Identifying artificially drained pasture soils using machine learning and Earth observation imagery

Article *in* Journal of Applied Remote Sensing · July 2020

DOI: 10.1117/1.JRS.14.034508

6 authors, including:

Rob O'Hara
TEAGASC - The Agriculture and Food Development Authority
8 PUBLICATIONS **32** CITATIONS
SEE PROFILE

Stuart Green
TEAGASC - The Agriculture and Food Development Authority
92 PUBLICATIONS **950** CITATIONS
SEE PROFILE

Conor Cahalane
National University of Ireland, Maynooth
24 PUBLICATIONS **342** CITATIONS
SEE PROFILE

Owen Fenton
TEAGASC - The Agriculture and Food Development Authority
242 PUBLICATIONS **3,571** CITATIONS
SEE PROFILE

Identifying artificially drained pasture soils using machine learning and Earth observation imagery

Rob O'Hara,[a,b,*] Stuart Green,[a] Tim McCarthy,[c] Conor Cahalane,[d] Owen Fenton,[e] and Pat Tuohy[f]

[a]Teagasc, Ashtown Food Research Centre, Rural Economy and Development Programme, Dublin 15, Ireland

[b]VistaMilk SFI Research Centre, Teagasc, Moorepark, County Cork, Ireland

[c]Maynooth University, National Centre for Geocomputation, County Kildare, Ireland

[d]Maynooth University, Department of Geography, County Kildare, Ireland

[e]Teagasc, Environmental Research Centre, Johnstown Castle, County Wexford, Ireland

[f]Teagasc, Animal and Grassland Research and Innovation Centre, Moorepark, County Cork, Ireland

Abstract. In many areas of the globe, the installation of artificial drains on naturally poorly drained soils is a necessary part of farm management. Identifying the location of artificially drained areas is an important step in achieving environmentally sustainable agricultural production. However, in many regions, data on the presence or the distribution of artificial drainage systems are rare. We outline an approach to identify artificially drained soils using Earth observation (EO) satellite imagery and digital elevation data. The method exploits the contrasting phenology of grass during a peak growth stage to identify artificially drained and undrained soils. Two machine-learning techniques, support vector machine and random forest, were tested. Classification accuracy up to 91% was achieved using photointerpreted accuracy points using higher resolution satellite imagery. Additional investigations would be required to establish whether the drained conditions identified were a result of artificial drainage or from naturally well-drained soils occurring within larger soil units. Herein, the Republic of Ireland is used as a test case. Based on our findings, the area of artificially drained grassland within the study area could be revised upward, with 44% (or ∼345, 000 ha) of pasture currently classed as "poorly drained" identified as "artificially drained." At one location, a change in the modeled drainage condition at field level was demonstrated following drain installation. The presented method demonstrates the ability of EO satellites to quickly and accurately map field drainage status at farm management scales over a wide area. This has the potential to improve management decisions at local scales, but also has implications in terms of national policy development and regulation in areas such as water quality and climate change mitigation. [DOI: 10.1117/1.JRS.14.034508]

Keywords: Landsat; heavy soils; artificial drainage; grassland.

Paper 200355 received May 8, 2020; accepted for publication Jul. 21, 2020; published online Jul. 30, 2020.

1 Introduction

Farms situated on poorly drained (wet) mineral soils can be less agriculturally productive and less environmentally sustainable than farms located on drained (drier) soils.[1] For example, previous studies have demonstrated how pasture farms on poorly drained soils have reduced grass growth and grass utilization, restricted access to fields for livestock and machine traffic, and lowered stocking rates and yields.[2–4] Sometimes referred to as "heavy" soils because they are sticky and difficult to work, poorly drained mineral soils are characterized by low porosity and compacted structure that reduces their ability to transport water. They can be found in lowland settings over shallow water tables, or where groundwater exudes from surface seepage or springs. Conversely, "drained" soils tend to have coarser texture (greater proportion of sand),

*Address all correspondence to Rob O'Hara, E-mail: rob.ohara@teagasc.ie

1931-3195/2020/$28.00 © 2020 SPIE

higher porosity and are not disadvantaged in terms of landscape position or groundwater breakout.

Heavy soils can be improved for intensive agricultural production by installing artificial drainage (also referred to as field drainage or land drainage) to control groundwater levels and/or improve soil permeability.[5] If the correct type of system is installed and correctly maintained, it should provide an economic benefit with improved yields and greater access for livestock and machinery with less risk of soil compaction.[6] However, the installation of artificial drainage can have significant negative environmental impacts. For example, artificial drainage is a major source of nutrient enrichment of surface water and groundwater, compromising water quality and destroying aquatic habitats.[7,8] The drainage of heavy soils also has complex impacts on greenhouse gases (GHG) emissions,[9,10] with drained soils potentially becoming a source[11] or a sink[12] for atmospheric CO_2, as well as potentially reducing nitrous oxide (N_2O) emissions from wet soils (denitrification).[13] Functioning land drainage is an important indicator of intensive agricultural management in some regions, with changes in drainage status and management intensity over time a major source uncertainty in current GHG inventory accounting.[14] Paul et al.[15] demonstrated how better definition of artificially drained soils could improve the accuracy of carbon emission budget estimates. However, without access to detailed maps of drained soils, the authors had to assume all poorly drained soils below 200 m above sea level on slopes ≤ 12 deg had artificial drainage in place. Accurate spatial data on artificially drained soils also have broader applications in agrometeorological forecasting[16] or in modeling variability in grass production.[3,17]

Globally, the extent of artificially drained soils is not well documented. Estimates suggest that ~167 million hectars (ha) of (rain-fed) farmland has some form of artificial drainage in place; however, few countries collect data on drained areas in any harmonized or systematic way.[18] General soil maps will often include a natural soil drainage classification based on field observations of water table depth, soil wetness, landscape position, and soil morphological characteristics.[19] However, these maps are typically produced at a very coarse scale and do not capture the high spatial and temporal heterogeneity of soil drainage regimes at the field level. They cannot capture dynamic events, for example, where poorly drained soils have been artificially drained, where drains exist but no longer function correctly, or where wetland conditions are being reinstated either through poor management or intentionally as a result of rewetting.

Remote sensing (RS) and in particular Earth observation (EO) satellites are increasingly used to supplement conventional soil mapping techniques by providing spatially continuous data on surface properties.[20] These data can be used to disaggregate existing coarse resolution soil maps.[21] Mapping soil drainage using EO data is based on the underlying principle that plants growing under persistently wet conditions will display signs of abiotic stress (lower leaf biomass and reduced photosynthetic activity) when compared with nonstressed plants.[22] The effect on individual plants can be detected at canopy level using multispectral EO imagery, and satellite imagery has been used in the past to predict soil drainage or artificial drainage over wide areas and to monitor vegetation recovery or agricultural improvement following drainage.[23–26] In other cases, soil drainage class has been predicted using topographic or soil data covariates[27,28] to understand how the permanent landscape can influence soil drainage properties. EO and RS data have also been used to map individual drainage systems at very high spatial resolution using geophysical sensors[29,30] or by optical or thermal imagery,[31,32] but such are not currently cost-effective over large geographical footprints. The spatially continuous nature of EO imagery compared with traditional soil survey methods has often resulted in higher overall accuracy EO-based maps being more accurate than conventional soil drainage maps when compared against ground truth data.[24,33] EO-based approaches can be constrained by the number of drainage classes that can be adequately resolved and are usually more successful in defining extreme classes[28] or very broad drainage classes.[34]

A range of image classification algorithms have been adopted to map soil drainage, including decision trees, logistic modeling, neural networks, and discriminant analysis.[24–26,28,34] There has been an accelerated adoption of machine learning (ML) image classification in the last decade driven by the need to process increasing volumes of EO data and facilitated by the wider availability of ML algorithms through free and open source software.[37] For example, Cho et al.[35] mapped artificial drainage using ML and Google Earth Engine to predict artificial

drainage (and its expansion over time) with an area of >10 million ha. Support vector machine (SVM) and random forest (RF) are two popular ML algorithms for land cover mapping. SVM is optimized for two-class data segmentation problems by identifying the broadest possible margin (or optimal hyperplane) between two linearly separable classes.[36] On the other hand, RF is an ensemble approach that assigns class labels based on the pooled results of randomly created decision tree classifiers. The algorithm uses bootstrapped training data to develop individual trees where each node split is on randomly selected predictor variables. This reduces the importance of individual trees or variables and reduces correlation between individual trees.[37] Both algorithms consistently achieve high overall accuracy and are shown to be quite robust with limited training in comparative studies.[38,39]

In this paper, we compare SVM and RF algorithms for mapping artificially drained heavy soils under pasture in the Republic of Ireland using satellite imagery and topographic variables. Initially, the two algorithms are compared for accuracy against photointerpreted validation points from high-resolution satellite imagery. Subsequently, a comparison of interannual normalized difference vegetation indices (NDVI) imagery is used to assess improvement following drainage at a particular site. Additionally, the ML approach is assessed to determine whether it can detect changes in drainage extent at field level following drain installation. This technique could be deployed on similar soils to identify drained grassland in similar regions and has the potential to optimize agricultural and environmental modeling in the range of research areas that can benefit from more accurate estimation of drained heavy soils.

2 Materials and Methods

2.1 *Study Area*

The study area was heavy grassland soils in the border, midlands, and western (BMW) region of the Republic of Ireland (Fig. 1). The total area of interest was 808,494 ha. Farm A was located within this region. This was a 3-ha field on an intensively managed pasture farm that had a groundwater drainage system installed in 2014 to reduce soil saturation and increase grass growth.

Geographically, the physical characteristics of the region are mixed. A highland zone stretches along much of the western seaboard. This gives way to a topographically diverse landscape in the east that is comprised of sediments and landforms deposited at the end of the last Ice Age. Bedrock geology over the region is varied but is dominated by Carboniferous period sedimentary rocks, particularly limestone, sandstone, and shale. Quaternary sediments are largely comprised of fine textured glacial tills and glaciolacustrine deposits.[40] Extensive tracts of blanket bog and raised bog are present within the study area. Drumlin belts occur within the region. These present a range of drainage characteristics, with the tops and sides typically well drained while the interdrumlin areas are wetter.

Poor soil drainage is compounded by high annual rainfall (<800 to 2500 mm/year). Precipitation is highly spatially variable, and is greatest in western coastal areas and upland regions. Evapotranspiration is relatively stable in the study area (~400 to 450 mm/year). The large volume of precipitation must, therefore, be recycled through drainage, either as surface runoff or within the soil matrix. As the soils within the region tend to have low infiltration rates, many remain waterlogged for prolonged periods and have developed histic, gleyic, or stagnic characteristics that can significantly constrain agricultural production.[1]

Land drainage is a common treatment within the study area to make farming a profitable enterprise. A previous analysis of land drainage grants issued to farms in the region in the period from 1940 to 1986 suggested an average of 23% of farmland across the region had been potentially drained within this period.[41] Drainage projects have traditionally been managed privately, with no legal requirement to register or map new drain locations. In the present day (and only since 2011), land drainage schemes in the region have required screening by the Irish Department of Agriculture only if they exceed 15 ha in area, or are expected to have a significant environmental impact. Screening determines whether an environmental impact assessment is required and it relates to new drainage operations, not the maintenance of existing drains.[42]

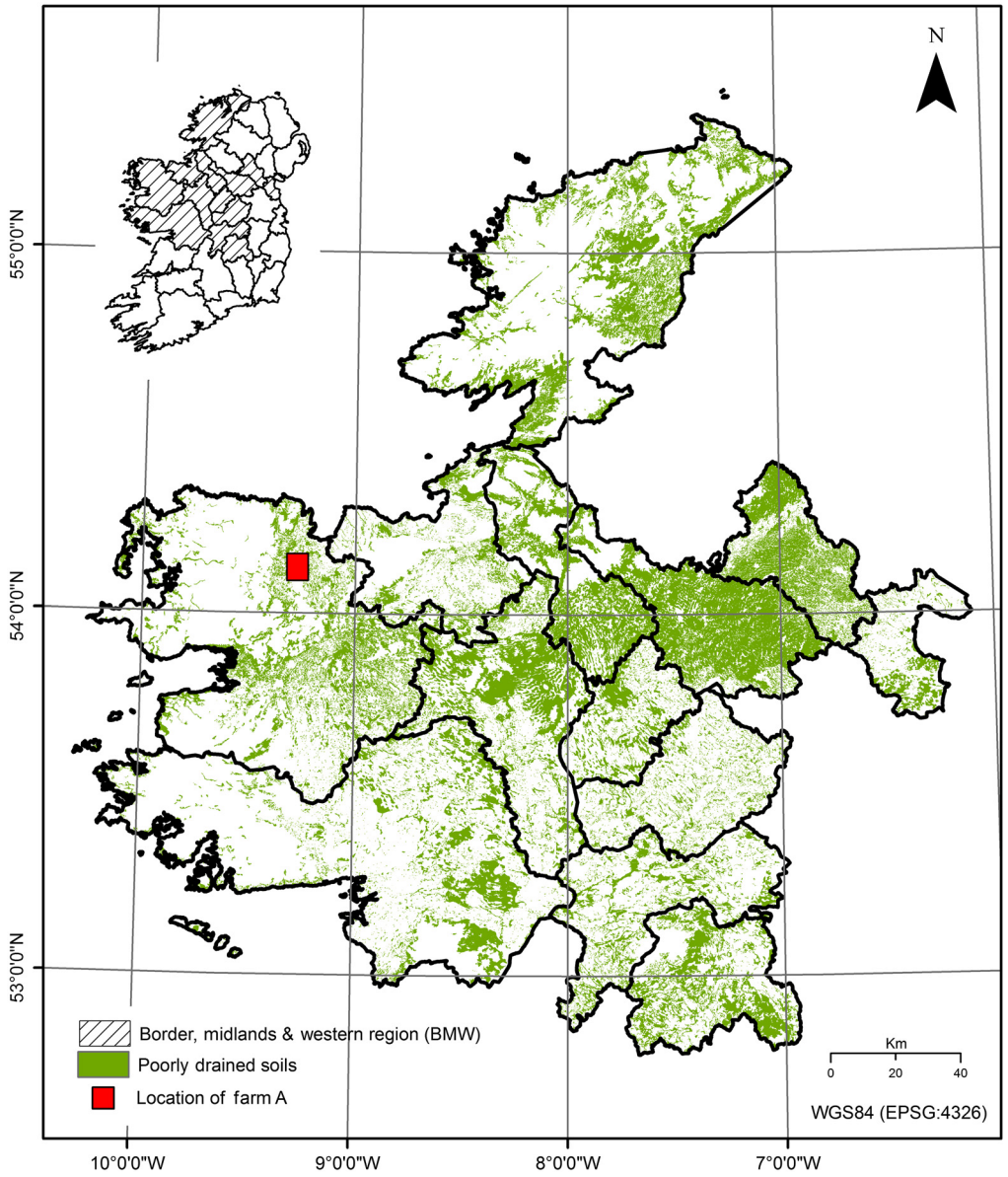

Fig. 1 Map of the study area with grassland on poorly drained soils indicated in the BMW region of the Republic of Ireland. Location of farm A indicated.

Furthermore, the 15-ha criterion applies to the area of drains and their immediate vicinity. Most farms in the region are small and fragmented holdings with the average farm size just ~27 ha.[43] Consequently, the majority of drainage schemes in the study area still occur without being registered or recorded.

2.2 *Datasets and Data Preprocessing*

The datasets used for this study were Landsat 8 optical bands (bands 2 to 7) and terrain attributes derived from a digital elevation model (DEM). All available Landsat 8 scenes covering the region (orbital paths 206 to 209 and rows 21 to 23) acquired during April or May 2014 to 2016 were selected. A total of 46 images (13, 16, and 17 images from 2014, 2015, and 2016, respectively). Spring images were chosen because unstressed grass canopies would be expected to be growing vigorously during this period. Previous research indicated that grassland types and grassland management systems in this region are most readily distinguishable in springtime imagery.[14] Satellite images were downloaded from the USGS Earth Explorer portal as Level

1TP products (radiometrically calibrated and terrain corrected). These were processed to surface reflectance using the LEDAPS algorithm.[44] Cloud and cloud shadow were removed using the F-mask algorithm.[45] Individual scenes were subsequently combined into a single multitemporal mosaic based on mean surface reflectance.

For the experiment at farm A, Landsat 8 was available for 2015 and 2016. Landsat 5 and Landsat 7 images were used in 2010 (April) and 2013 (May) when Landsat 8 imagery was not available. No images were available in 2014 due to extensive cloud cover over the target. Surface reflectance for the Landsat 5 and 7 images was also calculated using LEDAPS to minimize any inconsistencies in calculation of reflectance and reduce the potential for error in comparing reflectance from the three different sensors.

Vegetation indices (VI) sensitive to chlorophyll production (NDVI) and leaf water content [normalized difference water index (NDWI)] were used. These VI were chosen as they provide valuable information about the density, health, and leaf water content of vegetation. NDVI and NDWI are ubiquitous for most EO sensors and can be easily replicated with other sensors in other regions. Previous mapping projects within the study area demonstrated the ability of VI, and NDVI in particular, to distinguish contrasting grassland types.[14] VI were created according to their standard equations,[46,47] where

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + \mathrm{RED}} \tag{1}$$

and

$$\mathrm{NDWI} = \frac{\mathrm{NIR} - \mathrm{SWIR}}{\mathrm{NIR} + \mathrm{SWIR}}, \tag{2}$$

where RED (red), NIR, and SWIR are the red (band 4 0.64 to 0.67 μm), near-infrared (band 5, 0.85 to 0.88 μm), and shortwave infrared (band 6, 1.57 to 1.65 μm) bands, respectively, of Landsat 8.

Terrain attributes were selected based on a review of agrometeorological influences within the region[48] and through a review of previous EO-based soil drainage mapping.[24,25] Selected attributes were chosen for their potential influence on surface, or near surface, hydrology. They were included to assess whether including ancillary topographic data improved the identification of artificially drained soils. Euclidean distance to drainage, distance along slope to drainage, slope, aspect,[49] height above nearest drainage (HAND),[50] and topographic wetness index (TWI)[51] were derived from a 20-m DEM (vertical accuracy $\pm$2.5 m). The DEM was resampled (bilinear interpolation) to 30 m to match the spatial resolution of Landsat 8. The source DEM was a national elevation model supplied by the national mapping agency. Terrain attributes were created using SAGA GIS (v.4.1.0). Slope and aspect have a role in the magnitude and direction of surface runoff. HAND normalizes topography according to local relative heights above drainage networks. Distance to drainage quantifies how far surface runoff must travel to an outlet. TWI is a function of local slope and specific contributing area that is used widely to quantify topographical control on surface hydrological processes. HAND, Euclidean distance, and distance along slope were used with national mapping vector data, which included surface water (streams and rivers), lakes, and open drains.[52] The raster datasets were masked to poorly drained mineral soil using an existing soil map.[53] The conventional soil drainage classes making up the poorly drained class used in this study are listed in Table 1. In the absence of a national landcover dataset, grass land cover was masked using the coarse resolution CORINE 2012 land cover inventory.[54] Building footprints and existing forestry stands within area were masked using available vector datasets.[52,55]

2.3 *Algorithm Choice and Parameterization*

SVM and RF algorithms were available through R statistical software packages "e1071" (v.1.6-8) and "randomForest" (v.4.6-12). Three parameters are required for SVM: kernel type, kernel size γ (to transform linearly inseparable data), and a penalization term C (which defines the acceptable margin of error at class margins). The type of kernel used in this study was a radial

Table 1 Conventional drainage descriptions making up the "poorly drained" class in the current study. Based on Fealy et al.[53]

Drainage class	Description
Imperfectly drained	Partly reduced (gray color with some gray-brown and brown) below 30 cm with mottling. Above 30-cm natural colors (gray-brown and brown) with or without mottling
Poorly drained	Reduced gray throughout profile with many prominent mottles to the surface or a definite reduced layer at any depth below 30 cm and mottling to the surface
Very poorly drained	Reduced gray or gray-blue throughout profile with few mottles allowable; with or without organic surface layer

Table 2 Performance assessment of the effect of kernel type on SVM model accuracy (OA, overall accuracy; K, kappa statistic; PA, producer accuracy; and UA, user accuracy).

		SVM					
				Artificially drained		Poorly drained	
$n = 920$		OA	K	PA	UA	PA	UA
Kernel	RBF	0.838	0.65	0.674	0.893	0.946	0.814
	Polynomial	0.836	0.64	0.671	0.893	0.946	0.812
	Linear	0.826	0.62	0.659	0.871	0.936	0.806

basis function (RBF) kernel, which was found to have marginally better overall accuracy compared with other available types (Table 2). The broad suitability of the RBF kernel to land cover mapping has been reported in comparative studies.[41] Two parameters are required for RF: the number of variables to randomly sample at each split (ntree) and the total number of trees to grow (mtry). Values for the mtry parameter were determined using the tuneRF function within R. Optimal values were calculated for each model/training data combination and mtry values were automatically updated for each new calculation. Out-of-the-box (OOB) error rate was calculated for a range of ntree values from 101 to 2001 in incremental steps of 100. The lowest OOB error rate (4.8%) was recorded at the default value. Lower ntree values had higher error rates (≥5.6%). The default ntree value (501) was used as a higher ntree value would have increased computation time without reducing the error rate.

The RF algorithm also measures variable importance by calculating the mean decrease in accuracy (MDA) as random variables are omitted (see Sec. 3.3). Each algorithm was trained using three different levels of training with different combinations of variables (Table 3). Initially, variable combinations were mixed to identify the relative importance of satellite observations and VI of conditions during the study period with terrain attributes of the permanent landscape.

2.4 *Training Data*

Training data were based on higher spatial resolution satellite imagery (Google Earth). Both algorithms were trained on a binary model of poorly drained and artificially drained classes. Class labels were assigned through photointerpretation of high-resolution Google Earth images that were available in 2014. Three independent sets of training pixels were created using increasing numbers of pixels (60, 200, and 500 pixels) to test whether increasing pixel count had an impact on the outcome. At each training point, an assessment of drainage regime

Table 3 Variable combinations.

Model	Variables
1	All variables Landsat 8, NDVI, NDWI, slope, aspect, HAND, distance, and TWI
2	Landsat 8
3	Landsat 8 and NDVI
4	Landsat 8 and NDWI
5	Landsat 8, NDVI, and NDWI
6	NDVI, NDWI, slope, aspect, HAND, distance, and TWI
7	Landsat 8, slope, aspect, HAND, distance, TWI
8	NDVI, slope, aspect, HAND, distance, and TWI
9	NDWI, slope, aspect, HAND, distance, and TWI
10	Slope, aspect, HAND, distance, and TWI

(artificially drained or poorly drained) was made for a 30-m diameter buffer around each point (corresponding to the approximate area of a Landsat 8 pixel). Visual assessment of this nature favored the distinguishing of extreme drainage classes. If the drainage status at a point was ambiguous or unclear it was disregarded. Areas with extensive growth of water-tolerant vegetation (reeds and rushes) (in $\geq$10% of the buffer area) or areas where low infiltration was observed (water pooling at the surface) were classed as poorly drained. In contrast, artificially drained areas had verdant growth with no indication or water tolerant plants or surface water (see Fig. 2). Consideration was also made of surrounding conditions, for example, where surrounding farm infrastructure indicated good overall management practices (e.g., fenced off walkways, paddocks, and open ditches).

Fig. 2 Examples of drainage conditions used to train each model. Each square corresponds to the area of a single Landsat 8 pixel on heavy soils.

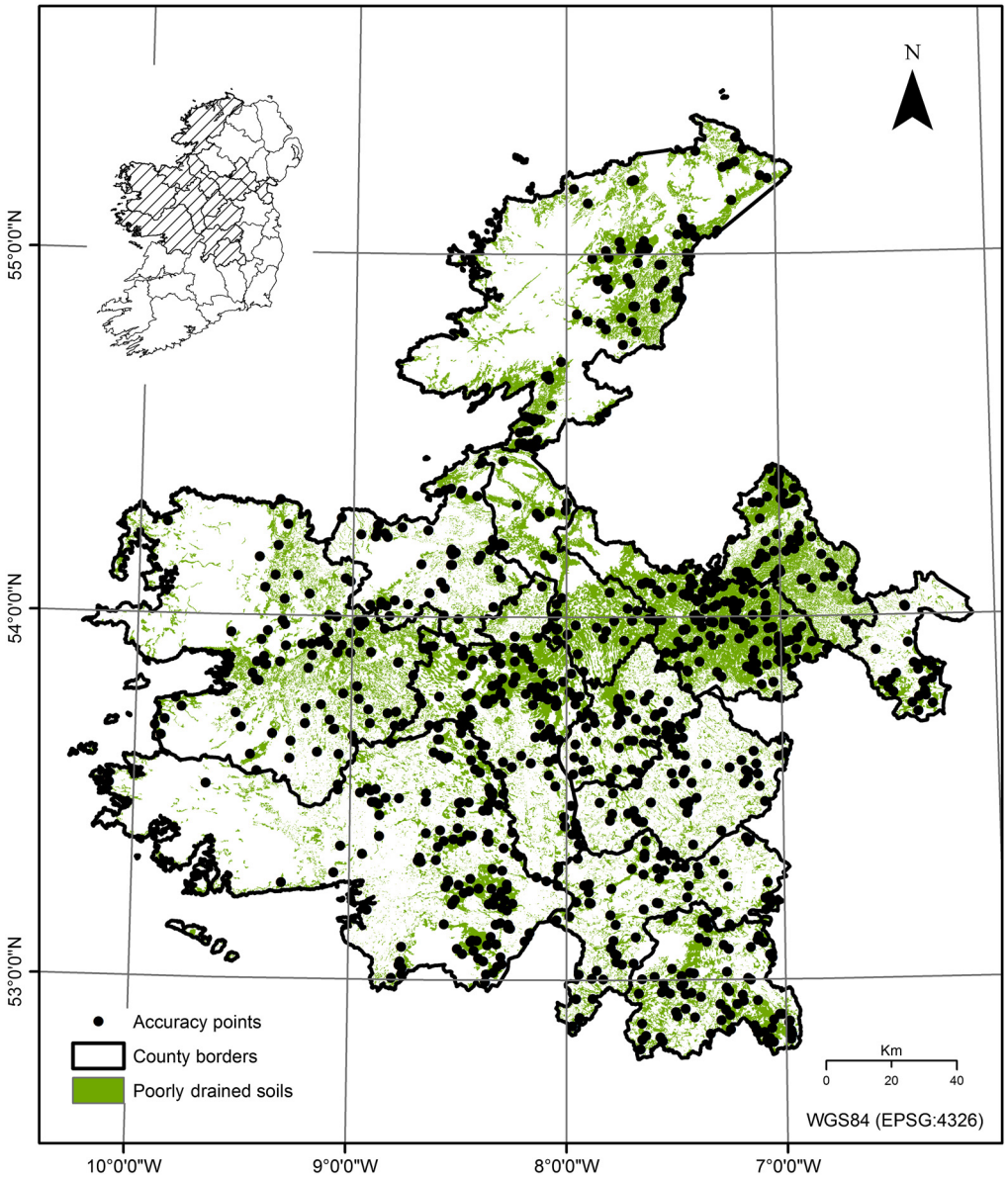

Fig. 3 Distribution of photointerpreted accuracy points.

2.5 *Accuracy Assessment Dataset*

High-resolution satellite imagery was also used as pseudoground truth to validate the map. Class labels were assigned to 920 randomly sampled points distributed across the region (Fig. 3) following the same method outlined in Sec. 2.4. A minimum distance of 30 m was maintained between accuracy points and field boundaries and/or structures to reduce potential sources error from mixed land cover or boundary shadow. Overall accuracy and producer/user accuracy (with 95% confidence intervals) were computed using an error matrix.[56] Kappa statistics are also reported. Relative accuracy between models are based on Z statistics and p values in the method described by Rossiter.[57]

3 Results

3.1 *Spectral Properties for Poorly Drained and Artificially Drained Classes*

Comparing the Landsat 8 spectral signals of training data for each drainage class, the artificially drained class exhibited lower reflectance (greater absorption) at RED and SWIR wavelengths,

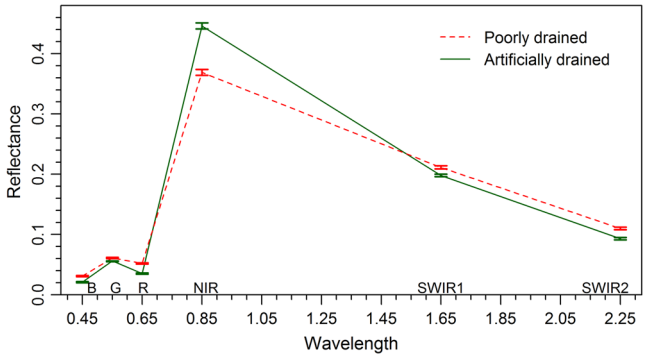

Fig. 4 Mean canopy reflectance for artificially drained and poorly drained classes using Landsat 8. Abbreviations: arrows (indicating 95% confidence intervals) placed at each band center wavelength.

with higher reflectance at the NIR wavelength (Fig. 4) This signal is consistent with greater production of chlorophyll in healthy grass canopy. The relatively lower NIR reflectance and higher SWIR reflectance signal in the poorly drained class was expected for canopy growing under a persistent stress. The separation in reflectance between classes was lowest in the green band and greatest at NIR wavelengths, but there were significant differences in mean reflectance values at each of the Landsat 8 bands, suggesting each band had some information to provide on canopy health and drainage status within the study area.

Anomalous growth patterns were compared for each drainage class by calculating month-on-month differences in mean NDVI between February and November 2014 to 2016 (217 images in total) (Fig. 5). There was a clear seasonal trend visible for each class. Each class exhibited positive, increasing NDVI values from March and April, peaking in May. The magnitude of increase was greatest for the artificially drained class. Previous studies also reported a disparity in spring grass yields on farms with contrasting drainage regimes.[2] A separation of classes was also observed in summer with marginally higher values for the drained class. The trend was reversed in the late summer period where higher values were observed for the poorly drained class during August/September. The reversal possibly corresponds to higher volumes of soil moisture on poorly drained soils offsetting the effect of drought on drained soils.

3.2 *Algorithm Performance*

Full accuracy metrics for each iteration of model/training data are presented in Table 4.

SVM performed strongly, with overall accuracies of 86.9% [84.7%, 89.1%], 89.6% [87.7%, 91.6%], and 87.3% [85.1%, 89.4%] at each respective training level. However, RF consistently achieved higher accuracies. At the lowest number of training points (60 pixels) per class,

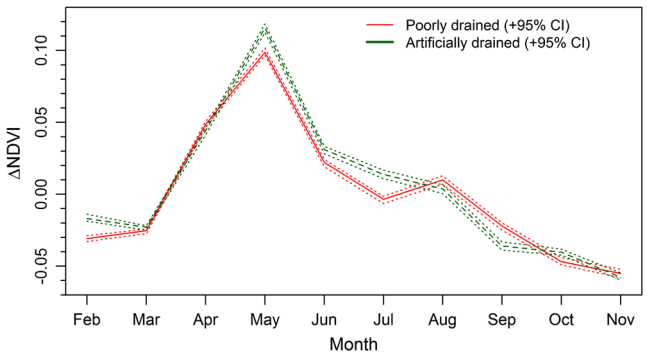

Fig. 5 Monthly differences in mean NDVI over a typical growing season (2014 to 2016) for artificially drained and poorly drained classes.

Table 4 Accuracy assessment for each algorithm (SVM and RF) by models at different levels of training data.

	SVM						RF					
			Artificially drained		Poorly drained				Artificially drained		Poorly drained	
Model	OA	K	PA	UA	PA	UA	OA	K	PA	UA	PA	UA
60 training points												
1	83.37	0.64	95.3	80.55	65.49	90.26	90.76	0.78	94.1	92.8	82.97	85.77
2	85.65	0.68	95.62	83.61	69.34	90.64	88.04	0.72	94	88.82	75.91	86.14
3	**86.96**	0.71	95.87	85.3	71.68	91.01	91.3	0.79	94.14	93.57	84.5	85.77
4	86.1	0.69	95.81	84.07	70.03	91.01	89.13	0.75	94.38	90.05	78.11	86.89
5	86.63	0.7	96.01	84.69	70.93	91.39	**91.41**	0.79	94.15	93.72	84.81	85.77
6	80.54	0.59	95.75	75.96	60.95	91.76	88.91	0.74	93.11	91.12	79.36	83.52
7	75.65	0.5	95.54	68.91	54.79	92.13	88.15	0.73	94.74	88.21	75.32	88.01
8	78.59	0.54	94.19	74.43	58.66	88.76	88.15	0.72	93.56	89.43	76.69	85.02
9	75.65	0.5	94.97	69.37	54.85	91.01	82.72	0.61	92.88	81.93	65.7	84.64
10	52.11	0.12	79.34	44.1	34.47	71.91	50.98	0.06	74.88	46..55	32.1	61.8
200 training points												
1	85.54	0.67	95.14	83.92	69.48	89.51	**91.2**	0.79	93.87	93.72	84.7	85.02
2	88.91	0.72	91.93	92.5	81.37	80.15	90.54	0.77	93.4	93.26	83.58	83.9
3	**89.67**	0.75	92.79	92.65	82.09	82.4	90.98	0.78	93.85	93.42	84.07	85.02
4	88.26	0.72	92.12	91.21	79.12	80.9	89.78	0.75	93.47	92.04	81.23	84.27
5	89.13	0.74	93.68	90.81	79.09	85.02	91.09	0.78	93.86	93.57	84.39	85.02
6	82.83	0.62	93.19	81.78	65.71	85.39	88.37	0.73	91.17	89..13	76.43	86.52
7	85.76	0.67	93.79	85.6	70.93	86.14	88.37	0.73	94.17	89.13	76.49	86.52
8	82.5	0.61	92.41	82.08	65.59	83.52	88.26	0.72	94.45	88.67	75.9	87.27
9	79.35	0.55	92.01	77.64	60.43	83.52	85.43	0.67	93.18	85.76	70.85	84.64
10	53.48	0.06	73.68	53.6	31.91	53.18	57.93	0.08	74.18	62.48	33.78	46.82
500 training points												
1	83.37	0.64	95.29	80.55	65.49	90.26	88.15	0.73	94.74	88.21	75.32	88.01
2	86.3	0.69	94.53	85.6	71.43	88.01	88.04	0.72	92.76	90.2	77.54	82.77
3	87.07	0.71	95.25	86.06	72.42	89.51	**89.02**	0.74	94.52	89.74	77.67	87.27
4	**87.28**	0.71	95.27	86.37	72.87	89.51	86.2	0.68	93.26	86.83	72.44	84.64
5	87.17	0.71	95.57	85.91	72.31	90.26	88.7	0.74	94.78	88.91	76.55	88.01
6	85.0	0.67	95.1	83.15	68.48	89.51	87.72	0.72	94.55	87.75	74.52	87.64
7	83.48	0.64	95.14	80.86	65.75	89.89	87.28	0.7	92.81	88.91	75.51	83.15
8	84.35	0.65	93.5	83.77	68.36	85.77	87.5	0.71	94.24	87.75	74.36	86.89
9	81.09	0.58	92.84	79.48	62.88	85.02	82.39	0.61	93.61	80.7	64.71	86.52
10	49.24	0.03	72.91	45.33	30.54	58.8	50.87	0.05	73.54	48.09	31.24	57.68

OA, overall accuracy (%); K, kappa statistic; PA, producer accuracy (%); and UA, user accuracy (%). Highest OA highlighted in bold.

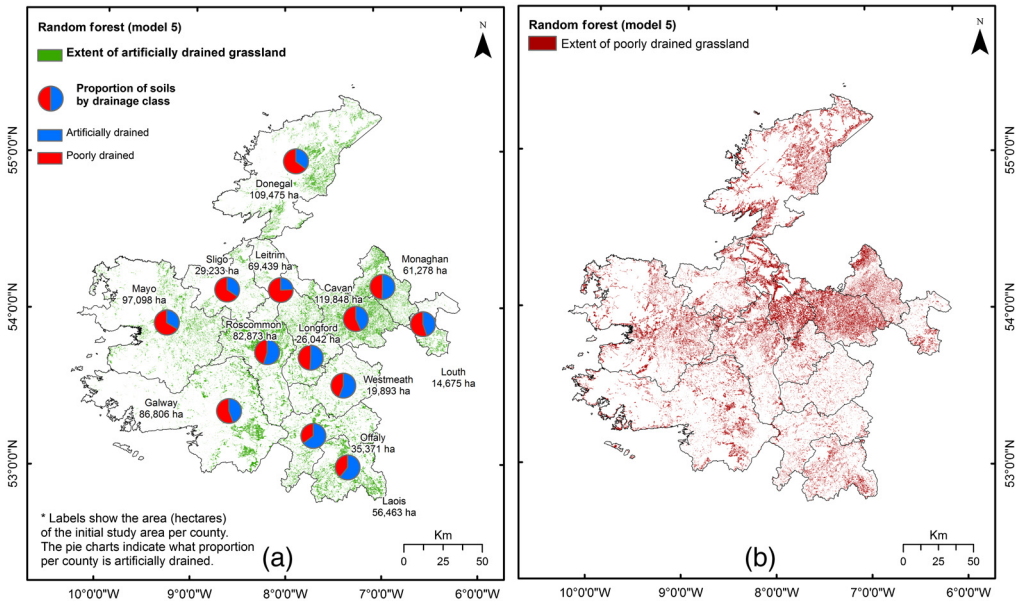

Fig. 6 (a) Extent and distribution of artificially drained soils in the study area and (b) revised poorly drained soils in the study area.

the model with the highest overall accuracy {91.4% (95% CI [89.6%, 93.2%])} was an RF classification of Landsat 8 optical bands NDVI and NDWI (model 5 in Table 3). Using this model, 44% (345,000 ha) of the initial study area (~808,495 ha) was classed as artificially drained. These were areas where grass canopy showed no indication of persistent stress [Fig. 6(a)]. The corresponding extent of the poorly drained class is illustrated in Fig. 6(b). At 200 training samples, an RF model applied to all spectral and topographical layers achieved an overall accuracy of 91.1% (95% CI [89.2%, 92.9%]) (model 1 in Table 3). At 500 training samples, an RF model applied to Landsat bands and NDVI had an overall accuracy of 89.0% (95% CI [87.0%, 91.0%]) (model 3 in Table 3).

Outcomes at each level of training between algorithms were compared using a two-tailed, binomial z-score test. There was a significant difference ($z = 3.08$, $p = 0.002$) in overall accuracy between RF and SVM algorithms at the lowest level of training. Although there was a marginal decrease in overall accuracy using 500 training pixels for both algorithms, there was no significant difference in accuracy between algorithms at 200 and 500 training pixels, respectively ($z = 1.03$, $p = 0.3$ and $z = 1.15$, $p = 0.2$). The decrease in accuracy using 500 training pixels was unexpected and was probably caused by an unidentified change in land cover or land use between the date of acquisition of the high-resolution satellite imagery and the Landsat images used for the study.

3.3 *Variable Importance*

Variable importance was based on the MDA function within the RF algorithm (Fig. 7). The spectral band with the greatest influence on map accuracy was NIR (Landsat 8 band 5). The VI (NDVI and NDWI) were also key variables. Both use NIR in their calculation. Individually, however, the two SWIR bands ranked low in overall importance. Terrain attributes had a minor impact on overall model accuracy, but elevation, aspect, and Euclidean distance to drainage were among the highest-ranking terrain attributes where included.

3.4 *Observing Changes in Drainage Class Following Drain Installation*

Following drain installation at farm A in 2014, increases to minimum, mean and maximum NDVI values were observed. Mean and maximum NDVI increased during the following spring (2015) (Table 5). The difference between mean NDVI in 2013 and 2015 was significant

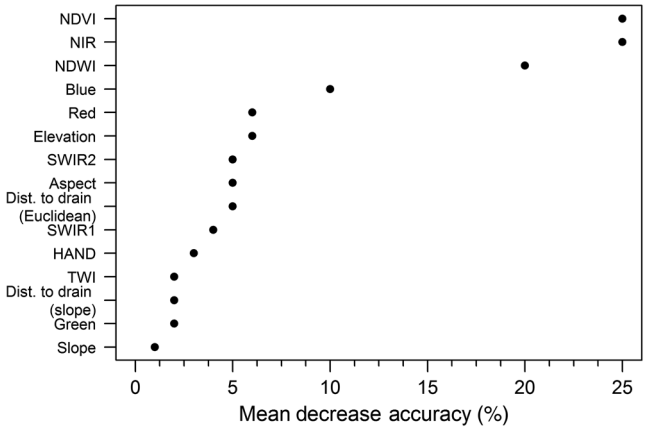

Fig. 7 Variable importance derived from RF MDA function.

Table 5 Descriptive statistics for spring NDVI values from farm A before and after drain installation.

Year	Min	Max	Mean	Standard deviation
2010	0.651	0.815	0.729	0.04
2011	No data available (cloud cover)			
2012	No data available (cloud cover)			
2013	0.713	0.866	0.798	0.04
2014	No data available (cloud cover)			
	Drainage installed July/August			
2015	0.685	0.914	0.849	0.09
2016	0.759	0.936	0.864	0.05

[$t(32) = 7.695$, $p < 0.0001$]. There was an initial drop in minimum NDVI values in 2015, possibly as a result of groundworks at the site where bare earth was exposed. Minimum NDVI values increased in 2016 relative to each of the previous years. Difference in mean NDVI for pre-2014 imagery and post-2014 imagery were statistically significantly [$t(32) = 5.166$, $p < 0.0001$]. The improvements in mean, minimum, and maximum values for amalgamated imagery before and after 2014 are visible in Fig. 8. The improvement in overall NDVI metrics was matched by concomitant increases in the area of the field characterized as artificially drained

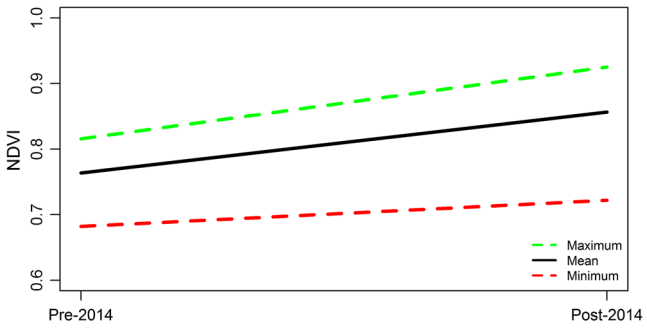

Fig. 8 Increases to minimum, mean, and maximum NDVI values at farm A following drain installation. The graph aggregates available image data into two groups before and after 2014.

by the RF algorithm. In 2010, the drained proportion of the field classed was 0.72 ha (25% of the area). Following drainage, this had increased to 1.8 ha (58%) in 2015 and 2.1 ha in 2016 (68%).

4 Discussion

4.1 *Extent of Artificially Drained Grassland*

Identifying current or dynamic drainage regimes at the field scale is challenging and not typically possible using traditional soil mapping approaches. Our method has shown how larger soil map units can be spatially disaggregated using EO imagery to identify discreet areas of artificially drained heavy soils at a farm management scale. The approach has a straightforward implementation based on the effect of soil saturation on grass yields[2] and photosynthetic activity[22] that can be observed using freely available EO imagery without the need for additional covariates.

Using RF, ~44% (345,000 ha) of the initial study area of 808,494 ha could be reclassified as artificially drained. This is the first time that the extent of drained heavy soils has been mapped in a consistent way across the region. Estimates of drained area for the current study and the earlier assessments are presented in Table 6. It is difficult to directly compare the accuracy of this model with earlier approaches as they are based on different criteria. Bruton and Convery[41] compiled drainage estimates using records of drainage grants. They reported the average area of drained farmland per county was 23% versus 44% for the current study. Although the area predicted in the current study was greater than that suggested by Bruton and Convery, there was similarity in the overall ranking of drained areas. Counties reporting the highest (Laois, Offaly, and Westmeath), and lowest (Sligo and Leitrim) area of drained land in each study were identical. Mockler et al.[58] assumed land drainage on all poorly drained soils below 200-m above sea level on slopes ≤ 12 deg, which represented 55% of the farmland in the region. This same approach was adopted by Paul et al.[15] Again, there was similarity between the predicted areas per county, but there was a large over prediction of drained area for some counties.

Table 6 County-by-county assessment of artificially drained area based on current study and previous research.

	Initial area	Current study			Bruton and Convery	Mocker et al.
	Poorly drained pasture (ha)	Poorly drained (ha)	Artificially drained (ha)	Grassland artificially drained (%)	Farmland artificially drained (%)	Farmland artificially drained (%)
Cavan	119,848	67,965	51,813	43	26	84
Donegal	109,475	72,066	39,206	36	24	73
Galway	86,806	48,300	38,459	44	22	36
Laois	56,463	21,679	33,445	59	37	53
Leitrim	69,439	51,678	16,552	24	13	89
Longford	26,042	12,162	12,700	49	17	49
Louth	14,675	7987	6455	43	19	38
Mayo	97,098	65,230	33,228	34	24	52
Monaghan	61,278	30,363	31,022	50	25	60
Offaly	35,371	12,527	23,117	65	31	48
Roscommon	82,873	38,265	46,134	56	15	61
Sligo	29,233	18,137	9,611	33	13	45
Westmeath	19,893	9,005	11,388	57	31	27

This overstates the significance of drain flow as a hydrological pathway in agricultural catchments in these counties, potentially inflating estimates of nutrient losses or GHG emissions from agricultural lands in these areas. The current study has not identified specific drainage types within the mapped area, though higher resolution EO techniques could be applied to that problem, for example, using low-altitude RS to identify thermal or spectral anomalies surrounding buried pipes.[31] Additional investigations are required to establish with certainty that the drained conditions identified through this method are a result of artificial drainage, or whether some are naturally draining soils within poorly drained soil units. An elaboration of this approach that could prove beneficial would be to consider the contrasting canopy reflectance over different seasons. As suggested by Fig. 4, drainage class could be distinguished in the later summer period when periods of droughts were less impactful on poorly drained soils.

It was necessary to use CORINE land cover data despite the known issues with this dataset in this region (coarse spatial resolution versus highly fragmented landscape).[14] The spatial resolution of the underlying land cover dataset will have included nongrass landcover. Every effort was made to mask nongrass landcover with available spatial datasets, but some nongrass classes will remain. Changes to land cover/land use will also have occurred over the period of the study. Most importantly, artificial drainage may have been installed in some areas. But changes in landcover also occurred. Afforestation records indicated ~9700 ha of new forestry was planted in the study area between 2014 and 2016.[59] Construction figures suggested >3000 new rural housing units were built in the same period.[60] Assuming the change in landcover in both of these cases was from poorly drained grassland, then 1.2% of the initial study underwent landcover change within the three years of the study. These changes would introduce a small amount of error into the overall result.

4.2 *Variable Importance*

The variables with the greatest impact on model accuracy were NDVI and the NIR band, followed by NDWI, blue, and red bands. The importance of NIR, blue, and red bands is not surprising. Grass growing without an underlying stress will preferentially absorb light at blue and red wavelengths to power photosynthesis and consequently reflect a higher proportion of NIR. NIR and NDVI were important in previous studies.[24] SWIR bands have also proven useful in the previous studies over bare soils[26] but were less important in this study. Other sensors with different spectral resolutions would have different variable amounts of importance. For example, the three red-edge bands (~0.68 to 0.74 μm) on Sentinel 2 would be expected to have increased sensitivity to canopy chlorophyll content.[61] Hyperspectral data have also been quite successful in distinguishing saturation and drainage class over several wavelengths.[62,63] Terrain attributes had little overall impact on accuracy, and when used without additional spectral data performed very poorly (see results for model 10 in Table 4). However, where included, elevation, aspect, and distance to drainage were consistently among the highest-ranking variables. Elevation, aspect influence of precipitation volume, and incoming solar radiation impact soil temperature, soil moisture content, and the number of degree days.[48] Elevation was important in the previous studies.[24,26] Slope had the lowest importance of all variables despite being used widely in the previous studies.[32,64]

4.3 *Observed Changes Following Drain Installation*

Identifying changes in drainage conditions over time, for example, as drainage regime improves or deteriorates, is critical for monitoring changes in land management that may impact the wider environment. The experiment at farm A demonstrated how multitemporal satellite imagery observed changes in drainage status within a 3-ha field following drainage. At farm A, the increase in mean NDVI in the years after 2014 are consistent with the removal of underlying stress and concomitant improvement in primary production. The increase in NDVI in 2016 was remarkable considering the region experienced record-breaking volumes of rainfall throughout the winter of 2015/16, which resulted in extensive waterlogging of soils that lasted for several weeks and depressed NDVI values over the region until midsummer 2016.[65] Improved growth characterized by increasing NDVI following drainage was observed by

Kobryn et al.,[23] who used Landsat 7 NDVI to map the areas surrounding deep drains that benefited from improved soil conditions.

Unfortunately, extensive cloud covers over farm A in the years on either side of 2014 required additional Landsat sensors to be used to complete the temporal series over the target at sufficient resolution. The marginally different sensor configurations between Landsat-5, -7, and -8 may have had some impact on the differences in NDVI between years. Despite this risk, it was necessary to use alternative sensors to monitor changes in NDVI within the field a at similar resolution to Landsat 8.

Although not specifically tested here, it is possible that deterioration in drainage status could also monitored, for example, where growth is impeded where drains stop functioning, or where formerly drained lands are reverting to wetlands, either through poor management or through intentional rewetting. Linking an EO-based model of land drainage extent and drainage function to soil moisture deficit (SMD) could have important implications for agrometereological modeling. Currently in Ireland, SMD is modeled by the national meteorological service from synoptic weather data and soil/terrain properties for three drainage classes (well drained, moderately drained, and poorly drained).[16] Integrating an up-to-date EO-based map of artificially drained soils could improve the current model, which is the existing SMD model, giving more precise spatial and temporal characterization of land drainage at farm management scale, supporting on-farm decision making for grass growth and utilization,[3] trafficability, and fertilizer management.[66]

4.4 *Misclassification Error*

Cloud masking of the initial Landsat 8 scenes created mosaics with extensive areas of missing data. There was less data available in these areas to calculate mean reflectance over the study period. Mean pixel values derived from fewer contributing images are potentially susceptible to antecedent meteorological conditions at the time of acquisition. For example, drained areas may appear poorly drained where the days preceding image acquisition were unusually wet. Similarly, poorly drained areas could appear to be drained if there has been a prolonged dry spell. Figure 9 illustrates how the number of contributing images possible influenced model accuracy. Misclassification error was greatest where there were fewer numbers of contributing images. At farm A, it was not possible to create multitemporal mosaics and single-date acquisitions were used. As we can see from the regional study, this will have increased the possibility of error in estimating drainage extent in all years.

Missing data as a result of cloud cover are unavoidable in the study area where cloud cover is frequently extensive. This presents considerable challenges for mapping land cover using optical sensors such as Landsat 8. At farm A, data from previous Landsat missions still in orbit were used to fill the in gaps, but as noted above, this can be problematic where there are different radiometric configurations, narrower bandwidths, and different wavelength centers. Such differences can lead to potentially significant differences in surface reflectance that propagate into band ratios or spectral indices.[67] The same principle applies to integrating

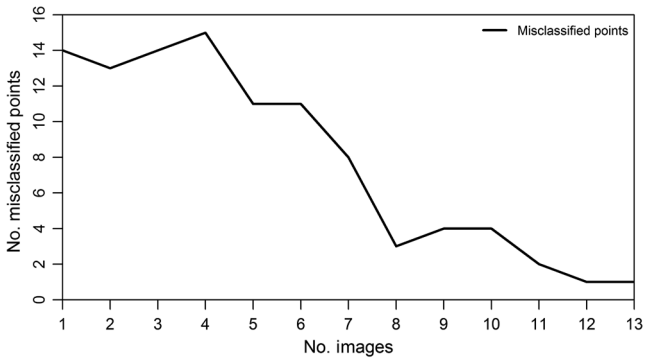

Fig. 9 A decrease in misclassification error was observed where the number of images contributing to the final mosaic increased.

Sentinel 2 imagery. Coarser resolution imagery from MODIS or Sentinel 3 imagery could be used to produce a more complete time series and would likely introduce further error due to the nature of the landscape in the region.[14] For more precise mapping, drones carrying multispectral sensors could map individual fields or farms at very high resolution. Drones are already used in some regions for mapping individual drainage systems for assessment of drain function.[31]

5 Conclusions

Accurate mapping of field drainage is essential to inform specific farm management decisions, national policy, and accounting in areas such as surface water quality and GHG emission estimates. This paper presents an EO-based method using multitemporal, multispectral Landsat 8 imagery to map artificially drained pasture on heavy soils by disaggregating larger soil drainage units from an existing soil map.

The key findings of this study were that artificially drained pasture on heavy soils could be distinguished in Landsat 8 images by lower reflectance of red and SWIR wavelengths and higher reflectance at NIR wavelengths. Furthermore, differences in the rate of grass growth for artificially drained soils were demonstrated using monthly increases in mean NDVI, where the magnitude of increase in NDVI in April/May was greatest on artificially drained soils (see Fig. 5).

In overall classification accuracy, RF outperformed SVM at each respective training level. The difference in accuracy between algorithms was significant ($p = 0.002$) at the lowest level of training, but not when further training pixels were introduced ($p = 0.3$ and $p = 0.2$ at 200 and 500 training pixels, respectively). The model with highest overall accuracy {91.4% (95% CI [89.6%, 93.2%])} was an RF classification of Landsat 8 optical bands and the two VI with 60 training pixels per class. Using this model, 44% (345,000 ha) of the initial study area was reclassed as artificially drained. Classification accuracy was related to the number of images contributing to the multitemporal mosaic.

At farm A, improved mean NDVI values were observed following installation of field drains in 2014 ($p < 0.0001$). An improvement in overall NDVI metrics (mean, minimum, and maximum) was mirrored by concomitant increases in the area of field characterized as artificially drained from 0.72 ha (25% of area) in 2010 to 2.1 ha (68%) in 2016. This approach has potential for providing mapping current local drainage conditions, which are a significant source of uncertainty in hydrological modeling or GHG estimations.[14] The described method can be extended to cover pasture or other crops on drained mineral and peat soils as a means of identifying hotspots for diffuse nutrient loss, for more precise accounting of GHG emissions, or for identifying degrading agricultural land. Although this study has shown that the combination of EO and ML is a feasible approach for mapping artificially drained areas, additional field observations validating the presence of artificial drainage would be desirable to confirm that the observed conditions are a result of artificial drainage and not from other soil or environmental factors.

Acknowledgments

This work was funded through the Teagasc Walsh Scholarship Program. Information on the drainage system installed at farm A was supplied by the Teagasc Heavy Soils Program. Disclosure: The authors declare no conflicts of interest.

References

1. R. P. O. Schulte et al., "A review of the role of excess soil moisture conditions in constraining farm practices under Atlantic conditions," *Soil Use Manag.* **28**(4), 580–589 (2012).
2. A. J. Brereton and M. Hope-Cawdery, "Drumlin soils: the depression of herbage yield by shallow water table depth," *Irish J. Agric. Res.* **27**(2/3), 167–178 (1988).
3. J. B. Fitzgerald, A. J. Brereton, and N. M. Holden, "Simulation of the influence of poor soil drainage on grass-based dairy production systems in Ireland," *Grass Forage Sci.* **63**(3), 380–389 (2008).

4. L. Shalloo et al., "Comparison of a pasture based system of milk production on a high rainfall, heavy clay soil with that on a lower rainfall, free draining soil," *Grass Forage Sci.* **59**, 157–168 (2004).
5. L. F. Galvin, "The drainage of impermeable soils in high rainfall areas," *Irish J. Agric. Res.* **22**(2/3), 161–187 (1983).
6. A. C. Armstrong and E. A. Garwood, "Hydrological consequences of artificial drainage of grassland," *Hydrol. Process.* **5**(2), 157–174 (1991).
7. R. W. Skaggs, M. A. Brevé, and J. W. Gilliam, "Hydrologic and water quality impacts of agricultural drainage," *Crit. Rev. Environ. Sci. Technol.* **24**(1), 1–32 (1994).
8. T. G. Ibrahim et al., "Spatial and temporal variations of nutrient loads in overland flow and subsurface drainage from a marginal land site in south east Ireland," *Biol. Environ. Proc. R. Irish Acad.* **113B**(2), 169–186 (2013).
9. N. Valbuena et al., "Greenhouse gas emissions from temperate permanent grassland on clay loam soil following the installation of artificial drainage," *Agric. Ecosyst. Environ.* **269**, 39–50 (2018).
10. E. Clagnan et al., "Influence of artificial drainage system design on the nitrogen attenuation potential of gley soils: evidence from hydrochemical and isotope studies under field-scale conditions," *J. Environ. Manage.* **206**, 1028–1038 (2018).
11. K. Leiber-Sauheitl et al., "High CO_2 fluxes from grassland on histic Gleysol along soil carbon and drainage gradients," *Biogeosciences* **11**(3), 749–761 (2014).
12. K. Byrne, G. Kiely, and P. Leahy, "Carbon sequestration determined using farm scale carbon balance and eddy covariance," *Agric. Ecosyst. Environ.* **121**, 357–364 (2007).
13. N. Valbuena-Parralejo et al., "Phosphorus and nitrogen losses from temperate permanent grassland on clay-loam soil after the installation of artificial mole and gravel mole drainage," *Sci. Total Environ.* **659**, 1428–1436 (2018).
14. F. Cawkwell et al., "The Irish Land Mapping Observatory: mapping and monitoring land cover, use and change," Environmental Protection Agency, Ireland (2017).
15. C. Paul et al., "Assessing the role of artificially drained agricultural land for climate change mitigation in Ireland," *Environ. Sci. Policy* **80**, 95–104 (2018).
16. R. P. Schulte et al., "A note on the hybrid soil moisture deficit model v2.0," *Irish J. Agric. Food Res.* **54**(2), 126–131 (2015).
17. P. Sharma, J. Humphreys, and N. M. Holden, "The effect of local climate and soil drainage on the environmental impact of grass-based milk production," *Int. J. Life Cycle Assess.* **23**(1), 26–40 (2018).
18. S. Feick, S. Siebert, and P. Doell, *A Digital Global Map of Artificially Drained Agricultural Areas*, Johann Wolfgang Goethe-Universität, Frankfurt, Germany (2005).
19. A. B. McBratney, M. L. Mendonça Santos, and B. Minasny, "On digital soil mapping," *Geoderma* **117**(1), 3–52 (2003).
20. V. L. Mulder et al., "The use of remote sensing in soil and terrain mapping: a review," *Geoderma* **162**, 1–19 (2011).
21. E. E. Frisbee et al., "Remote sensing for soil map unit boundary detection," *GSA Rev. Eng. Geol.* **22**, 119–129 (2014).
22. N. McFarlane, T. Ciavarella, and K. F. Smith, "The effects of waterlogging on growth, photosynthesis and biomass allocation in perennial ryegrass (*Lolium perenne* L.) genotypes with contrasting root development," *J. Agric. Sci.* **141**(2), 241–248 (2003).
23. H. T. Kobryn et al., "Remote sensing for assessing the zone of benefit where deep drains improve productivity of land affected by shallow saline groundwater," *J. Environ. Manage.* **150**(Suppl. C), 138–148 (2015).
24. A. Cialella et al., "Predicting soil drainage class using remotely sensed and digital elevation data," *Photogramm. Eng. Remote Sens.* **63**, 171–178 (1997).
25. P. Campling, A. Gobin, and J. Feyen, "Logistic modeling to spatially predict the probability of soil drainage classes," *Soil Sci. Soc. Am. J.* **66**(4), 1390–1401 (2002).
26. A. B. Møller et al., "Predicting artificially drained areas by means of a selective model ensemble," *Geoderma* **320**, 30–42 (2018).
27. J. C. Bell, R. L. Cunningham, and M. W. Havens, "Soil drainage class probability mapping using a soil-landscape model," *Soil Sci. Soc. Am. J.* **58**(2), 464–470 (1994).

28. Z. Zhao, M. I. Ashraf, and F.-R. Meng, "Model prediction of soil drainage classes over a large area using a limited number of field samples: a case study in the province of Nova Scotia, Canada," *Can. J. Soil Sci.* **93**(1), 73–83 (2013).
29. B. Allred et al., "Detection of buried agricultural drainage pipe with geophysical methods," *Appl. Eng. Agric.* **20**(3), 307–318 (2004).
30. M. Rogers, J. Cassidy, and M. Dragila, "Ground-based magnetic surveys as a new technique to locate subsurface drainage pipes: a case study," *Appl. Eng. Agric.* **21**, 421–426 (2005).
31. B. Allred et al., "Overall results and key findings on the use of UAV visible-color, multispectral, and thermal infrared imagery to map agricultural drainage pipes," *Agric. Water Manag.* **232**, 106036 (2020).
32. B. Tetzlaff et al., "Aerial photograph-based delineation of artificially drained areas as a basis for water balance and phosphorus modelling in large river basins," *Phys. Chem. Earth, Parts A/B/C* **34**, 552–564 (2009).
33. W. Peng et al., "Delineating patterns of soil drainage class on bare soils using remote sensing analyses," *Geoderma* **115**(3), 261–279 (2003).
34. A. Beucher, A. B. Møller, and M. Greve, "Artificial neural networks and decision tree classification for predicting soil drainage classes in Denmark," *Geoderma* **352**, 351–359 (2017).
35. E. Cho et al., "Identifying subsurface drainage using satellite Big Data and machine learning via Google Earth Engine," *Water Resour. Res.* **55**(10), 8028–8045 (2019).
36. C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.* **20**(3), 273–297 (1995).
37. L. Breiman, "Random forests," *Mach. Learn.* **45**(1), 5–32 (2001).
38. G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: a review," *ISPRS J. Photogramm. Remote Sens.* **66**, 247–259 (2011).
39. M. Pal, "Random forest classifier for remote sensing classification," *Int. J. Remote Sens.* **26**(1), 217–222 (2005).
40. A. Sleeman, B. McConnell, and S. Gately, *Understanding Earth Processes Rocks and the Geology of Ireland*, Geological Survey of Ireland, Dublin (2004).
41. R. Bruton and F. Convery, *Land Drainage Policy in Ireland*, Economic and Social Research Institute, Dublin (1982).
42. DAFM, *Environmental Impact Assessment (Agriculture) Regulations: Guide for Farmers*, DAFM, Dublin (2011).
43. CSO, "Farm structure survey 2016," Central Statistics Office, Cork (2016).
44. J. G. Masek et al., *LEDAPS Calibration, Reflectance, Atmospheric Correction Preprocessing Code*, Version 2, ORNL Distributed Active Archive Center, Oak Ridge, Tennessee (2013).
45. Z. Zhu, S. Wang, and C. Woodcock, "Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4-7, 8, and Sentinel 2 images," *Remote Sens. Environ.* **159**, 269–277 (2015).
46. J. W. Rouse et al., "Monitoring vegetation systems in the Great Plains with ERTS," in *3rd ERTS Symp.*, NASA, p. 8 (1973).
47. B. Gao, "NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space," *Remote Sens. Environ.* **58**(3), 257–266 (1996).
48. T. Keane and J. F. Collins, *Climate, Weather and Irish Agriculture*, 2nd ed., Joint Working Group on Applied Agricultural Meteorology, Dublin (2004).
49. L. W. Zevenbergen and C. R. Thorne, "Quantitative analysis of land surface topography," *Earth Surf. Process. Landforms* **12**(1), 47–56 (1987).
50. C. Rennó et al., "HAND, a new terrain descriptor using SRTM-DEM: mapping terra-firme rainforest environments in Amazonia," *Remote Sens. Environ.* **112**, 3469–3481 (2008).
51. K. J. Beven and M. J. Kirkby, "A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant," *Hydrol. Sci. Bull.* **24**(1), 43–69 (1979).
52. OSI, *PRIME 2: Data Concepts & Data Model Overview*, Ordnance Survey Ireland, Dublin (2015).
53. R. Fealy et al., "Teagasc EPA soil and subsoils mapping project," Environmental Protection Agency, Ireland (2009).

54. EPA, "CORINE land cover 2012 Ireland Final Report," Environmental Protection Agency, Ireland (2012).
55. DAFM, *The Second National Forest Inventory Republic of Ireland: Main Findings*, Department of Agriculture, Food and the Marine, Johnstown Castle (2013).
56. R. G. Congalton, "A review of assessing the accuracy of classification of remotely sensed data," *Remote Sens. Environ.* **37**(1), 35–46 (1991).
57. D. Rossiter, "Technical Note: Statistical methods for accuracy assessment of classified thematic maps," International Institute for Geo-information Science & Earth Observation, Enschede, Netherlands (2014).
58. E. Mockler et al., *Pathways Project Final Report Volume 4: Catchment Modelling Tool*, Environmental Protection Agency, Ireland (2014).
59. DAFM, "Integrated forestry information system," https://www.agriculture.gov.ie/forestservice/forestservicegeneralinformation/foreststatisticsandmapping/afforestationstatistics/ (accessed 5 July 2020).
60. DHPLG, "Building control management system," https://www.housing.gov.ie/housing/statistics/house-building-and-private-rented/construction-activity-starts (accessed 5 July 2020).
61. J. Delegido et al., "Evaluation of Sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content," *Sensors (Basel).* **11**(7), 7063–7081 (2011).
62. E. J. Emengini, G. A. Blackburn, and J. C. Theobald, "Discrimination of plant stress caused by oil pollution and waterlogging using hyperspectral and thermal remote sensing," *J. Appl. Remote Sens.* **7**, 073476 (2013).
63. J. Liu et al., "Mapping within-field soil drainage using remote sensing, DEM and apparent soil electrical conductivity," *Geoderma* **143**(3), 261–272 (2008).
64. B. S. Naz, S. Ale, and L. C. Bowling, "Detecting subsurface drainage systems and estimating drain spacing in intensively managed agricultural landscapes," *Agric. Water Manag.* **96**(4), 627–637 (2009).
65. R. O'Hara, S. Green, and T. McCarthy, "The agricultural impact of the 2015–2016 floods in Ireland as mapped through Sentinel 1 satellite imagery," *Irish J. Agric. Food Res.* **58**, 44–65 (2019).
66. A. Kerebel and N. M. Holden, "The relationship between farmer opinion of suitable conditions for nutrient application, soil moisture deficit and weather," *Soil Use Manag.* **32**(4), 613–622 (2016).
67. N. Flood, "Comparing Sentinel-2A and Landsat 7 and 8 Using Surface Reflectance over Australia," *Remote Sens.* **9**(7), 659 (2017).

Rob O'Hara is a postdoctoral researcher at Teagasc and VistaMilk SFI Research Centre. His PhD focused on EO methods of mapping artificially drained soils. Currently, he is researching EO mapping of land use intensity / habitats on managed grasslands. He is a member of SPIE.

Stuart Green is a senior research officer in Teagasc, the Irish agriculture and food development authority, researching the use of applied earth observation in grassland agriculture.

Owen Fenton is a principal research officer with Teagasc and works in the research area of soil hydrology with particular emphasis on pollutants losses from agricultural systems.

Pat Tuohy is a research officer at Teagasc specialising in soil management, soil physics, hydrology and land drainage.

Biographies of the other authors are not available.