

## Datenanalyse, statistische Methoden und Ensemble- Postprocessing zur Entscheidungsunterstützung bei der Feldbearbeitung

Jochen Fiedler<sup>1</sup> und Michael Burger<sup>2</sup>

**Abstract:** Im Rahmen des Fraunhofer Leitprojekts COGNAC arbeiten wir an der Integration von verschiedenen Datenquellen, um die Prädiktion und Analyse von landwirtschaftlichen Einflussfaktoren zu ermöglichen. Dabei beleuchten wir verschiedene Methoden der Datenaggregation und -analyse. Wir zeigen, dass Verfahren aus dem statistischen Ensemble-Post-Processing nutzbar sind für die Generierung von Pflanzenwachstumsprognosen. Ausgehend von einem Ensemble wird eine probabilistische Vorhersage gebildet, welche dann mittels etablierter statistischer Verfahren ausgewertet werden kann. Schließlich zeigen wir, wie solche Vorhersagen zur Optimierung von Feldbearbeitungsschritten genutzt werden können, um beispielsweise optimale, teilflächenspezifische Düngeempfehlungen abzuleiten. Übergeordnetes Ziel ist es, mithilfe von realistisch verfügbaren Daten und mathematischen Verfahren die z.T. hoch-komplexen Entscheidungsprozesse des landwirtschaftlichen Alltags zu unterstützen.

**Keywords:** Datenanalysen, Ertragsprognosen, Ensemble-Vorhersagen

### 1 Einleitung

Viele Landwirte haben bereits heute reichhaltige Datensätze über ihre Felder vorliegen. Dazu zählen beispielsweise (historische) Sensordaten über Ernteerträge und Biomasse oder mehr oder weniger detaillierte Bodendaten, welche durch die Entnahme von Proben gewonnen wurden. Hinzu kommen öffentlich generell gut verfügbare Daten, wie digitale Geländemodelle, Wettervorhersagen sowie historische Wetterdaten.

Ziel unserer Arbeit ist es, mit Hilfe dieser Daten die oft sehr komplexen Entscheidungsprozesse des Landwirts zu unterstützen. Besonderes Augenmerk liegt dabei zunächst auf Entscheidungen hinsichtlich Düngemittelapplikationen, da diese einerseits hinsichtlich Kosten-Nutzen-Abwägungen und andererseits gleichzeitig auch im Hinblick auf Umweltschutzaspekte zunehmend in den Fokus geraten, siehe bspw. [He20]. Im Rahmen des Fraunhofer Leitprojekts COGNAC, an dem insgesamt acht Fraunhofer-Institute aus verschiedenen Domänen beteiligt sind, forschen wir an der Nutzung und statistischen Analyse von diversen Schlag- und Ertragsdaten, deren Aggregation und visueller Aufbereitung sowie an deren Kombination mit Umgebungsinformationen wie

---

<sup>1</sup> Fraunhofer-Institut für Techno- und Wirtschaftsmathematik, Bereich MF, Fraunhofer-Platz 1, 67663 Kaiserslautern, jochen.fiedler@itwm.fraunhofer.de

<sup>2</sup> Fraunhofer-Institut für Techno- und Wirtschaftsmathematik, Bereich MF, Fraunhofer-Platz 1, 67663 Kaiserslautern, michael.burger@itwm.fraunhofer.de

z. B. Wetterdaten oder Feldtopographie. Wir entwickeln dabei insbesondere Methoden und Verfahren, historische Datensätze effizient zu analysieren und sie mit den genannten Umgebungsdaten in Bezug zu setzen. Dies dient u.a. dazu, dem Landwirt einfach und effizient vergleichende Analysen früherer Strategien zu ermöglichen, woraus er dann wiederum Schlüsse für anstehende Entscheidungen ziehen kann. Dem Landwirt soll unter anderem ermöglicht werden, den Einfluss von Maßnahmen oder Umweltfaktoren statistisch zu evaluieren und dies mit Erfahrungen abzugleichen. Zum anderen kombinieren wir alle verfügbaren Daten aus den unterschiedlichen Quellen und lassen diese in verfügbare Pflanzenwachstumsmodelle [PVS16] einfließen. Die Rolle von Pflanzenwachstumsmodellen beschränkt sich aktuell noch hauptsächlich auf den Bereich der Forschung. Aufgrund der Komplexität der Modelle ist eine gute Kalibrierung in der Praxis oft sehr herausfordernd, da diese eine gute Kenntnis aller relevanten Parameter erfordert.

An diesen Punkt setzen wir mit der Generierung und dem Postprocessing von Ensemble-Vorhersagen an. Diese ermöglichen es, Unsicherheiten in den Randbedingungen direkt in die Modellierung einfließen zu lassen und damit z. T. abzuschwächen oder auch zu kompensieren, mindestens aber Unsicherheiten und Streubänder zu quantifizieren. Gleichzeitig lassen sich Ensembles durch Beobachtungen aus unterschiedlichen Quellen wie Biomassesensoren oder Satelliten immer wieder aktualisieren und verbessern. Daher entsprechen sie auch seit Jahren dem Stand der Technik bei den Wettervorhersagen und haben geholfen, die Güte der Prognosen deutlich zu verbessern [GR05]. Im Kontext von Pflanzenwachstumsmodellen machten z. B. [Ma15] erste Beobachtungen, die nahelegen, dass Ensembles von Pflanzenwachstumsprognosen bessere Vorhersagen liefern, als es eine einzelne Prognose kann.

Wir möchten diese Erfahrungen umsetzen und auf Basis von Ensemble-Vorhersagen Entscheidungsunterstützung für den Landwirt geben. Dabei ist es uns sehr wichtig, dass die realistischen Randbedingungen des landwirtschaftlichen Alltags beachtet und berücksichtigt werden – so liegen Daten und besonders Modelle oft nur heterogen und lückenhaft oder gar nicht vor, Antworten und Entscheidungen müssen aber sehr schnell und möglichst genau gefunden werden. In diesem Beitrag berichten wir über einen methodischen Gesamtansatz, aus dem heraus gegenwärtig ein Software-Tool im Entstehen ist, bei dem hinreichend leistungsfähige mathematische Methoden zum Einsatz kommen, aber gleichzeitig die Darstellung von Ergebnissen und Analysen möglichst eingängig und intuitiv gehalten wird, um dem Landwirt umfänglich und verständlich Fakten sowie schnelle Hilfen für entsprechende Entscheidungen zu liefern.

## **2 Kompletter Workflow**

Das komplette System ist schematisch in Abb. 1 dargestellt. In dessen Kern steht die Datenverarbeitung sowie die Generierung von Ensembles, welche dann optimiert werden, um daraus Handlungsempfehlungen abzuleiten. In einem ersten Schritt werden verfügbare

Daten gesammelt und zusammengeführt. Dabei wird darauf geachtet, dass die Daten in einem für die statistische Analyse geeigneten Format vorliegen. Die Ergebnisse dieser Analyse werden dann für zwei Aspekte genutzt. Einerseits zur direkten Untersuchung empirischer Zusammenhänge: So kann ein Landwirt alle historischen Daten auf bestimmte Zusammenhänge hin untersuchen, beispielsweise ob und wo eine teilflächenspezifische Düngung im Feld einen Mehrwert beim Ertrag gebracht hat. Bei dieser Beurteilung wird der Landwirt durch geeignete Visualisierungen und Ausgabe statistischer Zusammenhänge unterstützt. Andererseits werden die empirischen Datenanalysen zur Kalibrierung von Pflanzenwachstumsmodellen genutzt. Diese sollten, gegeben der bekannten Randbedingungen wie Bodenzusammensetzung, die Spannweite der beobachteten Erträge abdecken.

Die auf diese Weise grob vorkalibrierten Modelle werden anschließend zur Erzeugung von Ensemble-Vorhersagen genutzt, die dann einem statistischen Postprocessing unterzogen werden, für welches zusätzliche Daten wie Beobachtungen der Biomasse genutzt werden. Zum Schluss werden aus dem so entstandenen Ensemble probabilistische Vorhersagen generiert, die dann hinsichtlich interessierender Größen, wie beispielsweise des Zeitpunkts der Düngung, optimiert werden können.

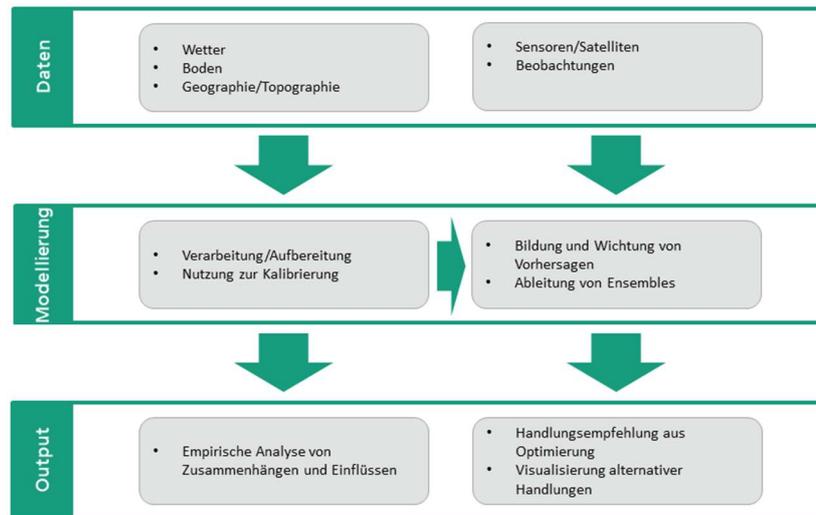


Abb. 1: Schematischer Workflow des Systems

### 3 Empirische Analyse vorhandener Daten

In Abb. 2 sind exemplarisch einige verfügbare Daten anonymisiert als Ausschnitte dargestellt. Visuell sind dort bereits starke Ähnlichkeiten zwischen Bodenbewertungszahl (erstellt nach dem Bodenschätzungsgesetz) und Stickstoffapplikation zu sehen, während

es zwischen Stickstoffapplikation und Ertrag wenig Korrelation zu geben scheint. Eine empirische Analyse zeigt: Eine einfache lineare Regression ohne räumliche Interaktionen genügt, um den Zusammenhang zwischen Bodenbewertungszahl und Applikation hinreichend zu beschreiben. Das lässt die Vermutung aufkommen, dass die Applikationskarte für die Düngung auf der Bodenbewertungszahl basiert und dementsprechend auf Bereichen mit einer höheren Bodenbewertungszahl größere Mengen an Stickstoff appliziert wurden. Die hohe Pearson-Korrelation von 0,82 unterstreicht diesen Schluss.

Wegen der starken Kollinearität zwischen Bodenbewertungszahl und Applikation wird in einem weiteren Analyseschritt nur der Einfluss der Düngung auf den Ertrag untersucht. Dazu zeigt eine interne Analyse in unserem Workflow, dass hier ein räumlicher, autoregressiver Prozess ein sinnvolles Modell darstellt, weswegen das System dann einen solchen Prozess zur Modellierung nutzt. Für die Interpretation der Ergebnisse der Modellierung wird der sogenannte *impact* als Maß genutzt. Dieser liefert negative Werte, was bedeutet, dass auf den Flächen, auf denen mehr Düngemittel appliziert wurde, weniger Ertrag entstand. Da diese Flächen gleichzeitig eine höhere Bodenbewertungszahl haben, ist dieses Ergebnis kontraintuitiv, spiegelt aber die visuellen Eindrücke wider (vgl. Abb. 2). Aus den hier genutzten Daten ist nun keine abschließende Interpretation dieses Ergebnisses möglich, so dass hier mehr Daten oder die Expertise des Landwirts notwendig sind. Die Auswertung eröffnet aber wichtige verdeckte Zusammenhänge, die dann mit agronomischer Expertise zusammengeführt werden können.

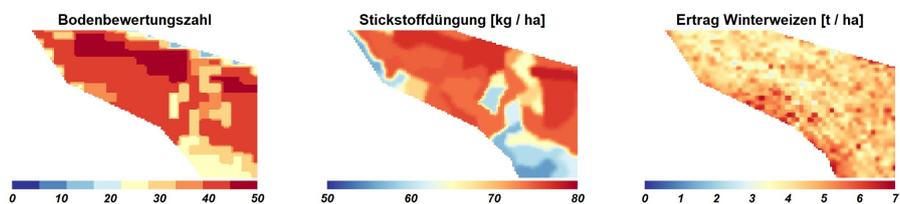


Abb. 2: Empirisch verfügbare Daten zu Ertrag, Bodenbewertungszahl und Stickstoffdüngung. Die Bilder zeigen anonymisierte Ausschnitte

## 4 Generierung und Postprocessing von Ensemble-Vorhersagen

Zur Generierung eines Ensembles haben wir beispielhaft das Pflanzenwachstumsmodell APSIM [Ke03] betrachtet und damit den Ertrag eines einzelnen Feldes simuliert. Exemplarisches Ziel ist dabei hier, die Auswirkungen zweier verschiedener Saatkulturen zu untersuchen. Als Pflanze wurde dabei auf Winterweizen kalibriert [Kn11]. Anschließend wurde für die zwei verschiedenen Saatkulturen von 50 kg/ha und 150 kg/ha jeweils eine Ensemble-Vorhersage generiert. Dabei entstand jedes Ensemble-Mitglied aus einer unterschiedlichen Kombination aus historischem Wetterverlauf und Boden-pH-Wert – alle anderen Randbedingungen wurden der Einfachheit gleich gelassen.

Abb. 3 zeigt exemplarisch, wie für eine Ensemble-Vorhersage eine Entscheidung entsteht und wie sich die beiden Ensembles für die beiden Saaddichten unterscheiden. In Plot (a) sehen wir eine relativ breite Streuung in den einzelnen Ertragsvorhersagen. Diese werden für jeden Zeitpunkt überführt in eine probabilistische Vorhersage (b), in der sich dann die Unsicherheit der Vorhersage direkt ablesen lässt. In unserem Falle wurde für jeden Zeitpunkt an eine Vorhersage des Ensembles eine Gammaverteilung angepasst, wobei jedes Mitglied im Ensemble dasselbe Gewicht erhält. In der Praxis wird so eine gleiche Gewichtung durch eine individuelle ersetzt, welche dann auf Beobachtungen der Biomasse reagieren kann. Somit werden im Laufe einer Saison verlässlichere Ensemble-Mitglieder stärker gewichtet als weniger verlässliche, was die Güte der Prognose verbessert. Diese Anpassung einer Gammaverteilung wurde für jeden Prognose-Zeitpunkt durchgeführt, was den Prognose-Schlauch in (c) ergibt. In (d) wurden dann beide Prognosen verglichen, wobei sich zeigt, dass die Entscheidung für die höhere Saaddichte einen leicht günstigeren Ertragsverlauf mit sich bringt, weshalb man sich in diesem Fall für die höhere Saaddichte entscheiden würde.

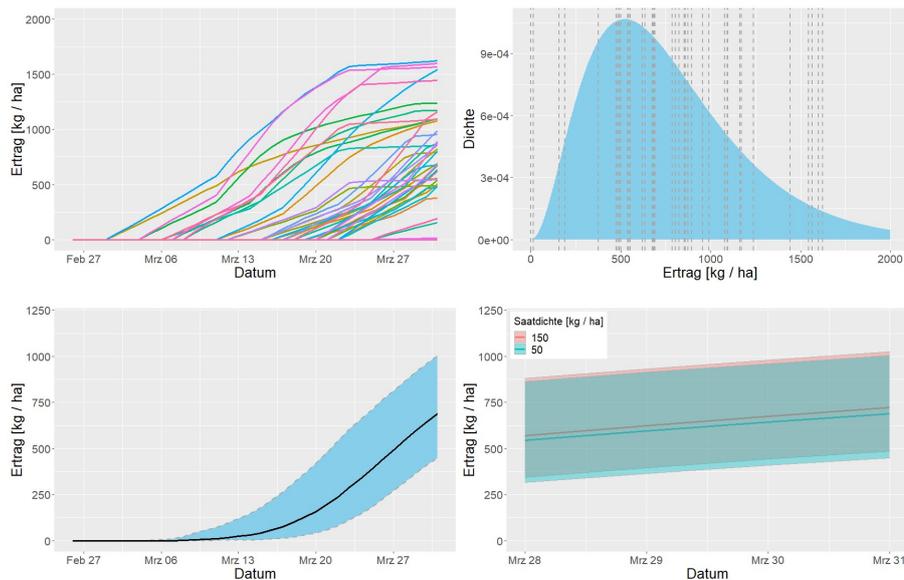


Abb. 3: Entstehung einer probabilistischen Vorhersage aus einer Ensemble-Vorhersage. In (a) sind die individuellen Prognosen zu sehen, welche aus unterschiedlichen, historischen Wetterverläufen in Kombination mit verschiedenen Boden-pH-Werten entstehen. In Graph (b) wurden die einzelnen Vorhersagen für einen einzelnen Tag zu einer Gammaverteilung kombiniert; die einzelnen Vorhersagen sieht man als vertikale gestrichelte Linien. Die zeitliche Entwicklung der Quartile und des Medians (schwarze Linie) sieht man in (c), während (d) einen Vergleich der beiden Prognosen zeigt. Für diesen Vergleich haben wir uns aufgrund der besseren Darstellbarkeit auf einen kleinen zeitlichen Ausschnitt beschränkt.

## 5 Diskussion und weitere Arbeiten

Im Rahmen des Projekts werden in nächsten Schritten obige Ansätze noch weiter ausgearbeitet. Ziel dabei ist es immer, dem Landwirt möglichst einfache und umfassende Hilfestellungen zu geben. Aus diesem Grund arbeiten wir weiter an möglichst guten Darstellungen und Analysen vorhandener Daten.

Diese Daten sollen später auch immer stärker für die Erzeugung der Ensemble-Prognosen genutzt werden, was zum aktuellen Zeitpunkt unseres Projekts noch sehr rudimentär erfolgt. Dennoch zeigt sich bereits jetzt, dass Ensembles bestehende Unsicherheiten abbilden und berücksichtigen können, was den Landwirt bei seinen Entscheidungen unterstützen kann. Die in diesem Beitrag gezeigten Analysen und Vorhersagen, Abhängigkeitsanalysen und räumliche Regressionsmodelle und ensemblebasierte Ertragsvorhersage, sind als Beispiele und Schlaglichter zu verstehen und sollen demonstrieren, wie (heterogene) Daten aus verschiedenen Quellen und verschiedener Güte kombiniert mit entsprechenden mathematischen Techniken Einsichten und Entscheidungshilfen bereitstellen können.

### Literaturverzeichnis

- [PVS16] Di Paola, A., R. Valentini, und M. Santini. „An overview of available crop growth and yield models for studies and assessments in agriculture.“ *Journal of the Science of Food and Agriculture* 96 (2016): 709-714.
- [GR05] Gneiting, T.; Raftery, A. E.: Weather Forecasting with Ensemble Methods. *Science* 310, S. 248-249, 2005.
- [He20] Heiß, A.; D. S. Paraforos; G. M. Sharipov; H. W. Griepentrop: Fuzzy Farmer – optimierte Stickstoffdüngung durch mehrparametrische Datenfusion und präzise Applikation in Echtzeit. In (Gandorfer, M. et al., Hrsg.): *Digitalisierung für Mensch, Umwelt und Tier, Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2020*, S. 109-114.
- [Ke03] Keating, B. A. et al. An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy* 18, S. 267-288, 2003.
- [Kn11] Knörzer, H.; Lawes, R.; Robertson, M.; Graeff-Hönninger, S., Claupein, W.: Evaluation and Performance of the APSIM Crop Growth Model for German Winter Wheat, Maize and Fieldpea Varieties in Monocropping and Intercropping Systems. *Journal of Agricultural Science and Technology B* 1, S. 698-717, 2011.
- [Ma15] Martre, P.; Wallach, D.; Asseng, S.; Ewert, F.; Jones, J. W.; Rötter, R. P.: Multimodel ensembles of wheat growth: many models are better than one. *Global Change Biology* 21, S. 911–925, 2015.