

1 Nonlinear parametric modelling to study how soil properties affect 2 crop yields and NDVI

3 Rebecca Whetton^a, Yifan Zhao^b, Sameh Shaddad^c, Abdul M Mouazen^{a,d*}

4 ^a *Cranfield Soil and AgriFood Institute, Cranfield University, Bedfordshire MK43 0AL, UK.*

5 ^b *Through-life Engineering Services Institute, Cranfield University, Bedfordshire MK43 0AL, UK*

6 ^c *Soil Science Department, Faculty of Agriculture, Zagazig University 44511- Zagazig, Egypt*

7 ^d *Precision Soil & Crop Engineering Group, Ghent University, Coupure 653, 9000 Gent, Belgium*

8 ^{*} *E-mail of corresponding author: Abdul.Mouazen@UGent.be*

9 **Abstract**

10 This paper explores the use of a novel nonlinear parametric modelling technique based on a Volterra Non-linear
11 Regressive with eXogenous inputs (VNRX) method to quantify the individual, interaction and overall
12 contributions of six soil properties on crop yield and normalised difference vegetation index (NDVI). The
13 proposed technique has been applied on high sampling resolution data of soil total nitrogen (TN) in %, total
14 carbon (TC) in %, potassium (K) in cmol kg⁻¹, pH, phosphorous (P) in mg kg⁻¹ and moisture content (MC) in %,
15 collected with an on-line visible and near infrared (VIS-NIR) spectroscopy sensor from a 18 ha field in
16 Bedfordshire, UK over 2013 (wheat) and 2015 (spring barley) cropping seasons. The on-line soil data were first
17 subjected to a raster analysis to produce a common 5 m by 5 m grid, before they were used as inputs into the
18 VNRX model, whereas crop yield and NDVI represented system outputs. Results revealed that the largest
19 contributions commonly observed for both yield and NDVI were from K, P and TC. The highest sum of the
20 error reduction ratio (SERR) of 48.59% was calculated with the VNRX model for NDVI, which was in line with
21 the highest correlation coefficient (*r*) of 0.71 found between measured and predicted NDVI. However, on-line
22 measured soil properties led to larger contributions to early measured NDVI than to a late measurement in the
23 growing season. The performance of the VNRX model was better for NDVI than for yield, which was attributed
24 to the exclusion of the influence of crop diseases, appearing at late growing stages. It was recommended to
25 adopt the VNRX method for quantifying the contribution of on-line collected soil properties to crop NDVI and
26 yield. However, it is important for future work to include additional soil properties and to account for other
27 factors affecting crop growth and yield, to improve the performance of the VNRX model.

28 **Keywords**

29 Yield limiting factors; proximal soil sensing; nonlinear parametric modelling; VNRX.

30 **1. Introduction**

31 Increasing crop yields requires the precision management of external farm resources (i.e.,
32 agrochemicals and fertilisers), which will help reduce input costs and detrimental
33 environmental impacts. Precision management of farm resources requires an understanding
34 and quantification of factors that limit crop yields, which is a research question yet to be
35 comprehensively answered. This currently prohibits precision management of farm resources
36 to be a routine activity. However, precision management of farm resources to correct existing
37 yield limiting factors require high sampling resolution data of variables impacting crop
38 growth and yield, which can then be incorporated within an analytical system. To realise this,
39 robust and reliable sensing platforms for soil and crop are needed. Due to the complexity and
40 high spatial variability of soils, the application of proximal soil sensors is still under active
41 research. Kuang et al. (2012) argue that the most favourable methods for on-line
42 measurement of key soil properties are visible and near infrared (VIS-NIR) spectroscopy and
43 electrochemical methods. The former is based on diffuse reflectance light collected from a
44 soil surface subjected to an external light source, whereas the latter uses ion selective
45 elements to produce a voltage output in a solution in response to the activity of the selected
46 ion (e.g., hydrogen, nitrate). Whilst VIS-NIR is most appropriate to soil properties have direct
47 spectral responses in the NIR spectral range, i.e., organic carbon (OC), moisture content
48 (MC), clay and clay mineralogy (Stenberg et al., 2010), electrochemical methods are capable
49 of quantifying mobile elements i.e., nutrients, mineral nitrogen, or pH (Adamchuk et al.,
50 1999). Since a soil solution is required for electrochemical sensors, their on-line use is
51 impeded. Although on-line VIS-NIR spectroscopy sensors are capable of collecting high
52 sampling resolution data (e.g., >500 samples per ha), they are limited to few research groups
53 (Christy, 2008; Shibusawa, et al. 2001; Mouazen et al., 2006a). Once key soil properties
54 needed in the analytical system are successfully collated using an on-line sensor, information

55 about crop growth (i.e., normalised difference vegetation index (NDVI) or leaf area index
56 (LAI)) can be obtained at high sampling resolution by means of earth observation utilising
57 satellite, airborne, drones or proximal crop sensing platforms.

58 Previous research has often assumed that the relationship between crop yield and growth
59 limiting factors is linear or approximately linear, which could be untrue for typically complex
60 agriculture systems. Mitscherlich (1909) proposed a model that simulates crop response to
61 growth factors increase. The model assumes that yield potential is constant, and isn't affected
62 by other factors that limit actual yields under field conditions; a further assumption that may
63 be false in complex agricultural systems. To reveal and characterise information hidden
64 within this complex system, a non-linear modelling approach is required to describe the
65 dependence among soil properties, NDVI and crop yield. Through these means, yield limiting
66 soil properties can be quantified.

67 Nonlinear methods include, among others, non-linear regression analyses and machine
68 learning. The Nonlinear Auto-Regressive Moving Average Model with eXogenous inputs
69 (NARMAX) is a parametric modelling method introduced by Billings et al. (1989). It is a
70 popular class of nonlinear system identification methods for a complex system, which
71 represents a typical input-output system with an unknown inner structure. Compared to
72 machine learning methods, an advantage of NARMAX is transparency. This means it can be
73 written down and easily understood and interpreted, related to known and existing models, as
74 well as being coupled with frequency domain or statistical analyses. These characteristics are
75 attractive for studying brain climatic change and agriculture systems that are typical input-
76 output systems with unknown inner structures. A Volterra Nonlinear Regressive with
77 eXogenous inputs (VNRX) is a special case of NARMAX that has more recently been
78 introduced. Although VNRX has had successful applications in brain signal analysis
79 (Sarrigiannis et al., 2014; Zhao et al., 2012), climate change (Bigg et al., 2014; Zhao et al.,

80 2016) and non-destructive tests (Zhao et al., 2017), its application in agriculture is novel due
81 to its capability to reveal hidden nonlinear information while other modelling methods cannot.
82 To our best knowledge no literature about the use of the VNRX model to predict NDVI and
83 crop yield based on on-line measured soil properties is available. This is important to
84 investigate, since on-line soil sensors provide high sampling resolution data (>500 sample per
85 ha), to enable accounting for variability over small spatial scales (e.g., few meters), which
86 cannot be efficiently achieved using traditional methods of soil sampling and laboratory
87 analyses that are tedious, time consuming and costly.

88 This study's aim is to implement a novel parametric VNRX model to quantify individual,
89 interaction and collective contribution of six soil properties (i.e., TN, total carbon (TC),
90 potassium (K), pH, phosphorous (P), and MC) on crop yield and NDVI. Soil data has been
91 collected at a high sampling resolution with an on-line VIS-NIR spectroscopy sensor.

92 **2. Materials and methods**

93 **2.1 Study site and data collection**

94 The study site is located on commercial farmland in Wilstead, Bedfordshire, United Kingdom
95 at coordinates 52°6'0.00"W latitude and 0°26'42.00"N longitude. The field is approximately
96 18 ha in area, with an average annual rainfall of 598 mm. The farms crop rotation consists of
97 barley, wheat and oil seed rape. The representative soil texture across the field to a depth of
98 0.20 m is non-homogeneous, including three textures of sandy loam, loam and sandy clay
99 loam in accordance with the United States Department of Agriculture (USDA) texture
100 classification system (Soil Survey Staff, 1999). Wheat and spring barley were cultivated over
101 the experiment timescale during the 2013 and 2015 cropping seasons, respectively. Soil
102 properties, yield and NDVI data were collected using an on-line VIS-NIR spectroscopy
103 sensor (Mouazen, 2006), on-board yield sensor from the farmer's combine harvester (New
104 Holland, CX8070 model), and a Crop Circle sensor (Crop Circle ACS 470, Holland Scientific,

105 Lincoln, NE USA), respectively. Wheat NDVI was measured in the booting (growth stage 43)
106 and heading (growth stage 52) stages, in May and June 2013, respectively. Spring barley
107 NDVI was measured during the stem extension (growth stage 37) and booting (growth stage
108 43) stages in April and May 2015, respectively. The growing stages are determined in
109 accordance with Zadok's decimal growth scale (Zadoks et al., 1974).

110 The on-line VIS-NIR soil sensor (Mouazen, 2006b) (Fig. 1) consisted of an AgroSpec mobile,
111 VIS-NIR spectrophotometer (tec5 Technology for Spectroscopy, Germany), with a
112 measurement range of 305-2200 nm. It has a differential global positioning system (DGPS)
113 (EZ-Guide 250, Trimble, USA) to record the position of the on-line measured spectra with
114 sub-metre accuracy. An optical probe fitted behind a subsoiler collected diffuse reflected

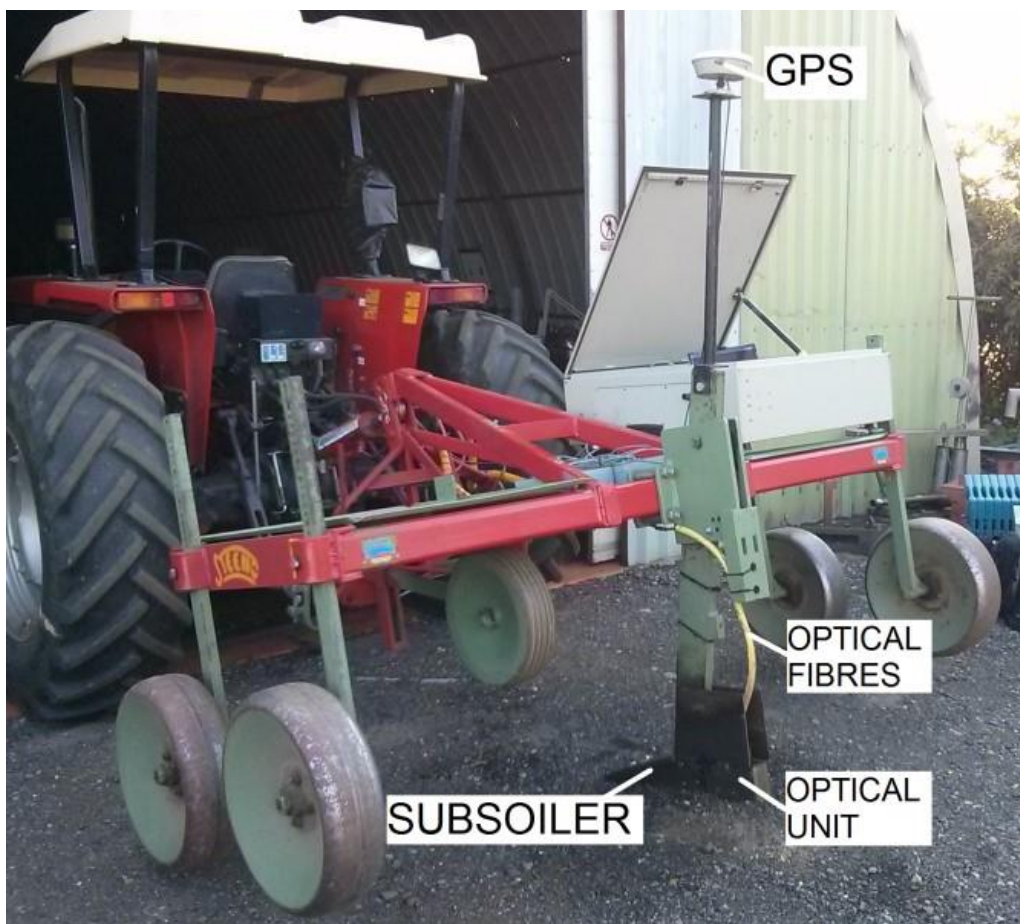


Fig. 1. The tractor mounted on-line visible and near infrared (VIS-NIR) spectroscopy sensor (Mouazen, 2006b).

115 spectra from the bottom of a smooth soil trench formed by the subsoiler. A 20 W halogen
116 lamp supplied by a tractor battery illuminated the base of the trench with artificial light. A
117 semi-rugged laptop was used for data logging and communication to the instrument.

118 On-line soil measurements were carried out in September 2012 and 2014, following crop
119 harvest, using the method reported in Mouazen et al. (2005). These measurements will be
120 referred to as 2013 and 2015 soil measurements, respectively, throughout the manuscript. The
121 on-line sensor produced measurement transects that were 6 cm wide, 15 cm deep, and were
122 distanced 15 m apart. The spectral measurements were collected with an average forward
123 speed of 2 km h⁻¹. Both on-line measurement and collection of soil samples were performed
124 prior to seed drilling (October in 2012 for wheat and February 2015 for spring barley) and
125 fertilisation (April to June in 2013 and 2015). Soil property changes during winter between
126 on-line measurement and next growing season are minimal, except for MC. However, the
127 spatial distribution of MC in a topographically uniform field like the study site may be
128 similar to the spatial distribution of clay (Mouazen et al., 2014), suggesting that the general
129 spatial pattern of MC would not significantly change from year to year. Only nitrogen
130 fertiliser was homogeneously applied in the 2013 and 2015 cropping seasons, whereas no K
131 or P fertilisers were applied.

132 **2.2 Laboratory analysis and development of calibration models of soil properties**

133 Ten soil samples per hectare (183 samples per 18 ha field area) based on a 30 by 30 m grid
134 (Fig. 2) were collected during the on-line measurement in 2012 from the bottom of the
135 subsoiler trenches. Sampling positions were recorded with a DGPS (Shaddad et al., 2016).

136 Approximately 700 g representing each soil sample was prepared as a composite of soil
137 collected over a 1.5 m travel distance at about 0.15 m depth. Soil samples were placed into
138 tightly sealed plastic bags to hold field moisture, and kept refrigerated at 4 °C, until

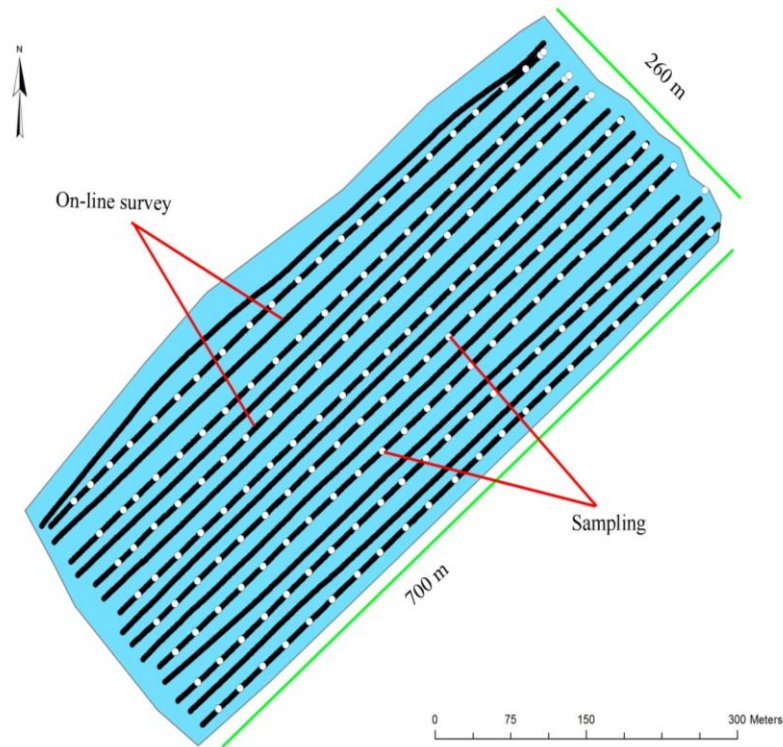


Fig. 2. On-line field survey transects and locations of 183 soil samples (Shaddad et al., 2016).

139 laboratory analyses to determine TN, TC, K, pH, P and MC. These analyses were based on
 140 the following procedures:

- 141 • TN and TC were determined with CHN 628 elemental analysis by combustion (LECO,
 142 USA) (British Standard Institute, 1995).
- 143 • Exchangeable K was determined using an atomic absorption spectrometer (AA
 144 analyst 200 Perkin Elmer Instruments, Shelton, Connecticut, USA).
- 145 • pH was measured using a glass electrode in a 1:5 (volume fraction) suspension of soil
 146 in distilled water (British Standard Institute, 1998).
- 147 • Extractable P was obtained in sodium hydrogen carbonate solution according to ISO
 148 11263:1994 (Olsen et al., 1954) and was determined by colorimetric approach using
 149 UV-VIS-NIR spectrophotometer (Murphy and Riley, 1962).
- 150 • Gravimetric MC was measured with oven drying at 105 °C for 24 h (British Standards
 151 Institute, 2007).

152 Results of laboratory analyses and spectral measurements were pooled together in one matrix.
153 The 183 soil samples were randomly split into calibration (70% of samples) and validation
154 (30%) sets. The calibration set was subjected to partial least squares regression (PLSR)
155 analysis to establish calibration models for the studied soil properties using Unscrambler
156 V9.8 software (Camo Software, Norway) (Shaddad et al., 2016). PLSR models were
157 validated using the 30% validation samples that were not included in the PLSR calibration
158 stage. Models were then used to predict the six soil properties using the on-line collected soil
159 spectra in September 2012 and 2014. The accuracy of these models are assessed by means of
160 the ratio of prediction deviation (RPD), which equals the standard deviation of laboratory
161 measured values divided by root mean square error of prediction (RMSEP). The following
162 RPD classes proposed by Viscarra Rossel et al. (2006) were adopted in this study: $RPD < 1.0$
163 indicates very poor model predictions and their use is not recommended; an RPD between 1.0
164 and 1.4 indicates poor model predictions, where only high and low values are distinguishable;
165 an RPD between 1.4 and 1.8 indicates fair model predictions, which may be used for
166 assessment and correlation; an RPD between 1.8 and 2.0 indicates good model predictions,
167 where quantitative predictions are possible; an RPD between 2.0 and 2.5 indicates very good
168 quantitative model predictions; and an $RPD > 2.5$ indicates excellent model predictions.

169 **2.3 Data processing**

170 Data processing begun with kriging of the on-line VIS-NIR predicted soil properties and
171 measured crop NDVI and yield. Kriged data layers were converted into a common 5 m^2 raster
172 grid in ArcGIS (Esri, USA) to aid data fusion (Frogbrook and Oliver, 2007). The 5 m^2 raster
173 grid was converted into a common grid of points that represented the value at the midpoint of
174 each raster pixel. These steps ensured that all layers consisted of common sets of 5 m^2 grid
175 points, essential for running the VNRX analysis. This method allowed fusion of data from a
176 diverse range of soil and crop property (e.g., NDVI, Yield, etc.) surveys, measured at

177 different resolutions (Khosla et al., 2008). However, it is worth noting that converting data
 178 from 5 m² raster squares to point locations introduced unavoidable errors to the data's spatial
 179 distribution. Finally, the different soil and crop layers of 5 m² grid were subjected to the
 180 VNRX non-linear parametric modelling, which is explained further in the following section.

181 **2.4 Non-linear parametric modelling**

182 The VNRX model, also known as nonlinear finite impulse response (NFIR) model, is used in
 183 this paper to represent a multi-inputs and single-output system. The model can be expressed
 184 as:

$$185 \quad y(k) = f(u_1^{[k-1]}, u_2^{[k-1]}, \dots, u_R^{[k-1]}) + \varepsilon(k) \quad (1)$$

186 where $k(k = 1, 2, \dots)$ is a time index, R is the number of system inputs, f is some unknown
 187 linear or non-linear mapping, which links the system output y to the system
 188 inputs u_1, u_2, \dots, u_R and $\varepsilon(k)$ denotes the model residual. The symbol $u_i^{[k-1]}(i = 1, 2, \dots, R)$
 189 denotes the past information of the input u_i , which can be expanded as:

$$190 \quad u_i^{[k-1]} = \bigcup_{j=0}^{n_i} u_i(k-j) \quad (2)$$

191 where n_i is the maximum temporal lag to be considered for the input u_i .

192 The Volterra series is a model for nonlinear behaviour that has similarities to the Taylor
 193 series. However, it differs from the Taylor series in its ability to capture 'memory' effects.
 194 The Taylor series can be used to approximate the response of a nonlinear system to given
 195 inputs if the output of this system depends strictly on the inputs at that particular time. In the
 196 Volterra series the output of the nonlinear system depends on the input to the system at all
 197 previous times. This provides the ability to capture the 'memory' effect of devices like
 198 capacitors and inductors (Tashev, 2009).

199 A commonly employed model type to specify the function f in Eq. (1) is a polynomial
200 function (Chen and Billings, 1989; Wei et al., 2004), which can be expressed as:

$$201 \quad y = \theta_0 + \sum_{m=1}^N \theta_m \phi_m + \varepsilon \quad (3)$$

202 where ϕ_m is the m^{th} model term generated from all input vectors, θ_m is the corresponding
203 unknown parameters, and N is the total number of potential model terms. It is worth noting
204 that ϕ_m is, in general, non-linear. Considering a system with two inputs u_1 and u_2 , a second
205 order polynomial function can be written as:

$$206 \quad y = \theta_0 + \theta_1 u_1 + \theta_2 u_2 + \theta_3 u_1^2 + \theta_4 u_2^2 + \theta_5 u_1 u_2 + \varepsilon \quad (4)$$

207 The next step is to estimate the parameters $\theta_m (m = 0, 1, \dots, 5)$ based on the
208 observations $\{y, u_1, u_2\}$. The procedure begins by determining the structure, or the important
209 model terms, using the orthogonal least squares (OLS) estimation procedures. It determines
210 which dynamics and nonlinear terms should be included in the model by computing the
211 contribution that each potential model term makes to the variation of the system output. The
212 model is to be built up term by term in a manner that exposes the significance of each new
213 term that is added. Once the structure of the model has been determined, the unknown
214 parameters can be estimated, and the procedure of model validation can ensure the model is
215 adequate. In this paper, a routine called adaptive-forward-orthogonal least squares (AFOLS)
216 was employed, not only to determine the model structure, but also to estimate unknown
217 parameters. The forward model selection scheme adopted consisted of a greedy optimisation
218 algorithm that progressively includes additional terms into the model, starting from an empty
219 structure, on the basis of the error reduction ratio (ERR) criterion (Cantelmo and Piroddi,
220 2010). It is a well-tested strategy for parsimonious modelling of data due to its effectiveness
221 and merit to reduce ill conditioning and overfitting problems (Zhao et al., 2012).

222 The on-line measured soil properties (i.e., pH, MC, TN, P, K and TC) were normalised by
223 removing the mean of each property, after which they were used as inputs to the VNRX
224 model, whereas the model output was mean normalised crop yield and NDVI. The analysis
225 also included the interaction between pairs of soil properties and their contribution to crop
226 yield and NDVI. The aim was to investigate the contribution of each soil property and their
227 pairwise interaction on crop NDVI and yield on one hand and to understand how the
228 contribution varied amongst different cropping seasons on the other hand. To calculate the
229 contribution of soil properties on yield, the NVRX analysis was carried out once in each
230 cropping season in 2013 and 2015. However, for NDVI, the NVRX model was run twice per
231 cropping season (e.g., May and June in 2013, and April and May in 2015). The VNRX
232 modelling was carried out utilising Microsoft visual studio code written with C++
233 programming language.

234 Finally, the performance of the VNRX model in the prediction of NDVI and yield was
235 evaluated by means of the of error reduction ratio (ERR) for each selected term calculated
236 from AFOLS that measures the percentage this term contributes to the system output. Values
237 of ERR always range between 0% and 100%. A higher ERR represents a greater dependence
238 between this term and the output. Therefore, it is an important index for indicating the
239 importance of each term to the output. To calculate the contribution of each input variable to
240 the output, the sum of ERR values of all selected terms, denoted by $SERR$, and calculated by

$$241 \quad SERR = \sum_{i=1}^N [err]_i \quad (5)$$

242 was used to describe the percentage explained by the identified model to the system output,
243 where N denotes the number of the selected terms. If the considered inputs can fully explain
244 the variation of system output, the value of $SERR$ is equal to 100%. It is an indicator of
245 model performance and uncertainty. The contribution of the i^{th} input variable to the variation

246 of the system output, denoted as $ERRC_i$, is defined as the sum of ERR values of the terms
 247 that include this input variable. Because some selected terms may involve more than one
 248 input variable due to nonlinearity, the sum of $ERRC_i$ for all input variables can be greater
 249 than $SERR$. To overcome this problem, the value of $ERRC_i$ is used, which can be written as:

$$250 \quad ERRC_i = \frac{\sum_{j=1}^N ([err]_j | u_i \in \emptyset_j)}{\sum_{p=1}^r \sum_{j=1}^N ([err]_j | u_p \in \emptyset_j)} \times SERR \quad (6)$$

251 The value of $ERRC_i$ should be always between 0% and 100%.

252 **2.5 Mapping**

253 Similar spatial distributions of measured versus predicted NDVI and yield were evaluated by
 254 comparing the corresponding maps. Maps were produced through interpolation with an
 255 inverse distance weighing (IDW) method, using ArcGIS software (ESRI, USA). The
 256 interpolation grid size of all maps had a radius of 12.5 m and a power of 2. The map cell size
 257 was 2.5 m² with 254 rows and 282 columns. Similarity between maps was assessed by visual
 258 comparison. In addition, Pearson's correlation (r) coefficient was calculated between each
 259 pair of data sets used to produce maps.

260 **3. Results and discussion**

261 **3.1 Accuracy of on-line measured soil properties**

262 The best independent validation of PLSR calibration models using on-line spectra (Table 1)
 263 is obtained for pH and TN with RPD values of 2.06 (very good prediction) and 1.85 (good
 264 model prediction), and RMSEP values of 0.434 and 0.013 (%), respectively (Shaddad, 2013).
 265 These results are better than those reported by Mouazen et al. (2007). Although both TC and
 266 MC have direct spectral responses in the NIR spectral range, they have not resulted in the
 267 best prediction accuracy in this study (Table 1). RPD values of P, TC and MC are rather
 268 small with values of 1.77, 1.50 and 1.49, respectively, which are classified as fair model

269 predictions. The lowest RPD value of 1.31 is calculated for K, indicating poor prediction
 270 accuracy (Viscarra Rossel et al., 2006).

Table 1: Range of on-line measured soil pH, phosphorous (P), total nitrogen (TN), total carbon (TC), moisture content (MC) and exchangeable potassium (K) used in the Volterra Nonlinear Regressive with eXogenous inputs (VNRX) models.

Year	Range	pH	P (mg kg ⁻¹)	TN (%)	TC (%)	MC (%)	K (cmol kg ⁻¹)
2013	Min	5.31	20.33	0.11	1.46	11.98	0.18
	Max	7.83	56.21	0.18	2.40	17.41	0.31
2015	Min	5.87	26.83	0.05	0.92	4.53	0.22
	Max	6.44	43.34	0.16	1.81	9.79	0.47

271

272 3.2 Influences of soil properties on yield

273 Statistics for on-line measured soil properties used as input in the VNRX model are provided
 274 in Table 2. Observed ranges (e.g., minimum and maximum values) are different between
 275 2013 and 2015, which can be attributed to farm practices (e.g., fertilisation) and different
 276 weather conditions affecting soil MC in particular.

Table 2: Validation of partial least squares regression (PLSR) models to predict soil pH, phosphorous (P), total nitrogen (TN), total carbon (TC), moisture content (MC) and exchangeable potassium (K) using on-line collected spectra of the prediction set.

Statistics	Soil properties					
	pH	P (mg kg ⁻¹)	TN (%)	TC (%)	MC (%)	K (cmol kg ⁻¹)
Sample no	48	23	22	24	45	24
Min	5.16	4.80	0.11	1.30	13.41	0.12
Max	8.17	50.00	0.20	2.46	24.28	0.40
Mean	6.46	22.50	0.15	1.79	18.03	0.23
SD	0.90	15.23	0.02	0.28	2.16	0.08
RMSEP*	0.43	8.61	0.01	0.18	1.45	0.06
R ²	0.73	0.69	0.72	0.57	0.56	0.44
RPD	2.06	1.77	1.85	1.50	1.49	1.31
Model quality**	A	B	A	B	B	C

*RMSEP: Root mean square error of prediction; **Model quality is categorized according to Viscarra Rossel et al. (2006) (A: residual prediction deviation (RPD)>1.8; B: RPD=1.4–1.8; C: RPD<1.4).

277

278 Based on Eq. (3), the polynomial model to express the relationship between the six input soil
279 variables and the output NDVI and yield is of quadratic terms, written as:

$$280 \quad y = \theta_0 + \sum_{i=1}^6 \theta_i u_i + \sum_{i=1}^6 \sum_{j=i}^6 \theta_{ij} u_i u_j + \varepsilon \quad (7)$$

281 This model includes 28 terms consisting of 7 linear terms $\{\theta_0, \theta_i u_i | i = 1, 2, \dots, 6\}$ and 21
282 nonlinear terms $\{\theta_{ij} u_i u_j | i = 1, 2, \dots, 6; j = i, i + 1, \dots, 6\}$. The main reason for selecting
283 quadratic instead of cubic terms is to balance the number of candidate terms and number of
284 samples. If cubic terms are considered, there are 84 candidate terms, which require a large
285 memory and high computational cost to implement the algorithm based on 7187 sampled
286 points. A model with cubic terms was tested and revealed no significant differences in results,
287 hence, the quadratic terms' model was adopted.

288 Values of ERRC calculated by Eq. (6), explaining the contribution of individual soil
289 properties to the crop yield in 2013 and 2015 cropping seasons, are shown in Table 3. To
290 evaluate the change of model uncertainty, the value of *SERR* needs to be examined. It is
291 observed that the *SERR* value in 2013 (21%) was larger than the corresponding value in 2015
292 (12.51%), which could be attributed to varying weather conditions that exert a big impact on
293 crop growth and yield (Renouf et al., 2010; Boone et al., 2016), or to errors in the estimation
294 models (both kriging and PLSR models). Other affecting factors that vary through cropping
295 seasons are pests, which are similarly associated with different weather conditions, but are
296 strongly linked to crop variety (Eberhart and Russell, 1966; Paveley et al., 2012). Finally, the
297 different crops grown throughout the experiment (in 2013 and 2015) represents one of the
298 major factors that explaining why contribution of soil properties to yield differ through the
299 two cropping seasons.

Table 3: Calculated individual contribution (ERRC) of normalised on-line soil properties on wheat and spring barley yields in 2013 and 2015, respectively.

Input	ERRC	
	2013	2015
K (cmol kg ⁻¹)	7.66	0.23
P (mg kg ⁻¹)	4.28	1.96
TC (%)	3.99	3.23
pH	3.51	1.45
TN (%)	1.56	4.46
MC (%)	0.00	1.18
Total (SERR)	21.00	12.51

TC is total carbon, TN is total nitrogen, K is exchangeable potassium, P is extractable phosphorous and MC is moisture content.

300

301 Observations show that K, P and TC contribute most to wheat yield in 2013, whereas TC and
302 TN contribute most to spring barley yield in 2015. The largest contributor to wheat yield in
303 2013 is K (*ERRC* = 7.66%) followed by P, which represent key nutrients to crop growth and
304 development (Baligar et al., 2001). This is the reason why P and K together with nitrogen are
305 applied annually. However, this is not the case for the spring barley in 2015, at least for K. It
306 seems that TC retains almost the same contribution to wheat yield in 2013 and spring barley
307 yield in 2015, which may be explained by nutrient demands varying between crops. For
308 example, wheat requires about 200 kg N, 55 kg P₂O₅ and 252 kg K₂O/ha (Roy et al., 2006),
309 whereas the UK national averages are 175 kg, 69 kg and 212 kg ha⁻¹, respectively.
310 Furthermore, nutrient requirements for the same crop vary between seasons and have to be
311 checked every 3 to 5 years (Nicholls, 2015). Depending on the cropping system, carbon in the
312 form of organic fertilisers is frequently added to agriculture fields, as it supports
313 photosynthesis (Ravikumar, 2013) and improves soil structure and hydraulic conductivity.
314 Therefore, it is unsurprising to observe that TC is a strong contributor to yield in both study
315 years. pH was a persistently strong contributing soil property, particularly in 2013. An acidic
316 or basic soil can prevent nutrient uptake and thus impede plant production (Schubert et al.,

1990). Farmer's guides commonly argue that the optimum pH for soils under continuous arable cropping (wheat and barley) is between 6 and 7 with 6.5 being optimum. Since the soil pH range in this study is wider than the optimum range (Table 2), pH is considered to have an influence on nutrient availability and subsequently yield. Although TN contribution ranks 5th in 2013, it has the largest contribution to yield in 2015. The narrow variation range over two sampled years: 2013 (0.05 to 0.16%), and 2015 (0.11 to 0.18%) (Table 2) may explain the fluctuated *ERRC* value of TN to yield (Table 3). MC has low yield contribution, where the *ERRC* value in 2013 is null. This could be explained by the time difference between MC and yield measurements. However, this time difference has only a minor influence on the remaining five soil properties considered in this study, as they are much less dynamic compared to MC.

3.2 Influences of soil properties on NDVI

Table 4 shows calculated *ERRC* and *SERR* values for NDVI in 2013 and 2015 based on the on-line measured soil properties in 2013 and 2015, respectively. *SERR* values, indicating the total contribution of soil properties to NDVI (Table 4), are much larger than the corresponding figures for yield (Table 3). These are 30.92% and 35.42% for May and June 2013, and 48.59% and 11.35% for April and May 2015, respectively. However, *SERR* value in May 2015 is notably low (11.35%), which can be attributed to a drought period occurring mid growing season; where the combination of a dry March and the sunniest April on record with little rainfall was recorded (UK Meteorological Office). This is because NDVI measurement took place at the booting growing stage, at which point the crop is particularly susceptible to drought and certain diseases. Elsewhere, a decrease in growth rate has been attributed to drought imposed at various growth stages in wheat, among which booting was listed (Ashraf, 1998).

Table 4: Calculated individual contribution (*ERRC*) of normalised on-line measured soil property to normalised difference vegetation index (NDVI) of wheat and spring barley in 2013 and 2015, respectively.

ERRC				
Input	2013		2015	
	May	June	April	May
TC (%)	10.25	16.46	5.86	3.52
K (cmol kg ⁻¹)	9.82	3.19	5.90	4.12
P (mg kg ⁻¹)	6.00	12.33	31.31	0.00
pH	2.69	0.91	3.21	0.00
MC (%)	1.71	1.39	2.31	2.83
TN (%)	0.45	1.14	0.23	0.88
Total (SERR)	30.92	35.42	48.59	11.35

TC is total carbon, TN is total nitrogen, K is exchangeable potassium, P is extractable phosphorous and MC is moisture content.

341
342 Similar to crop yield (Table 3), K (*ERRC* = 9.82% and 3.19%) and P (*ERRC* = 6% and
343 12.33%) are the largest contributors to NDVI after TC (*ERRC* = 10.25% and 16.46%) in
344 2013 (Table 4). A similar trend can be observed for NDVI response in 2015, where K, P and
345 TC are again the largest contributors to NDVI (Table 4), except P in May. It is important to
346 note that P contribution to NDVI surges in April 2015, with a *ERRC* value of about 6 times
347 of those of K and TC. Phosphorus is an essential nutrient for both plant structural compounds
348 and energy conversion (Ozanne et al., 1980). P availability is essential for crop growth during
349 spring. For example, Grant et al. (2001) reported for a barley crop that during the period from
350 March to May, 70% of phosphate is taken up. This may explain the surge in P contribution to
351 NDVI in the April 2015 measurement. pH, MC and TN have low contributions to NDVI in
352 both years.

353 **3.3 Prediction of NDVI and yield based on on-line measured soil properties**

354 To evaluate the performance of the proposed model for predicting NDVI and yield based on
355 on-line measured soil properties, the first five terms ranked by *ERRC* were selected, and the
356 corresponding parameters were estimated, to establish the following model:

357
$$y = \theta_0 + \sum_{m=1}^5 \theta_m \phi_m \quad (8)$$

358 The input variables used are the normalisation values obtained with mean normalisation.
359 Table 5 shows the first five (largest contributors) individual and interaction terms to describe
360 the relationship between on-line measured soil properties and NDVI and yield in 2013 and
361 2015 cropping seasons. Generally, TC, P, K and MC are the most influential variables on
362 NDVI in the two experimental years. Among the interaction terms, TC * K appears first in
363 June 2013 and April 2015 measurements, whereas MC * TC appears first in May 2013 and
364 2015. Interaction P * K is the second most contributing term to NDVI in June, 2013. Once
365 again this confirms the high individual and interaction contributions of TC, P and K on crop
366 NDVI for the two studied cereal crops.

367 Examining interaction effects of soil properties on yield reveals almost a similar trend to that
368 of NDVI, where K, P and TC are the most influential individual factors in 2013 only,
369 whereas no individual influence for K and P can be observed in 2015 (Table 5). However,
370 MC is not part of the most influential interactive terms anymore, and is instead replaced by
371 pH. This may be attributed to the drought impact during spring in 2015, according to the UK
372 Meteorological Office. Both pH * K and TN * P are the most influential interaction terms on
373 yield in 2013, whereas TN * K and pH * P are the most important interaction terms in 2015.
374 Although N, P and K are key nutrients for crop growth and yield, pH is important for nutrient
375 availability to plants (Schubert et al., 1990).

376 NDVI and yield can now be predicted, by substituting values of on-line measured soil
377 properties into Eq. (8) using coefficients shown in Table 5. A map showing the spatial
378 distribution of predicted versus measured NDVI and yield is shown in Fig. 3 and Fig. 4,
379 respectively.

380

Table 5: Individual and interaction relationship between on-line measured soil properties and normalised difference vegetation index (NDVI) and yield for data collected in 2013 and 2015. The order of the terms was ranked by the calculated individual contribution (ERRC) of each soil property.

		Month	0	1	2	3	4	5
NDVI	May-13	ϕ_m^*	constant	P	TC	K	K²	TN*MC
		θ_m^{**}	0.45	-0.003	0.095	0.7	16.45	0.804
	Jun-13	ϕ_m	constant	TC	P	TC*K	P*K	pH*K
		θ_m	0.488	0.1	-0.002	2.159	-0.036	-0.397
	Apr-15	ϕ_m	constant	P	K*TC	MC	pH	TC
		θ_m	0.345	0.014	3.798	0.024	-0.162	-0.068
May-15	ϕ_m	constant	K	TC	MC	MC*TC	TN²	
	θ_m	0.485	0.297	-0.069	0.006	-0.032	13.589	
Yield	2013	ϕ_m	constant	K	pH*K	TN*P	P	TC
		θ_m	6.711	13.712	-21.677	1.902	-0.036	1.226
	2015	ϕ_m	constant	TN	TC	TN*K	pH²	pH*P
		θ_m	8.449	35.843	-2.457	-153.7	6.971	-0.279

* ϕ_m : Model term; ** θ_m : coefficient of the model term

381

382 Observations show that similarities exist between the spatial distributions between each pair
383 of maps, particularly for NDVI. Interestingly, there is a distinct similarity between NDVI and
384 yield maps in 2013 (high r values in Table 6) for both measured and predicted maps (Figs. 3
385 and 4). Conversely, similarities in 2015 for both measured and predicted maps are not clear
386 (low r values in Table 6). The poor similarities shown in the 2015 cropping season may be
387 attributed to deterioration in PLSR prediction accuracy for the on-line collected soil data in
388 2015, since PLSR calibration models were developed on the basis of samples collected in
389 2013. This could also explain the drop in the total contribution of soil properties to yield
390 (SERR = 12.51) and low r (Table 6). Another potential explanation is that external factors not
391 accounted for in this study (e.g., fungi diseases) have a stronger influence on crop yield in
392 2015 than in 2013.

393 Pearson correlation coefficient values shown in Table 6 demonstrate that the prediction
394 performance for NDVI is more successful than yield in three out of four occasions. This
395 observation is supported by the fact that SERR values for NDVI are consistently higher than

396 the corresponding values for yield (Tables 3 and 4). The highest SERR value of 48.59% is
 397 calculated for NDVI prediction in April 2015 (Table 4), which is in line with the highest r
 398 value of 0.71 calculated between measured and predicted values (Table 6).

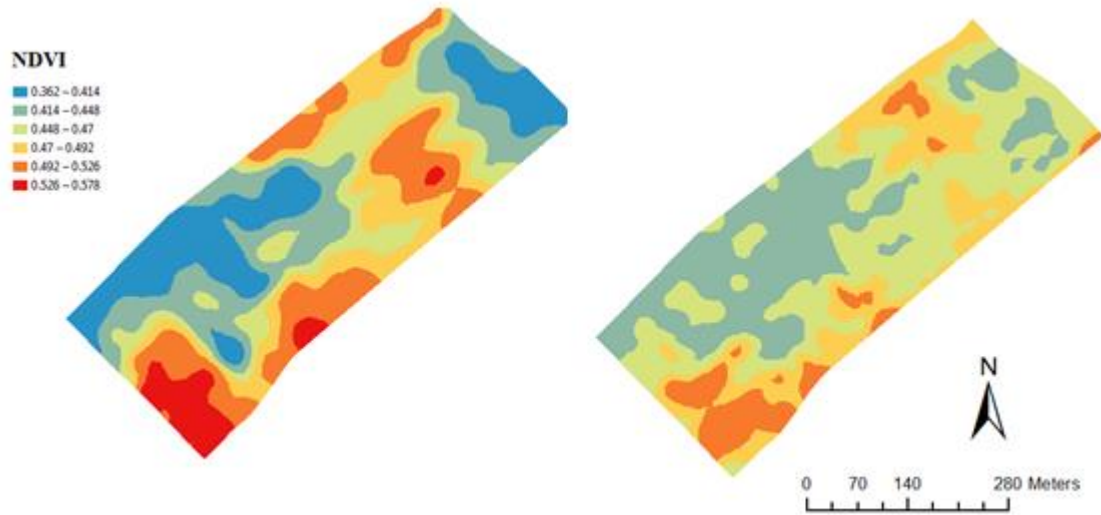
Table 6: Pearson correlation coefficient (r) values between measured and predicted crop yield and normalised difference vegetation index (NDVI), based on Eq. (8) and identified terms and coefficients shown in Table 5. The crops were wheat and spring barley in 2013 and 2015, respectively.

Year	Output	Correlation coefficient
2013	Yield	0.48
2015	Yield	0.38
2013	NDVI May	0.56
2013	NDVI June	0.60
2015	NDVI April	0.71
2015	NDVI May	0.40
2013	NDVIm (May) vs Yieldm	0.60
2013	NDVIp (May) vs Yieldm	0.40
2013	NDVIp (May) vs Yieldp	0.83
2015	NDVIm (April) vs Yieldm	0.12
2015	NDVIp (April) vs Yieldm	0.10
2015	NDVIp (April) vs Yieldp	0.25

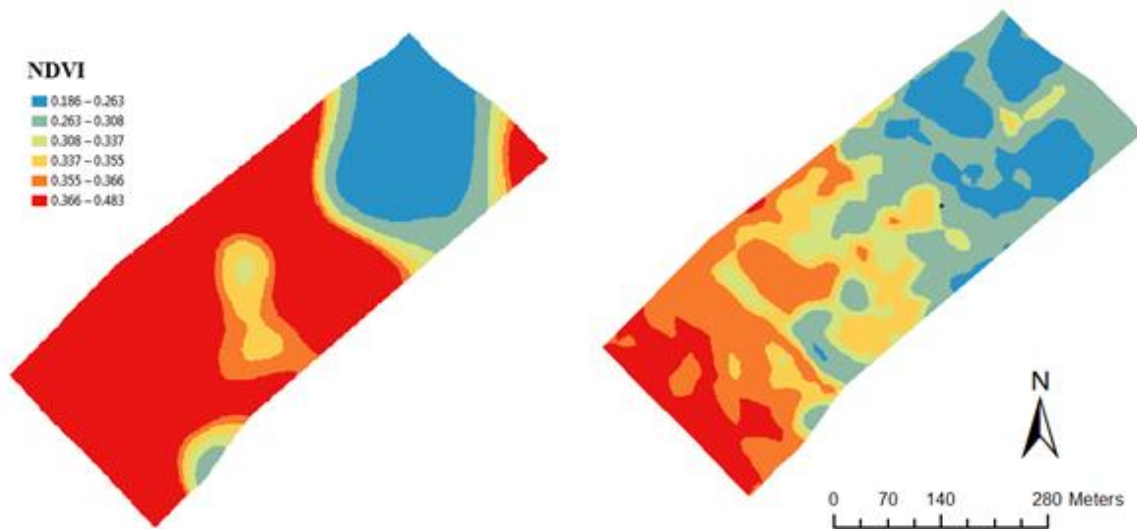
The significance threshold is 0.062 at 95% confidence level, NDVIm is measured NDVI, NDVIp is predicted NDVI, Yieldm is measured yield and Yieldp is predicted yield

399

400 Although r values are larger for NDVI than for yield during the two cropping seasons (Table
 401 6), it is not recommended to rely on only the six on-line measured soil properties to predict
 402 yield. Low r values for yield could be attributed to the exclusion of other factors affecting
 403 yield and encountered at late growing stages such as soil compaction, inter plant competition,
 404 fungal disease and insect pressures (Donald, 1963; Cannell et al., 1980; Coakley, 1988;
 405 Paveley et al., 2012). The latter factors have high spatial variability and can reduce yields by
 406 up to 7 tonne ha⁻¹ (Bravo et al., 2003). Therefore, it is suggested to expand the current work
 407 by accounting for other soil properties and diseases, alongside weather conditions, which is
 408 considered the most influential factor controlling the distribution and severity of fungal
 409 infections (Dammer, 2006).



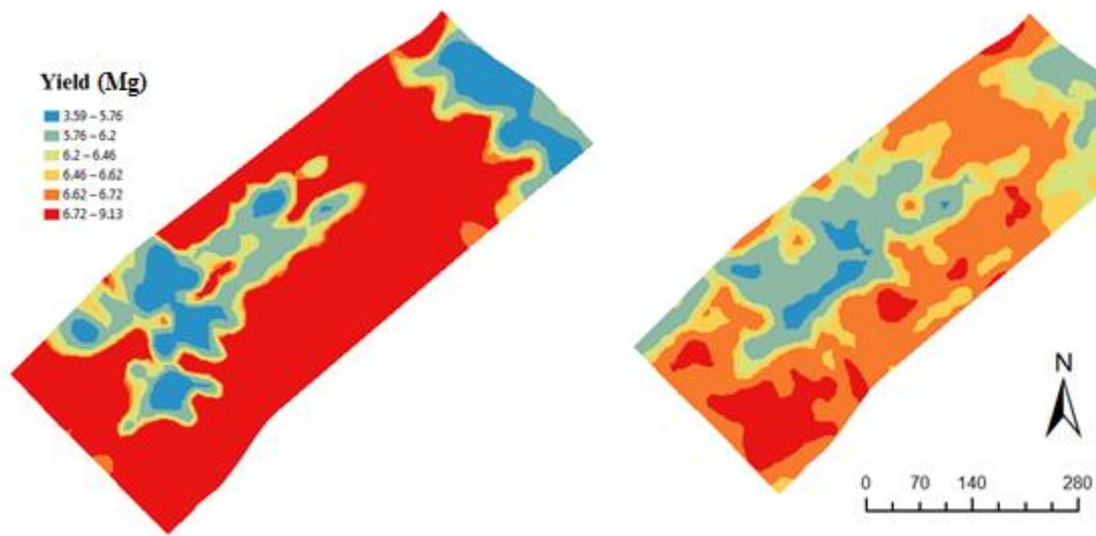
(a)



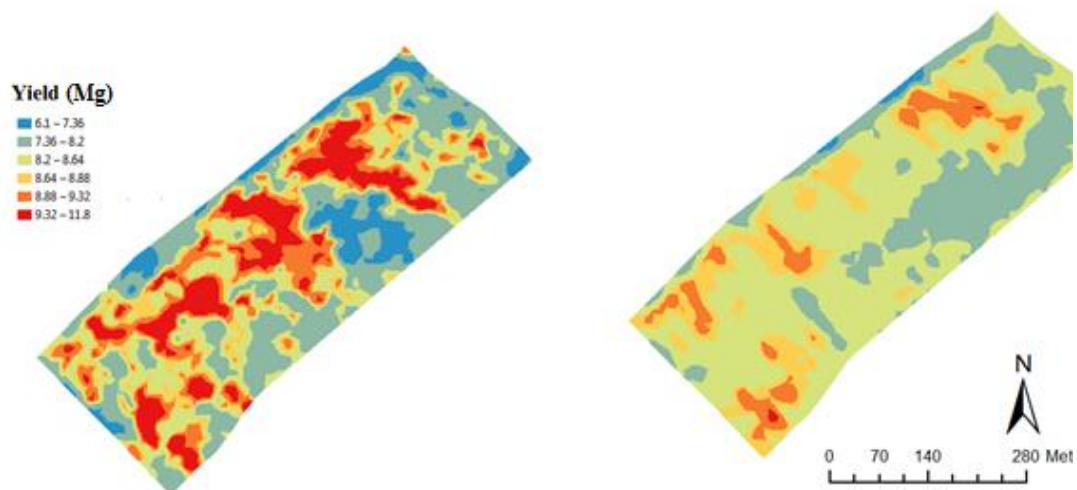
(b)

Fig. 3. Comparison between measured (right) and predicted (left) normalised difference vegetation index (NDVI) based on Eq. (8) and corresponding terms and coefficients shown in Table 5 for May 2013 (a) and April 2015 (b).

410 This study has presented a novel application of nonlinear parametric modelling technique
 411 based on a Volterra Nonlinear Regressive with eXogenous inputs (VNRX) model to study the
 412 influences of six soil properties: total nitrogen (TN), total carbon (TC), moisture content
 413 (MC), potassium (K), phosphorous (P) and pH. Soil properties were collected at high
 414



(a)



(b)

Fig. 4. Comparison between measured (right) and predicted (left) crop yield using the model of Eq. (8) and corresponding terms and coefficients shown in Table 5 for wheat in 2013 (a), and spring barley in 2015 (b).

415 sampling resolution with an on-line soil sensor on crop yield and normalised difference
 416 vegetation Index (NDVI). The analysis was carried out in two cropping seasons in 2013
 417 (wheat) and 2015 (spring barley). The results provided for the following conclusions:

- 418 1. The performance of the VNRX model in the prediction of yield evaluated with the
419 error reduction ratio contribution (ERRC) indicated that different soil properties have
420 different influences on yield. K, P and TC were the highest contributors to wheat yield
421 and TN, TC and P to spring barley.
- 422 2. Sum of error reduction ratio (SERR) showed soil property contributions to NDVI to
423 be higher than those to yield, with TC, K and P being the most influencing factors.
424 The highest SERR value of 48.59% was calculated for NDVI, which was in line with
425 the highest Pearson correlation coefficient (r) of 0.71 calculated between measured
426 and predicted NDVI.
- 427 3. The highest influential interaction terms of soil properties on NDVI were TC * K, and
428 MC * TC, whereas the most important terms for yield were pH * K and TN * P for
429 wheat and TC * K and pH * P for spring barley. These contributions may vary among
430 different fields, crops, weather conditions and soil fertility status.
- 431 4. Although VNRX models allowed the prediction of yield and NDVI to a given degree
432 of success, relatively low correlations between measured and predicted yield
433 necessitate a need to understand other influencing factors (i.e., weather conditions,
434 disease and other soil properties), to improve VNRX model predictions.

435 **Acknowledgements**

436 We acknowledge the funding received for FarmFUSE project from the ICT-AGRI under the
437 European Commission's ERA-NET scheme under the 7th Framework Programme, and the
438 UK Department of Environment, Food and Rural Affairs (contract no: IF0208).

439 **References**

440 Adamchuk, V.I., Morgan, M.T., and Ess, D.R., 1999. An automated sampling system for measuring
441 soil pH. *Transactions of the ASABE*. 42(4), 885-891.

- 442 Ashraf, M.Y., 1998. Photosynthetic efficiency of wheat under water stress conditions. *Journal of*
443 *Scientific & Industrial Research*. 41, 151-163.
- 444 Baligar, V.C., Fageria, N.K. and He, Z.L., 2001. Nutrient use efficiency in plants. *Communications in*
445 *Soil Science and Plant Analysis*. 32(7-8), 921-950.
- 446 Bigg, G. R., Wei, H. L., Wilton, D. J., Zhao, Y., Billings, S. A., Hanna, E., and Kadirkamanathan, V.,
447 2014. A century of variation in the dependence of Greenland iceberg calving on ice sheet
448 surface mass balance and regional climate change. *Proceedings of the Royal Society A:*
449 *Mathematical, Physical and Engineering Sciences*. 470(2166), 20130662.
- 450 Billings, S.A., Chen, S. and Korenberg, M.J., 1989. Identification of MIMO nonlinear systems using a
451 forward regression orthogonal estimator. *International Journal of Control*. 49, 2157–2189.
- 452 Boone, L., Van linden, V., De Meester, S., Vandecasteele, B., Muylle, H., Roldán-Ruiz, I., Nemecek,
453 T. and Dewulf, J., 2016. Environmental life cycle assessment of grain maize production: An
454 analysis of factors causing variability. *Science of The Total Environment*. 553, 551-564.
- 455 Bravo, C., Moshou, D., West, J., McCartney, A. and Ramon, H., 2003. Early disease detection in
456 wheat fields using spectral reflectance. *Biosystems Engineering*. 84(2), 137-145.
- 457 British Standard Institute, (BSI), 1995. Determination of organic and total carbon after dry
458 combustion (elementary), BSI 389. Chiswick High Road, London W4 4AL, UK..
- 459 British Standard Institute, (BSI), 1998. Determination of particle size distribution in mineral soil
460 material–method by sieving and sedimentation, BSI 389. Chiswick High Road, London W4
461 4AL, UK.
- 462 British Standards Institute, (BSI), 2007. Sample preparation for chemical and physical tests,
463 determination of dry. UK: British Standards Institute.
- 464 Cannell, R.Q., Belford, R.K., Gales, K., Dennis, C.W. and Prew, R.D., 1980. Effects of waterlogging
465 at different stages of development on the growth and yield of winter wheat. *Journal of the*
466 *Science of Food and Agriculture*. 31(2), 117-132.
- 467 Cantelmo, C. and Piroddi, L., 2010. Adaptive model selection for polynomial NARX models. *IET*
468 *Control Theory & Applications*. 4(12), 2693-2706.
- 469 Chen, S. and Billings, S.A., 1989. Representation of non-linear systems: the NARMAX model.
470 *International Journal of Control*. 49, 1013-1032.
- 471 Christy, C.D., 2008. Real-time measurement of soil attributes using on-the-go near infrared
472 reflectance spectroscopy. *Computers and Electronics in Agriculture*. 61, 10-19.
- 473 Coakley, S.M., 1988. Variation in climate and prediction of risk in balanced fertility, control of weeds,
474 field scouting, re-plants. *Annual Review of Phytopathology*. 26, 163–181.

475 Dammer, K.H. and Ehlert, D., 2006. Variable-rate fungicide spraying in cereals using a plant cover
476 sensor. *Precision Agriculture*. 7(2), 137-148.

477 Donald, C.M., 1963. *Competition among crop and pasture plants*. Academic Press.

478 Eberhart, S.T. and Russell, W. A., 1966. Stability parameters for comparing varieties. *Crop science*.
479 6(1), 36-40.

480 Frogbrook, Z.L and Oliver, M.A., 2007. Identifying management zones in agricultural fields using
481 spatially constrained classification of soil and ancillary data. *Soil Use and Management*. 23(1),
482 40-51.

483 Grant, C.A., Flaten, D.N., Tomasiewicz, D.J. and Sheppard, S.C., 2001. The importance of early
484 season phosphorus nutrition. *Canadian Journal of Plant Science*. 81(2), 211-224.

485 Khosla, R., Inman, D., Westfall, D. G., Reich, R. M., Frasier, M., Mzuku, M., Koch, B. and Hornung,
486 A., 2008. A synthesis of multi-disciplinary research in precision agriculture: site-specific
487 management zones in the semi-arid Western Great Plains of the USA. *Precision Agriculture*.
488 9, 85-100.

489 Mitscherlich, E.A., 1909. The law of the minimum and the law of diminishing soil productivity.
490 *Landwirtschaftliche Jahrbuecher*. 38, 537-552.

491 Mouazen, A. M., Maleki, M.R., De Baerdemaeker, J. and Ramon, H., 2007. On-line measurement of
492 some selected soil properties using a VIS-NIR sensor. *Soil & Tillage Research*. 93, 13-27.

493 Mouazen, A. M., Alhwaimel, S. A., Kuang, B. and Waine, T., 2014. Multiple on-line soil sensors and
494 data fusion approach for delineation of water holding capacity zones for site specific
495 irrigation. *Soil & Tillage Research*. 143, 95-105.

496 Mouazen, A., De Baerdemaeker, J. and Ramon, H., 2006. Effect of wavelength range on the
497 measurement accuracy of some selected soil constituents using visual-near infrared
498 spectroscopy. *Journal of Near Infrared Spectroscopy*. 14, 189-199.

499 Mouazen, A.M., 2006. Soil survey device. BE Patent WO/2006/015463.

500 Mouazen, A.M., De Baerdemaeker, J. and Ramon, H., 2005. Towards development of on-line soil
501 moisture content sensor using a fibre-type NIR spectrophotometer. *Soil & Tillage Research*.
502 80(1-2), 171-183.

503 Murphy, J. and Riley, J. P., 1962. A modified single solution method for determination of phosphate
504 in natural waters. *Analytica Chimica Acta*. 27, 31-36.

505 Nicholls, C. *Soils 5: pH and Nutrients*. HGCA. 26 11 2015.
506 <http://www.fwi.co.uk/academy/lesson/soils-5-ph-and-nutrients> (accessed 04 22, 2016).

507 Olsen, S. R., Cole, C.V., Watanabe, F.C. and Dean, L.S., 1954. Estimation of available phosphorus in
508 soil by extraction with sodium bicarbonate. Circular / United States Department of
509 Agriculture, 939.

510 Ozanne, P. G., 1980. Phosphate nutrition of plants - A general treatise. In: Khasawneh, F.E., Sample,
511 E. C. and Kamprath, E. J. Editors. The Role of Phosphorus in Agriculture. American Society
512 of Agronomy. Madison, pp. 559-589.

513 Paveley, N., Blake, J., Gladders, P. and Cockerell, V., 2012. *Wheat disease management guide*.
514 HGCA.

515 Ravikumar, P., 2013. Evaluation of nutrient index using organic carbon, available P and available K
516 concentrations as a measure of soil fertility in Varahi River basin, India. *Proceedings of the*
517 *International Academy of Ecology and Environmental Sciences*. 330.

518 Renouf, M.A., Wegener, M.K. and Pagan, R.J., 2010. Life cycle assessment of Australian sugarcane
519 production with a focus on sugarcane growing. *The International Journal of Life Cycle*
520 *Assessment*. 15(9), 927–937.

521 Roy, R.N., Finck, A., Blair, G.J. and Tandon, H.L.S., 2006. *Plant nutrition for food security*. Rome,
522 Italy: Food and Agriculture Organization of the United Nations.

523 Sarrigiannis, P.G., Zhao, Y. Wei, H.L., Billings, S.A., Fotheringham, J. and Hadjivassiliou, M., 2014.
524 Quantitative EEG analysis using error reduction ratio-causality test; validation on simulated
525 and real EEG data. *Clinical Neurophysiology*. 125(1), 32–46.

526 Schubert, S., Schubert, E. and Mengel, K., 1990. Effect of low pH of the root medium on proton
527 release, growth, and nutrient uptake of field beans (*Vicia faba*). *Plant and Soil*. 124(2), 239-
528 244.

529 Shaddad, S., 2013. *Proximal soil sensors and geostatistical tools in precision agriculture applications*.
530 Sassari, Italy: Universita Degli Studi Di Sassari.

531 Shaddad, S.M., Madrau, S., Castrignanò, A. and Mouazen, A.M., 2016. Data fusion techniques for
532 delineation of site-specific management zones in a field in UK. *Precision Agriculture*. 17(2),
533 200-217.

534 Shibusawa, S., Imade Anom, S.W., Sato, A., Sasao, A. and Hirako, S., 2011. *Soil mapping using the*
535 *real-time soil spectrophotometer*. Vol. 1, in *Third European Conference on Precision*
536 *Agriculture*, by G. Grenier and S. Blackmore, pp. 497–508. Montpellier, France: ECPA.

537 Soil Survey Staff, 1999. Soil Taxonomy - A basic system of soil classification for making and
538 interpreting soil surveys; second edition. Agricultural Handbook 436; Natural Resources
539 Conservation Service, USDA. Washington DC, USA.

- 540 Tashev, I.J., 2009. *Sound Capture and Processing: Practical Approaches*. John Wiley & Sons.
- 541 Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J. and Skjemstad, J.O., 2006
542 Visible, near infrared, mid infrared or combined diff use. *Geoderma*. 131, 59–75.
- 543 Wei, H.L., Billings, S.A. and Liu, J., 2004. Term and variable selection for nonlinear system
544 identification. *International Journal of Control*. 77, 86-110.
- 545 Zadoks, J.C., Chang, T.T. and Konzak, C.F., 1974. A decimal code for the growth stages of cereals.
546 *Weed Research*. 14(6), 415–421.
- 547 Zhao, Y., Bigg, G.R., Billings, S.A., Hanna, E., Sole, A.J., Wei, H., Kadirkamanathan, K. and Wilton,
548 D.J., 2016. Inferring the variation of climatic and glaciological contributions to West
549 Greenland iceberg discharge in the twentieth century. *Cold Regions Science and Technology*.
550 121, 167–178.
- 551 Zhao, Y., Wei, H.L. and Billings, S.A., 2012. A New Adaptive Fast Cellular Automaton
552 Neighborhood Detection and Rule Identification Algorithm. *IEEE Transactions on Systems,
553 Man, and Cybernetics—Part B: Cybernetics*. 42(4), 1283-1287.
- 554 Zhao, Y., Mehnen, J., Sirikham, A. and Roy, R., 2017. A novel defect depth measurement method
555 based on Nonlinear System Identification for pulsed thermographic inspection. *Mechanical
556 Systems and Signal Processing*. 85, 382–395.
- 557 Zhao, Y., Billings, S.A., Wei, H. and Sarrigiannis, P.G., 2012. Tracking time-varying causality and
558 directionality of information flow using an error reduction ratio test with applications to
559 electroencephalography data. *Physical Review E*. 86, 051919.