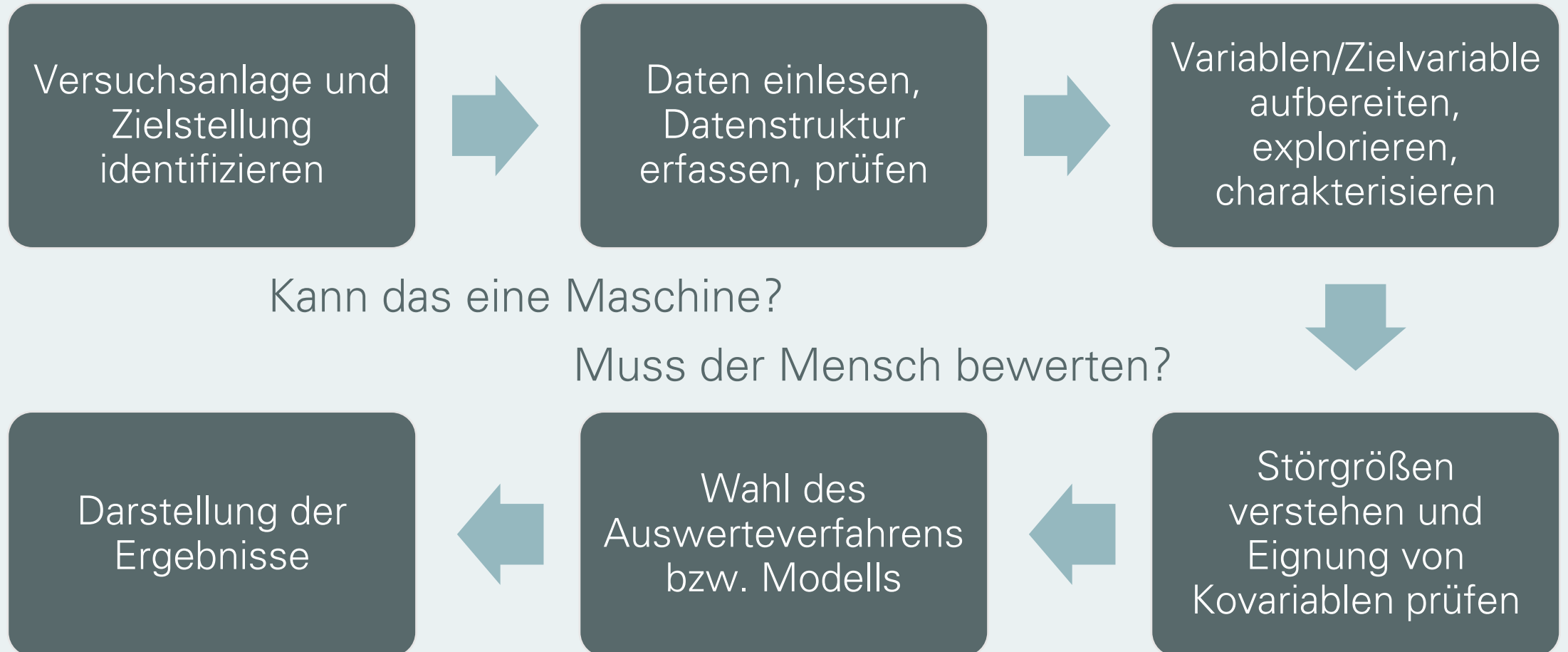
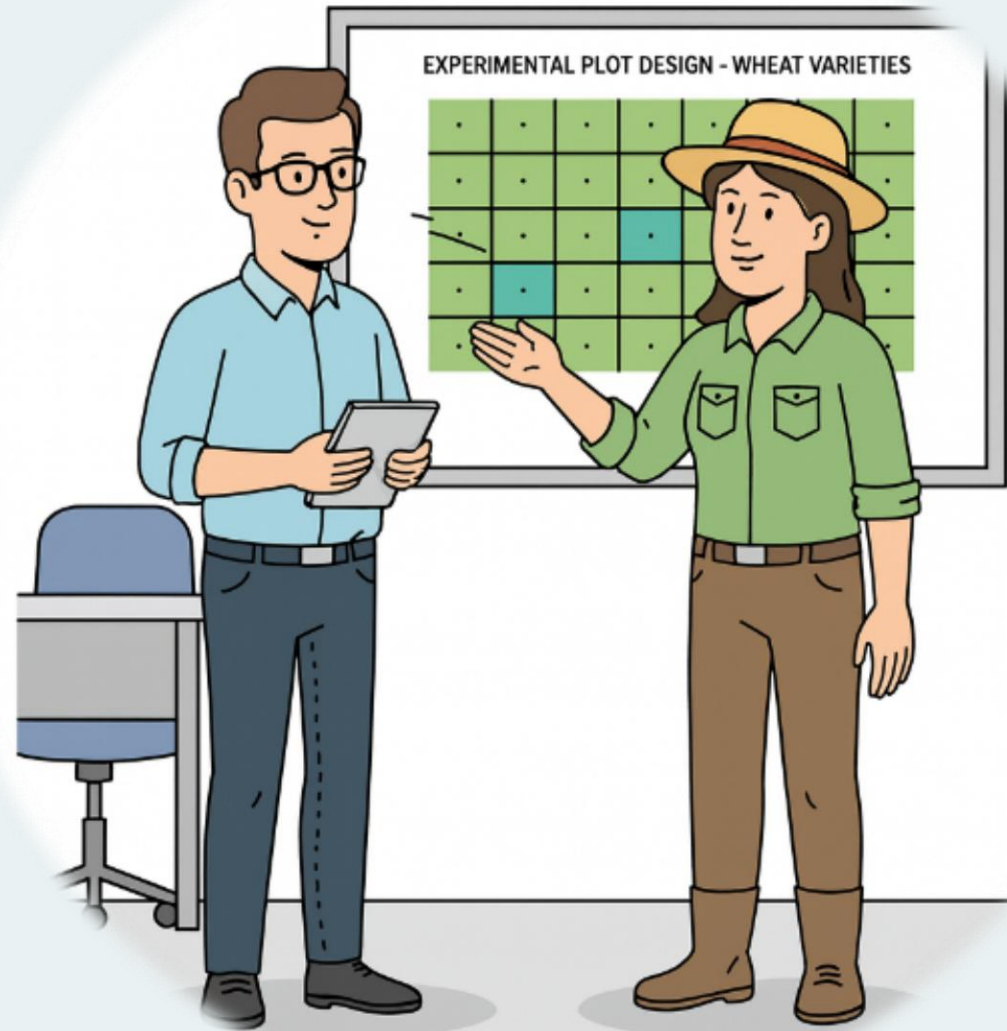


Workflow Statistische Datenanalyse

Möglichkeiten und Grenzen einer
automatisierten Datenanalyse für OFE



Versuchsanlage
und Zielstellung
identifizieren



Versuchsplanung und Struktur verstehen



Teilweise
automatisierbar

- Was ist die Versuchsfrage?
- Wie war die ursprüngliche Planung?
- Feldplan anschauen, Besichtigung vor Ort
- Faktoren, Faktorstufen identifizieren
- Kreuzklassifiziert? Hierarchisch geschachtelt?
- Gibt es (echte) Wiederholungen?
- Wurde randomisiert? Einschränkungen bei Randomisation?
- Blockstruktur: Vollständig/Unvollständig?
- Kleinteilstücke, Großteilstücke?
- Messwiederholung innerhalb der Kleinteilstücke?

*OFE oft nur wenige Faktoren und Faktorstufen
d.h. Fehlerfreiheitsgrade begrenzt*

*Messwiederholung bei OFE die Regel;
große Parzellen; Sensorik verfügbar*

Daten einlesen,
Datenstruktur
erfassen, Daten
prüfen,
Harmonisieren



Daten einlesen, Datenstruktur erfassen, prüfen

Teilweise
automatisierbar

- Verfügbarkeit? Dateitypen (CSV, EXCEL, Datenbank, xml, json, ...)
- Datenstruktur: Datentypen, Spalten, Zeilen
- ID für Parzelle verfügbar?
- Räumliche Koordinaten verfügbar? (→ Geostatistik)
- Messwiederholung in breitem oder langem Format?

Fehlende Werte bei OFE häufig!?

Praxistechnologie verwendet anstatt spezieller Versuchstechnik

Tendenziell bei OFE komplexer und mehr, selbst nach vorangegangener Datenaufbereitung

Daten einlesen, Datenstruktur erfassen, prüfen

Teilweise
automatisierbar

- Verfügbarkeit? Dateitypen (CSV, EXCEL, Datenbank, xml, json, ...)
- Datenstruktur: Datentypen, Spalten, Zeilen
- ID für Parzelle verfügbar?
- Räumliche Koordinaten verfügbar? (→ Geostatistik)
- Messwiederholung in breitem oder langem Format?

Fehlende Werte bei OFE häufig!?

Praxistechnologie verwendet anstatt spezieller Versuchstechnik

Tendenziell bei OFE komplexer und mehr, selbst nach vorangegangener Datenaufbereitung

Variablen
explorieren und
Datenqualität
prüfen

&

Zielvariablen
charakterisieren



Variablen explorieren und Datenqualität prüfen

**Teilweise
automatisierbar**

*Kein Unterschied zu
Kleinparzellenversuch auf Station*

- Faktornamen identisch mit Versuchsplan?
- Fehlwerte, leere Zellen, Nullen
- Informative Missing? Fehlen Teile der Versuchsfläche oder ganze Parzellen/Varianten? Warum?
- Generelle Filter anwenden; Werte im erwarteten biologischen Bereich?
- Ausreißer identifizieren
- Tippfehler, Einheitenfehler, Dezimalfehler?
- Konsistenz zwischen verwandten Variablen

Zielvariable charakterisieren

- Datentyp: Stetig, diskret (Count), binär, ordinal, Anteilswerte?
- Verteilungsannahme: Normal, Poisson, binomial, negativ-binomial, Beta?
- Wertebereich: Grenzen, Nullwerte, negative Werte möglich?
- Transformationsbedarf erkennbar (z.B. log, sqrt, arcsin)?

**Teilweise
automatisierbar**

*Kein Unterschied zu
Kleinparzellenversuch auf Streifen*

Kovariablen und
Störgrößen
verstehen und
Eignung prüfen



Störgrößen verstehen und Eignung von Kovariablen prüfen

Teilweise
automatisierbar

- Welche Störgrößen wurden erwartet/vermutet? Gradienten (Boden, Feuchtigkeit, Gefälle) ?
- Warum wurde dieser Teil des Feldes gewählt? Warum wurden Blöcke so angeordnet? Welche Störgrößen wurden antizipiert?
- Welche Störgrößen wurden tatsächlich gemessen/quantifiziert?
 - Bodeneigenschaften (pH, Nährstoffe, ...)?
 - Anfangszustände (Anfangsgewicht, Pflanzdichte, ...)
 - Umweltfaktoren (Temperatur, Niederschlag, ...)
 - Fernerkundungsdaten (Satelliten-, Drohnenbilder, ...)
- Auf welcher Ebene wurden Kovariablen gemessen? Pro Parzelle? Oder mehrere Messungen pro Parzelle? Oder pro Block, pro Feldabschnitt)?

*Wichtig bei OFE!
Einfluss Windrichtung, Relief,
Schädlinge; Fehler bei
Pflegemaßnahmen
Wichtig für Versuchsserie.
Sind Umweltbedingungen
vergleichbar? Ceteris-Paribus.*

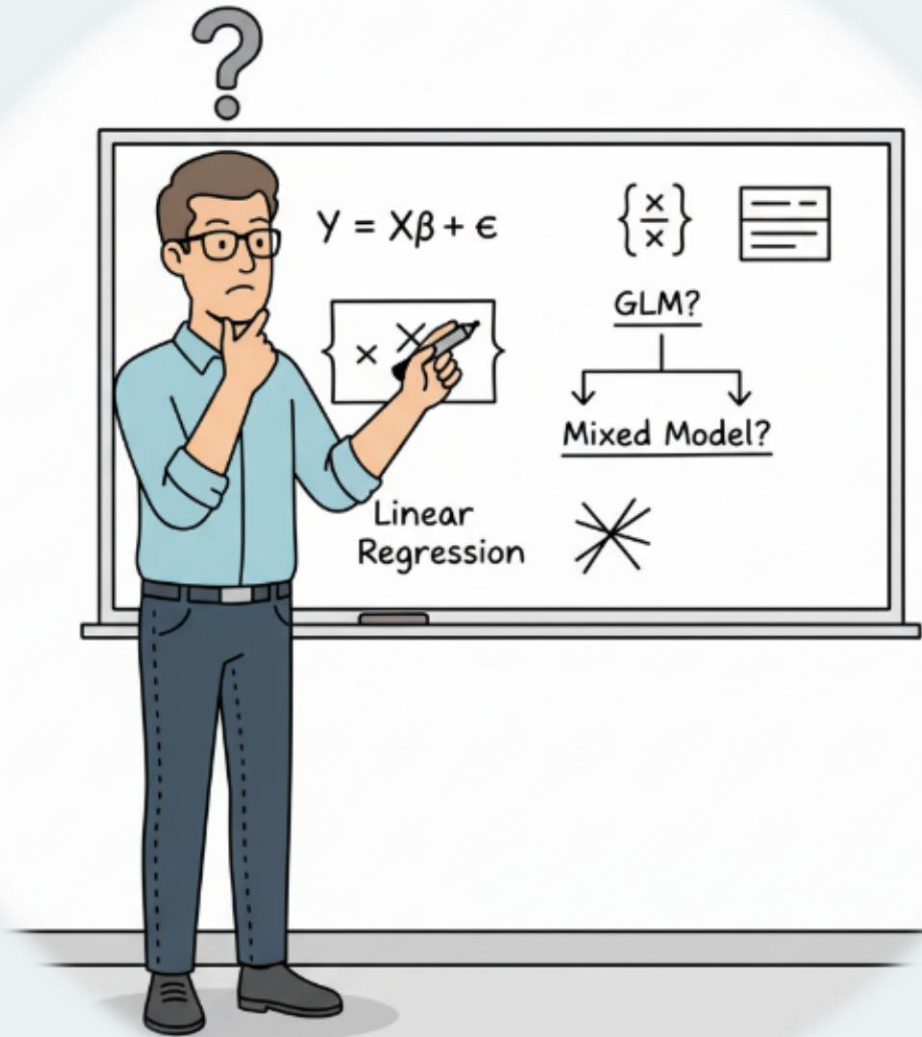
Welche Kovariablen sollen wie ins Modell?

Teilweise
automatisierbar

- Ist die Kovariable unabhängig von den Behandlungen?
- Hat die Kovariable ausreichend Variabilität über Parzellen hinweg?
- Gibt es eine plausible biologische/physikalische Beziehung?
- Ist die Beziehung linear oder nicht-linear?
- Ist die Kovariable vollständig gemessen oder gibt es NAs?
- Als fixe Effekte (ANCOVA)?
- Zur Adjustierung vor der eigentlichen Auswertung (z.B. Ertrag auf einheitliche Pflanzdichte korrigieren)? Zur Modellierung räumlicher Trends (Splines, Koordinaten)? Zur Definition von Heterogenität (unterschiedliche Varianzen pro Kovariablen-Gruppe)? Zum Post-Blocking oder zur Bildung von Kategorien (insbesondere bei Precision Farming)

*Alle Variablen in der Auflösung
auf das Ertragsgrid beziehen?*

Wahl des Auswerteverfahrens bzw. Modells





Analyze as randomized! Oder?

- Designeffekte : Blocks, Klein- und Großteilstücke
- Treatmenteffekte (Haupteffekte, Wechselwirkungen, stetig vs. diskret)
- Was ist fix, was zufällig?
 - Zufällig: Stufen sind zufällige Stichprobe aus einer größeren Population, Generalisierung auf Population gewünscht; Fest: Stufen wurden spezifisch ausgewählt, Interesse an genau diesen Stufen
 - Faustregel: Mindestens 4 (oder mehr?) Stufen für zuverlässige Schätzung zufälliger Effekte
 - Alles, was eine Randomisationseinheit ist und nicht alle Treatmentstufen enthält, sollte zufällig sein (z.B. Großteilstück bei Split-Plot, Incomplete Block)
 - Interaktion zwischen fest und zufällig ist zufällig
- Ist Lineares Modell / ANOVA / Linear Mixed Modell überhaupt angemessen?
Modellvoraussetzungen prüfen
- Alternativen prüfen, je nach Art der Zielvariable:
 - Generalized Linear Model (GLM); Generalized Linear Mixed Models (GLMM); Poisson-Modell, Negativ-Binomial (für Zählwerte); Logit, Probit für Binäre Daten; Anteilsdaten: Prozentwerte über Transformation oder Quasi-Binomial?
- Varianz- und Korrelationsstrukturen

Darstellung der
Ergebnisse

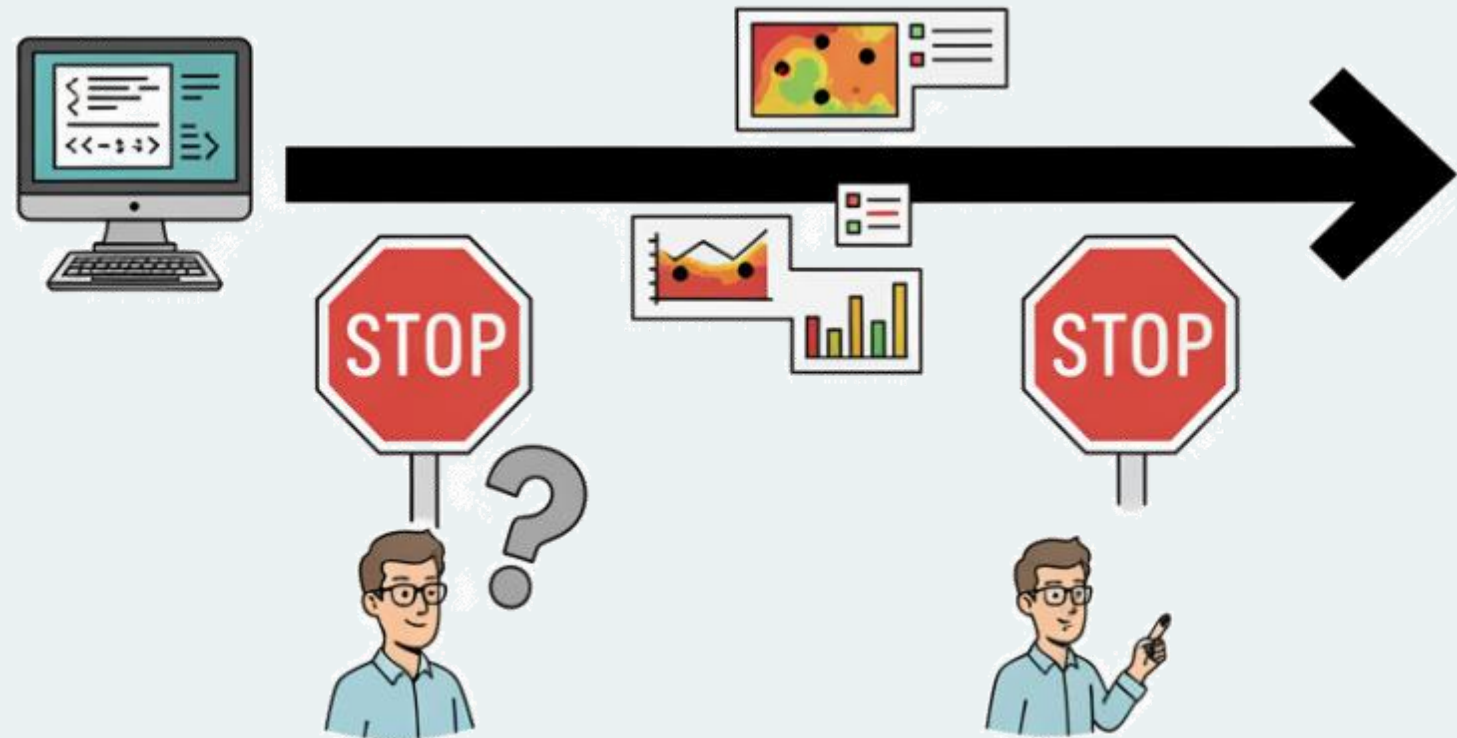


Darstellung der Ergebnisse

- Welche Tabellen und Grafiken sollen publiziert werden?
- Text, Abschlussbericht(e)



- Viele Arbeitsschritte zumindest teilweise automatisierbar.
- **Bewertung und Entscheidung durch Menschen dennoch unverzichtbar.**
- Welche Expertise ist notwendig? Benötigt OFE die Unterstützung durch den erfahrenen Statistiker? immer? manchmal? wer entscheidet, ob Unterstützung notwendig ist?
- 80:20 Lösungen möglich?
Wieviel menschlichen / interaktiven Eingriff möchte man haben?
- Mit „hohem Aufwand“ wäre ein Expertensystem programmierbar was statt 80:20 eine 95:5 Lösung anbietet. **Aber lohnt sich das? Ist das effizient?**



AgDoIT

IT & Datenmanagement im
Agrarbereich
mit Fokus auf Versuchsthemen

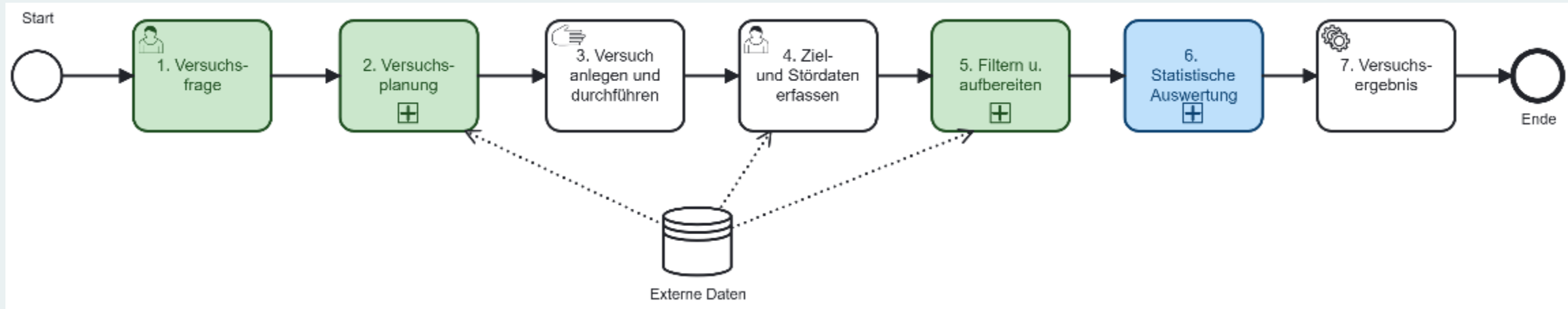


Motivation ISOFarmResearch

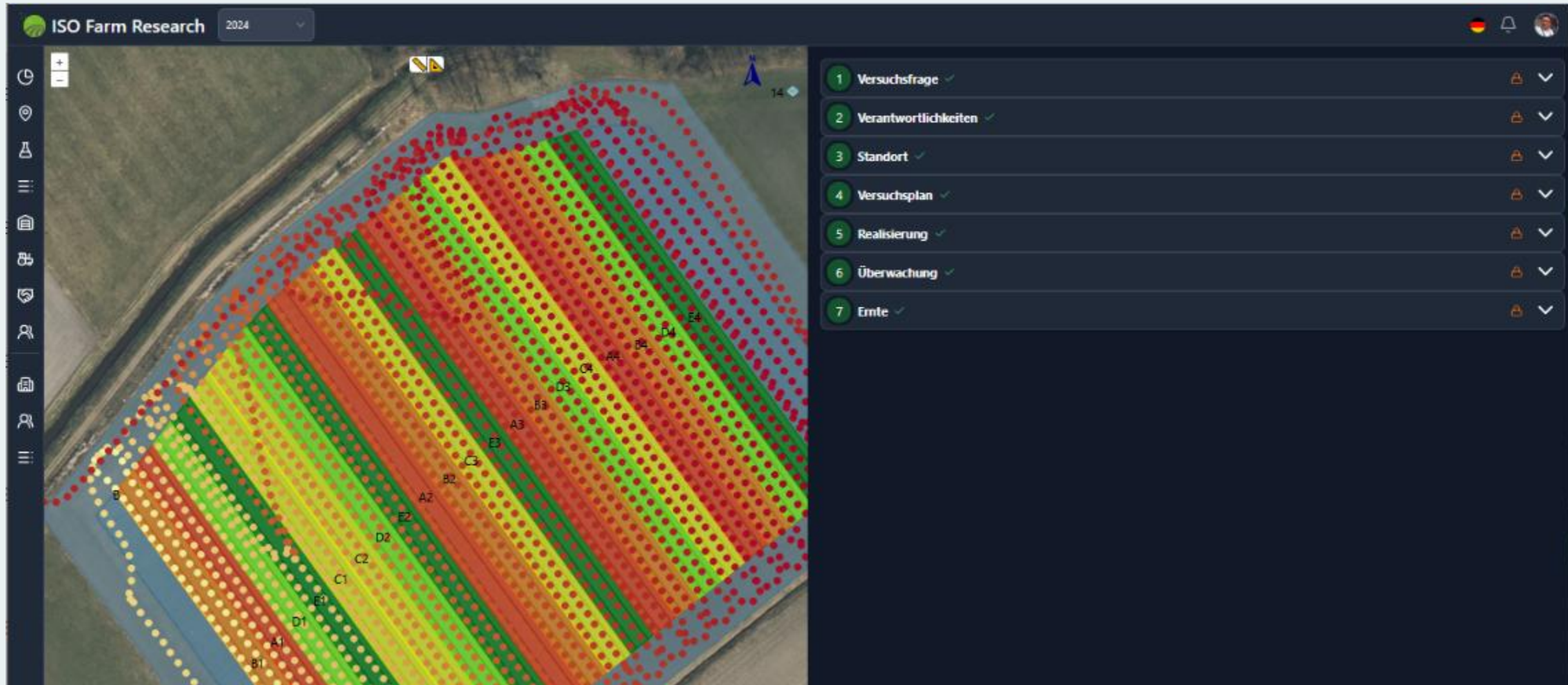
- Mehr pflanzenbauliche Fragestellungen
- „KI als Berater“? ML- und KI-Systeme probabilistisch: mal so mal so...
- Mehr Evidenz aus der Fläche sinnvoll – nicht als Ersatz sondern Ergänzung
- Einsatz von Precision farming immer mehr für Versuchsfragestellungen geeignet
- Mehr Vergleichbarkeit – auch international
- Bisher kein durchgängiges System verfügbar
- OFE-Leitfaden & Checkliste



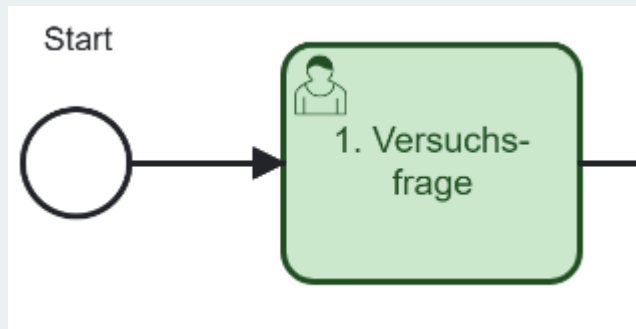
Ablauf ISOFarmResearch



Überblick Programm



Fragestellung



The screenshot shows the ISO Farm Research software interface. The main view is an aerial map of a field with a grid of colored rectangles representing experimental plots, labeled A1 through E4. The right sidebar contains the following information:

- 1 Versuchsfrage** (checked)
- Versuchstyp**: Düngerversuch
- Fragestellung***: Kann eine portionierte und punktuelle Düngung mit Amazone Fertisport den Ertrag bei geringerem Nährstoff-Input auf ähnlichem...
- Beschreibung**:
 - Randomized complete Block design
 - 5 Versuchsglieder:
 - A: Keine Unterfußdüngung (UFD), 0% UFD*
 - B: 150kg/ha DAP UFD Bandablage „100% Band“
 - C: 75kg/ha DAP UFD Bandablage „50% Band“
 - D: 150kg/ha DAP UFD Fertisport „100% Spot“
 - E: 75kg/ha DAP UFD Fertisport „50% Spot“
 - vier Wiederholungen
 - Ansonsten Gleichbehandlung (Ceteris paribus)
- Titel***: Fertisport 2024
- Varianten***:
 - A 0% UFD 0 kg/ha
 - B 100% Band 150 kg/ha
 - C 50% Band 75 kg/ha
 - D 100% Spot 150 kg/ha
 - E 50% Spot 75 kg/ha
- Wiederholungen***: 4
- Anzahl Jahre***: 1
- Buttons: Speichern, Alle speichern, Abbrechen

2 Verantwortlichkeiten (checked)

- Alex Trukhin (trukhin@agdoit.com) - Technician
- Christoph Ratke Ratke (christoph.ratke@agdoit.com) - Trial Responsible

Von der Frage zu den Varianten

- Am Anfang steht die Fragestellung - Pflicht
- Rahmen im gesamten Prozess beibehalten
- Fragestellung + Kontext
- Sprachsupport für qualifizierte Fragestellung
- Ableitung von Faktor(en) und Varianten mit LLM – hier macht der Einsatz von KI Sinn, um den Prozess zu unterstützen, nicht das Ergebnis

1 Versuchsfrage ✓

Versuchstyp: Düngerversuch

Fragestellung*: Kann eine portionierte und punktuelle Düngung mit Amazone Fertisplot den Ertrag bei geringerem Nährstoff-Input auf ähnlichem

Beschreibung: - Randomized complete Block design
- 5 Versuchsglieder:
A: Keine Unterfußdüngung (UFD) „0% UFD“
B: 150kg/ha DAP UFD Bandablage „100% Band“
C: 75kg/ha DAP UFD Bandablage „50% Band“
D: 150kg/ha DAP UFD Fertisplot „100% Spot“
E: 75kg/ha DAP UFD Fertisplot „50% Spot“
- vier Wiederholungen
- Ansonsten Gleichbehandlung (Ceteris paribus)

Titel*: Fertisplot 2024

Varianten planen

- Fokus auf einfaktorielle Versuche, dann auf 2 Faktoren erweitern (Splitplot), 3 Faktoren ermöglicht werden
- Faktorstufen (Varianten) festlegen
- Betriebsübliche Variante einbeziehen
- Einheiten einheitlich festlegen
- Wichtig: Varianten praxistauglich sein – nicht automatisierbar
- Grundsätzlich Varianten mit KI-Sprachsupport möglich

The screenshot shows a software interface for planning experimental variants. A red stamp in the top right corner reads "Teilweise automatisierbar". The interface is divided into several sections:

- Fragestellung***: Kann eine portionierte und punktuelle Düngung mit Amazone Fertisplot den Ertrag bei geringerem Nährstoff-Input erhöhen, im Vergleich zu einer gleichmäßigen Bandablage?
- Beschreibung**:
 - Randomized block design
 - 5 Versuchsglieder
 - A: Keine Unterfußdüngung (UFD) „0% UFD“
 - B: 150kg/ha DAP UFD Bandablage „100% Band“
 - C: 75kg/ha DAP UFD Bandablage „50% Band“
 - D: 150kg/ha DAP UFD Fertisplot „100% Spot“
 - E: 75kg/ha DAP UFD Fertisplot „50% Spot“
 - vier Wiederholungen
 - Ansonsten Gleichbehandlung (Ceteris paribus)
- Titel***: Fertisplot 2024
- Varianten***:

Varianten	Düngung	kg/ha	Icon
A 0% UFD	0 kg/ha	Icon	
B 100% Band	150 kg/ha	Icon	
C 50% Band	75 kg/ha	Icon	
D 100% Spot	150 kg/ha	Icon	
E 50% Spot	75 kg/ha	Icon	

At the bottom, there is a plus sign (+) and a refresh icon.

Beispiel mit KI-Planung

Teilweise
automatisierbar

2.

Variante bearbeiten ✕

Variantenname*

A

Variante

50 kg N/ha

Einheit

kg (Kilogram) / ha (Hectare)

☒ Apply trial type filter

✓ DDI Export Supported

Menge

50

Farbe*

#33ff66

Abbrechen Speichern

1.

1 Versuchsfrage

Versuchstyp

Fragestellung*

Düngeversuch

Wie wirkt sich eine reduzierte N-Gabe im WW auf den Ertrag aus?

Erzeuge 3 Varianten A, B, C. Variiere von A: 50kg/ha bis C: 85 kg/ha

Düngeversuch

3.

Düngeversuch

● A 50 kg N/ha

50 kg/ha

● B 75 kg N/ha

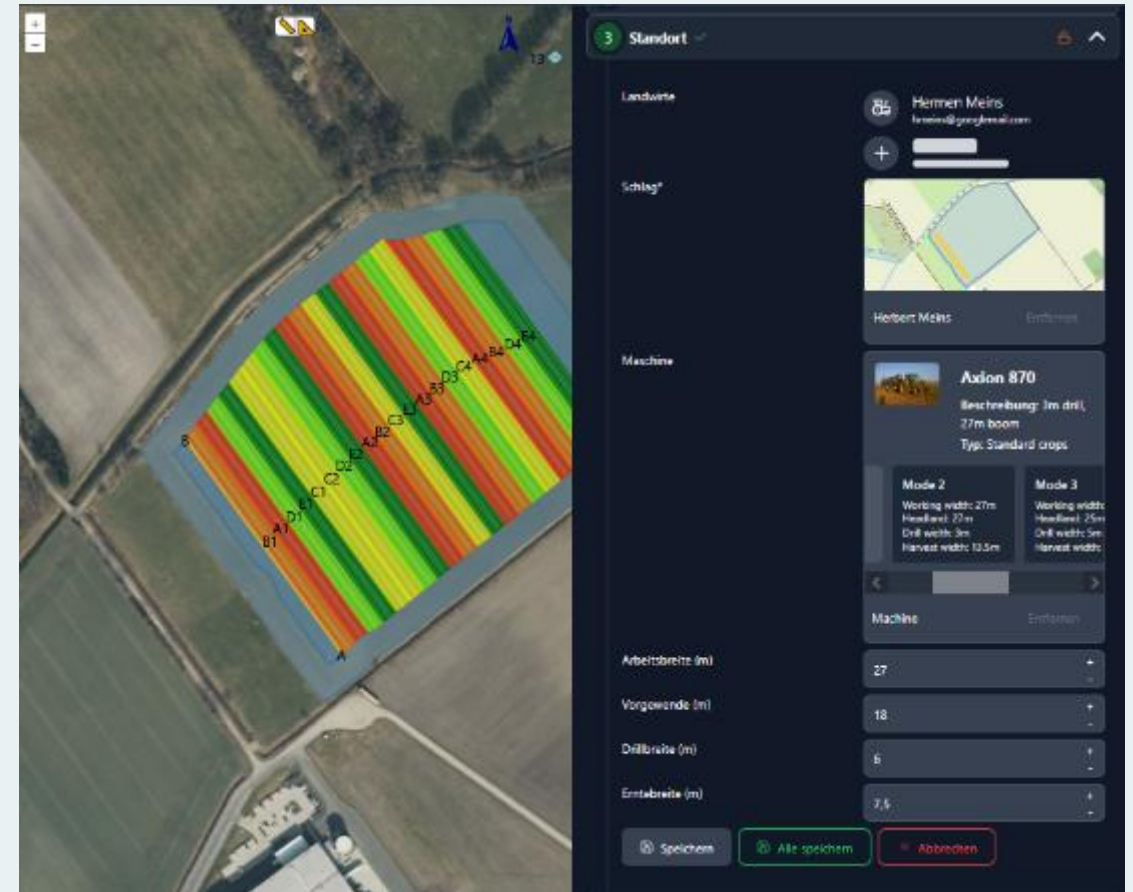
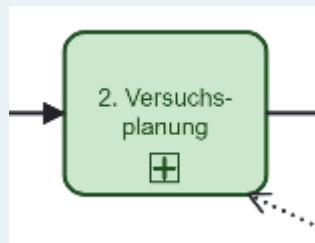
75 kg/ha

● C 80 kg N/ha

80 kg/ha

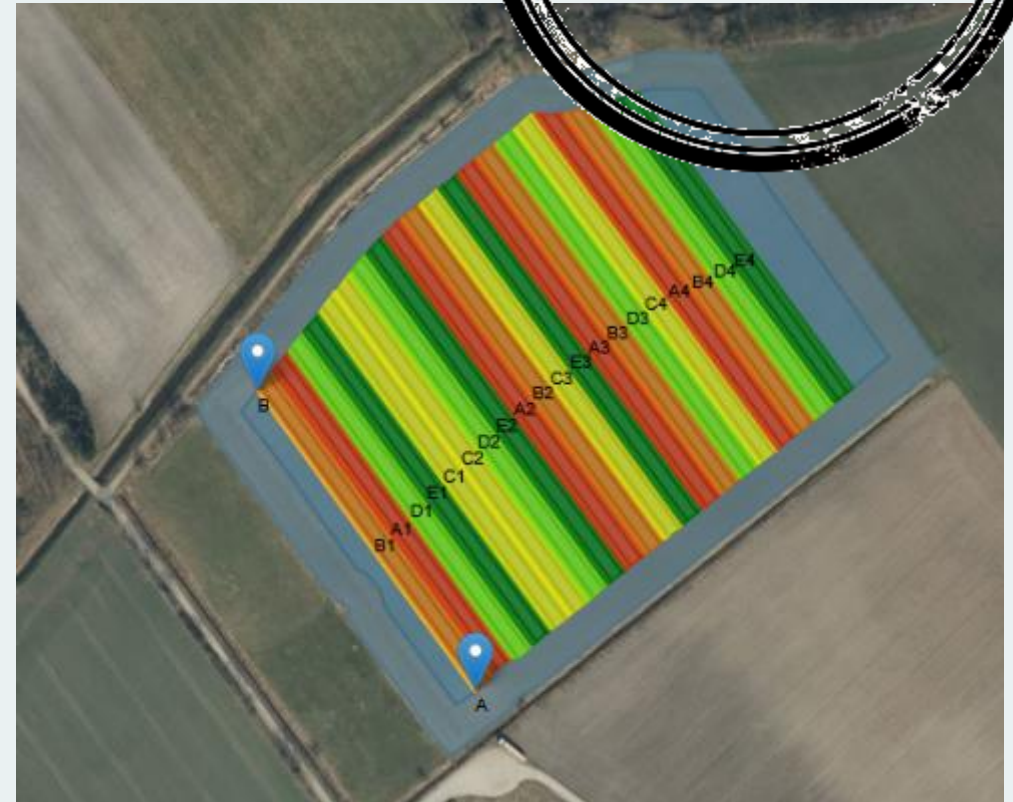
+

Versuchsplanung

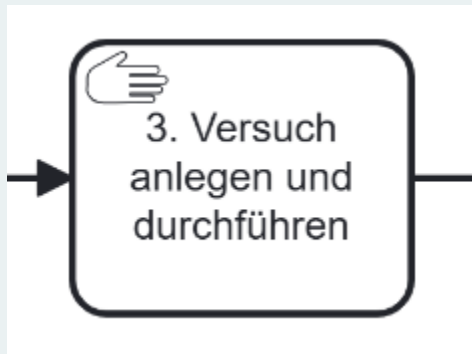


Planung: immer randomisierte, vollständige Blockanlage

- Wiederholungen
- Blockanlage: vollständig
- Immer Randomisierung!
- Exakte Planung auf Grund von real existierenden Lenkspuren!
- Planung auf Fuhrpark optimieren, Aufwand für Praktiker minimieren
- Teilflächen berücksichtigen
- Versuchsprinzipien möglichst einfach einhalten!

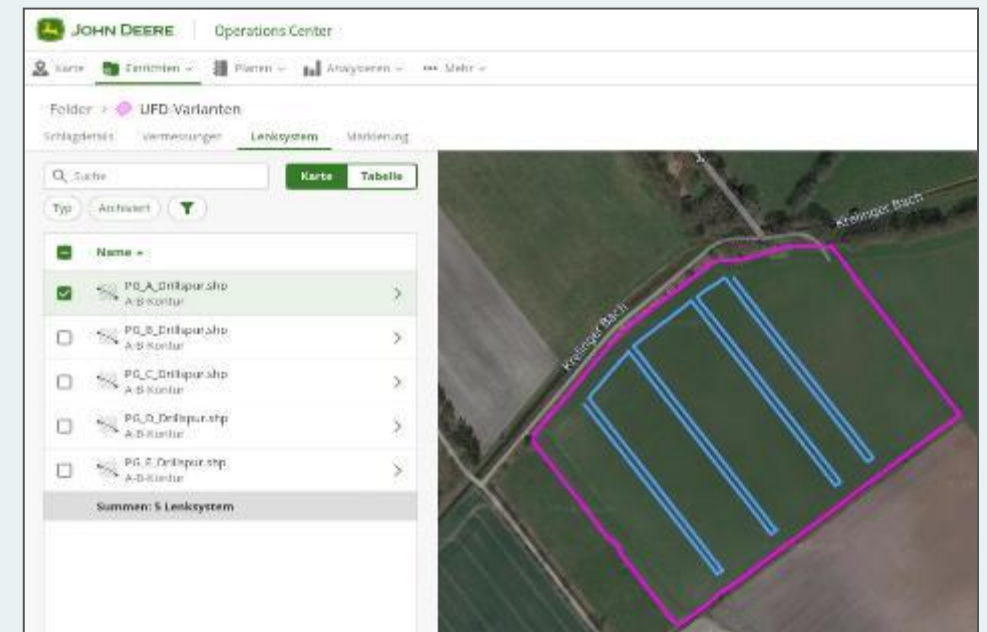


Versuchsanlage

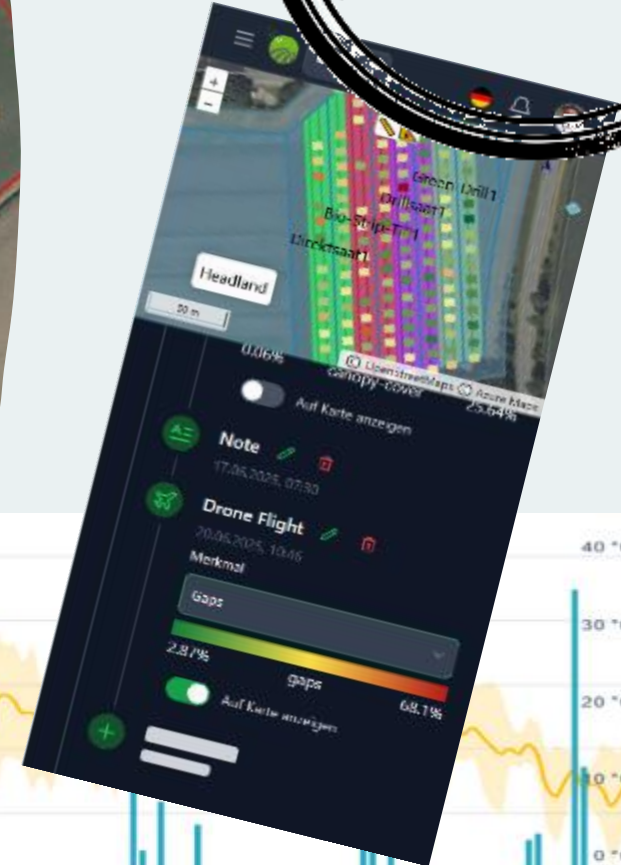
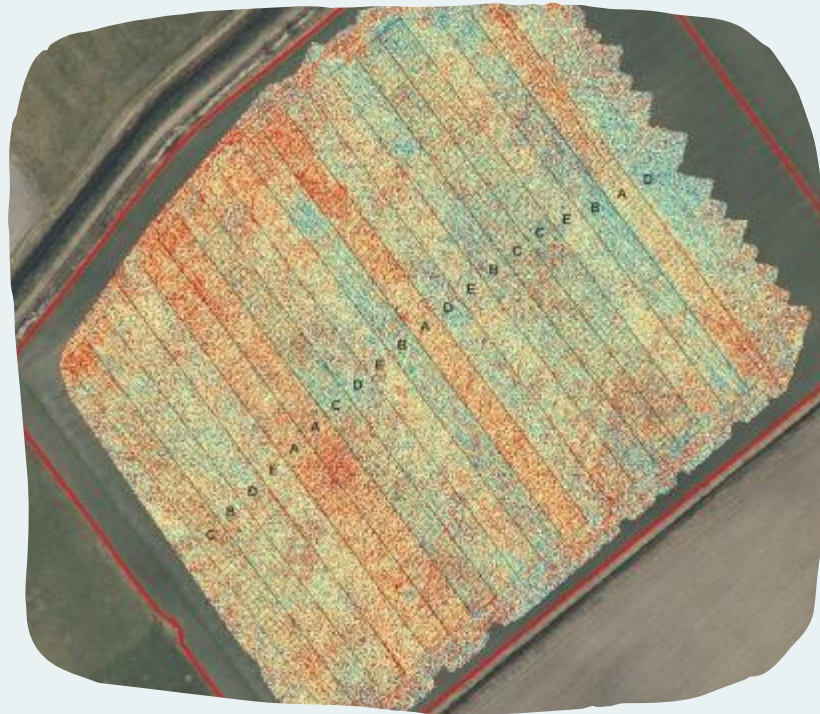


Versuchsanlage

- Ziel: möglichst einfache Versuchsanlage mit vorhandenen Maschinen
- Exakte Versuchsanlage durch genaue Verortung
- Parzellen und Lenkspuren exportieren
- Agrar-Schnittstellen bedienen, sowohl für Zugmaschine, als auch Arbeitsgerät
- Applikationskarte als Standard (ISOXML)
- Automatisierte Datenübertragung z.B. mit **Agrirouter** oder JD OpCenter



Bonituren



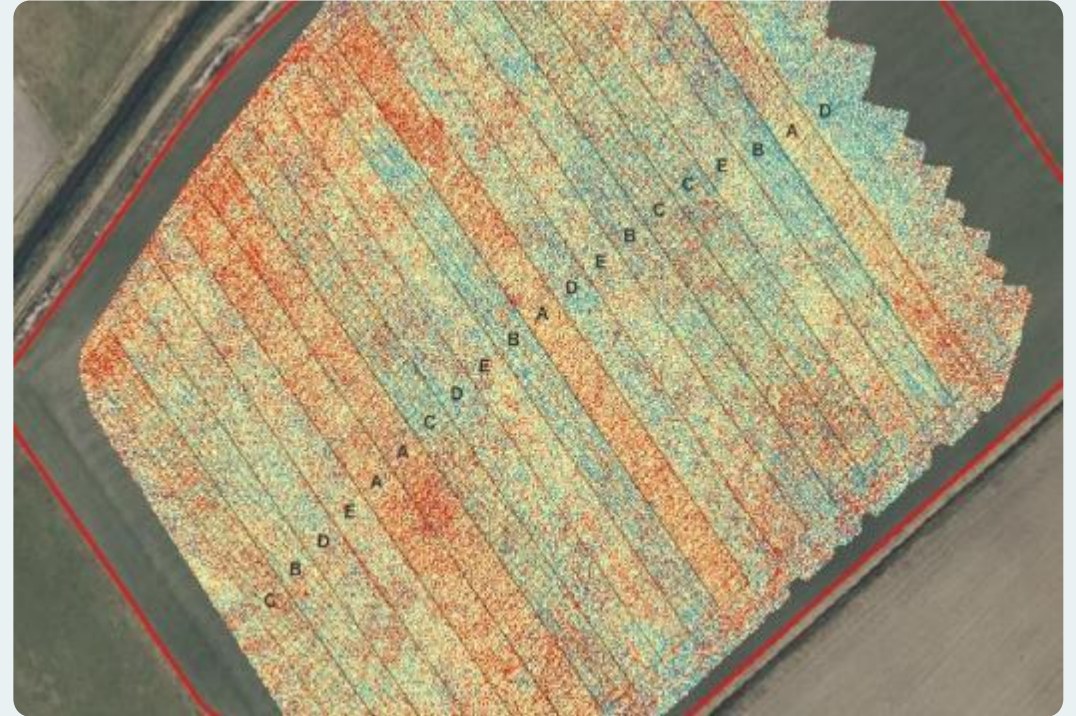
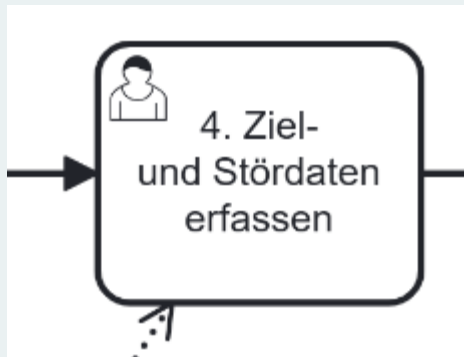
Dokumentation

- Ziel: Schlagdokumentation einbinden
- Beleg Ceteris-Paribus
- Bestätigung durch As-Applied-Karten (ISOXML)
- Kontext & Metadaten
- Bei Cloud-Anbindung gut automatisierbar



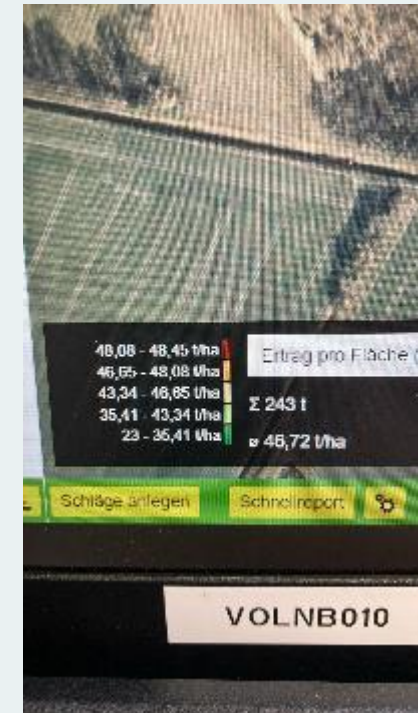
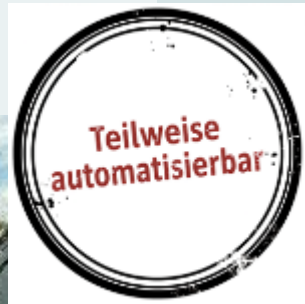
```
><DVC A="DVC-529" B="Horsch Sprayer TC-SC/GEO" C="1.15.1" D="a00c8100374ad1b"
...
</DVC>
<TSK A="TSK-529" B="Epso Profi und Bortop 03.2025" C="CTR-74" D="FRM-75" E=
<CTR A="CTR-529" B="Broder Preuß-Driessen"/>
<FRM A="FRM-529" B="Gut Grünholz" I="CTR-529"/>
<PFD A="PFD-529" C="Brandholz" E="CTR-529" F="FRM-529"/>
><TSK A="TSK-530" B="Sprühen" C="CTR-529" D="FRM-529" E="PFD-529" G="2" H="0"
...
</TSK>
<PFD A="PFD-530" C="11 Tonnen" E="CTR-74" F="FRM-75"/>
><TSK A="TSK-531" B="Naehrstoffe" C="CTR-74" D="FRM-75" E="PFD-530" G="2" H="0" I="0">
...
</TSK>
<TSK A="TSK-532" B="Nährstoffe" C="CTR-74" D="FRM-75" E="PFD-87" G="2" H="0" I="0"/>
><TSK A="TSK-533" B="Pisan" C="CTR-74" D="FRM-75" E="PFD-130" G="2" H="0" I="0">
...
</TSK>
><TSK A="TSK-534" B="Piasan" C="CTR-74" D="FRM-75" E="PFD-79" G="2" H="0" I="0">
  <TLG A="TLG00033" C="1"/>
  <TLG A="TLG00034" C="1"/>
  <TLG A="TLG00035" C="1"/>
  <TLG A="TLG00036" C="1"/>
  <TLG A="TLG00037" C="1"/>
  <TLG A="TLG00038" C="1"/>
  <TLG A="TLG00039" C="1"/>
  <TLG A="TLG00040" C="1"/>
  <TLG A="TLG00041" C="1"/>
  <TLG A="TLG00042" C="1"/>
  <TLG A="TLG00043" C="1"/>
  <TLG A="TLG00044" C="1"/>
  <TLG A="TLG00045" C="1"/>
  <TLG A="TLG00046" C="1"/>
</TSK>
<TSK A="TSK-535" B="Refine " C="CTR-74" D="FRM-75" E="PFD-88" G="2" H="0" I="0"/>
><TSK A="TSK-536" B="Refine" C="CTR-74" D="FRM-75" E="PFD-530" G="2" H="0" I="0">
  <TLG A="TLG00049" C="1"/>
</TSK>
<TSK A="TSK-537" B="Upton 0,1/ ha." C="CTR-74" D="FRM-75" E="PFD-87" G="2" H="0" I="0"/>
<TSK A="TSK-538" B="TEBU25" C="CTR-74" D="FRM-75" E="PFD-97" G="2" H="0" I="0"/>
<TSK A="TSK-539" B="TEBU 25" C="CTR-74" D="FRM-75" E="PFD-76" G="2" H="0" I="0"/>
><TSK A="TSK-540" B="Tribenuron /Fungizid" C="CTR-74" D="FRM-75" E="PFD-530" G="2" H="0" I="0">
  <TLG A="TLG00047" C="1"/>
  <TLG A="TLG00050" C="1"/>
  <TLG A="TLG00051" C="1"/>
</TSK>
</ISO11783_TaskData>
```

Datenmangement: Ziel- und Stördaten erfassen

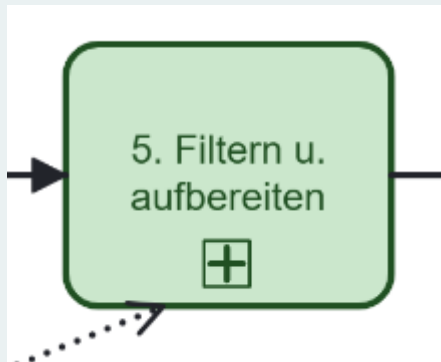


Zielerfassung

- In den meisten Fällen Beerntung
- Unterstützend z.B. Multispektral-Indizes
- Manuell oder Cloud-to-cloud
- Maschinen-Hersteller-Formate vs. Standards (ISOXML Timelog)
- Datenbesonderheiten? Erfahrung!
- Datenverfügbarkeit? EU Data Act!
- Kalibrierung? Datenqualität?
- **Metadaten!**

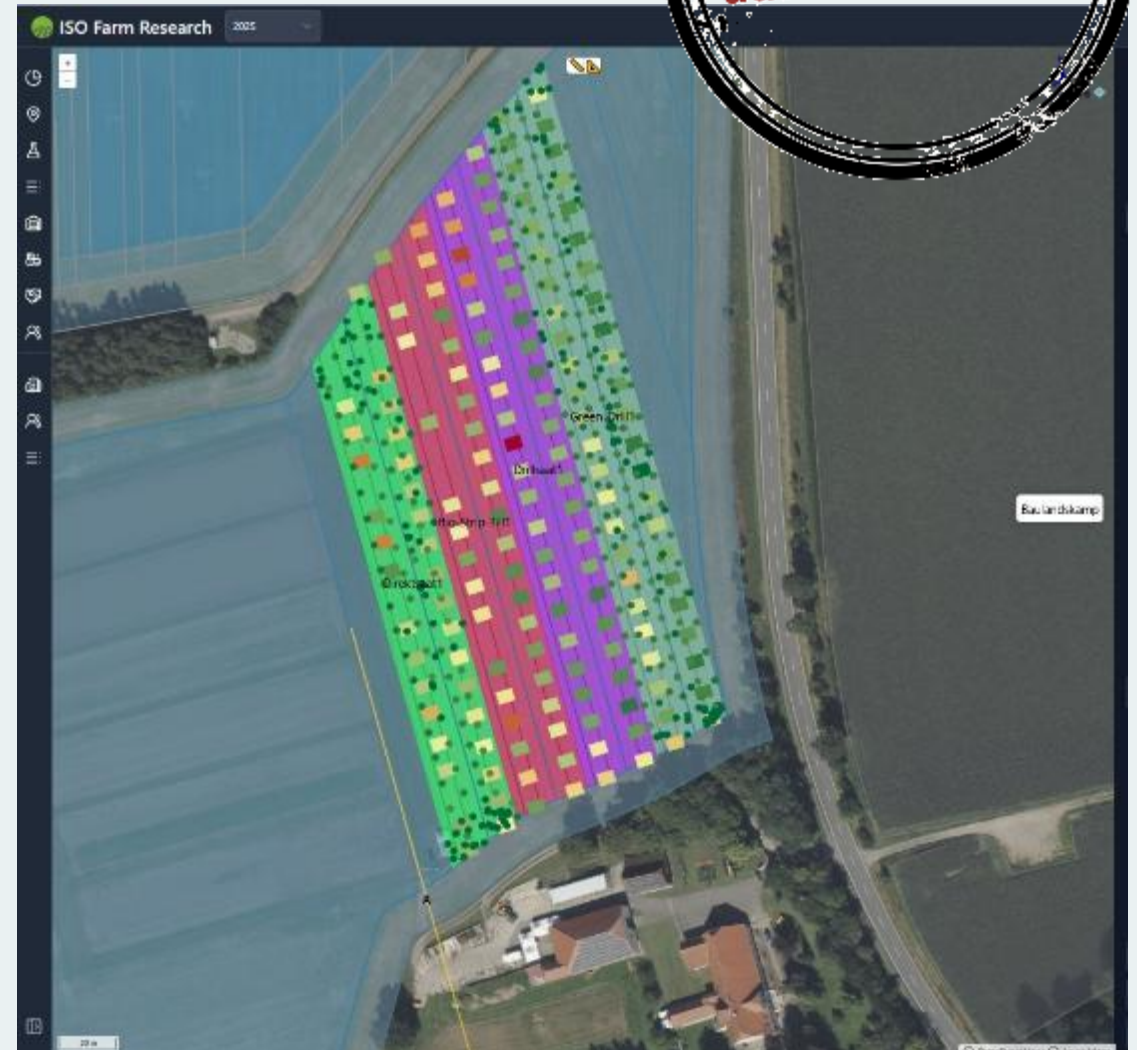


5. Filtern und aufbereiten



Datenaufbereitung

- Voranzeige
- Filtern: Standardfilter
- Plausibilisieren manuell
- Standardisiertes Verknüpfen
- Ko-Variablen nach Bedarf
- Standardvorgehen erleichtert Automatismen
- Ziel: **Gesamtmatrix** inklusive Versuchsdaten, Parzellen, Zieldaten und Störgrößen



- 🕒 Übersicht
- 📍 Karte
- 🧪 Versuche
- ☰ Schläge
- 🏠 Betriebsprofil
- 🚜 Maschinen
- 🤝 Vertragspartner
- 👤 Mitarbeiter
- 🏢 Unternehmen
- 👤 Benutzer
- ☰ Kulturen



1 Zielgröße(n)

Ertragskartierung

1 selected

Load

Definition

[DRYMATTER]

2 Filtern

Based on*

☒ Trial ☐ Block ☐ Plot ☐ Variant

Histogram

Yield Range	Frequency
0-5	3
5-10	6
10-15	12
15-20	20
20-25	14
25-30	9
30-35	5
35-40	2
40-45	3
45-50	1
50-55	1
55-60	0

Download results

Download CSV

3 Kovariablen

4 Modellauswahl

5 Varianzanalyse

6 Ergebnis



1 Zielgröße(n)

Ertragskartierung

1 selected

Load

Definition

[DRYMATTER]

2 Filtern

Based on*

☐ Trial ☐ Block ☒ Plot ☐ Variant

Plots

2 selected

- Block 1 - Ohne (betr.)1
- Block 1 - Mit Naturamin (300g/ha)1
- Block 2 - Mit Naturamin (300g/ha)2
- Block 2 - Ohne (betr.)2
- Block 3 - Mit Naturamin (300g/ha)3
- Block 3 - Ohne (betr.)3

0 0-5 5-10 10-15 15-20 20-25 25-30 30-35 35-40 40-45 45-50

Download results

Download CSV

3 Kovariablen

Zusätzliche Ko-Variablen laden

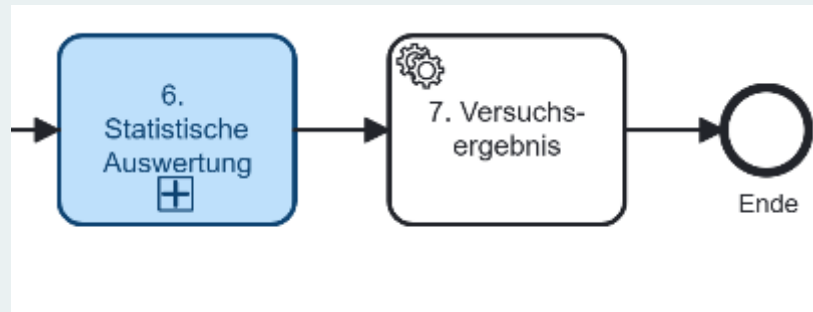
Zum Hochladen klicken oder per Drag & Drop

zip, xml

Bonituren laden

Select samplings

5. Auswertung und Darstellung

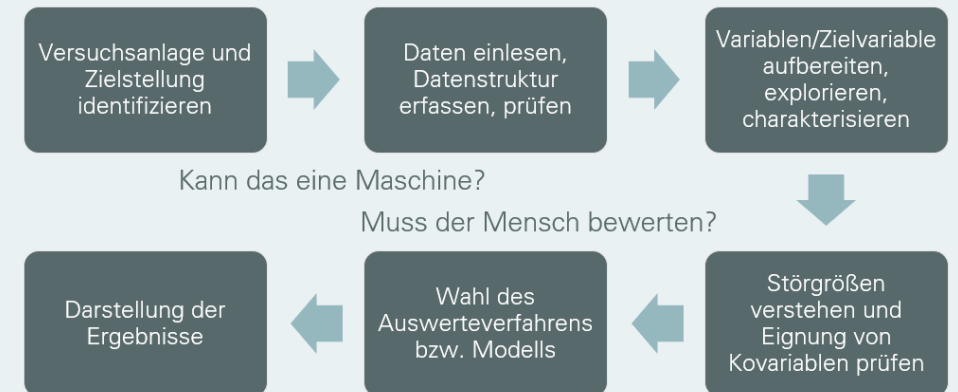


Effiziente Auswertung der Gesamtmatrix



Fazit

- ISOFarmResearch ein Werkzeug für den OFE-Einsatz
- Vor allem Planung und Datenmanagement
- Auch Kommunikation und Übersicht leichter
- Räumlich-zeitliche Daten ermöglichen saubere geostatistische Verfahren
- Viele Arbeitsschritte zumindest teilweise automatisierbar.
- Standards?
- Vorteil: Prozessschritte analysieren



Gesamtfazit

- Ein standardisiertes Vorgehen und eine saubere Datenvorbereitung erleichtert die Auswertung enorm
- Möglichkeiten: Einzelschritte sind gut automatisierbar (zumindest teilweise)
- Grenzen: Immer wieder ist manueller Eingriff und Expertenwissen gefragt. Bewertung und Entscheidung durch Menschen bleibt unverzichtbar.
- „Totalautomatisierung“ nicht machbar – aber auch nicht nötig. Es geht um mehr Effizienz!



Anhang - Diskussionspunkte

- Kostenrechnung Versuch
- Wieviel Bonitur braucht ein Versuch (Kosten/Nutzen)
- Brutto- vs. Nettoparzelle
- Offene Themen Digitalisierung
- Standardisierung