

FAIRagro *use case*

„Learning from incomplete data“

Prof. Dr. Gunnar Lischeid

Leibniz Institute for Agricultural Landscape Research (ZALF)

lischeid@zalf.de

Photo: Petair / Fotolia

Generelles Forschungsziel:

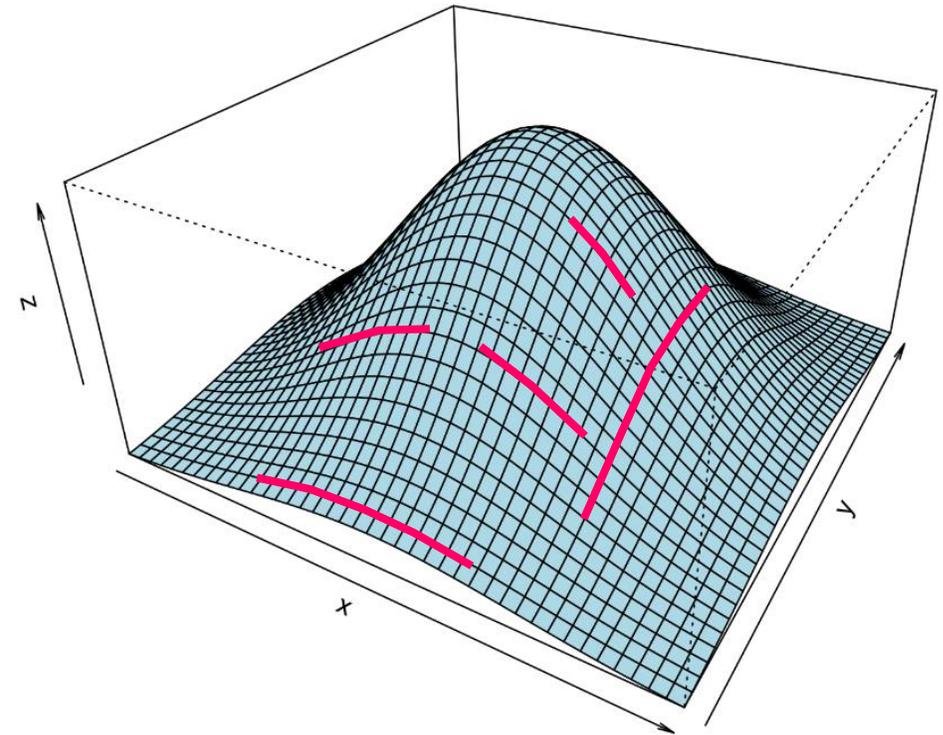
- Weiterentwicklung von Ertragsmodellen unter Einbeziehung zusätzlicher Faktoren und Abhängigkeiten

Forschungsfrage konkret:

- Wie können die „Informationssplitter“ **unvollständiger Datensätze** (im Sinne fehlender Messgrößen) in wiss. Repositorien systematisch genutzt werden?

Verschneidung zweier Ansätze:

- **Qualitativ**: Erfassung weiterer, über die Metadaten hinausgehender Kontext-Informationen durch automatisiertes Text-Mining der mit den Daten verknüpften Publik
- **Quantitativ**: Schrittweise Berücksichtigung unvollständiger Datensätze und deren Kontext-Informationen für die Ertragsmodellierung (z.B. mittels Kalman-Filter)



Disziplinen:

- Pflanzenproduktion
- Bodenkunde
- Pflanzenernährung
- Phytomedizin
- Züchtung
- Agrar-Klimatologie
- Ertragsmodellierung
- ...

Skalen > Einzelpflanze:

- Feld
- Betrieb
- Landschaft
- Region

- Weiterentwicklung von Ertragsmodellen unter Einbeziehung zusätzlicher Faktoren und Abhängigkeiten
- Rückkopplung zu Forschungsdateninfrastrukturen hinsichtlich der Bereitstellung von Metadaten und sonstigen Kontextinformationen (Maschinen-lesbare Schnittstellen, ...)
- Methode zur effizienten Nutzung unvollständiger Datensätze (= typisch für Daten in Repositorien) für Ertragsmodelle, und somit zur Nachnutzung der Daten über die ursprünglichen Fragestellungen hinaus
- Ansatz zur systematischen Verschneidung von Daten und Fragestellungen verschiedener Disziplinen im Bereich der Pflanzenproduktion

Was existiert bereits?

- Vielzahl publizierter, oft kleiner Datensätze mit entsprechenden Metadaten
- Publikationen zu den publizierten Datensätzen
- Modelle zur Ertragsbildung (konzeptionelle, biophysikalische, statistische, ML/KI-Modelle)

Was wird im Use Case realisiert (outcomes)?

- (Verfahren für eine automatisierte Bereitstellung von) weitergehenden Kontextinformationen der publizierten Daten (über die Metadaten hinaus)
- Klassifizierung/Bewertung publizierter Datensätze hinsichtlich der jeweils wirkenden „Constraints“ der Ertragsbildung
- Ergebnisse zu den Ursachen der räumlichen und zeitlichen Variabilität der Erträge => Hinweise auf kritische Wechselwirkungen verschiedener Effekte
- Methodische Basis für eine engere Zusammenarbeit unterschiedlicher Disziplinen

Vorgesehen:

- Ertragsmodellierer(innen) (ZALF, ...)
- Versuchsanstalten / Versuchsbetriebe
- Global Long-Term Agricultural Experiment Network (**GLTEN**; <https://glten.org>)
- European Joint Programme Cofund on Agricultural Soil Management (**EJP Soil**; <https://ejpsoil.eu>)
- ...
- UC1, UC2, UC3, UC6

Defizite auf Seiten der Wissenschaft:

- Bisher: Modelle definieren die Ansprüche an die Daten
- Demnächst: Neue Ansätze ermöglichen die Nutzung vorhandener, nicht vollständiger Daten durch Verschneidung von weitergehender Kontextinformation und Modellierung

Defizite auf Seiten der Datenbereitstellung:

- Nicht genügend leistungsfähige Portale zur **Recherche** von Daten
- Unzureichende **Metadaten** über den jeweiligen **Kontext** der Daten