SMIS: Soil Management Information System

Machine learning application in soil management case study

16 January 2025

Dr Tomasz Kurowski
t.j.kurowski@cranfield.ac.uk

www.cranfield.ac.uk

SMIS - Soil Management Information System

Jane Rickson (PI),
Rob Simmons (Sustainable Soil Management),
Fady Mohareb (Data scientist),
Steve Hallett (Information Systems Management),
Caroline Keay (LandIS data),
Jason Carvalho (Geospatial developer),
Lynda Deeks (BBSRC/NERC/HIP Horticulture Knowledge Exchange Fellow),
Tomasz Kurowski (Postgraduate Researcher),
Jo Niziolomski (Postdoctoral Researcher)

The Soil Management Information System (SMIS) is a software platform which aims to improve the assessment and management of soil in horticulture, thereby supporting sustainable crop production and environmental protection.

Funded by the AHDB Horticulture and PGRO

Primarily funded by AHDB

Levy board funded by farmers and growers.
Deals in R&D and farm-level knowledge transfer.

Statutory goals:

- *Increasing efficiency or productivity in the industry.*
- *Improving marketing in the industry.*
- *Improving or developing services that the industry provides or could provide to the community*
- *Improving the ways in which the industry contributes to sustainable development.*

SMIS SOIL MANAGEMENT INFORMATION SYSTEM

SMIS ANALYTICS TOOLKIT

AHDB

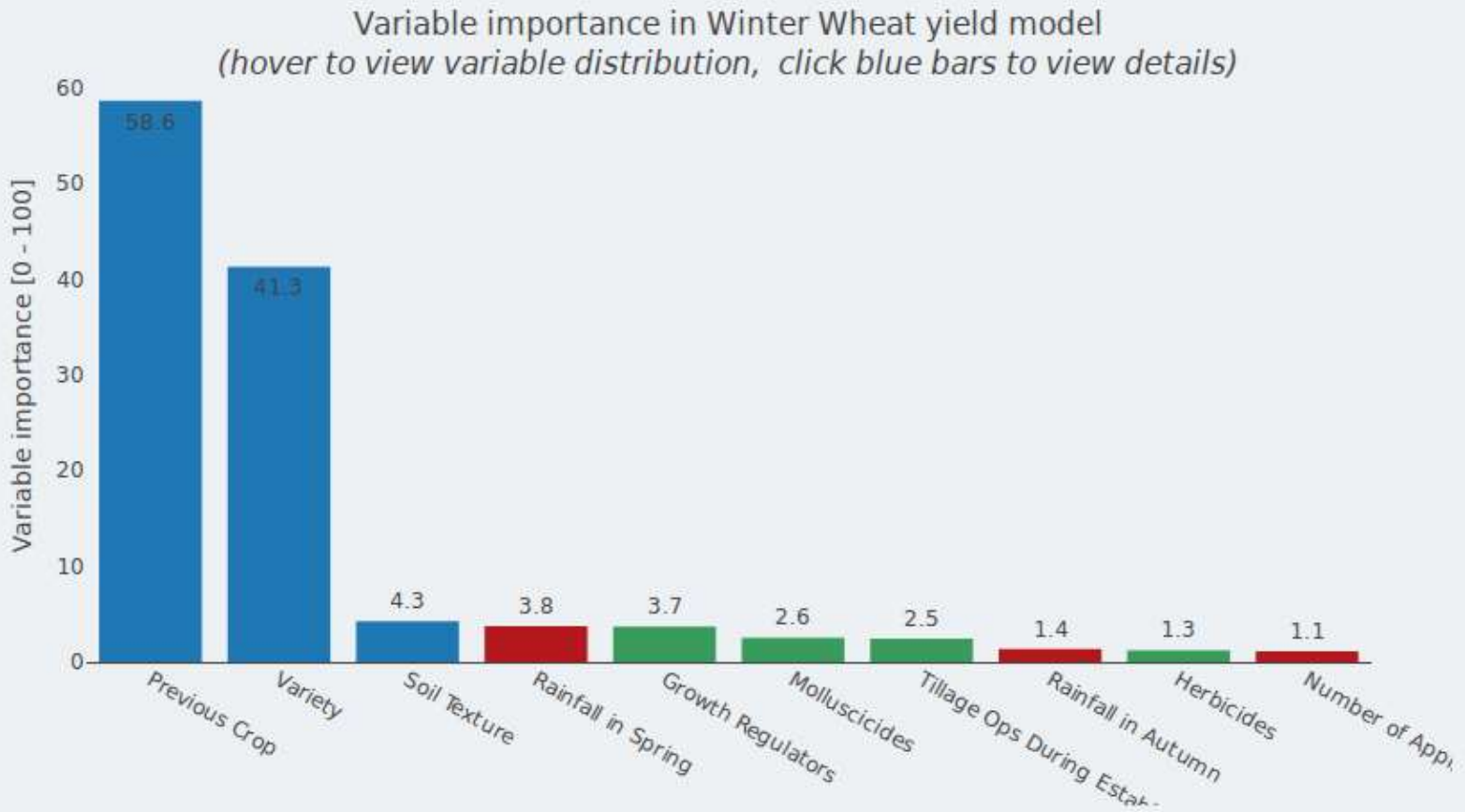
- Browse Database
- Rule Bases
- Established Queries
  - Factors affecting yield
  - Compaction risk
  - Foot rot index
  - PCN level
  - Cavity spot

Factors influencing yield

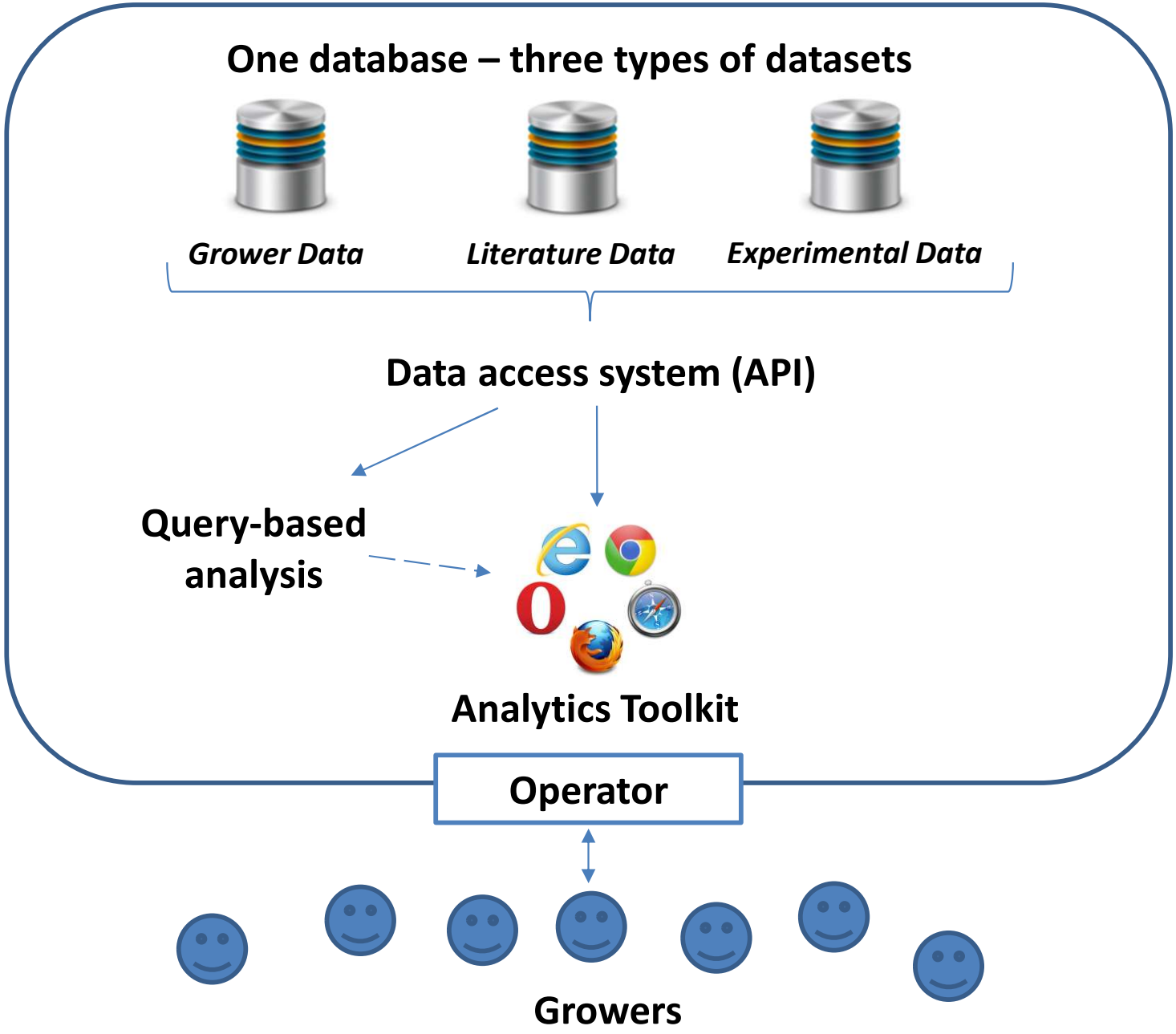
Query Constructor

Crop: Winter Wheat | Add filter | Execute

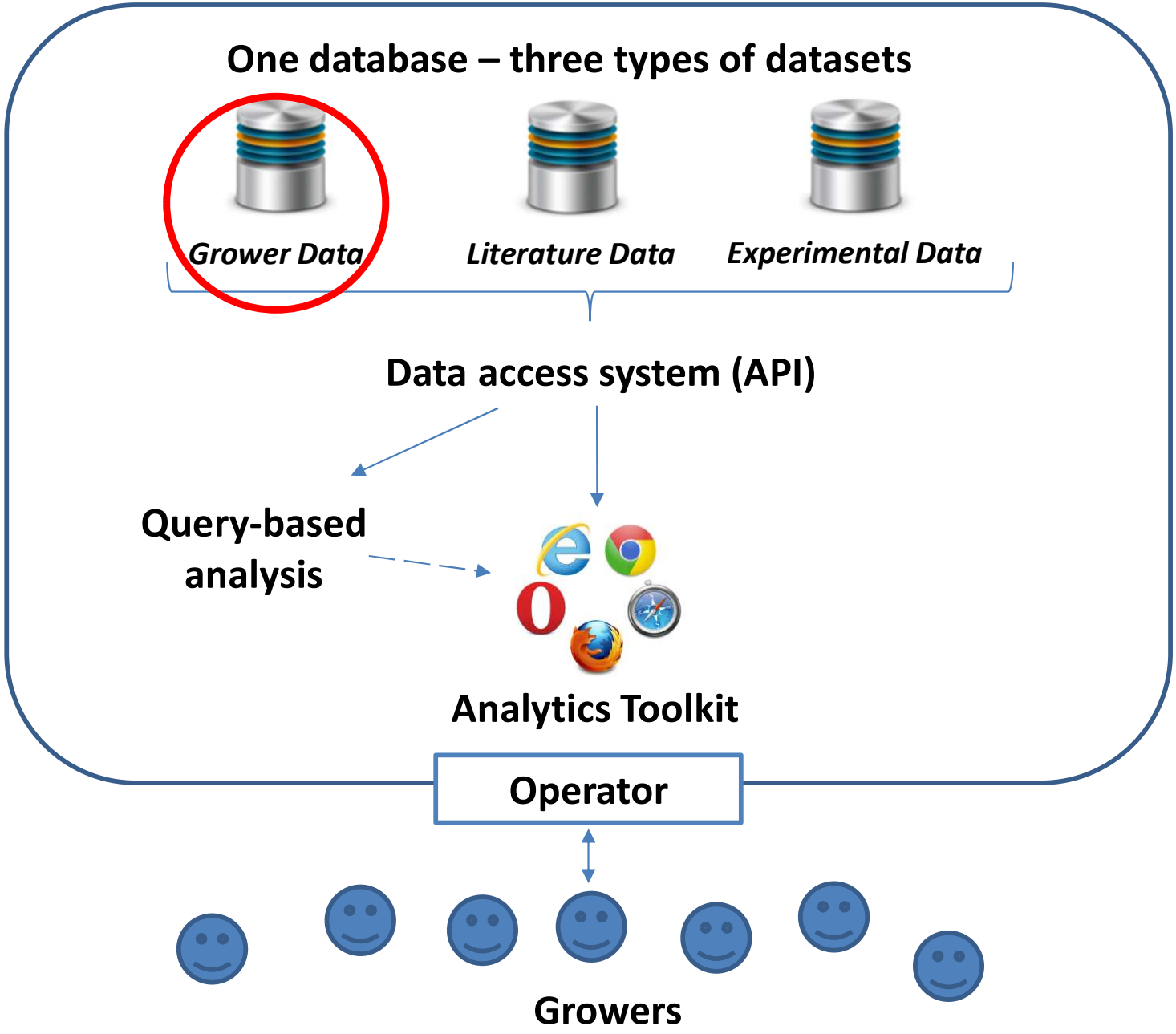
Yield model



General system overview



General system overview



Research data vs Grower data

Research data:

- Many variables recorded and controlled (e.g. soil characteristics), quantitative
- Set up and intended to investigate scientific questions (hypothesis-driven)
- Entered by scientists
- Mostly custom spreadsheets
- Limited volume and coverage

ROTHAMSTED RESEARCH

e.g. Rothamsted Research

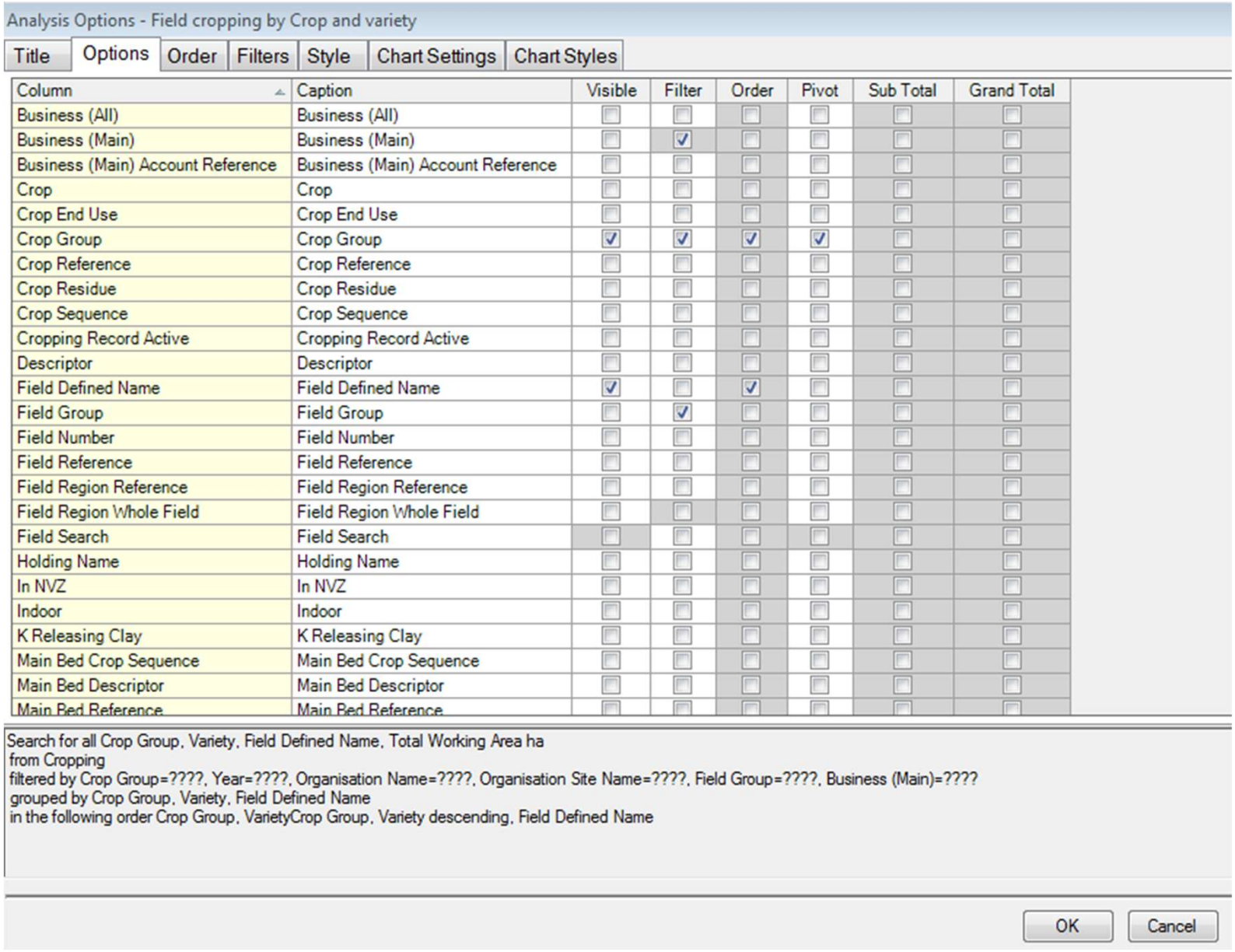
Founded in 1843, over 800 hectares in multiple locations

Research data vs Grower data

Grower data:

- Few variables recorded
- Mostly operations / field applications stored for recordkeeping
- Entered by growers
- Mostly stored using specialised, commercial software (e.g. GateKeeper used by 40% of growers in the UK)
- Very large untapped volume of data!

Grower records



Research data vs Grower data

Research data:

- Many variables recorded and controlled (e.g. soil characteristics), quantitative
- Set up and intended to investigate scientific questions (hypothesis-driven)
- Entered by scientists
- Mostly custom spreadsheets
- Limited volume and coverage

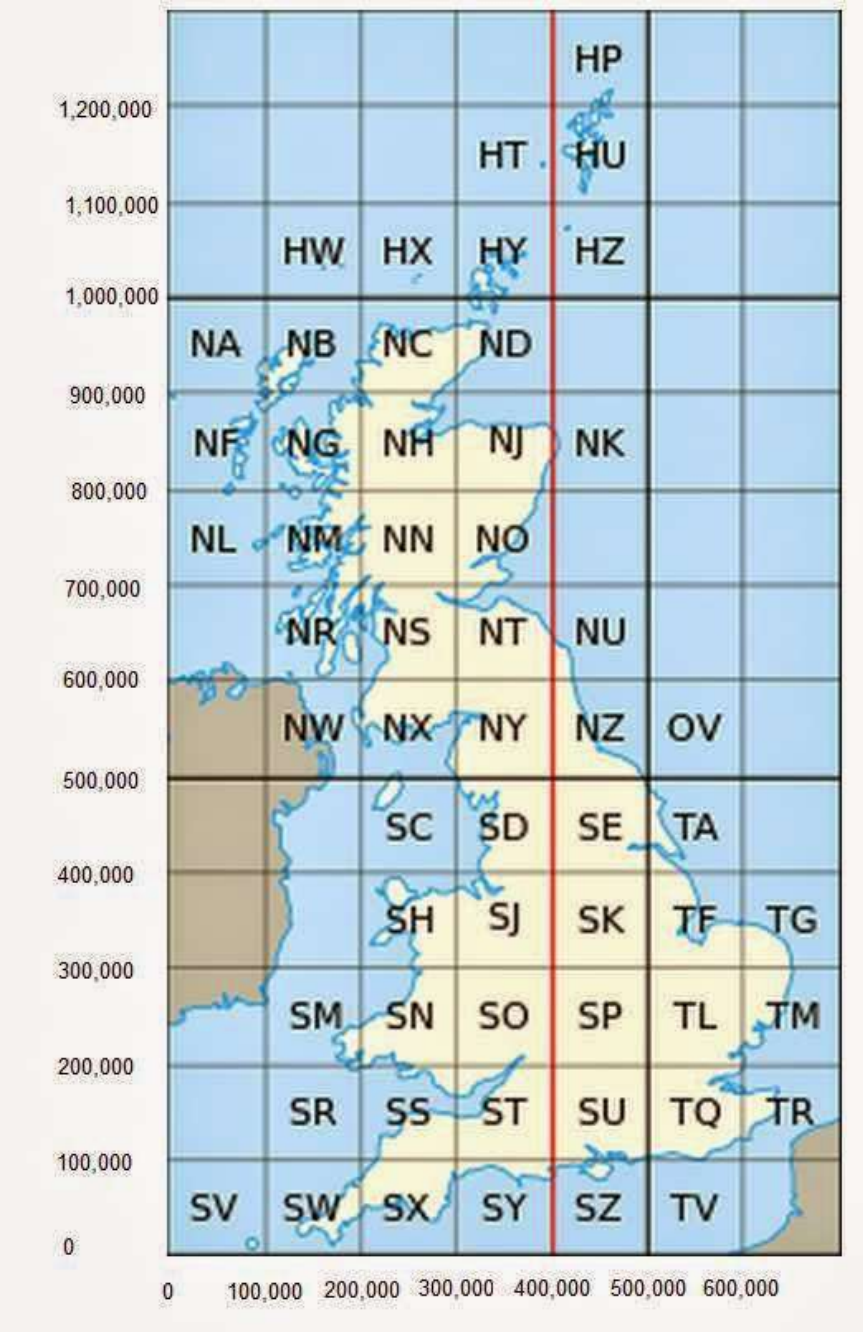
Grower data:

- Few variables recorded
- Mostly operations / field applications stored for recordkeeping
- Entered by growers
- Mostly stored using specialised, commercial software (e.g. GateKeeper used by 40% of growers in the UK)
- Very large untapped volume of data!

Enriching grower data

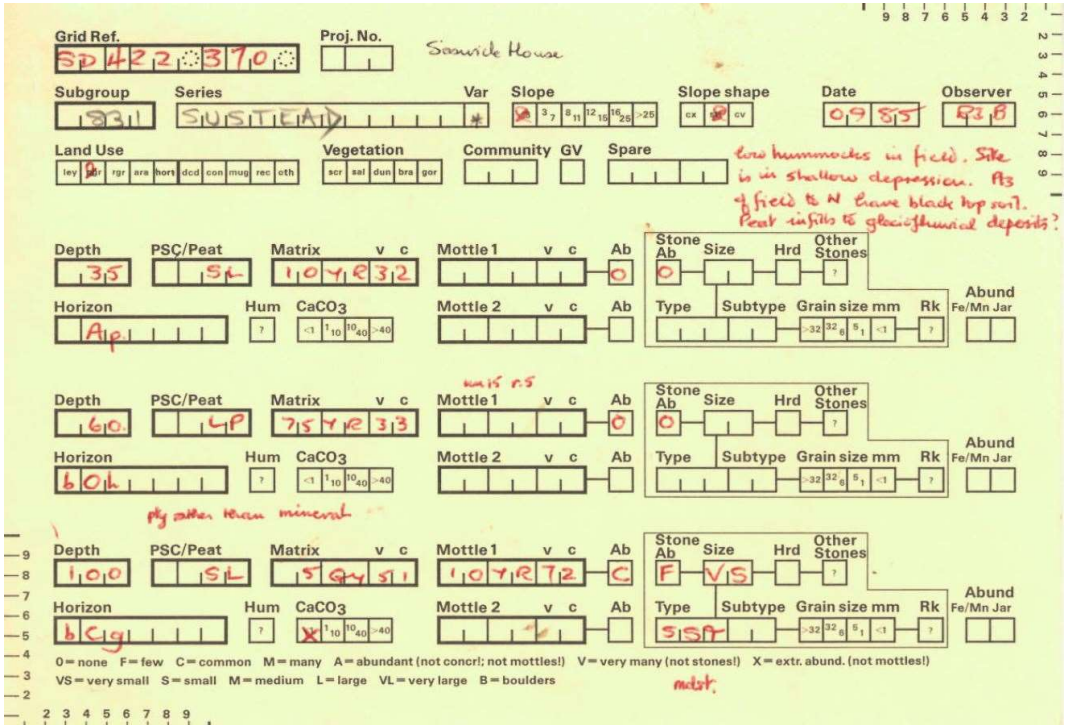
LandIS & Met Office Data

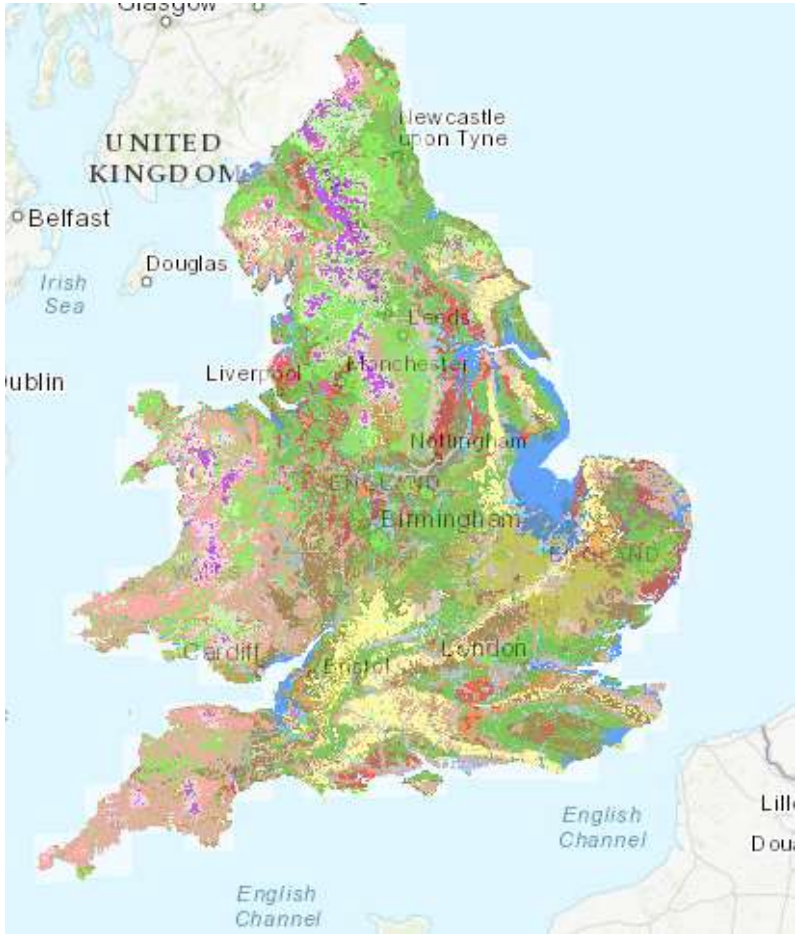
- Cranfield maintains LandIS (Land Information System) containing detailed soil data for the entirety of England & Wales.
- A 1km x 1km grid subset of LandIS has been made available for internal use by SMIS.
- Met data on weather (temperature, rainfall) by time and location is also available and can be matched to fields.



LandIS

- Maintained by Cranfield University (CSAI, Cranfield Soil and AgriFood institute)
- Uses the 1:250,000 scale National Soil Map of England and Wales. (NATMAP)
- Soil Survey of England and Wales (1939 – 1987)





http://www.landis.org.uk/soilscapes/

Grid Ref. SD 42 2 37 0 | Proj. No. | Seamside House

Subgroup 931 | Series SUSTEAD | Var * | Slope | Slope shape | Date 09 85 | Observer RJB

Land Use: ley | arr | rgr | ara | hort | dcd | con | mug | rec | oth

Vegetation: scr | sal | dun | bra | gor | Community | GV | Spare

two hummocks in field. Site is in shallow depression. N3 of field is N have black hop soil. Peat infills to glaciofluvial deposits?

Depth	PSC/Peat	Matrix	v	c	Mottle 1	v	c	Ab	Stone Ab	Size	Hrd	Other Stones
35	SL	10YR	3	2				0	0			?

Horizon: Ap | Hum: ? | CaCO3: <1 | 1 10 | 10 40 | >40 | Mottle 2 v c Ab

Type | Subtype | Grain size mm: >32 | 32 6 | 5 1 | <1 | Rk: ? | Abund Fe/Mn Jar

Depth	PSC/Peat	Matrix	v	c	Mottle 1	v	c	Ab	Stone Ab	Size	Hrd	Other Stones
60	LP	7.5YR	3	3				0	0			?

was no.5

Horizon: bOh | Hum: ? | CaCO3: <1 | 1 10 | 10 40 | >40 | Mottle 2 v c Ab

Type | Subtype | Grain size mm: >32 | 32 6 | 5 1 | <1 | Rk: ? | Abund Fe/Mn Jar

dry other than mineral

Depth	PSC/Peat	Matrix	v	c	Mottle 1	v	c	Ab	Stone Ab	Size	Hrd	Other Stones
100	SL	5GY	5	1	10YR	7	2	C	F	VS		?

Horizon: bCg | Hum: ? | CaCO3: X | 1 10 | 10 40 | >40 | Mottle 2 v c Ab

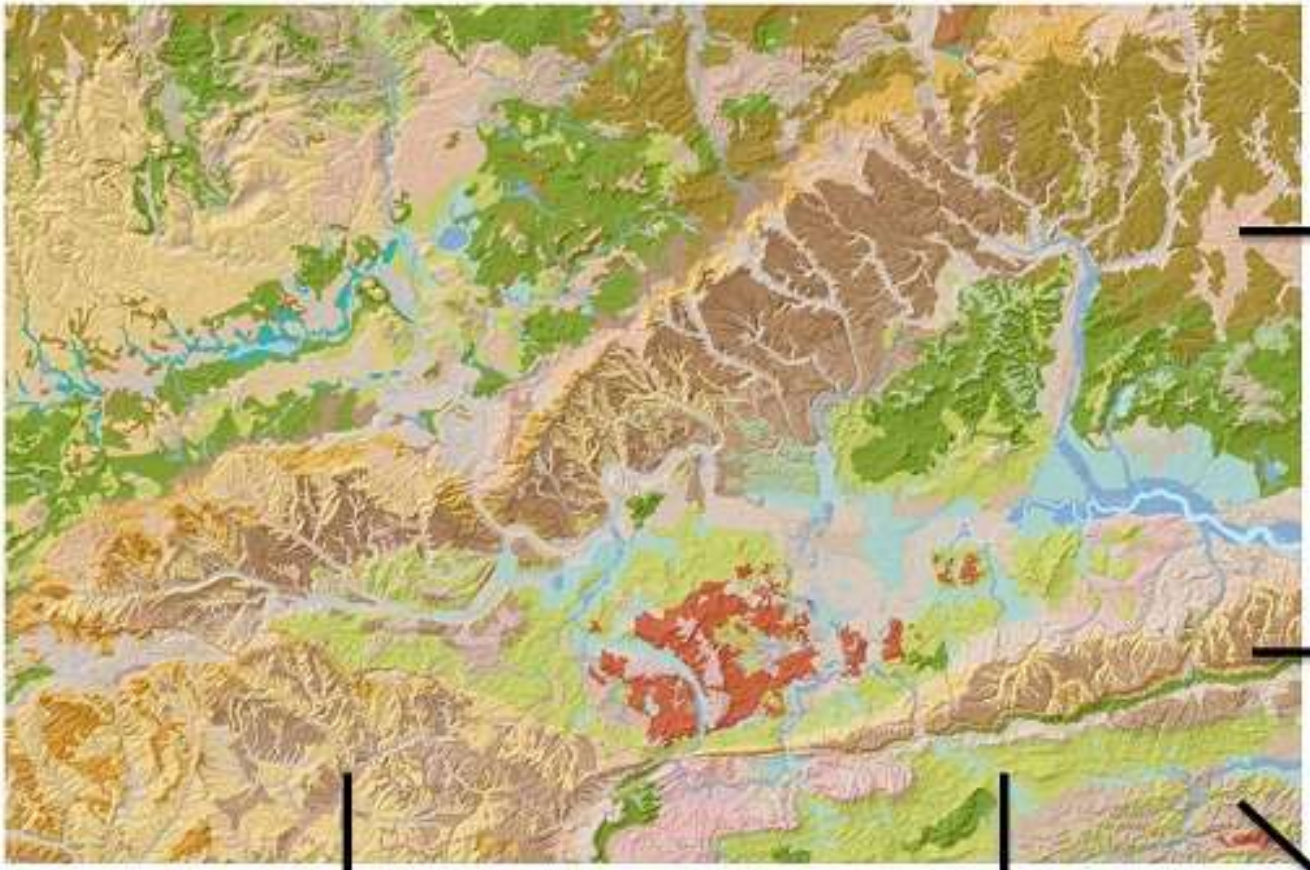
Type: SISA | Subtype | Grain size mm: >32 | 32 6 | 5 1 | <1 | Rk: ? | Abund Fe/Mn Jar

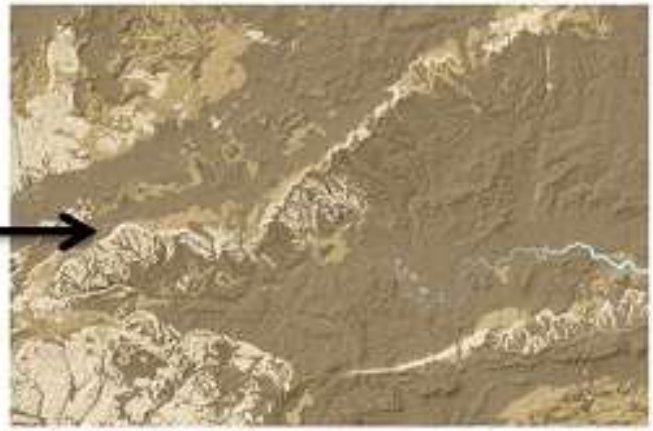
0 = none F = few C = common M = many A = abundant (not concr!; not mottles!) V = very many (not stones!) X = extr. abund. (not mottles!)

VS = very small S = small M = medium L = large VL = very large B = boulders

moist

The National Soil Map

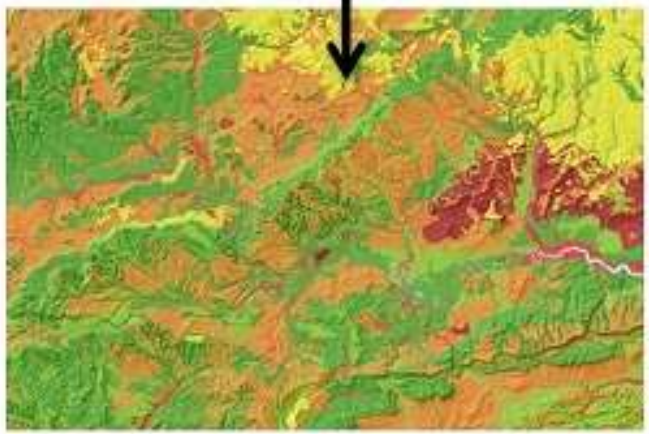




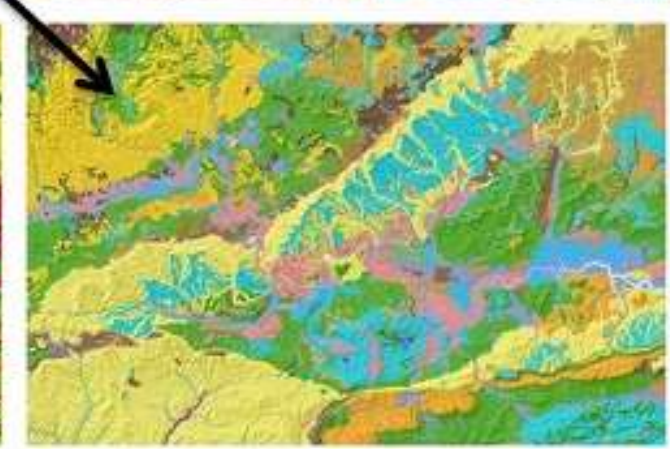
depth

wetness

corrosivity to iron



ground movement



hydrology

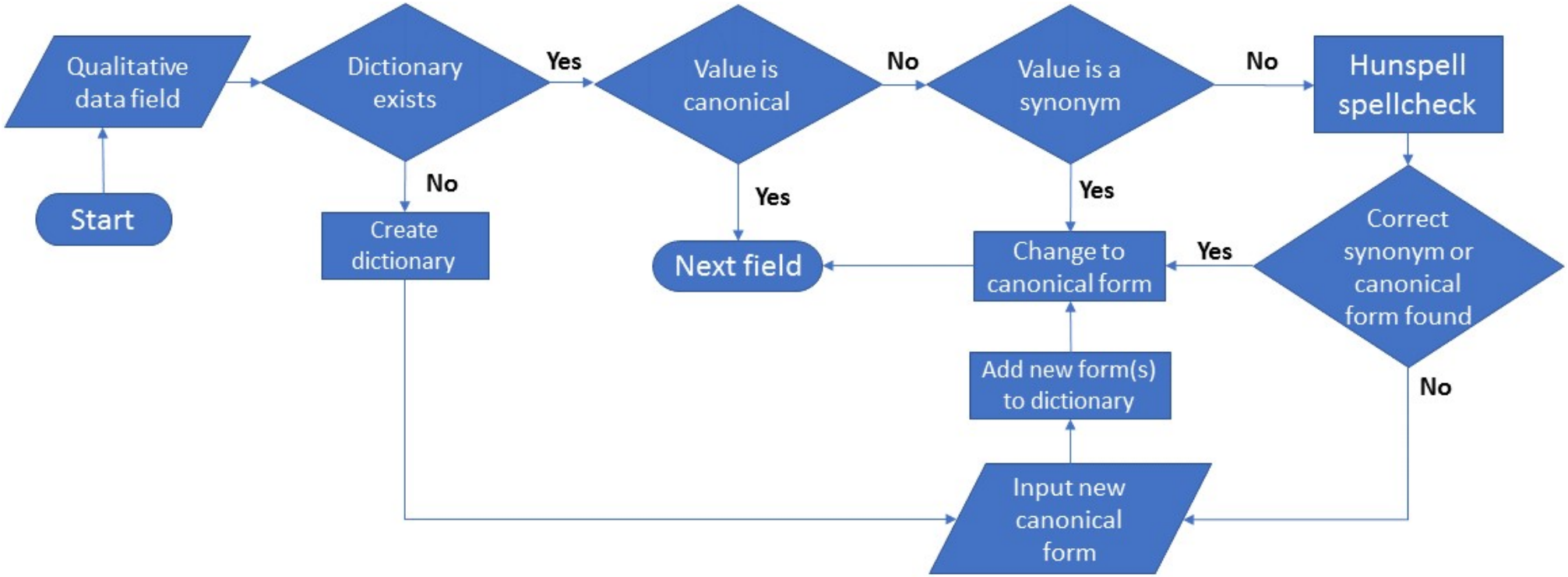
Agri-Informatics

Enriching grower data

Soil Management Issue Inference

- Soil management issues (e.g. compaction) are not recorded, but standard preventive measures (e.g. deep tillage) are!
- These could be used to identify (based on expert knowledge) the occurrence of soil management issues.
- Not necessarily reliable.

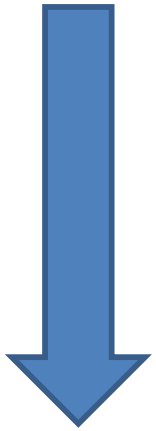
Parsing pipeline



Data anonymisation

Grower	Field	Field Reference	NG Code	Date	Crop	Operation	…
Kurowski Farms	Main Field	B	OV 243179	25/08/2018	Winter Wheat	Harvest	…

Anonymisation (pseudonymisation)



LandIS

Meteorological data

Subfield ID	Date	Crop	Operation	Rainfall	Soil Type	Slaking risk	…
53cd215e6c832fj	25/08/2018	Winter Wheat	Harvest	77.2 mm	Deep silt	C	…

Compiling machine learning datasets

Table | Hectarage overview | Yield overview

Date	Crop	Variety	Product Name	Field Operations	Subfi
	All Crops	All Varieties		All Field Operations	a1b
29/04/2009	Fennel	Orion Fennel	Fennel Drill		0490ca04491
29/04/2009	Fennel	Orion Fennel	Fennel Drill	Establishment	0490ca04491
29/04/2009	Fennel	Orion Fennel	Power Harrow Shallow 4m	Establishment	0490ca04491
29/04/2009	Fennel	Orion Fennel	Orion Fennel	Seed / Plants	0490ca04491
02/05/2009	Potatoes	Melody-Scotch	Basta		97a370bd4d(
02/05/2009	Potatoes	Melody-Scotch	Bateman Sprayer		97a370bd4d(
02/05/2009	Potatoes	Melody-Scotch	Bateman Sprayer	Application	97a370bd4d(
02/05/2009	Potatoes	Melody-Scotch	Basta	Herbicides	97a370bd4d(
10/05/2009	Potatoes	Melody-Scotch	Bateman Sprayer		97a370bd4d(
10/05/2009	Potatoes	Melody-Scotch	Shark		97a370bd4d(

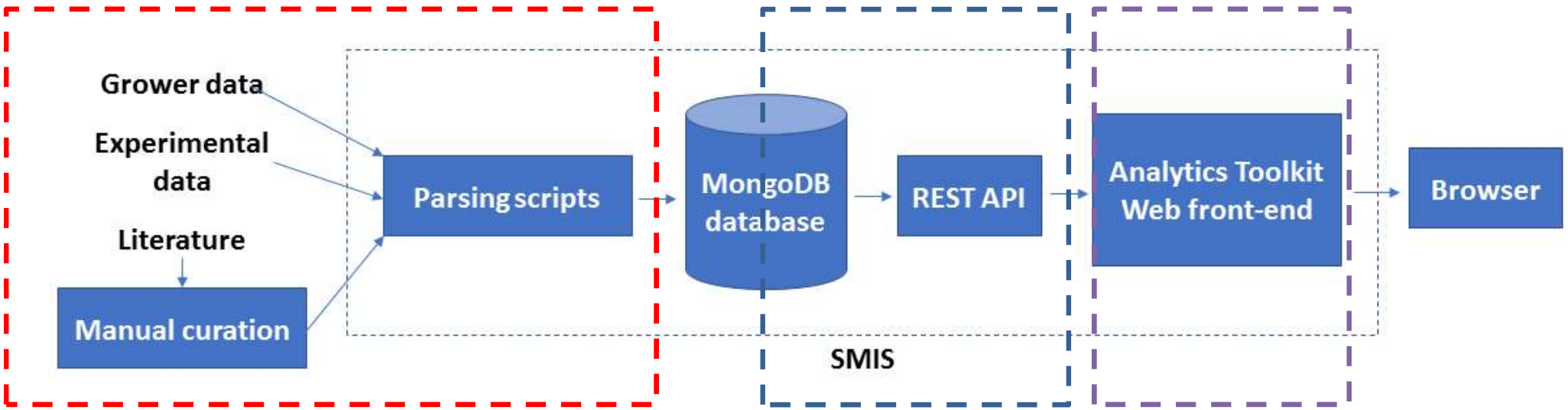
8 9 10 11 12 10

Summarising (e.g. count number of herbicide applications) and analysing (operations outside MWD) records

(Some data recorded for filtering rather than ML)


```
{
        "_id" : ObjectId("5be5adde93a71e2d13d7dc1e"),
        "Herbicides" : 2,
        "Molluscicides" : 0,
        "Number of Applications" : 14,
        "Variety" : "Broccoli",
        "Insecticides" : 11,
        "Rainfall in Autumn" : 206.2,
        "Rainfall in Spring" : 154.9,
        "Soil Texture" : "Medium (Eng)",
        "Subfield" : "55bb5a796c5d60e9bbd61620785ff1a0ef457400
        "Trace Elements" : 5,
        "Rainfall in Summer" : 183.9,
        "Growth Regulators" : 0,
        "Operations outside MWD" : 1,
        "Yield" : 2.6930131004366813,
        "Fungicides" : 0,
        "Fertiliser Applications" : 3,
        "Crop" : "Broccoli",
        "Year" : 2015,
        "Tillage Ops During Establishment" : 6,
        "Dependent Variable" : "Yield",
        "Previous Crop" : null
```

Information flow



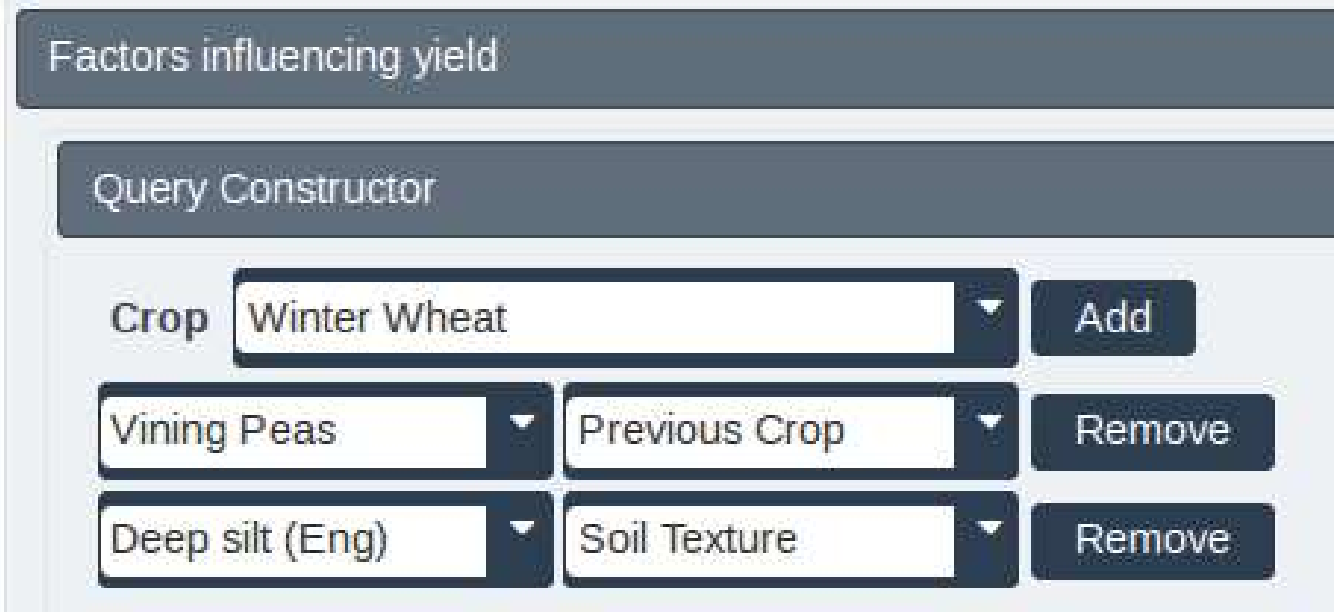
SMIS PS (Parsing Suite)

SMIS API (Application Programming Interface)

SMIS AT (Analytics Toolkit)

Analytics – “Rule base” queries

“Queries” can constructed for the rule base by grouping data restriction (e.g. limiting to particular crop or soil type) terms.



Rule bases depend on machine learning, some elements of queries (e.g. excluding compaction inference based on Grower data) require the rule base to be recreated, which can take several minutes.
Rule bases are stored so that the second time the same query is requested, it will not have to be recalculated.

Machine learning approaches

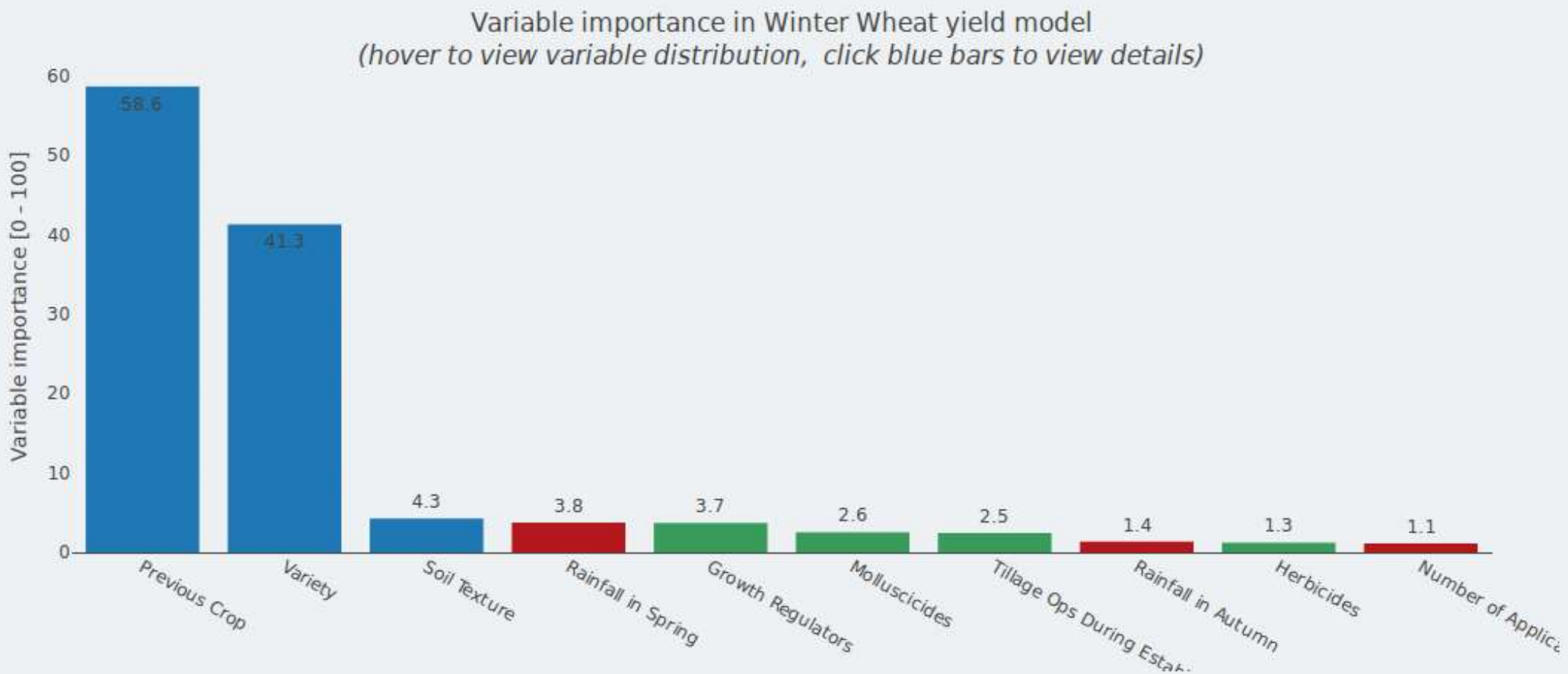
After the user selects filters in the front end, filtered data is fed to an R pipeline.

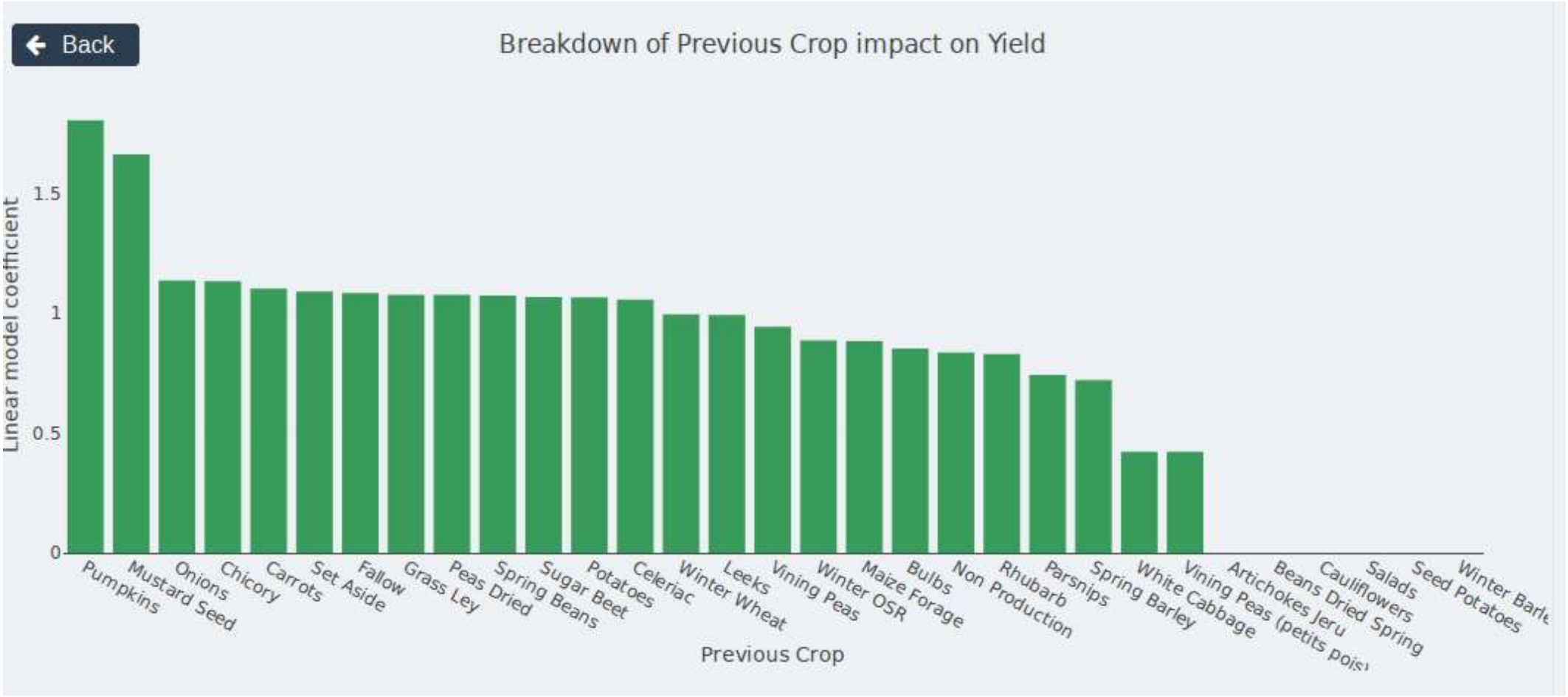
Regression methods used depend on the query:

- Random forest / general linear model for continuous numeric
- Random forest / logistic regression for true/false problems

The primary output are the variable importance values and relative coefficient values.

Yield model


Variable importance in Winter Wheat yield model
(hover to view variable distribution, click blue bars to view details)
Variable importance [0 - 100]
60
50
40
30
20
10
0
58.6
41.3
4.3
3.8
3.7
2.6
2.5
1.4
1.3
1.1
Previous Crop
Variety
Soil Texture
Rainfall in Spring
Growth Regulators
Molluscicides
Tillage Ops During Estab
Rainfall in Autumn
Herbicides
Number of Applic


Back
Breakdown of Previous Crop impact on Yield
Linear model coefficient
1.5
1
0.5
0
Pumpkins
Mustard Seed
Onions
Chicory
Carrots
Set Aside
Fallow
Grass Ley
Peas Dried
Spring Beans
Sugar Beet
Potatoes
Celeriac
Winter Wheat
Leeks
Vining Peas
Winter OSR
Maize Forage
Bulbs
Non Production
Rhubarb
Parsnips
Spring Barley
White Cabbage
Vining Peas (petits pois)
Artichokes Jeru
Beans Dried Spring
Cauliflowers
Salads
Seed Potatoes
Winter Barle
Previous Crop

Machine learning approaches

CRANFIELD UNIVERSITY

Jakub Porc

Assessment of machine learning approaches in soil management

School of Water, Energy and Environment
Applied Bioinformatics

MSc.
Academic Year: 2017 - 2018

Supervisors: Dr Fady Mohareb, Mr. Tomasz Kurowski
August 2018

Assessment of machine learning approaches in soil management

Introduction

As a result of increased commercial pressures on growers and extreme weather events, soils are increasingly subject to degradation processes which can harm soil health, leading to poorer crop establishment, impaired root growth and losses in yield quantity, quality and reliability. Determining which factors influence the soil and the yield is a complex task that may provide the answer on how to prevent these negative processes and be a source of guidelines for the growers to help them maintain their crops in good condition.

Objectives

The aim of this project was to investigate and compare the usefulness of multiple machine learning methods in identifying relationships related to yield and soil health in the collected data.

Pipeline

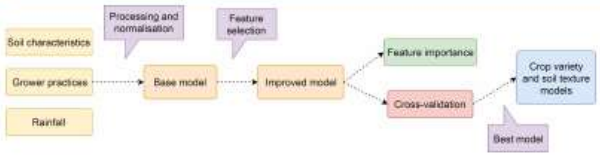


Figure 1. Project's pipeline of machine learning approaches testing.

A set of models was created for the chosen crops: linear regression, random forest, logistic regression, support vector machine, decision tree and neural network. The importance of certain factors on the crop growth was assessed.

Results

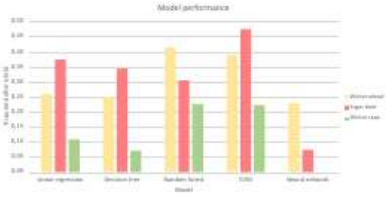


Figure 2. Model performance results after cross-validation.

Figure 3. Significant factors for each of the tested crops.

A set of machine learning models was created with medium to low predictive capabilities. Overall, random forest and support vector machine provided the best results. The most important factors affecting the crop growth and yield were found, allowing to create guidelines for the growers regarding management strategies.

Jakub Porc, BSc
Cranfield University, College Rd., Cranfield MK43 0AL, UK
Email: jakpor199426@gmail.com

Supervision:
Mr Tomasz Kurowski[1]
Dr Fady Mohareb[1]
[1] Cranfield University
College Rd, Cranfield MK43 0AL

www.cranfield.ac.uk

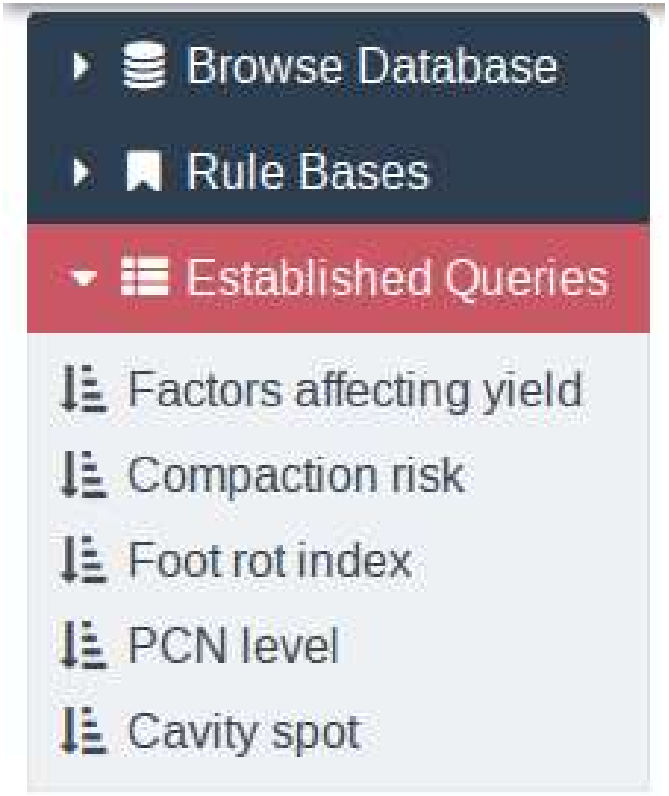

Model performance
R squared after k-fold
0,50
0,45
0,40
0,35
0,30
0,25
0,20
0,15
0,10
0,05
0,00
Linear regression
Decision tree
Random forest
SVM
Neural network
Model
Winter wheat
Sugar beet
Winter rape

Established queries

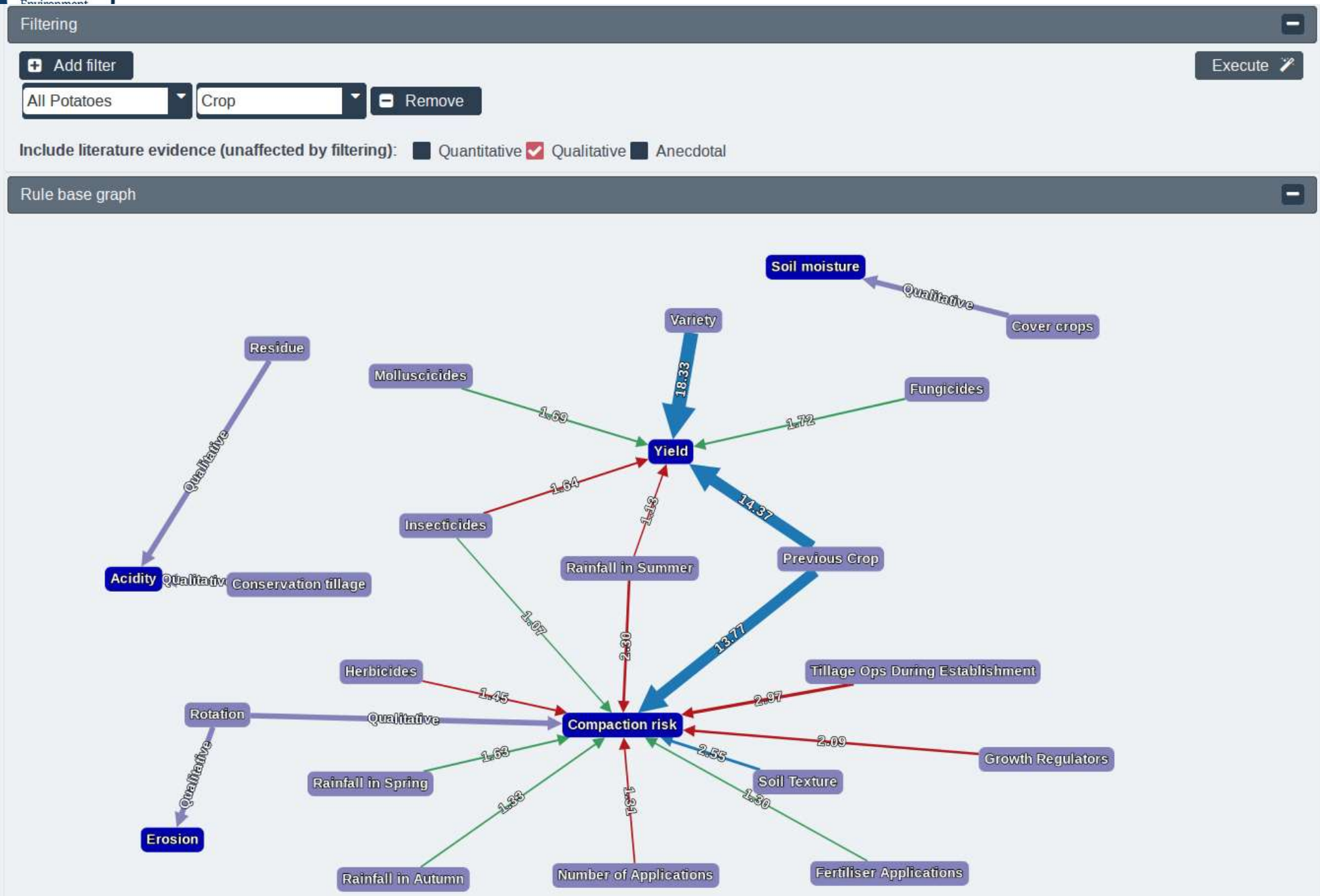
Established queries are “special case” visualisations of rule bases, with their own interfaces.

What queries can be isolated in such a way will depend on:

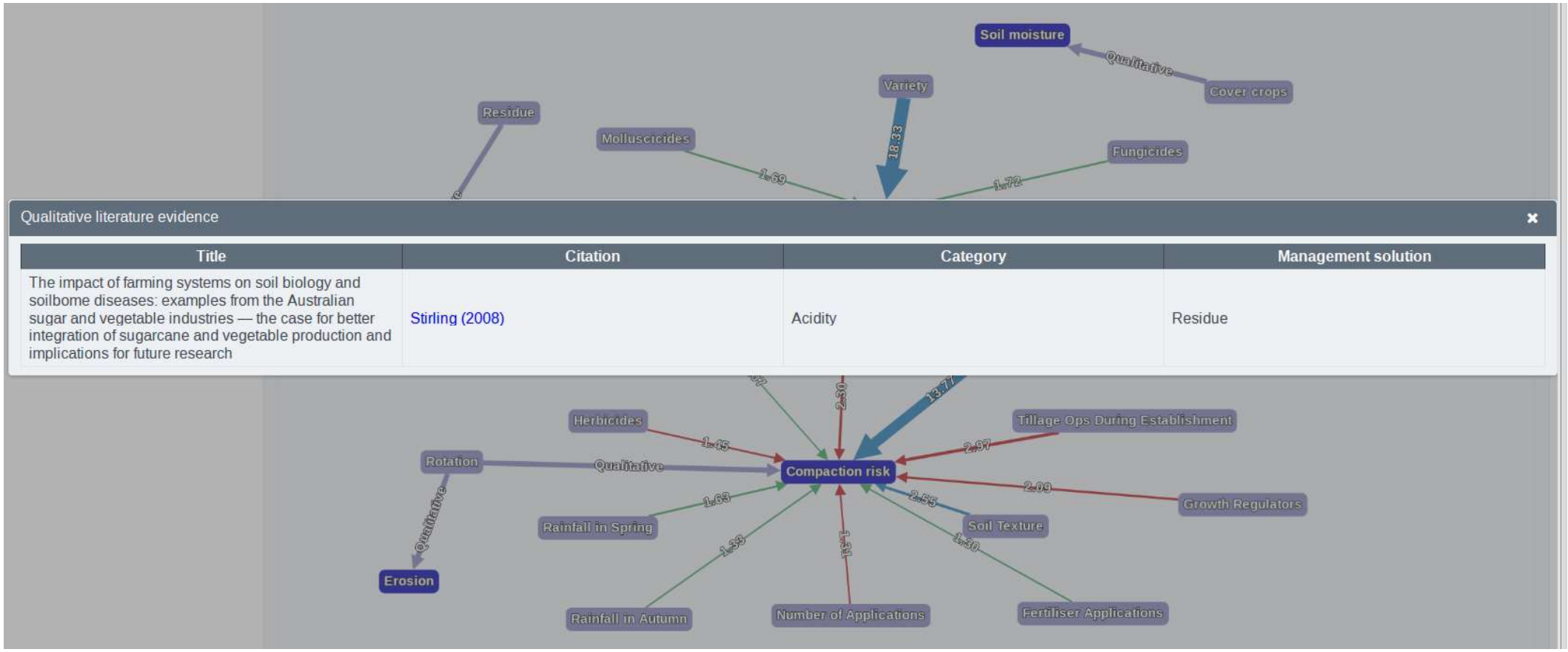
- Volume and type of data available
- Results of machine learning approaches (currently exploratory), i.e. what relationships can be reliably identified from the data



Rule base browsing – views

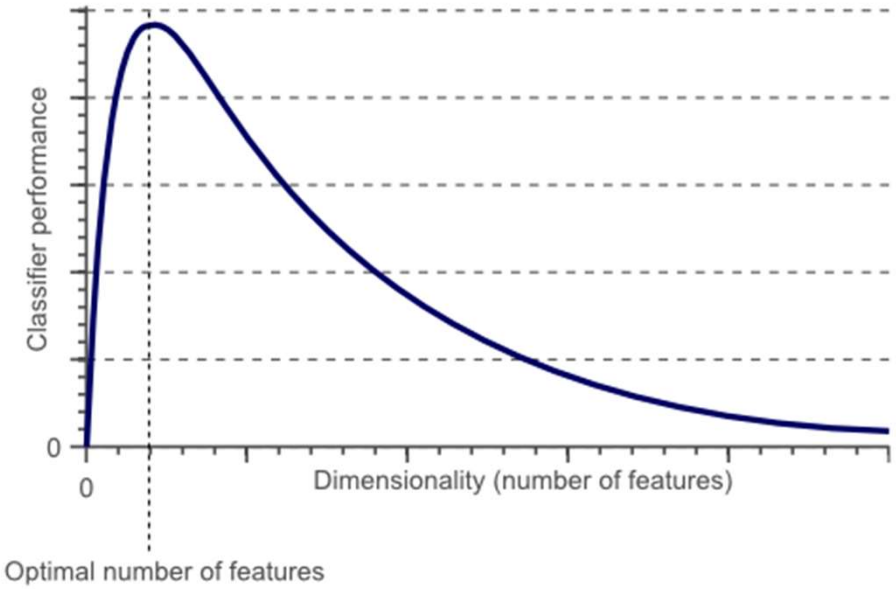


Rule base browsing – literature integration



Limitations

- Curse of dimensionality (Hughes phenomenon) – too many variables vs number of observations!



- Proliferation of best practice approaches – variables need to vary meaningfully for their effect to be studied! (positive deviance)
- Inconsistent data management practices

Future of SMIS

Project concluded in 2019, deployed internally at AHDB.

Potential follow-up in partnership with Soil Benchmark (likely Innovate UK funding).

The current system was made for experts; there are plans to develop it as a commercial product.

SMIS
SOIL MANAGEMENT INFORMATION SYSTEM