Sharp, Ryan T.; Henrys, Peter A.; Jarvis, Susan G.; Whitmore, Andrew P.; Milne, Alice E.; Coleman, Kevin; Mohankumar, Sajeev Erangu Purath; Metcalfe, Helen. 2021. **Simulating cropping sequences using earth observation data.**

The definitive version was published in *Computers and Electronics in Agriculture*, 188, 106330. https://doi.org/10.1016/j.compag.2021.106330

The definitive version is available at https://www.elsevier.com/

Simulating cropping sequences using earth observation data

Ryan T. Sharp [a,] *, Peter A. Henrys [b], Susan G. Jarvis [b], Andrew P. Whitmore [a], Alice E. Milne [a], Kevin Coleman [a], Sajeev Erangu Purath Mohankumar [a], Helen Metcalfe [a]

[a] Sustainable Agricultural Sciences, Rothamsted research, Harpenden, UK

[b] UK Centre for Ecology and Hydrology, Lancaster Environment Centre, Lancaster, UK

*Corresponding author email: ryan.sharp@rothamsted.ac.uk

Highlights

- *Remotely sensed data provides an opportunity to characterise crop sequences.*
- *Crop transition matrices describe the probability of one crop following another.*
- *Data and agronomic rules used to derive realistic crop sequences for Great Britain.*
- *Derived crop transition matrices characterise the business-as-usual state.*
- *Crop transition matrices and code are provided for use in future modelling studies.*

ABSTRACT

Model-based studies of agricultural systems often rely on the analyst defining realistic crop sequences. This usually involves relying on a few 'typical rotations' that are used in baseline scenarios. These may not account for the variation in farming practices across a region, however, as farmer decision making about which crops to grow is influenced by a combination of economic, environmental and social drivers. We describe and test an approach for generating random realisations of plausible crop sequences based on observed data as quantified by earth observation. Our approach combines crop classification data with a series of crop management rules that reflect the advice followed by farmers (e.g. to reduce the chance of crop-pests and disease). We adapt the approach to generate crop sequences specific to regions and soil type. This demonstrates how the method can be adapted to generate crop sequences typical of a study area of interest.

Keywords:

Crop rotations; Land Cover® *plus*: Crops; Modelling; Crop management; baseline scenario modelling.

1. Introduction

Simulation is an important tool in agricultural research as it allows for the investigation of scenarios that are infeasible to study through experimental trials due to, for example, time or resource constraints. Agricultural systems models have been widely used to investigate management scenarios with a view to identifying ways to improve production efficiency (Reidsma *et al.*, 2009), limit impacts on the environment (Schoumans and Groenendijk, 2000) or investigate trade-offs and synergies between these two (Milne *et al.*, 2020). These types of models, however, can be limited in their success by a scarcity in data needed not only for their parameterisation but also to generate useful scenarios for investigation (Jones *et al.*, 2017). One key aspect of the agricultural system, which is often incorporated into simulation models, is the concept of crop rotation.

Crop rotations are a short sequence of crops that are repeated over time. They are used in many agricultural systems across the globe for several reasons, including improvement of soil health and pest management. Crop rotations tend to vary across regions, developed to account for the characteristics of each region (e.g. due to climate, soil, topography, pests and diseases, etc.). In Great Britain, this has led to a prevalence of arable farming in the east and pastoral farming in the west.

Due to its importance in several key metrics, which are often captured as outputs from simulation models, the chosen crop rotation for a simulation can have important consequences on the outcome of a study. Indeed, many studies focus on choosing the optimal crop rotation to maximise one or more chosen outputs. For example, Smith *et al.* (2016) looked at the effect of changing the rotation design on soil N, P, and K balances. In many studies, it is common to rely on “typical rotations” for use in the baseline scenarios as Smith *et al.* (2016) did in their study on soil nutrients. Alternatively, a model system is chosen, and data on the cropping history from a typical study site or region is used as the baseline for simulations (e.g. Metcalfe *et al.*, 2020). However, the power of modelling for testing agricultural scenarios comes from the ability to model many scenarios in a short space of time and to capture some of the variation in farming practices across a region in a way which cannot be replicated when simply selecting one case study or crop rotation scenario.

Some efforts have been made to develop models for determining crop rotations. However, these have largely been built to address technical constraints to do with reconciling the timing of agronomic events (You and Hsieh, 2017). Castellazzi *et al.* (2008) developed a general method for simulating crop rotations where the transition from one crop to another is represented by a transition matrix where the allocation of a crop in a given year depends on the crop allocated in the previous year. They developed this general methodology to provide a mathematical way to describe predefined crop rotations based on expert knowledge of those rotations. However, the choice of crop grown in a given field is often much more complex than simply following a fixed rotation repeatedly. Indeed, in practice it is unusual for farmers to maintain a predefined crop rotation and use it in succession (I. Shield, pers. comm.). Instead, they will often choose crop sequences according to several factors and their decision over which crop to grow will be based not only on the previous cropping history of the field but other more stochastic elements such as weather, variable costs, crop price and pressure from pests, weeds and diseases. Increasingly, to avoid the build-up of pests, weeds and pathogens and to maintain soil health, farmers are encouraged to maintain diversity in their cropping and to be flexible in their choices (AHDB, 2019).

If cropping history could be used to derive sequences of crops commonly grown and to populate the probability matrix, then it would be possible to use this generic method to simulate current cropping practices. Some efforts have been made to try and derive crop rotation history from remotely sensed data. For example, Mueller-Warrant *et al.* (2017) quantified cropping histories for an 11-year period in north-western Oregon and south-western Washington and used that to understand the cropping sequences that farmers chose to adopt between the end of one grass seed stand and the start of the next. Cropping histories derived from remotely sensed data offer a sound means to quantify how decisions made by individuals manifest as crop sequences at the landscape level. It implicitly accounts for a host of factors influencing crop choice, for example crop prices, regional constraints, synergies between crops or pest pressure. However, to our knowledge, such an approach has not been used to characterise the sequences in current use and predict future ones.

To simulate cropping sequences in lieu of fixed rotations, the decision processes of the farmers need to be reflected in the simulations. In this paper, we present a method to generate crop sequences that characterise the business-as-usual state by implicitly accounting for the environmental constraints (weather, soil, topography, etc.) of a given region and explicitly accounting for constraints related to the control of pests, weeds and diseases. We demonstrate our method using data from Great Britain and as such provide the reader with the necessary data to generate realistic crop sequences for regions across Great Britain.

2. Methods

2.1 Data

Earth observation data have proven to be a useful resource for predicting which crops are grown in field parcels across landscapes (Graesser and Ramankutty, 2017). If these predictions are available across sequential years, then they can be used to estimate the probability of transitioning from one crop to another. The most comprehensive data available indicating crop choices across Great Britain are the UKCEH Land Cover® *plus*: Crops maps. These maps are produced using satellite data and indicate the crop grown in each parcel of agricultural land in Great Britain. Data are currently available for 2016, 2017 and 2018, with more limited information available for 2015. The crops included within the maps are winter wheat (including oats), spring wheat, winter barley, spring barley, oilseed rape, field beans, potatoes, sugar beet, maize, and improved grass. Other cereals, root crops, early potatoes, and vegetables are grouped in a class called 'other'.

2.2 Crop Transition Matrices

For each land-parcel in the UKCEH Land Cover® *plus*: Crops dataset, the crop transitions between seasons were determined (2016 to 2017 and 2017 to 2018). Land parcels where the crop was grass across all years (2015–2018) were excluded. This allowed us to consider only grass crops that are grown as a ley within an arable rotation in isolation from those in continuous grassland. We were then able to describe the probability of transition from any crop in the data set to any other crop in a transition matrix (as an example, see Table 1). These matrices, as described by Castellazzi *et al.* (2008), typically have as many rows and columns as there are distinct crops in the data set and describe the probability of transitioning from one crop (row) to another (column).

Table 1

Example of a crop transition matrix (medium soil subregion of NUTS region H) giving the probability of transitioning from one crop (row) to another (column).

	Next Crop										
Previous Crop	Sugar beet	Field beans	Grass	Maize	Oilseed rape	Other	Potato	Spring barley	Spring wheat	Winter barley	Winter wheat
Sugar beet	0.008	0.007	0.005	0.051	0.002	0.072	0.022	0.105	0.176	0.041	0.512
Field beans	0.012	0.011	0.027	0.01	0.01	0.053	0.009	0.022	0.027	0.065	0.755
Grass	0.033	0.021	0.563	0.044	0.018	0.09	0.018	0.031	0.037	0.041	0.105
Maize	0.067	0.009	0.119	0.172	0	0.129	0.035	0.038	0.049	0.059	0.323
Oilseed rape	0.005	0.002	0.01	0.002	0	0.023	0.003	0.005	0.013	0.058	0.878
Other	0.051	0.033	0.132	0.038	0.05	0.177	0.029	0.044	0.063	0.078	0.306
Potato	0.039	0.004	0.017	0.026	0.001	0.071	0.008	0.027	0.039	0.058	0.711
Spring barley	0.07	0.079	0.044	0.021	0.166	0.085	0.033	0.09	0.082	0.133	0.195
Spring wheat	0.075	0.078	0.054	0.022	0.1	0.094	0.038	0.052	0.097	0.142	0.248
Winter barley	0.118	0.06	0.048	0.02	0.345	0.086	0.035	0.029	0.035	0.123	0.102
Winter wheat	0.118	0.079	0.027	0.022	0.124	0.082	0.053	0.05	0.05	0.141	0.255

2.3 Additional considerations

To generate typical cropping sequences, we must address the two main concerns in a farmer's decision-making process: the environment; and, controlling pest, weed, and disease pressure.

2.3.1 The environment

To address the first factor (the environment), we decide to split Great Britain by region. The reason for regional differences in farming systems in Great Britain is primarily due to climate and topology; with the wetter hillier regions in the west being favoured for grazing systems whilst the drier, flatter parts of Great Britain tending to be favoured for arable crops. In addition to climate and topography, soil type also determines the suitability of a field for growing a particular crop. To address this, we first split Great Britain into NUTS1 regions, which are a Eurostat geocode that references the subdivisions of the United Kingdom of Great Britain and Northern Ireland for statistical purposes (Fig. 1) and capture the course-scale regional differences across the UK. We then used data on soil type to classify the regions further. We divided each region into three subregions according to the soil clay content (FAO *et al.*, 2012): light soils were classified as having less than or equal to 18% clay; heavy soils as having greater than 35% clay; and medium soils as anything in between. We then assigned each land parcel within the Land Cover® *plus*: Crop maps to one or more of our subregions, with a land parcel overlapping a subregion boundary assigned to both subregions.

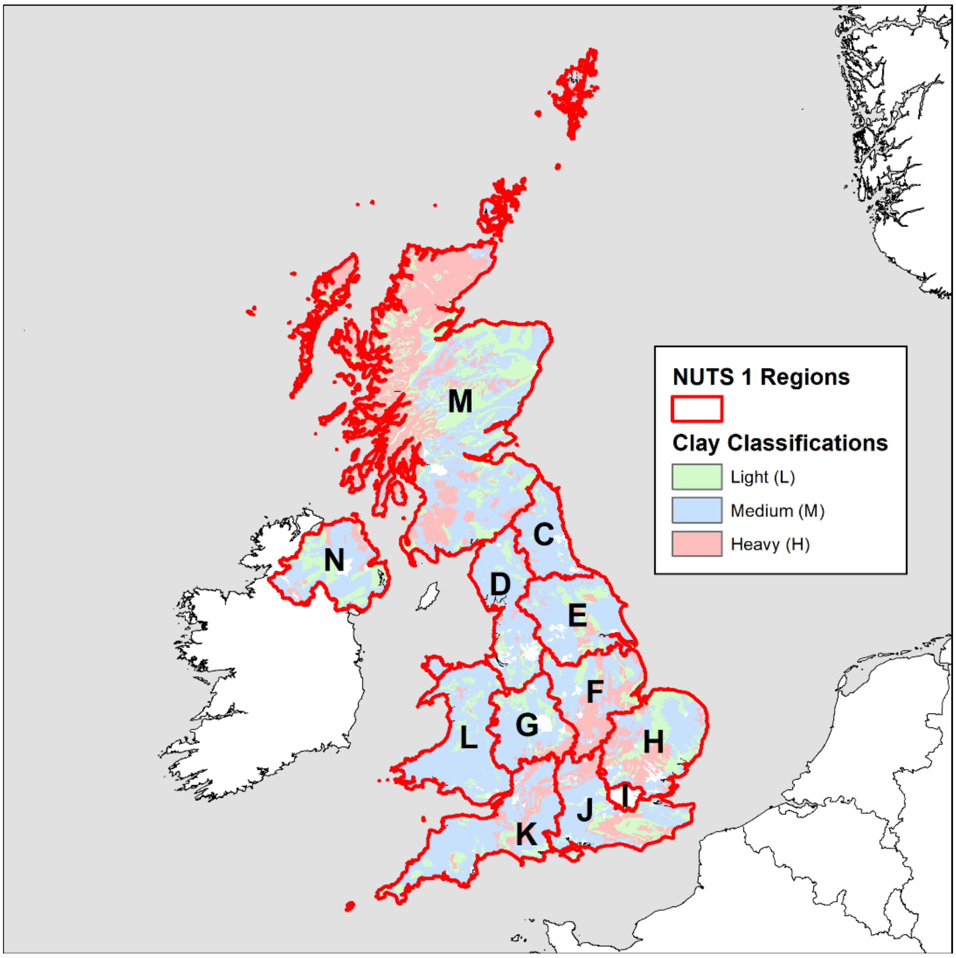


Fig. 1. A map of the UK showing subdivision according to NUTS1 regions and soil clay content.

2.3.2 Controlling pest, weed, and disease pressure

To account for the second aspect of a farmer's decision-making process (controlling pest, weed, and disease pressure) we incorporated a set of rules either directly into the transition matrices or, for rules that cannot be easily incorporated, by adding a rule-based step to our simulations.

Because of the investment required to plant a grass ley and in order to reduce pest, weed and disease pressure, it is recommended that grass leys are kept for at least two years (I. Shield, pers. comm.). We therefore represented grass as two crop types in our matrix (bringing the total number of crops to 12 in this case). We refer to our grass crops as G_1 and G_S, which represents the grass-ley in its first year and the grass-ley in subsequent years, respectively. By our rule-based definition that there must be at least two years of grass: (i) G_1 can only transition to G_S (which can be expressed as $P(G_S|G_1) = 1$, i.e. the probability that the next crop is G_S given that our current crop is G_1 is 1); (ii) G_1 is the only crop that can transition to G_S (i.e. other crops, C, cannot transition to it; $P(G_S|C) = 0$); and, (iii) G_S cannot transition to G_1 (i.e. $P(G_1|G_S) = 0$). From our data, we can directly calculate the transition probability from each non-grass crop, C, to G_1 in the same way we calculate other transition probabilities between crops. To calculate the $P(G_S|G_S)$ we must adjust our observed probability of grass remaining from one year to the next to account for our restriction that $P(G_S|G_1) = 1$. This is calculated from

$$P(G_S|G_S) = \frac{P(G)P(G|G) - P(G_1)P(G_S|G_1)}{P(G_S)}, \tag{1}$$

where: $P(G_S|G_1) = 1$ (as defined above); $P(G|G)$ is defined from our data; the probability of being in a grass crop, $P(G)$, is defined in the grass component of the steady state vector, v, such that the transition matrix, M, derived from the data satisfies $M'v = v$ where v is a vector that records crop proportions at steady state and M' is the transpose of M; the probability of being in the first year of a grass ley, $P(G_1)$, is estimated from $N'y = y$ where N' and y are the modified versions of M' and v , respectively, in which the grass component is split into first and subsequent grass; and, the probability of being in a subsequent year of a grass ley, $P(G_S) = P(G) - P(G_1)$. These modified transition matrices, N, are given in Metcalfe and Sharp (2021).

Using transition matrices calculated from the remotely sensed data, we accounted for the agronomic "rules" that depend on the previously grown crop. For example, constraints on cropping transitions due to overlap in the crop harvest and sowing times such as the incompatibility between following an early-winter-harvested sugar beet crop with an early-autumn-sown cereal crop will be avoided as they will not be observed in the data. However, some crop rules require consideration of more than one year of cropping history. These types of rules generally either place limits on the crop frequency or on continuous cropping. Such constraints generally arise to break the cycle or to prevent build-up of pests, weeds or diseases. For example, a requirement of four years between sugar beet crops is common to minimise the effect of beet cyst nematode (Wibberley, 1996). Similarly, it is unusual for a farmer to grow more than two wheat crops consecutively to avoid the build-up of fungal diseases such as take-all (Castellazzi *et al.*, 2008). Crop rules limiting both the frequency of crops within a sequence and the maximum length of a continuous sequence of certain crops were implemented according to standard agronomic practice in England and Wales (Table 2).

Table 2

Crop rules not captured by the transition matrices that are implemented within our model to simulate standard agronomic practice aimed at preventing the build-up of pest, weeds, and diseases.

Crop	Rule	Reference
Limit on continuous cropping		
Wheat	No more than two consecutively	Castellazzi *et al.* (2008)
Maize	No more than five consecutively	AHDB (2014)
Grass	Grass ley lasting no more than four years	Defra (2019)
Limit on crop frequency		
Potato	maximum one crop in four years	Wolny (1992)
Beet	maximum one crop in four years	Wibberley (1996)
OSR	maximum one crop in four years	Hilton *et al.* (2013)
Grass	Break of two years between grass leys	I. Shield, pers. comm.

2.4 Simulation

We simulated crop sequences in MATLAB (MATLAB, 2018). For each subregion type we initialised a model with 1000 fields. The number of fields starting in each crop type was determined by calculating the steady state proportions v for each transition matrix M, that is the vector v such that $M'v = v$, where M' is the transpose of M. We then simulated 200 years of crop choices in each of those fields. In each year of simulation, the next crop to be sown in the field was drawn from a probability distribution according to the row of the transition matrix corresponding to the crop currently in that field. If a crop rule enforced that a certain crop could not be grown in a given field (e.g. potatoes were grown the previous year so are not allowed again this year) then the crop was removed from the transition matrix and other values rescaled so that the row summed to 1. If, after all rules have been enforced, there is no crop left to choose from, we default back to the steady state proportions (this only occurs in regions with limited available data, e.g. NUTS region I, London). The code is available from Zenodo (Metcalfe and Sharp, 2021).

2.5 Analysis

We analysed the resulting crop sequences using the R software package (2018) for any commonalities. We removed the first ten years of simulated crops as there were large fluctuations in some regions. For the simulated time period of 11-200 years we counted the number of unique crop sequences of lengths three to five crops across the 1000 fields and determined the most frequently observed crop sequences for each subregion. To validate these results, half of the satellite data were used as a training data set to produce the crop transition matrices with the other half used as a validation set. Crop sequences were simulated from these crop transition matrices. The probability of a given three-crop sequence (e.g. a winter wheat – winter barley – winter wheat sequence) were calculated from the simulated sequences and the observed sequences from the validation data and compared. We performed this analysis both with and without our additional crop rules from Table 2. By comparing the validation set with sequences generated without the additional crop rules, we can assess how well the transition matrices are able to predict the validation data. While our additional crop rules aren't enforced on the validation set sequences, we also compare these sequences with the simulated sequences that were generated with the additional crop rules.

3. Results

During the simulated 200 years of cropping, crop proportions generally fluctuated for a short initial period, due to our additional crop rules, before stabilising, e.g. with the NUTS L region in Fig. 2. Each subregion followed its own unique distribution of crop types and there were distinct differences in the simulated cropping frequencies across both NUTS1 regions and soil types.

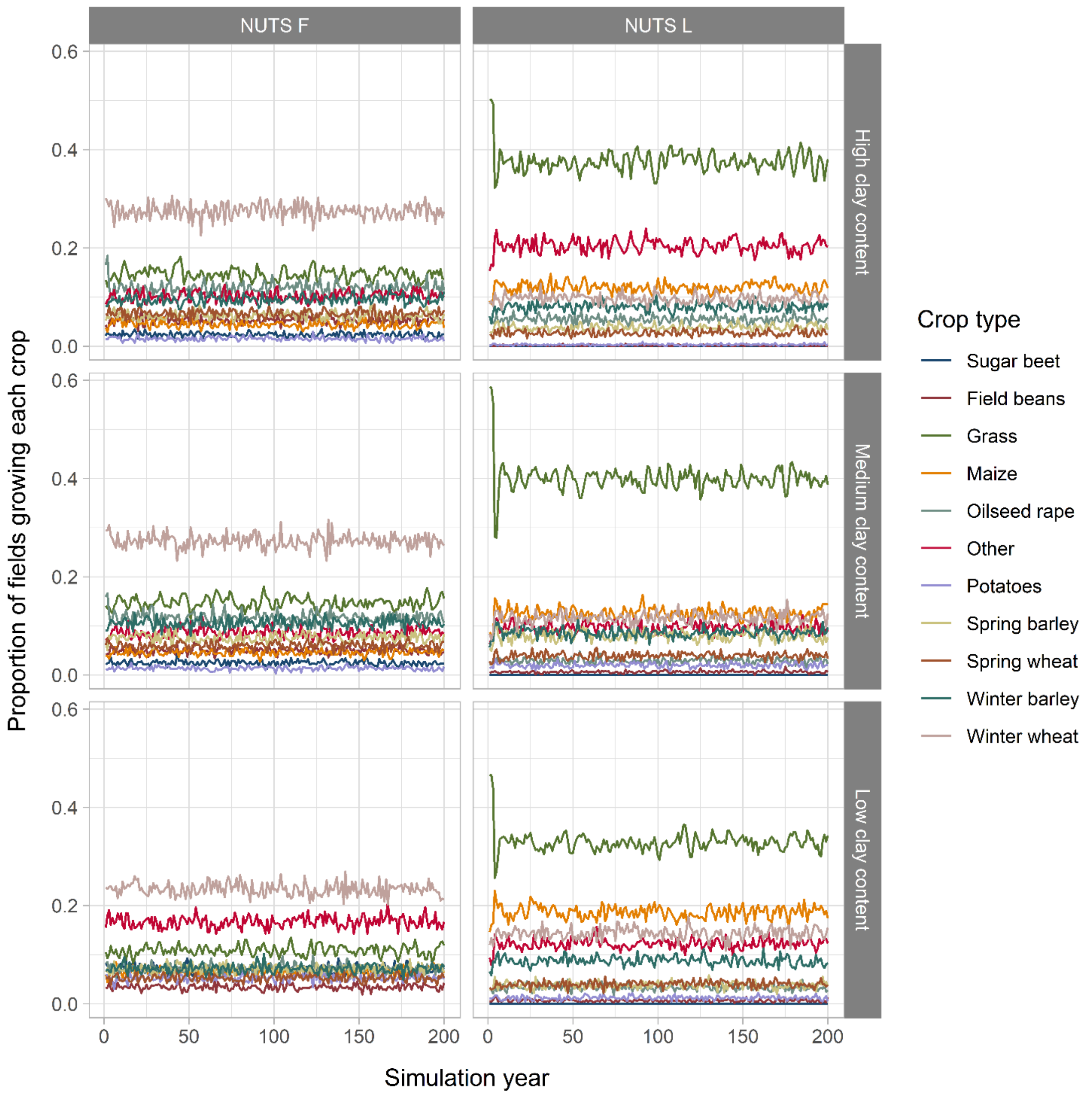


Fig. 2. Example simulations for six subregions of Great Britain, encompassing two NUTS regions (East Midlands, F and Wales, L) and the three soil types. Here the proportion of fields growing each crop is shown over 200 years of simulation.

The number of unique crop sequences of lengths three to five crops varied between subregion (Table 3). Some regions had very conserved crop sequences. In the heavy soil subregion of NUTS region I (London) only 10 different three crop sequences were simulated, accounting for only 0.58% of the total possible permutations. This is due to the limited number of fields in this subregion, however. Outside of London, the region with the fewest sequences was the heavy soil subregion of NUTS region L (Wales) where there were 413 different three-crop sequences observed, accounting for 23.9% of the total possible permutations. The NUTS regions E (Yorkshire and the Humber), F (East Midlands), and H (East of England) showed much greater diversity in the crop sequences being simulated with the light soil subregion of NUTS region F displaying 64% of the total possible three crop sequences. When we look at the generated crop sequences of length five there is much less variation and we see a much smaller subset of the total possible crop sequences generated. However, region M still shows very low diversity (only 2.39% of all total five-crop sequences on the heavy soil) and region F still shows very high diversity (15.75% of all total five-crop sequences on the light soil).

Table 3

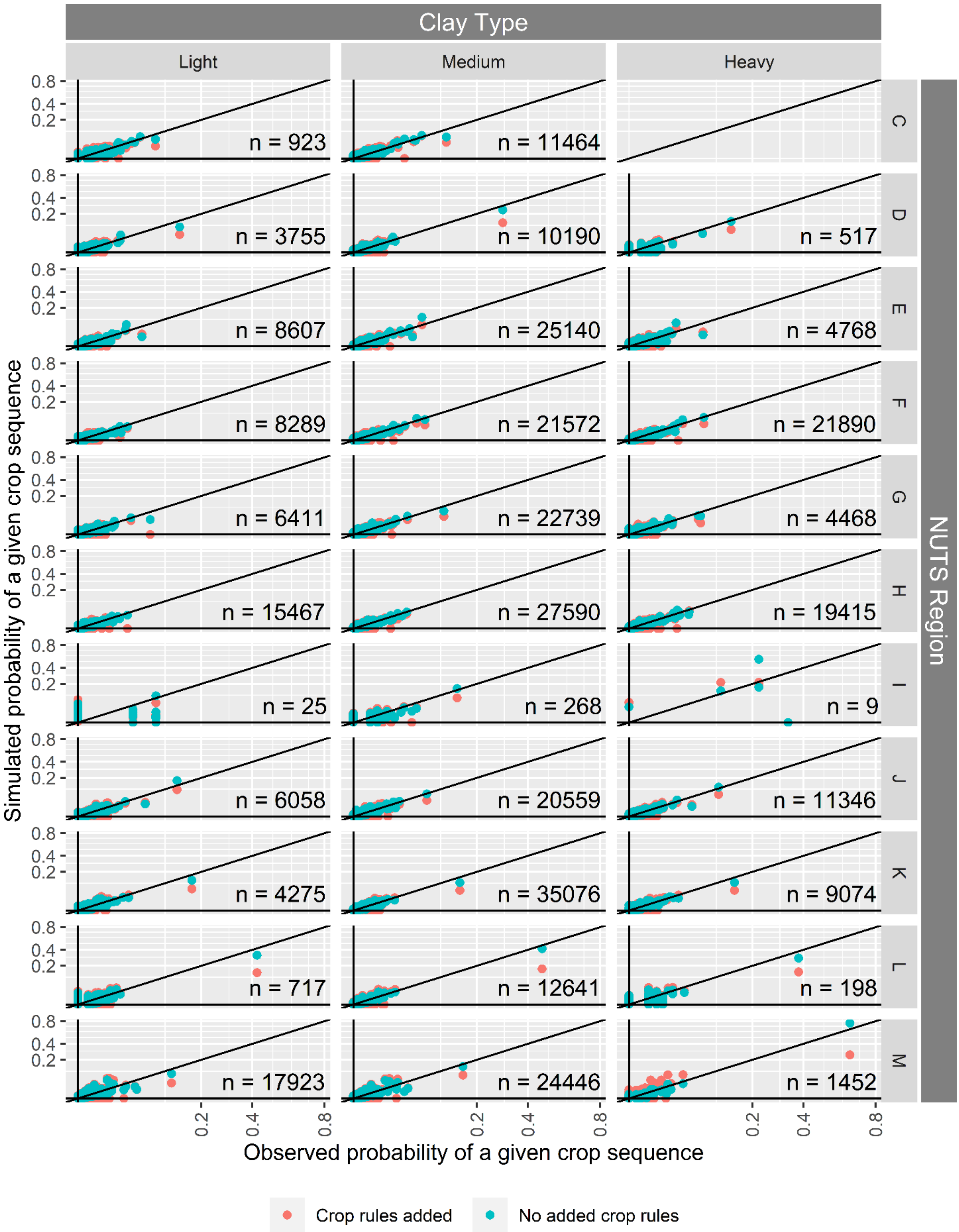
The number of unique crop sequences of length three to five crops simulated in each subregion across 1,000 fields for 200 years (maximum possible permutations of twelve crops is 1,728 three-crop sequences, 20,736 four-crop sequences, 248,832 five-crop sequences).

Sequence Length:		three-crops		four-crops		five-crops	
NUTS-1	*Clay Classification*	*Count*	*% of total possible*	*Count*	*% of total possible*	*Count*	*% of total possible*
C	Heavy	*	*	*	*	*	*
	Medium	748	43.29	4216	20.33	15288	6.14
	Light	721	41.72	4495	21.68	17580	7.07
D	Heavy	698	40.39	4695	22.64	22200	8.92
	Medium	833	48.21	5635	27.17	23858	9.59
	Light	819	47.4	5933	28.61	27968	11.24
E	Heavy	1043	60.36	7293	35.17	30000	12.06
	Medium	1030	59.61	6559	31.63	24631	9.9
	Light	1053	60.94	7847	37.84	33818	13.59
F	Heavy	1058	61.23	7407	35.72	29453	11.84
	Medium	1069	61.86	7424	35.8	29562	11.88
	Light	1106	64	8704	41.98	39200	15.75
G	Heavy	784	45.37	5123	24.71	21443	8.62
	Medium	804	46.53	5391	26	22636	9.1
	Light	822	47.57	5788	27.91	26340	10.59
H	Heavy	1024	59.26	6551	31.59	25027	10.06
	Medium	1057	61.17	7306	35.23	28992	11.65
	Light	1050	60.76	7778	37.51	34182	13.74
I	Heavy	10	0.58	17	0.08	28	0.01
	Medium	543	31.42	3382	16.31	14039	5.64
	Light	160	9.26	557	2.69	1834	0.74
J	Heavy	800	46.3	5105	24.62	21448	8.62
	Medium	888	51.39	5440	26.23	22035	8.86
	Light	865	50.06	5552	26.77	21541	8.66
K	Heavy	764	44.21	4868	23.48	20626	8.29
	Medium	809	46.82	5556	26.79	25382	10.2
	Light	790	45.72	5312	25.62	22762	9.15
L	Heavy	413	23.9	2335	11.26	9877	3.97
	Medium	811	46.93	5037	24.29	19488	7.83
	Light	616	35.65	3968	19.14	16164	6.5
M	Heavy	429	24.83	2034	9.81	5950	2.39
	Medium	810	46.88	4964	23.94	19103	7.68
	Light	786	45.49	4863	23.45	19003	7.64

* the large majority of this subregion was permanent grassland and so too few transitions remained to make it viable for simulation.

The sequences most commonly generated vary by region, however this is in line with expectation with more arable-type rotations in the east and pastoral or mixed systems in the west. In NUTS region F (East Midlands) the most commonly predicted three-crop sequence was winter wheat – oilseed rape – winter wheat on all three soil types. As the crop sequences are lengthened, either field beans, winter barley or potatoes are introduced. In NUTS region K (South West), however, we most commonly predict three- to four-year grass leys. In a Defra project (Defra, 2010) that previously looked at determining crop rotations in different regions of England they predicted crop rotations very similar to ours in the east of England but they did not include grass in their predictions and so their results differ for the south west. However, their predicted rotation of winter wheat – winter wheat – winter barley – oilseed rape is very similar to our most frequent four-crop sequence (after removing any grass leys) of winter wheat – winter barley – oilseed rape – winter wheat generated by our model in NUTS region K on the medium soil. Even in regions where the cropping sequences are very conserved, such as the heavy soil subregion of NUTS region L (Wales) the most commonly predicted three-crop sequence only occurred 8.4% of the time. In other regions where there was more variety in cropping such as the light soil subregion of NUTS region F (East Midlands) the most commonly predicted three-crop sequence occurred just 2% of the time. This suggests that farmers are indeed not repeating typical rotations and are in fact actively adapting the crops produced according to the current conditions.

Fig. 3 plots the results from our validation analysis. Here we compare our simulated sequences generated both with and without the additional crop rules from Table 2 to the sequences from the validation data. We first compare the sequence simulated without the additional crop rules. These simulations performed well, with the simulated probability of a given of crop sequence closely matching the probability observed in the validation data. This demonstrates the goodness of fit of the crop transition matrices that were produced using the training data and shows that crop transition matrices are an effective way of replicating regional crop choice. The main source of error occurred with the heavy soil subregion of region C. This is likely due to the lack of data in this subregion, however. Enforcing the additional crop rules on the simulated sequences results in a poorer fit with the validation data, this is to be expected however as the validation sequences do not necessarily comply with the additional crop rules of Table 2. For instance, there were sequences in the validation set that are strictly prohibited by our crop rules, e.g. a grass-crop-grass sequence. The largest discrepancy that we consistently found across subregions was in the prevalence of all-grass sequences, generally indicated by the points to the farthest right of the plots, e.g. the heavy soil subregion of region M in which approximately 64% of the validation sequences were all-grass sequences compared to 25% of our sequences simulated with the additional crop rules. This is again expected as our crop rules control both the maximum consecutive number of years a grass ley in an arable rotation can be planted for, as well as the minimum time to wait between planting grass leys, all of which serves to reduce the amount of grass in the simulated sequences. Possible reasons why our crop rules aren't reflected in the data could include unexpected farmer behaviour, rented land in which new tenants don't base their decisions on the actions of prior tenants or due to crop misclassifications in the data (UKCEH, 2018). This therefore highlights the need to include these additional crop rules.



257

258 **Fig. 3.** Plot of the probability of a given three-crop sequence appearing in the validation data and the

259 simulations generated using the training data. An identity line is included to assess the accuracy of the

260 simulated sequences. The heavy soil subregion of region C (North East England) was not simulated due to

261 insufficient data. The number of fields in the validation set is given by n.

4. Discussion

Simulation of crop sequences using transition matrices in combination with agronomic rules allows us to generate realistic sequences of crops reflecting the business-as-usual state at the subregional level in Great Britain. The decision as to which crop a farmer will grow is a two-stage process taking into account: the environment; as well as, pest, weed and disease pressure. Lawes and Renton (2010) took some steps to incorporating some of these decisions into their land use sequence optimiser accounting for the need to place break crops into the crop sequence to provide relief from soil pathogens, to minimise populations of herbicide resistant weeds or to increase soil nitrogen reserves. In this article we incorporated these decisions in two ways. By populating transition matrices at the subregion level, we accounted for decisions related to the environment as soil and climatic conditions tend to vary at this scale. These transition matrices only account for agronomic rules that depend on the previously grown crop, however. By applying additional crop rules from within the simulation, that depend on crops planted prior to the previously grown one, we can better account for the second part of the decision-making process.

An alternative approach to including crop-based rules explicitly in the method would be to look at transition probabilities conditional on more than just the previous year (e.g. if one observes winter wheat twice in a rotation then what is most likely to come next?). Whilst this approach is attractive because it is purely data driven and so captures a more realistic picture of crop sequences, it is limited by the fact it requires a large data set over a number of years. When this study was performed, there were only three reliable years of data available from Land Cover® *plus*: Crops, and so a more detailed data driven approach over more than three years was not possible.

In defining the crop sequence transitions in subregions according to NUTS1 areas and soil types, we have assumed homogeneity within these regions with respect to crop patterns. Whilst this is a reasonable assumption to make, as the combination of NUTS1 regions and soil type capture the key geographical and environmental characteristics, it is possible that potentially large within-subregion variation is ignored. Furthermore, the use of discrete subregions results in discontinuities at the boundaries, which is an artefact of the aggregation and not likely in reality. Localised sequences smoothed across space would overcome these issues, however it may suffer other problems including interpretation, lack of data used to estimate sequences and subjective choice of smoothing parameter. Such approaches still rely on some defined region from which sequence patterns are derived and hence, although these are likely to be much smaller regions to capture the localised sequences, issues around homogeneity assumptions may still exist.

To further overcome potential issues with high within subregion variability in crop sequences, a more detailed classification could be undertaken. For example, we also implemented a hierarchical cluster analysis to form classes according to soil and topographic variables considered important in crop choice (clay, silt, organic carbon and bulk density). For ease of exposition and to keep the number of transition matrices to a manageable level for reporting, we chose not to present those results here, but highlight this method as an alternative for defining regions.

Previous work on the simulation of crop rotations has focused on finding the best rotation to optimise a given objective, be it economic (Pakawanich *et al.*, 2020), environmental (dos Santos *et al.*, 2011), or regulatory (Mauri, 2019). Some studies have even considered combining multiple objectives such as maximising economic returns whilst reducing nitrogen loss and soil erosion (Watkins and Lu, 1998). However, objective-oriented crop rotations do not necessarily reflect the decision-making process considered by farmers, as the "best" crop sequence will not only depend on cropping constraints but also a number of stochastic factors, such as weather, pest pressure, and the environment. For example, we found that the most common three-crop sequence for each region occurred between 2-8.4% of the time, supporting the idea that there is no typical crop rotation that is used repeatedly. By taking a data-driven probabilistic approach we hope to be able to capture some of these stochastic processes when the cropping sequences are scaled up to a landscape scale.

It is important to consider that Land Cover® *plus*: Crops contains uncertainty that could influence the observed transition probabilities. Overall accuracy of the product is around 86%, but this varies between crop types with grass, oilseed rape and winter wheat having over 90% accuracy and field beans and spring wheat having below 50% accuracy (UKCEH, 2018). Misclassifications could lead to incorrect transition probabilities and therefore spurious crop sequences. For example, the originally calculated $P(G_S|G_S)$ was negative in the light soil subregions of NUTS regions G (West Midlands) and H (East of England), these were set to zero and the other probabilities in the row were rescaled to sum to one. We believe these negative values are due to our assumption that $P(G_S|G_1) = 1$ is not necessarily what is found in the Land Cover® *plus*: Crops data. This is likely due to a combination of misclassifications in the Land Cover® *plus*: Crops data, the small time-window of data available, and farmers planting grass for just a single year. If farmers are planting grass for a single year, then we expect the number of cases in which this happens to be small because of the investment required to plant a grass ley and due to the recommendation to plant grass for at least two years to help reduce pest, weed and disease pressure (I. Shield, pers. comm.).

Here we have described a method of simulating crop sequences that reflect the business-as-usual state at the subregional level for Great Britain. The use of our method to generate realistic cropping sequences will allow agricultural systems modellers to move away from simulating crop rotations, an agronomic practice rarely implemented rigidly on farms, and instead mimic more accurately the decision processes undertaken by a farmer when making crop choices. This method is well suited to simulating realistic crop sequences and so will support business-as-usual scenarios. Not only this, but the method offers a sound way to investigate future scenarios. These could be simulated by incorporating a greater diversity of crops into the transition matrix or by including more agronomic rules on the frequency of cropping. Simulation in this way could also allow gradual transitions to new cropping regimes over time rather than abrupt changes at arbitrary points in a simulation when the crop rotation is changed to some "future scenario". As more data become available these "business-as-usual" transition matrices should be updated to reflect the changing patterns in farming. For example, recent changes in pesticide regulations have left many farmers defenceless against the cabbage stem flea beetle, a serious pest in oil seed rape, resulting in a dramatic reduction in the area that this crop is grown (Dewar, 2017). Shocks of this type, including climate change, changes in crop prices, new invasive pests and diseases and further loss of chemical control methods will continue to impact observed crop sequences and so it is important to update transition matrices on a regular basis. In addition, as methods for identifying crops are developed and improve, the satellite predictions of crop types will become more accurate and able to detect a wider variety of crops leading to more complex transition matrices that reflect this.

Acknowledgements

This work was supported by: Institute Strategic Programme Grant 'Achieving Sustainable Agricultural Systems' (ASSIST) [grant number NEC05829]; Institute Strategic Programme Grant 'Soils to Nutrition' (S2N) [grant number BBS/E/C/000I0330]; 'Transforming and growing relationships within regional food systems for Improved nutrition and sustainability' (TGRAINS) [grant number BB/S014292/1]; and, Strategic Priorities Fund 'Landscape Decisions: Towards a new framework for using land assets' programme 'New Science to Enable the Design of Agricultural Landscapes that Deliver Multiple Functions – AgLand' [grant numbers NE/T001178/1 (Rothamsted Research) and NE/T000244/2 (UKCEH)].

References

AHDB, 2014. Growing and feeding forage maize – a review. Work Package 3b: Alternative forages. Research Partnership: Grasslands, Forage and Soil.

AHDB, 2019. How resilient is your rotation? [Online]. Available: https://cereals.ahdb.org.uk/publications/2018/june/07/how-resilient-is-your-rotation.aspx [Accessed 17/09/2019].

Castellazzi, M.S., Wood, G.A., Burgess, P.J., Morris, J., Conrad, K.F., Perry, J.N., 2008. A systematic representation of crop rotations. Agricultural Systems 97 (1-2), 26-33. https://doi.org/10.1016/j.agsy.2007.10.006.

Defra, 2010. Soil carbon: studies to explore greenhouse gas emissions and mitigation, Project code SP1106 [Online]. Available: http://sciencesearch.defra.gov.uk/Default.aspx?Menu=Menu&Module=More&Location=None&Completed=0&ProjectID=17323

Defra, 2019. June Survey of Agriculture and Horticulture [Online]. Available: https://www.gov.uk/government/statistical-data-sets/structure-of-the-agricultural-industry-in-england-and-the-uk-at-june

Dewar, A.M., 2017. The adverse impact of the neonicotinoid seed treatment ban on crop protection in oilseed rape in the United Kingdom. Pest Manage. Sci. 73 (7), 1305-1309. https://doi.org/10.1002/ps.4511.

dos Santos, L.M.R., Michelon, P., Arenales, M.N., Santos, R.H.S., 2011. Crop rotation scheduling with adjacency constraints. Annals of Operations Research 190 (1), 165-180. https://doi.org/10.1007/s10479-008-0478-z.

FAO, IIASA, ISRIC, ISSCAS, JRC, 2012. Harmonized World Soil Database (version 1.21). FAO, Rome, Italy and IIASA, Laxenburg, Austria.

Graesser, J., Ramankutty, N., 2017. Detection of cropland field parcels from Landsat imagery. Remote Sens. Environ. 201, 165-180. https://doi.org/10.1016/j.rse.2017.08.027.

Hilton, S., Bennett, A.J., Keane, G., Bending, G.D., Chandler, D., Stobart, R., Mills, P., 2013. Impact of Shortened Crop Rotation of Oilseed Rape on Soil and Rhizosphere Microbial Diversity in Relation to Yield Decline. PLOS ONE 8 (4), e59859. https://doi.org/10.1371/journal.pone.0059859.

Jones, J.W., Antle, J.M., Basso, B., Boote, K.J., Conant, R.T., Foster, I., Godfray, H.C.J., Herrero, M., Howitt, R.E., Janssen, S., Keating, B.A., Munoz-Carpena, R., Porter, C.H., Rosenzweig, C., Wheeler, T.R., 2017. Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. Agricultural Systems 155, 269-288. https://doi.org/10.1016/j.agsy.2016.09.021.

Lawes, R., Renton, M., 2010. The Land Use Sequence Optimiser (LUSO): A theoretical framework for analysing crop sequences in response to nitrogen, disease and weed populations. Crop and Pasture Science 61 (10), 835-843. https://doi.org/10.1071/CP10026.

MATLAB, 2018. 9.5.0 (R2018b). The MathWorks, Inc., Natick, Massachusetts, United States.

Mauri, G.R., 2019. Improved mathematical model and bounds for the crop rotation scheduling problem with adjacency constraints. European Journal of Operational Research 278 (1), 120-135. https://doi.org/10.1016/j.ejor.2019.04.016.

Metcalfe, H., Sharp, R.T., 2021. CropSequenceGenerator (Version 1.0.4). Zenodo. https://doi.org/10.5281/zenodo.5001978.

Metcalfe, H., Milne, A.E., Deledalle, F., Storkey, J., 2020. Using functional traits to model annual plant community dynamics. Ecology 101 (11), e03167. https://doi.org/10.1002/ecy.3167.

Milne, A.E., Coleman, K., Todman, L.C., Whitmore, A.P., 2020. Model-based optimisation of agricultural profitability and nutrient management: a practical approach for dealing with issues of scale. Environ. Monit. Assess. 192 (11), 730. https://doi.org/10.1007/s10661-020-08699-z.

Mueller-Warrant, G.W., Trippe, K.M., Whittaker, G.W., Anderson, N.P., Sullivan, C.S., 2017. Spatial methods for deriving crop rotation history. International Journal of Applied Earth Observation and Geoinformation 60, 22-37. https://doi.org/10.1016/j.jag.2017.03.010.

Pakawanich, P., Udomsakdigool, A., Khompatraporn, C., 2020. Robust production allocation model for an agricultural cooperative with yield uncertainty and similar revenue constraints. Comput. Electron. Agric. 168. https://doi.org/10.1016/j.compag.2019.105090.

R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Reidsma, P., Ewert, F., Boogaard, H., van Diepen, K., 2009. Regional crop modelling in Europe: The impact of climatic conditions and farm characteristics on maize yields. Agricultural Systems 100 (1-3), 51-60. https://doi.org/10.1016/j.agsy.2008.12.009.

Schoumans, O.F., Groenendijk, P., 2000. Modeling Soil Phosphorus Levels and Phosphorus Leaching from Agricultural Land in the Netherlands. Journal of Environmental Quality 29 (1), 111-116. https://doi.org/10.2134/jeq2000.00472425002900010014x.

Smith, L.G., Tarsitano, D., Topp, C.F.E., Jones, S.K., Gerrard, C.L., Pearce, B.D., Williams, A.G., Watson, C.A., 2016. Predicting the effect of rotation design on N, P, K balances on organic farms using the NDICEA model. Renew. Agric. Food Syst. 31 (5), 471-484. https://doi.org/10.1017/S1742170515000381.

UKCEH, Land Cover® *plus*: Crops © UKCEH. © RSAC. © Crown Copyright 2007.

UKCEH, 2018. CEH Land Cover® plus Crop Map: Quality Assurance [Online]. Available: https://www.ceh.ac.uk/ceh-land-cover-plus-crop-map-quality-assurance [Accessed 27/03/2020].

Watkins, K.B., Lu, Y.C., 1998. Economic and Environmental Tradeoffs Among Alternative Seed Potato Rotations. J. Sustainable Agric. 13 (1), 37-53. https://doi.org/10.1300/J064v13n01_05.

Wibberley, J., 1996. A brief history of rotations, economic considerations and future directions. Aspects of Applied Biology (United Kingdom).

Wolny, S., 1992. The threat of parasitic nematodes to farm crops grown in various rotations and monoculture. Acta Academiae Agriculturae ac Technicae Olstenensis, Agricultura (55), 103-113.

You, P.S., Hsieh, Y.C., 2017. A computational approach for crop production of organic vegetables. Comput. Electron. Agric. 134, 33-42. https://doi.org/10.1016/j.compag.2016.11.003.