Yield prediction in winter wheat using machine learning; improving implemented farm management tool

By Mette K. Langgaard, special consultant

SEGES Innovation

05-07-2023

STØTTET AF

Promilleafgiftsfonden for landbrug

SEGES INNOVATION

SEGES INNOVATION

- non-profit research and development organization

We deliver

INNOVATION, RESEARCH & KNOWLEDGE

for future farming and food production

Yield prognosis in winter wheat

The Objectives

The objectives of this study were to:

1) Increase the accuracy and robustness of the yield prediction model in winter wheat by adding more data and new features to the model.

2) Implement the new model in the web-based management tool CropManager used by Danish Farmers.

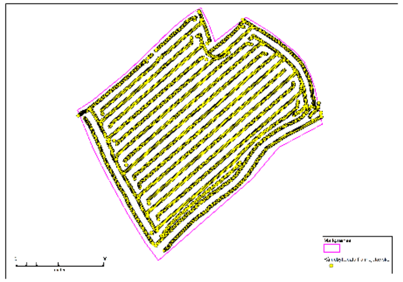
Goal: Mean Absolute Error (MAE) < 1 t ha^{-1}

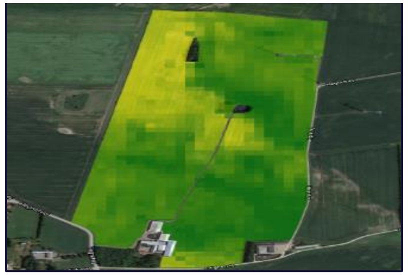
Materials and methods

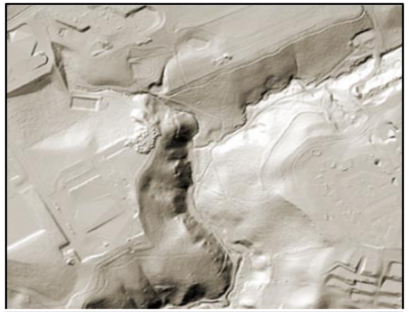
STØTTET AF
Promilleafgiftsfonden for landbrug
SEGES INNOVATION

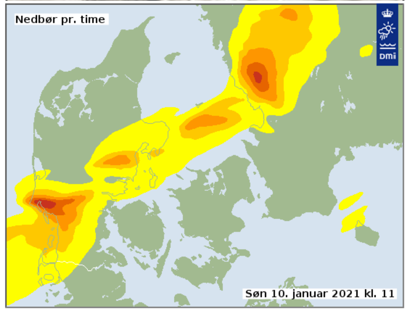
Data layers used for modelling

1) Yield maps from combine harvesters of farmers
2) Satellit data (L1C Sentinel data)
3) Terrain Elevation (The Danish Terrain Elevation model)
4) Weather data (DMI)
5) Soil texture
6) Crop variety
7) Crop rotation







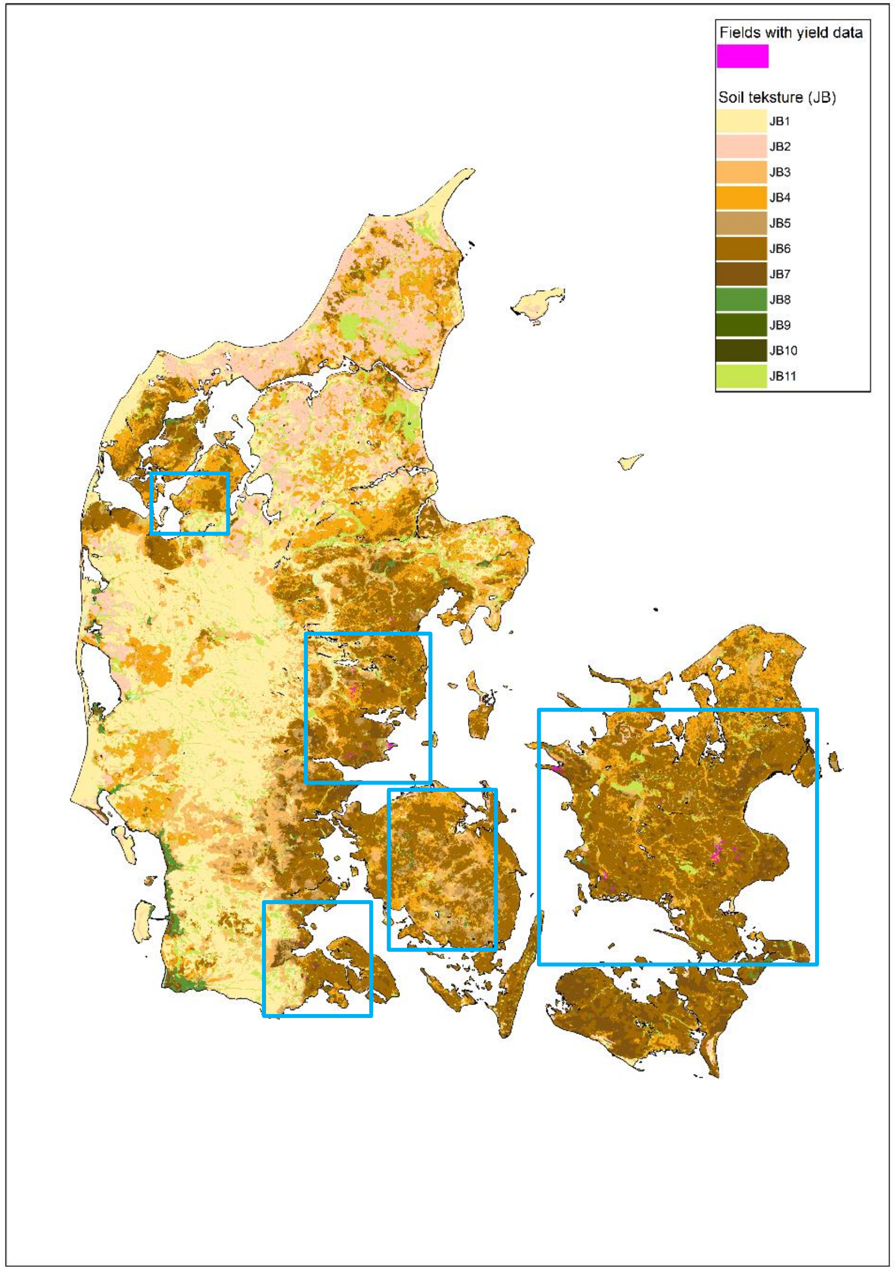


SEGES
INNOVATION

Distribution of yield data

Year	Yield data				
	Number of fields	Number of farmers	Hectare	Pixels[1]	Avg. yield, t ha^{-1}
2016	33	7	289	28,898	10.4 (1.8)
2017	95	15	856	85,611	9.6 (2.2)
2018	35	6	356	35,580	6.8 (1.5)
2019	26	5	221	22,062	7.2 (1.3)
2020	29	4	233	23,322	7.3 (1.6)
2021	69	5	984	98,356	7.9 (1.3)
Sum:	287		2,938	293,829	

1) Pixels of 10 x 10 m.



Models

791 features in the model

ML algorithm:

Gradient Boosting Regressor

Prediction dates:

April 6th, May 4th, June 1st and July 27th

The prediction performance:

$$\mathrm{MAE} = \sum_{i=1}^{n} \frac{|h_i - p_i|}{n}$$

h is the measured yield,
p predicted yield and
n the number of observations

R^2

4 model experiments: varies in prediction date, features, number of observation and spilt of data between training and validation.

SEGES INNOVATION

Results

STØTTET AF
Promilleafgiftsfonden for landbrug

SEGES INNOVATION

Experiments	Prediction date	Features	Observations	Split of data
1	April 6th	All	293,829 pixels	Field level (approx. 40 fields in validation data)
	May 4th			
	June 1st			
	July 27th			

SEGES INNOVATION

Experiments	Prediction date	Features	Observations	Split of data	MAE, $t\ ha^{-1}$ Validation	R^2
1	April 6th	All	293,829 pixels	Field level (approx. 40 fields in validation data)	0.67	0.74
	May 4th				0.62	0.79
	June 1st				0.59	0.82
	July 27th				0.56	0.83

SEGES INNOVATION

Experiments	Prediction date	Features	Observations	Split of data	MAE, t ha^{-1} Validation	R^2
1	April 6th	All	293,829 pixels	Field level (approx. 40 fields in validation data)	0.67	0.74
	May 4th				0.62	0.79
	June 1st				0.59	0.82
	July 27th				0.56	0.83
2	July 27th	Aggregated to field level + feature elimination	287 fields	Field level (approx. 40 fields in validation data)		

SEGES INNOVATION

Experiments	Prediction date	Features	Observations	Split of data	MAE, t ha^{-1} Validation	R^2
1	April 6th	All	293,829 pixels	Field level (approx. 40 fields in validation data)	0.67	0.74
	May 4th				0.62	0.79
	June 1st				0.59	0.82
	July 27th				0.56	0.83
2	July 27th	Aggregated to field level + feature elimination	287 fields	Field level (approx. 40 fields in validation data)	0.41	0.91

SEGES INNOVATION

Experiments	Prediction date	Features	Observations	Split of data	MAE, t ha^{-1} Validation	R^2
1	April 6th	All	293,829 pixels	Field level (approx. 40 fields in validation data)	0.67	0.74
	May 4th				0.62	0.79
	June 1st				0.59	0.82
	July 27th				0.56	0.83
2	July 27th	Aggregated to field level + feature elimination	287 fields	Field level (approx. 40 fields in validation data)	0.41	0.91
3	May 4th	Aggregated to field level + feature elimination	287 fields	Cross-validation with years as fold		
	July 27th					

SEGES INNOVATION

Experiments	Prediction date	Features	Observations	Split of data	MAE, t ha^{-1} Validation	R^2
1	April 6th	All	293,829 pixels	Field level (approx. 40 fields in validation data)	0.67	0.74
	May 4th				0.62	0.79
	June 1st				0.59	0.82
	July 27th				0.56	0.83
2	July 27th	Aggregated to field level + feature elimination	287 fields	Field level (approx. 40 fields in validation data)	0.41	0.91
3	May 4th	Aggregated to field level + feature elimination	287 fields	Cross-validation with years as fold	0.90	0.69
	July 27th				0.88	0.68

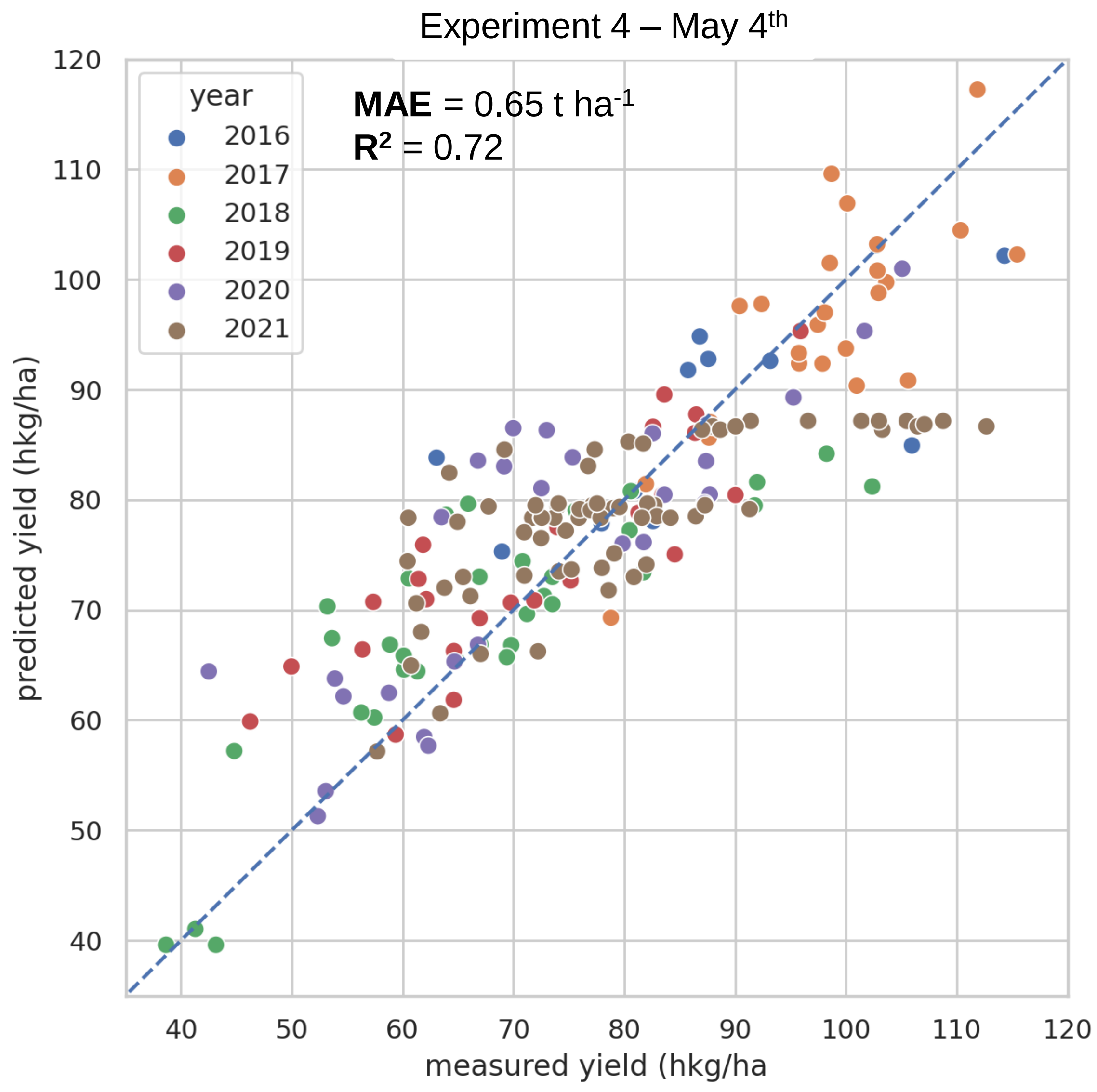
Experiments	Prediction date	Features	Observations	Split of data	MAE, t ha^{-1} Validation	R^2
1	April 6th	All	293,829 pixels	Field level (approx. 40 fields in validation data)	0.67	0.74
					0.62	0.79
	May 4th				0.59	0.82
	June 1st					
	July 27th				0.56	0.83
2	July 27th	Aggregated to field level + feature elimination	287 fields	Field level (approx. 40 fields in validation data)	0.41	0.91
3	May 4th	Aggregated to field level + feature elimination	287 fields	Cross-validation with years as fold	0.90	0.69
	July 27th				0.88	0.68
4	May 4th	Aggregated to field level + feature elimination	195 fields	Cross-validation with years as fold (only data collected in 2022)		
	July 27th					

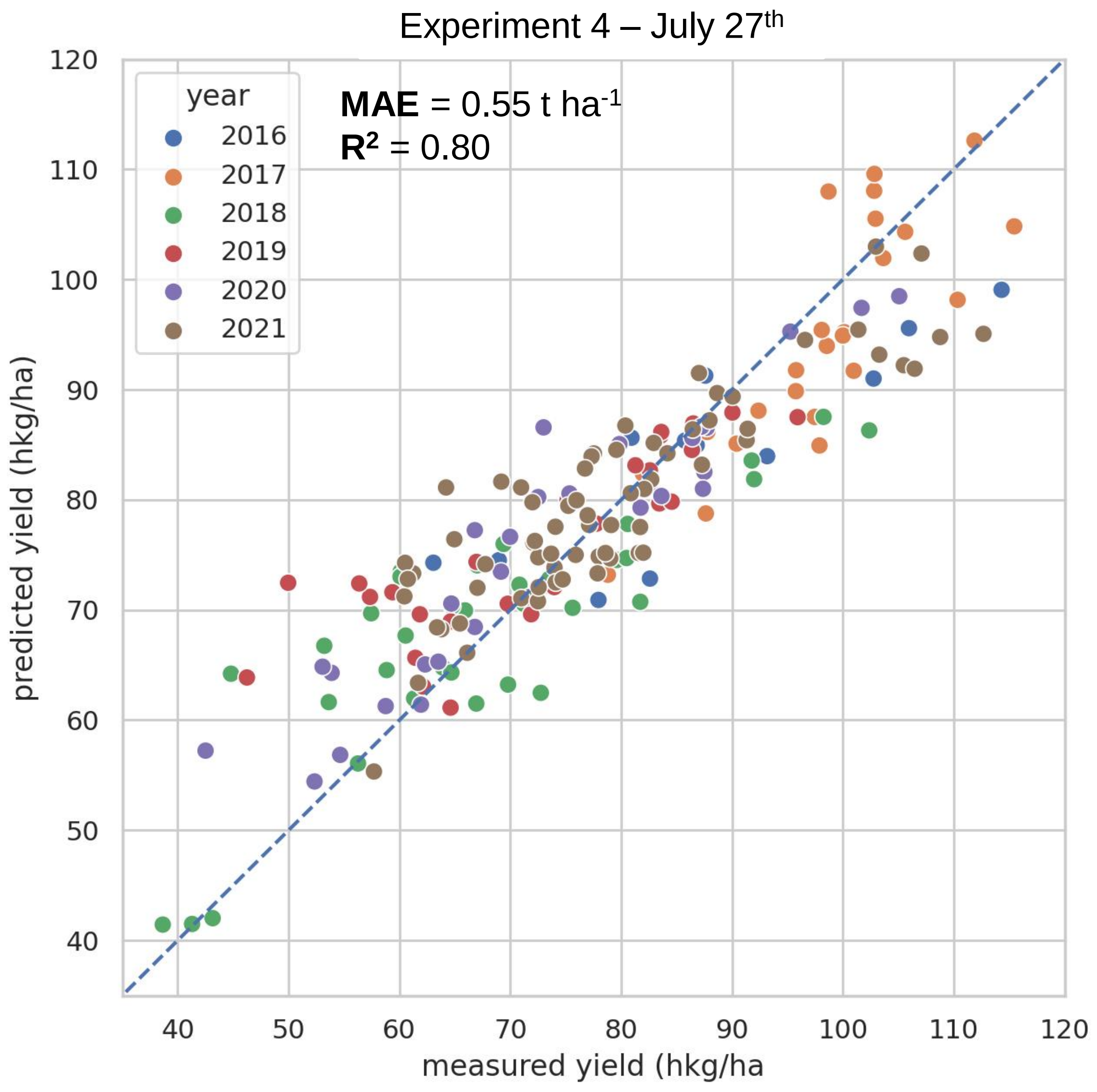
SEGES INNOVATION

Experiments	Prediction date	Features	Observations	Split of data	MAE, t ha^{-1} Validation	R^2
1	April 6th	All	293,829 pixels	Field level (approx. 40 fields in validation data)	0.67	0.74
	May 4th				0.62	0.79
	June 1st				0.59	0.82
	July 27th				0.56	0.83
2	July 27th	Aggregated to field level + feature elimination	287 fields	Field level (approx. 40 fields in validation data)	0.41	0.91
3	May 4th	Aggregated to field level + feature elimination	287 fields	Cross-validation with years as fold	0.90	0.69
	July 27th				0.88	0.68
4	May 4th	Aggregated to field level + feature elimination	195 fields	Cross-validation with years as fold (only data collected in 2022)	0.65	0.72
	July 27th				0.55	0.80

SEGES
INNOVATION

Yield prediction at field level





Conclusion

- We were able to predict winter wheat yield on field level with a MAE of 0.65 and 0.55 t ha^{-1} on May 4th and July 27th respectively when cross-validating with years.

- The prediction accuracy on May 4th (field level) is acceptable to regulate nitrogen application to crop demand in third application in growth stage 37 (BBCH).

- The models are incorporated into CropManager used by Danish farmers.

Thanks to the team

Lasse Rose Malskær, lead data scientist, SEGES Innovation
Jacob Hein, data scientist, SEGES Innovation
Mike Jørgensen, special consultant, SEGES Innovation
Rita Hørfarter, special consultant, SEGES Innovation
Finn M. Andreasen, special consultant, SEGES Innovation
Dianna Hjorth Kristensen, consultant, SEGES Innovation
Andreas Jydebjerg

<u>Former colleagues</u>
Peter Fogh
Christian Oxvig
Donjeta Runjeva
Ashley Montcalm