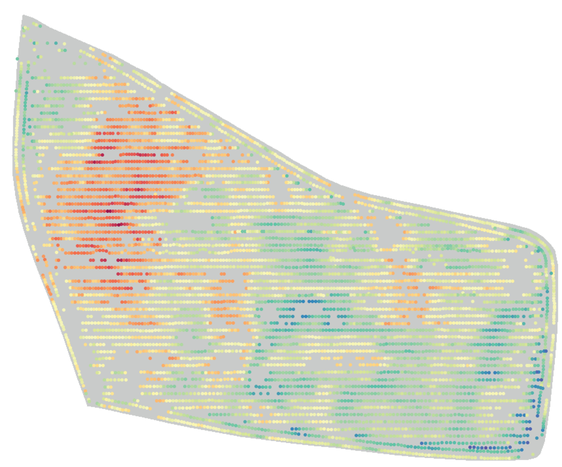


Ursprünglich für die Identifikation von räumlichen Hotspots gedacht, werden einzelne Beobachtungen yi mit ihrer Nachbarschaft yj verglichen. Positive Werte deuten auf ähnliche, sich clusternde Werte hin. Negative Werte entstehen, wenn sich der aktuelle Punkt von dessen Nachbarschaft unterscheidet - dann kann dieser als Ausreißer betrachtet werden. Das Diagramm stellt dabei den Feldertrag für yi dem gemittelten Feldertrag umliegender Punkte yj gegenüber. Ausreißer werden dabei im zweiten und vierten Quadrant dargestellt.

Error Removal Protocol
(Vega, 2019)



Averaged Difference Algorithm
(Kou, 2006)



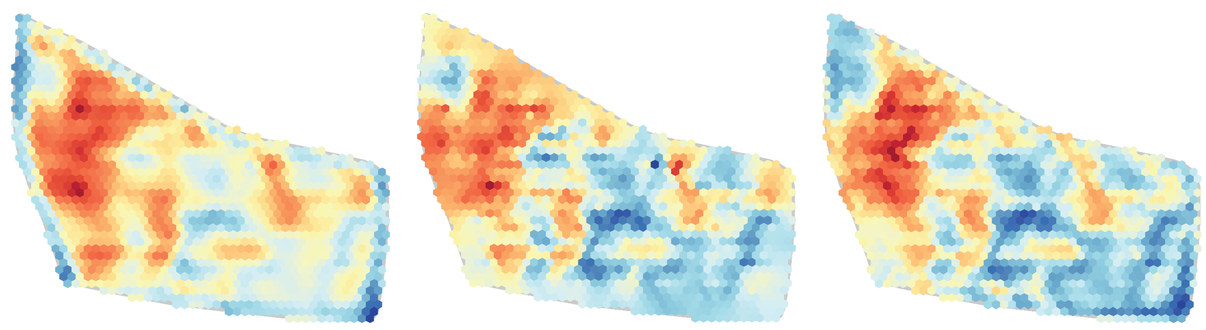
Der Algorithmus nach Kou berechnet für jede Punktbeobachtung die Abweichungen gegenüber benachbarten Punkten. Abweichungen zu unmittelbaren Nachbarn werden dabei höher gewichtet als Abweichungen zu mittelbaren Nachbarn. Das Ergebnis ist eine Rangliste der Punkte nach Größe der Abweichung.

$$AvgDiff_{x_i} = |x_i - \sum_{j=1} w_{i,j} x_j|$$

Die Formel gilt für reihen-standardisierte Gewichte in der Kontiguitätsmatrix wij, welche die Nachbarschaft jedes Punktes i definiert.

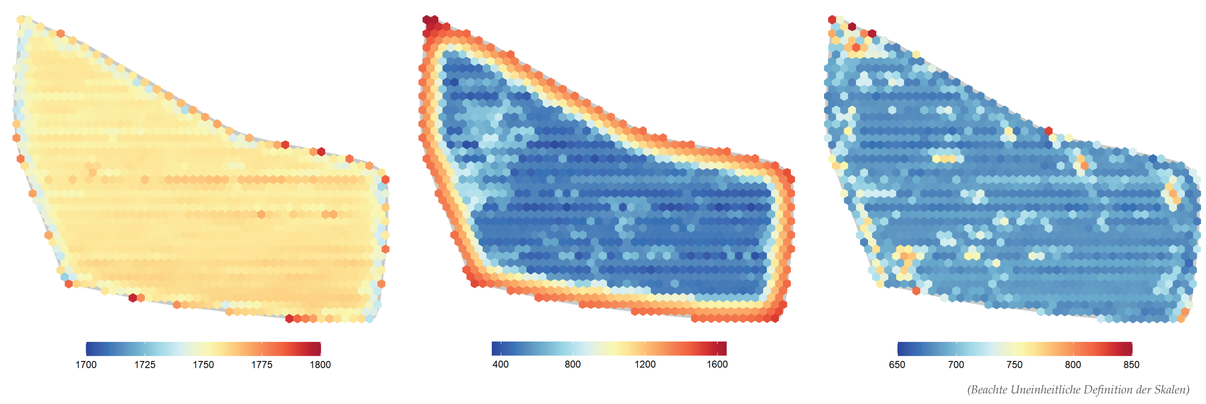
Nachverarbeitung der Daten

Die bereinigten Daten werden mittels Ordinärem Kriging auf ein einheitliches Raster interpoliert, um Analyseergebnisse vergleichbar zu machen:



ohne Datenbereinigung — Vega, 2019 — Kou, 2006

Neben der eigentlichen Interpolation werden auch die Unsicherheiten der Interpolationsergebnisse als Standardabweichung berechnet. Die Unsicherheit steigt mit abnehmender Punktdichte:



(Beachte Uneinheitliche Definition der Skalen)

Bewertung der Methoden

Bei der Karte für die unbereinigten Originaldaten sind die Ungenaugigkeiten insgesamt am höchsten, die fehlerhaften Feldrandmessungen schlagen durch. Der erste Algorithmus nach Vega liefert für das Feldinnere die verlässlichsten Ergebnisse, das kategorische Entfernen der äußeren Punkte macht aber eine Analyse des gesamten Feldes unmöglich. Der zweite Algorithmus nach Kou erzeugt kleinflächige Unsicherheiten, erlaubt aber die Analyse der gesamten Feldfläche.

Ausblick: Datenanalyse

Verfahren zur Identifikation von Hotspots können, mithilfe von Ertrags- und Fernerkundungsdaten, Aufschluss über besonders (un-)produktive Feldabschnitte geben. Die Ergebnisse helfen dem Landwirt Dünge- und Bewässerungsarbeiten zu koordinieren.
Mithilfe weiterer Beobachtungen lassen sich räumliche Regressionsmodelle erstellen, die Auffälligkeiten in den Daten identifizieren können.
Raumzeitliche Verfahren vergleichen zusätzlich Beobachtungen zwischen zwei Zeiträumen. Damit lassen sich beispielsweise Effekte von Düngung und Bewässerung räumlich differenziert (intra-)saisonal analysieren.

Innerhalb des Forschungsprojekts FuzzyFarmer erfordert das Zusammenspiel aus Fuzzy Expertensystem (Heiß, 2020) der Universität Hohenheim und der Optimierung der Mineraldünger-Applikation durch die TU Chemnitz, neue Ansätze zur Datenanalyse. Dabei kann die Anwendung klassischer KI-Verfahren durch geoanalytische Methoden erweitert werden, um Ergebnisse teilflächenspezifisch zu präzisieren.
Die Analyseergebnisse können die agrarwirtschaftliche Beratungsarbeit der Hanse Agro unterstützen und Landwirten zusätzlich als Grundlage für die landwirtschaftliche Planung dienen.

Andreas Heiß et al. (2020). Fuzzy Farmer – optimierte Stickstoffdüngung durch mehrparametrische Datenfusion u. präzise Applikation in Echtzeit
Andrés Vega et al. (2019). Protocol for Automating Error Removal from Yield Maps.
Yufeng Kou , Chang-Tien Lu (2006). Spatial Outlier Detection.
David A. Belsley et al. (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.
Luc Anselin (1995). Local Indicators of Spatial Association - LISA.

Gefördert durch:
Bundesministerium für Ernährung und Landwirtschaft
aufgrund eines Beschlusses des Deutschen Bundestages