Assessment of multi-temporal, multi-sensor radar and ancillary spatial data for grasslands monitoring in Ireland using machine learning approaches

Authors: Brian Barrett\textsuperscript{a}, Ingmar Nitze\textsuperscript{a1}, Stuart Green\textsuperscript{b}, Fiona Cawkwell\textsuperscript{a}

Addresses:
\textsuperscript{a} School of Geography \& Archaeology, University College Cork (UCC), Rep. of Ireland
\textsuperscript{b} Teagasc, Ashtown, Dublin 15 – Irish Agriculture and Food Development Authority, Rep. of Ireland
\textsuperscript{1} Present address: Alfred Wegener Institute, Telegrafenberg A43, D-14473 Potsdam, Germany

* Corresponding author at: School of Geography \& Archaeology, University College Cork (UCC), Rep. of Ireland

Ireland. Tel: +353(0) 21 490 2391
Email address: bbarrett@ucc.ie

Keywords: Extremely Randomised Trees, Random Forests, Support Vector Machines, Grasslands, Classification, Radar
Abstract

Accurate inventories of grasslands are important for studies of carbon dynamics, biodiversity conservation and agricultural management. For regions with persistent cloud cover the use of multi-temporal synthetic aperture radar (SAR) data provides an attractive solution for generating up-to-date inventories of grasslands. This is even more appealing considering the data that will be available from upcoming missions such as Sentinel-1 and ALOS-2. In this study, the performance of three machine learning algorithms; Random Forests (RF), Support Vector Machines (SVM) and the relatively underused Extremely Randomised Trees (ERT) are evaluated for discriminating between grassland types over two large heterogeneous areas of Ireland using multi-temporal, multi-sensor radar and ancillary spatial datasets. A detailed accuracy assessment shows the efficacy of the three algorithms to classify different types of grasslands. Overall accuracies ≥ 88.7% (with kappa coefficient of 0.87) were achieved for the single frequency classifications and maximum accuracies of 97.9% (kappa coefficient of 0.98) for the combined frequency classifications. For most datasets, the ERT classifier outperforms SVM and RF.
1. Introduction

Detailed knowledge on land cover and land use are key parameters for sustainable planning and management of resources, and are crucial data components for many aspects of global change studies (Verburg et al., 2011). Globally, grasslands cover 37% of the total land area (O’Mara, 2012) and are valuable ecosystems that support a multitude of roles, most importantly food security, biodiversity conservation and greenhouse gas mitigation. In Ireland, grassland is the dominant land cover, occupying approximately 60% of the country’s terrestrial area (Eaton et al., 2008), and represents over 90% of all agricultural land (pasture, grass silage or hay, and rough grazing) (~4,000,000ha). The dominance of grassland in Ireland stems largely from the favourable climatic conditions that allow for a prolonged grass growing season (Lafferty et al., 1999). Given their extent, there is considerable potential to increase carbon sequestration in grasslands through improved land management and restoration of degraded grasslands (Soussana et al., 2004; O’Mara, 2012). This has a particular importance for Ireland, given the dominance of grassland and expected trends for intensification of grassland management through implementation of the *Food Harvest 2020* strategy (DAFM, 2011) by the Irish Government and the abolition of milk quotas across the EU-28 in 2015 (Läpple & Hennessy, 2012). Milk production is expected to increase by 50 per cent by 2020 compared to the 2007-2009 average and beef output by 40 per cent under the *Food Harvest 2020 strategy*. These are the agri-sectors (dairy and beef) of most importance in relation to GHG emissions (Casey & Holden, 2005, 2006). While these transitions will affect the sequestration and emission processes, they also present significant implications for biodiversity (Benton et al., 2003; Walker et al., 2004) and water quality. Intensification and expansion of production, fertiliser use and land improvement (drainage) will impact surface and groundwater and it is essential that Ireland complies with legislation (Water Framework Directive (2000/60/EC), Groundwater Directive (2006/118/EC) and Nitrates Directive (91/676/EEC)) to maintain (and/or restore) good water quality and avoid fines from the European Commission.

In light of the major changes now undergoing in Irish grassland practices, it is crucial that more detailed and more precise inventories of grassland are obtained so that sustainable grassland
management can be achieved. Under EU legislation (Decision No 529/2013/EU), mandatory accounting on greenhouse gas emissions and removals resulting from activities related to grazing land management is to be phased in between 2013 and 2021. From 2016 to 2018, Member States will be required to report to the Commission on the inventory systems in place or being developed to estimate emissions and removals from grazing land management. Collection of such data by traditional means (e.g. through field work) can be cost prohibitive and resource intensive. The use of earth observation (EO) data can provide routine coverage over large and remote areas and readily be incorporated into operational mapping or monitoring programs to provide a cost effective means of replacing or complementing field data collection. As agricultural areas are characterised by temporal differences due to plant phenology and farm management (e.g. grazing and mowing activities), the use of a multi-temporal image series is more effective than using a single date acquisition (Ranson & Sun, 1994; Kasischke et al., 1997; Pierce et al., 1998). Given the difficulty of acquiring multiple acquisitions throughout a growing season with optical sensors limited by frequent cloud cover, synthetic aperture radar (SAR) data can be an important alternative, or complementary data source, to provide the best opportunity for generating a multi-temporal data set as they are almost independent of weather and illumination conditions, making their use for operational purposes especially appealing. Semi-automatic analysis of these data for land cover applications usually consists of them being classified into a pre-determined number of individual classes.

A number of different approaches exist for semi-automated classification of EO data (e.g. see Lu & Weng, 2007 for a comprehensive review). The maximum likelihood classifier (ML) is one of the most commonly used supervised classification techniques due to its simplicity and implementation in almost all standard image processing software packages (Waske & Braun, 2009). This technique assumes a normal Gaussian data distribution and computes the statistical probability of a given pixel belonging to a particular class. The assumption of a Gaussian distribution is not ideal in the context of SAR data due to the interference of speckle. Machine learning approaches are non-parametric and therefore do not rely on any assumption about data distribution. Support Vector Machines (SVM) and Random Forests (RF) are two of the most sophisticated machine learning approaches available. Both methods have become increasingly popular and used frequently in land cover classifications using
multi- and hyper-spectral satellite imagery (e.g. Melgani & Bruzzone, 2004; Gislason et al., 2006; Waske & Braun, 2009; Dalponte et al., 2009; Kavzoglu & Colkesen, 2009; Loosvelt et al., 2012a).

Although these classifiers have produced promising results with optical data, only relatively few applications to SAR data are known (e.g. Waske & Benediktsson, 2007; Waske & van der Linden, 2008; Waske & Braun, 2009; Walker et al., 2010; Longépé et al., 2011; Loosvelt et al., 2012a; Rodriguez-Galiano et al., 2012; Deschamps et al., 2012). Similarly, most studies that have applied RF or SVM using SAR data have focused on relatively small study areas, e.g. Lardeux et al., (2009); Waske & Braun, (2009), and Loosvelt et al., (2012b) all focused on an area size of 5x5km. The backscatter and environmental variability associated with different land cover classes rises with increasing study area size, and therefore the performance of the classifiers strongly depends on the spatial scale of the study area. Furthermore, most of these studies are concerned with general land cover classification or crop classification, and relatively few studies focusing on the use of SAR for grasslands monitoring or parameter retrieval exist (e.g. Hill et al., 1999; Pairman et al., 2008; Smith & Buckley, 2011; Dusseux et al., 2012; Satalino et al., 2012; Voormansik et al., 2013; Wang et al., 2013).

In this study, the RF and SVM classifiers are applied to multitemporal C & L-band SAR datasets covering two different regions in Ireland of 1091km$^2$ and 1837km$^2$ in size. In addition, the use of a relatively new machine learning classifier - Extremely Randomised Trees (ERT) is also investigated. ERT has not found widespread application within the EO community to date, and has been used primarily in computer vision (Moosmann et al., 2008) and bio-medical imaging applications (Marée et al., 2013). This paper sets out to evaluate their potential for creating grassland inventories over large heterogeneous areas, making the key distinction between grasslands that are improved, and those that are semi-improved. Improved grasslands are intensively managed and include pastures and grassland harvested for hay and silage. They occur across a range of soil and site conditions and usually have high external inputs of fertiliser and herbicides and disturbance (e.g. reseeding and drainage). Alternatively, semi-improved grasslands are characterised by relatively low external inputs of fertiliser and herbicides and low disturbance relative to improved grasslands. They are threatened largely by the abandonment of all management (and the subsequent reversion to scrub) or the
intensification of management (and a loss of plant species diversity). Their survival is of increasing concern given the targets set out in Food Harvest 2020 and the decline of rural populations, increased costs in farming marginal lands and improved off-farm opportunities that are resulting in the abandonment of marginally productive lands. Some areas may be abandoned as a result of intensification in other areas (Joyce, 2013).

2. Study Areas & Datasets

2.1 Study sites

The Republic of Ireland has a land area of 70,282 km$^2$ and is divided into 26 counties for administrative purposes. The study areas are located in central and north western Ireland, encompassing the counties of Longford and Sligo respectively (see Fig. 1). Longford has an area of 1091 km$^2$ and is predominantly lowland, with the highest altitude reaching 278m (Cairn Hill) in the north of the county. It is a rural county with the second lowest population in Ireland (~36,000 people) and where pastures are the dominant land cover. The west of the county is dominated by extensive tracts of peatlands whereas lakelands are found in the south, west and north-east. The south-east represents the predominantly agricultural land in the county, interspersed with plantations of deciduous and mixed woodland. The north and northwest is characterised by drumlin topography. Annual average rainfall is between 900 and 1000 mm. Sligo has an area of 1837 km$^2$ and shares a substantial part of its border with the North Atlantic Ocean. The county has two distinctive upland areas; the Dartry mountains to the north of the county and the Ox mountains to the south-west. Due to its proximity to the ocean, Sligo has a maritime climate with generally low annual temperature amplitudes and humid conditions. Annual average rainfall varies between 1000-1200 mm for the low-lying areas and between 1600-2000 mm for the highest mountains in the Dartry and Ox ranges.

2.2 SAR Dataset

A total of 12 ENVISAT ASAR IM scenes, 15 ERS-2 scenes and 12 ALOS PALSAR scenes from 2008 have been acquired for this study. ERS-2 was launched on 20$^{th}$ April 1995 by the European Space Agency (ESA) and operated at C-band with a single mode (VV polarisation and ~23° incidence
angle) until it was retired on 5\textsuperscript{th} September 2011. ENVISAT, the successor to ERS-2 was launched by ESA on 1\textsuperscript{st} March 2002 and provided continuous data from a range of sensors until operations ceased in April 2012. The Advanced Synthetic Aperture Radar (ASAR) instrument operated at C-band and provided data in various modes with differing spatial resolutions and coverage. In this study, data in Image Mode (IM) were obtained to coincide with the ERS-2 beam mode. The ALOS satellite was launched on 19\textsuperscript{th} January 2006 and operated until 12\textsuperscript{th} May 2011. The PALSAR instrument on board ALOS operated at L-band and provided fully polarimetric capabilities. Table 1 shows the acquisition dates and associated data product characteristics for both Longford and Sligo. All PALSAR data were acquired from ascending orbits, ENVISAT ASAR data was acquired from descending orbits and ERS-2 data acquired from both ascending and descending orbits. The ERS-2 and ENVISAT ASAR scenes were acquired in VV polarisation at an incidence angle of ~23° and a 100km swath width. For PALSAR data, there are six scenes per county – two scenes in Fine Beam Single (FBS) mode (HH polarisation) and four scenes in Fine Beam Dual (HH/HV polarisation). The PALSAR scenes were acquired with an incidence angle of ~38° and have a swath width of 70km. Two frames were required to cover each county which results in a total of three separate acquisition dates per county.

2.3 Ancillary data

In addition to the SAR data, various ancillary datasets were used during the classification process. The Irish Land Parcel Identification System (LPIS), developed by the Irish Department of Agriculture, Food & the Marine, provides an alphanumeric identification system for all agricultural parcels receiving Area Aid funding from the European Union (EU). The dataset contains vector boundaries of the parcels with attribute information including area, crop type and stocking densities. The 2008 LPIS dataset was used to facilitate the identification of arable lands and improved grassland. The National Parks and Wildlife Service (NPWS) semi-natural grasslands field survey dataset for Longford (O’Neill et al., 2009) and Sligo (O’Neill et al., 2010) were used to provide training data for the semi-improved grassland classes. A total of 193 2m x 2m relevés from 49 sites (covering an area of 1306ha) in Longford were surveyed for the NPWS field survey. In Sligo, 322 2m x 2m relevés from 52 sites (area coverage of 1546ha) were surveyed. The Forest Inventory &
Planning System (FIPS), maintained by the Forest Service (Ireland) aided the collection of forest reference points and further datasets used included the Teagasc-EPA Soils and Subsoils dataset (Fealy et al., 2009), an Ordnance Survey Ireland (OSi) 10m spatial resolution DEM and OSi orthophotography.

3. Methodology

3.1 Training Data

Ten landcover classes were considered according to the specific needs of the project: dry humic semi-improved grassland, dry calcareous semi-improved grassland, wet semi-improved grassland, reclaimed improved grassland, dry improved grassland, forest, peatland, cropland, water, and settlement. Table 2 displays these ten classes along with their definitions and the associated number of training and validation samples per class. They were largely defined according to the classification scheme developed by the Irish Heritage Council (Fossitt, 2000) and the Intergovernmental Panel on Climate Change (IPCC) Guidelines for National Greenhouse Gas Inventories (IPCC, 2006). The dominant species composition and management regime of the grassland classes is detailed in Table 3. Improved grasslands are typically species-poor and usually dominated by ryegrasses (*Lolium perenne*, *Lolium multiflorum*) and clover (*Trifolium repens*). They are intensively managed for grazing and/or cutting and have high external inputs of fertiliser and disturbance (reseeding and biomass removal). Species composition varies considerably in the semi-improved classes (GSw, GSdc, and GSdh). GSw usually contains abundant rushes (*Juncus* spp.) and sedges (*Carex* spp.), while GSdc and GSdh both comprise of a wide range of grasses and broadleaved herbs. These grasslands are often unenclosed and not managed intensively. The mean number of species recorded at each site in Longford was 92 and 117 for Sligo.

A stratified random sampling approach was adopted for the selection of training and validation data for each class. A 20m x 20m grid was overlaid on the datasets and point values extracted from the cell centroid locations. User interpretation of OSi orthophotography, Bing Imagery, LPIS and FIPS data facilitated the distinction of the non-grassland classes, whereas the grassland classes were distinguished using additional data from the NPWS semi-natural grasslands field survey datasets,
LPIS and the Teagasc/EPA Soil dataset. All samples were selected independently of the SAR data. A 5-fold cross validation was performed to minimise the impact of the training data selection. The 5-fold cross validation partitions the dataset randomly and uses 5-1 folds for training and the remaining one for validating the classifier. The main advantage is that all the samples are eventually used for both training and validating the classifier. The same strategy was used for all three classifiers.

3.2 Image Processing

All SAR data were delivered as single look complex (SLC) data products from the European Space Agency (ESA) and processed using SARscape® software within an ENVI® environment. Auxiliary orbit and calibration information for each image acquisition was used to generate true backscattering coefficients ($\sigma^0$). The most recent external calibration files (XCA) along with precise satellite orbital data (_VOR) provided by the DORIS (Doppler Orbitography and Radiopositioning Integrated by Satellite) instrument on board ENVISAT were used for the processing of the ASAR scenes, while ERS PRARE (Precise Range And Range-Rate Equipment) Precise Orbit Products were used for processing the ERS-2 data. As a first step, the SLC images acquired with similar track and frames were co-registered. Multi-looking factors of 2 (in range) and 6 (in azimuth) for the PALSAR FBS data and 1 and 6 for the FBD data were then applied to create 20 x 20m pixels. ERS-2 and ASAR data were multi-looked by a factor of 1 and 5, and 1 and 6 in range and azimuth respectively. The multi-looked data were subsequently speckle filtered using a 5x5 kernel size Frost filter and radiometrically and geometrically calibrated and converted to dB. An OSi 10m spatial resolution DEM with a vertical accuracy of 0.5m and several ground control points (GCPs) were used to geometrically correct the SAR scenes to the Irish Transverse Mercator (ITM) projection using a Range-Doppler approach.

Several texture features derived from the SAR data were investigated to evaluate their contributions to the classification accuracies. The common Grey Level Co-occurrence Matrix (GLCM) (Haralick et al., 1973) was applied to all SAR data to extract textural information, as carried out in studies by Longépé et al. (2011) and Li et al. (2012). Four textural parameters (homogeneity, contrast, entropy and second moment) were computed using a 3x3 sliding window, and the relation
between neighbouring pixels was considered around the four main directions (0°, 45°, 90° and 135°).

Multi-temporal features (gradient, standard deviation, median, maximum, minimum, span ratio, maximum increment, span difference, min ratio and max ratio) were also extracted from the time-series intensity data to enable structures and/or temporal changes to be detected.

No terrain distortions were present in the Longford dataset as a result of its low-lying topography. Conversely, the scenes for Sligo needed to be masked for certain terrain-induced distortions due to a more varying topography. The total area masked ranged from 0.2 - 1.4% of the total county area depending on acquisition characteristics. These areas were subsequently masked from all the SAR intensity data and SAR-derived texture and multitemporal measures and excluded from the classifications.

3.2.1 Grassland phenology

The Normalised Difference Vegetation Index (NDVI) was calculated to provide an approximate phenological stage for the Level 1 grassland classes at each SAR acquisition date. NDVI is a strong indicator of the photosynthetic capacity of a canopy and is used in this study as a proxy for the approximate phenological stage of grasslands throughout 2008. Sixteen day composites of Moderate Resolution Imaging Spectroradiometer (MODIS) Terra (MOD13Q1) (250m spatial resolution) acquisitions were used to calculate NDVI values over the study areas. Only pure improved and semi-improved grassland pixels were collected. In total, 124 and 122 pixels were identified for Longford and 301 and 167 pixels for Sligo for the improved and semi-improved classes respectively. Fig. 2 displays a smoothed time series (Hodrick-Prescott-(HP) filter (Hodrick & Prescott, 1997)) of MODIS NDVI values for both classes for the 2007 – 2009 period.

3.3 Classifiers

3.3.1 Random Forest

Three supervised classification algorithms; Random Forests (RF), Support Vector Machines (SVM) and Extremely Randomised Trees (ERT) have been applied to the SAR data on a per-pixel basis in this study. The RF classifier (Breiman, 2001) is an advanced form of bagging (bootstrap aggregating) that forms an ensemble of classification and regression tree (CART)-like classifiers
RF can handle high-dimensional datasets, accepts a variety of measurement scales for both numeric and categorical variables, is less sensitive to noise, can handle many input variables and does not suffer from overfitting (Rodriguez-Galiano et al., 2012). The algorithm randomly selects a subset of samples (2/3 of data samples) for the training of each individual decision tree with the remaining samples (1/3 of data samples) assigned as out-of-bag (oob) samples which are used to test the classification and estimate the error. For each individual tree, the Gini index (a measure of class homogeneity) is used to perform the best split of a random set of input features at each node. Using a majority vote, a final class is assigned from the multiple outputs of all constructed decision trees.

The algorithm can also provide an estimate of the relative variable importance and allow for a quantification of the classification probabilities. This is of particular value for studies which involve multi-source variables, as it allows for the contribution of each of the different source variables to the classification accuracy to be evaluated (Gislason et al., 2006). By reducing the data dimensionality (thereby minimising sensitivity to the Hughes phenomenon (Hughes, 1968)) and identifying the optimum input variables, an enhanced classification performance (in terms of accuracy and efficiency) may be obtained. To do this, RF switches one input variable while keeping the remaining variables constant and measures the decrease in accuracy, if any, by means of the oob error estimation and Gini Index decrease (Breiman, 2001). The RF classifier requires only two input parameters; the number of trees ($N$), and the number of variables to split at each node ($m$). In this study, $N$ was set to 200 and $m=\sqrt{p}$, where $p$ is the number of predictor variables.

### 3.3.2 Support Vector Machines

Support Vector Machines are based within the framework of the statistical learning theory developed by Vapnik (1995) and use the principles of structural risk minimisation (SRM). The aim of SVMs is to fit an optimal separating hyperplane between different classes by using only the training samples that lie at the edge of the class distributions (i.e. the so-called support vectors). As the hyperplane maximises the distance between itself and the two classes, it can generalise accurately on unknown samples (Foody & Mathur, 2004), and minimises overfitting – a problem common to
classifiers such as neural networks and decision trees. As SVMs were initially developed to perform
binary classifications, various methods such as one-against-all, one-against-one and directed acyclic
graph have been proposed to extend SVMs to multiclass problems (Hsu & Lin, 2002). To separate
classes with non-linear boundaries, kernel functions are used to transform training samples into a
higher dimensional space where linear class separation can be performed (Huang et al., 2002). This
allows SVMs to perform well even when relatively few training samples are available (Pal & Mather,
2005).

In this study, the Radial Basis Function (RBF) kernel and one-against-one strategy were adopted.
The SVM is optimised for appropriate values for C (penalty or regularisation term) and \( \gamma \) (controls the
width of the radial basis function kernel) by a grid search method using cross-validation. The grid
search method tests different pairs of parameters before finally selecting the set with the highest cross
validation accuracy. A full mathematical formulation of the SVM can be found in Vapnik (1995),
Burges (1998), and Huang et al. (2002) while a more in-depth summary of the use of SVMs in
remote sensing is provided in Mountrakis et al. (2011).

### 3.3.3 Extremely Randomised Trees

Extremely Randomised Trees or Extra-Trees is a relatively new tree-based ensemble classifier
method, introduced by Geurts et al. (2006) and similar in manner to Random Forests. The main
difference is that Random Forests finds the best split among a random subset of variables when
constructing a tree whereas Extremely Randomised Trees selects a node split randomly and uses the
same input training set to train all individual trees (i.e. no bagging). The primary benefit of increasing
the randomisation is to reduce the variance among trees. Using the full training set rather than a
bootstrap sample minimises the bias. Furthermore, ERT is much faster than Random Forests, as the
computational load of the training stage is simplified as the algorithm doesn’t search exhaustively for
the optimal split as in RF. ERTs have been widely used in biomedical imaging applications but very
few applications to EO data exist (Zou et al., 2010). To the author’s knowledge, this is one of the first
known applications of ERT to multi-temporal SAR data. As with the RF classifier, \( N \) was set to 200
and \( m = \sqrt{p} \) for the ERT classifications. All three classifications were implemented using the open-source Scikit-learn module (Pedregosa et al., 2011) in Python v2.7.3.

4. Results

4.1. Temporal backscatter analysis

The distribution of backscatter coefficient (\( \sigma^0 \)) values extracted for the training samples representing the different grassland classes are displayed in Fig. 3. The boxes represent the inter-quartile range and the whiskers extend to within 1.5 times the inter-quartile range from the upper and lower quartiles. The C-band VV polarisation signatures display similar dynamic ranges for acquisitions in both counties (Fig. 3(a) & (b)), though the overall variation in the distribution of \( \sigma^0 \) values for each class is higher for the Sligo acquisitions. The difference between median \( \sigma^0 \) values varies between classes for all acquisition dates. In general, the GSdh and GSdc classes display the greatest separability for the Longford dataset, with GSdh \( \sigma^0 \) values on average 3dB lower than GSdc. A similar pattern is not observed for all Sligo C-band acquisitions. In general, the autumn and winter acquisitions display the greatest separation (approximately 2 – 3dB) for the GSdh and GSdc classes. The GAr class tends to have the highest median \( \sigma^0 \) values for the C-band Sligo dataset, while the GSdc class generally has the highest median \( \sigma^0 \) values for Longford. The degree of overlap between the classes is stronger for the Sligo acquisitions, although it is not consistent throughout the time series.

The L-band data display a better separation between the GSw and GAd classes, with median GSw \( \sigma^0 \) values approximately 3dB higher than GAd for both HH and HV polarisations. Wet grasslands typically have a higher soil moisture content compared to improved grasslands which may be responsible for the increased backscatter. As expected, the HV signals are weaker than the HH but the variations in median \( \sigma^0 \) of the grasslands display a similar pattern. As attenuation of microwave signals is greater in areas with increased canopy cover, the contribution from these scattering mechanisms is progressively reduced during the vegetation growth stages. This is observed in the lower median \( \sigma^0 \) values obtained for the summer L-band HH datasets (Fig. 3(c) & (e)) and less clearly for some of the summer C-band acquisitions (Fig. 3(a) & (b)). The analysis of the temporal
backscatter profiles confirms the possibility to separate between the grassland classes using a combination of C- and L-band acquisitions and different polarisations.

4.2 Feature Importance

RF has the ability to generate a feature or variable importance ranking which allows for the reduction of the number of input variables used in the classification process. A RF model with 5000 trees, as suggested by Díaz-Uriarte & De Andres (2006) was used to calculate the importance of the contribution of each variable to the classifications. The accumulation of all feature scores equals one. All SAR intensity, SAR-derived and ancillary data features were initially included in the classifications. In all cases, the GLCM texture measures were found to have little measureable influence on the classification accuracies and were subsequently removed from any further analysis. Separate RF models were considered for the three sets of data: those derived from C-band acquisitions only, those derived from L-band acquisitions only, and those derived from C and L-band acquisitions combined. The feature importance scores for the C-band, L-band and merged C- and L-band datasets for both study areas can be seen in Fig. 4. In Fig. 4(a), the first 12 variables are the C-band backscatter intensities in chronological order, the following 30 variables are the multitemporal features for the different track and frames (in order - standard deviation, span ratio, span difference, min ratio, minimum, max ratio, maximum increment, maximum, gradient, median), and the last four are the soils, sub-soils, elevation and slope ancillary data. A similar structure is adopted for Fig. 4(b)-(f). It is interesting to note that the ancillary variables; soils, sub-soils, elevation, slope are the most important variables in each dataset where all SAR and SAR-derived variables are relatively less important.

For the C-band Longford intensity data (Fig. 4 (a)), three acquisitions (20080408, 20080722 and 20080914) stand out as having the highest importance. Interestingly, these are all ASAR acquisitions. In Sligo, the intensity data with the highest importance are two ERS-2 acquisitions (20080729 and 20080817). Considering the L-band intensity data, the summer acquisitions (20080613 – Longford and 20080601/20080717 – Sligo) have the highest importance. Due to the low temporal density of observations, it is more challenging to determine the optimum acquisition timing. The choice of polarisation also varies according to each study site, where the HH polarisation data has a higher
importance than the HV polarisation data in Sligo (Fig. 4(d)) and where there is no real difference between polarisation importance for Longford (Fig. 4(c)). For both frequencies, the contribution of the multitemporal minimum, maximum, and median backscatter intensity for both polarisations also display high importance. In the combined C- and L-band classifications (Fig. 4(e) – (f)), the dominant influence of the L-band measurements is apparent for both study areas.

From this analysis, the number of input features was subsequently reduced based on the lowest importance ranking scores and in an effort to minimise data redundancy. A series of RF, SVM and ERT classifications were carried out to assess the performance of the classifiers based on the reduced feature inputs until the optimum variables were selected that led to the highest accuracy classifications (v1 classifications in Table 4). For the final classifications, only the backscatter intensity measurements and ancillary data variables were included (i.e. all multitemporal-features were excluded). The importance of the contribution of each feature to the reduced input RF classifications for the five grassland classes is displayed in Fig.5. For both Longford and Sligo, it is shown that different variables have the strongest influence for specific classes. Overall, it can be seen that there is no clear dominance of either C- or L-band intensity data in terms of importance scores that satisfies all grassland classes. For example, in Longford the C-band intensity measurements have a much higher importance for dry humic semi-improved grasslands (GSdh), whereas the L-band measurements have a higher importance for the wet semi-improved grasslands (GSw). In Sligo, the L-band intensity measurements are most important for discriminating between improved (GAd and GAr) and wet semi-improved grasslands (GSw).

4.3 Classification Accuracy Assessment

Several measures derived from the confusion matrix were used to evaluate the classifier performance and their associated uncertainties. These included the overall accuracy (OA), user’s accuracy (UA), producer’s accuracy (PA) (Congalton, 1991; Congalton & Green, 2009) and the Kappa statistic (κ) (Cohen, 1960). The training and validation of the classifiers was performed on 5202 and 6442 samples from the Longford and Sligo datasets respectively. Two different datasets were used for the classifications; v0 represents classifications that considered all input variables, and
v1 represents classifications that considered only the backscatter intensity measurements along with the soils, sub-soils, elevation and slope data (see Table 4). The effect of reducing the number of input features can be seen across both study areas in all classifications where OA increases of between 0.8 – 5.8% were obtained.

From Table 4, it can be seen that the L-band dataset classifications are only marginally outperformed by the C-band dataset (generally on the order of ~2%). This is significant given the fact that the L-band time-series is comprised of just three separate acquisitions for both study areas while the C-band dataset is made up of 12 and 15 acquisitions for Longford and Sligo respectively. The positive impact of a synergy between frequencies during classification is clearly demonstrated in Table 4 where an increase in overall accuracy (average of 1.5%) is observed for each classifier in both study areas. In all three sets of classifications (C-band, L-band and combined C and L-band), the Longford results outperform the Sligo results. Notwithstanding this, the final (v1) Sligo classifications have an average overall accuracy of 91%. In contrast, the Longford datasets produce average classification accuracies of >96%. The user’s and producer’s accuracy of individual classes for both areas were also high, with Longford having considerably lower class standard deviations compared to Sligo.

The classification output maps in Fig. 6 display the distribution of the different land cover categories in the two study counties. These maps were created using SAR intensity measurements only and excluded the ancillary variables. In Sligo, the large-scale effect of the regular rainfall and resulting high humidity can be observed in the manner in which blanket peat bogs cover the mountainous regions of the Ox (southwest) and Dartry (north) ranges. It can be seen that the majority of improved grassland occurs along the coastline and in a central corridor in the county. The extensive peatlands (raised bogs) in the west of Longford are also clearly distinguished with the majority of improved grassland appearing to occur in the southern and north eastern parts of the county. Some areas of confusion are present, namely the mixing of settlement and forest classes due to their similar temporal backscatter signal, nonetheless the distribution of the different classes appears to have been captured reasonably well.
4.4 Prediction Probabilities

An important output of the ERT (and RF) classifier is the class probabilities (Eq. 1). This is the probability \( p \) of an observation being classified into class \( i \), where \( k \) is the total number of trees in the ensemble and \( k_i \) is the total number of trees classifying the observation as class \( i \).

\[
p(i) = \frac{k_i}{k}
\]  

This can be of particular use as a measure of quantifying the level of uncertainty in the generated classification maps (Fig. 6). For example, low probabilities in Fig. 7 represent pixels that are unlikely to be dry improved or wet semi-improved grasslands while intermediate probabilities indicate possible confusion between another or more classes. High probabilities indicate pixels that have limited uncertainty about the assigned class. As can be seen in Fig. 7, areas along the coastline of Sligo that were classified as improved grasslands have a very high probability of being assigned to the correct class. The semi-improved grasslands are more noticeable to the northwest and south of the Ox mountains and towards the east of the county. In Longford, the area dominated by extensive tracts of commercial peatlands interspersed with vegetation is clearly distinguishable in the south west of the county with corresponding low probabilities. At the same time, the areas of dry improved grassland with the highest probabilities are observed in the southwest and east of the county while wet semi-improved grasslands in the centre and northwest of the county have the highest probability of having been correctly classified.

5. Discussion

5.1 Backscatter analysis

Microwave scattering from grasslands is complex and has been investigated by several model experiments (Chauhan et al., 1992; Saatchi et al., 1994; Stiles & Sarabandi, 2000; Stiles et al., 2000). In practice, determining and isolating the scattering mechanisms from grasslands is difficult due to the various influencing factors. The backscattering coefficient is a function of the radar system parameters (frequency, polarisation and incidence angle) and of the surface parameters (dielectric and geometric properties) (Ulaby et al., 1982). Each of these parameters influences the backscatter.
response from the imaged surface. For vegetated surfaces, penetration depth depends on the moisture, density and geometric structure of the plants and soil. In general, the longer wavelength L-band signal (26cm) partly penetrates through the grass canopy and the return signal predominantly contains information about the underlying soil properties. Shorter wavelength C-band (5.6cm) backscatter is influenced more by the plant canopy (Stolz & Mauser, 1997; Schieche et al., 1999; Moreau & Le Toan, 2003). Similarly, the incidence angle and polarisation dependence of radar backscatter is influenced by the vegetation cover (Skriver et al., 1999). The higher incidence angle of the PALSAR sensor (~38°) leads to a higher vegetation influence on the backscattered signal, compared to the steeper 23° incidence angle of ERS-2 and ENVISAT ASAR, which results in the returning signal having a higher dependence on the dielectric properties of the ground surface. Concerning polarisation, several studies have also reported stronger backscatter from grasses at VV polarisation, compared to HH polarisation which interacts more strongly with broad-leaved canopies (Macelloni et al., 2001; Hill et al., 2005). Changes in the geometric and dielectric properties of the different grasslands over the calendar year significantly alter the backscatter signal response. For each acquisition date, the grasslands across the study areas are in various conditions, due to differences in grazing rotation, grazing intensity, and moisture conditions. As a result, it is not possible to determine a single wavelength, polarisation and season that are best able to separate the five grassland classes based on their $\sigma^0$ values.

5.1.1 Soil moisture influence on the backscatter signal

Soil moisture is a significant factor influencing grass growth and the radar backscatter signal from grasslands (e.g. Hill et al. (1999); Barrett & Petropoulos (2013)). Temporal patterns in soil moisture impact on the grass growth rate, nutrient uptake and on the length of the grazing season. When the soil is very wet, grass growth and nutrient uptake is low and it can affect the agronomic management of the farm (e.g. timing of fertiliser spreading). Soil moisture variations usually follow precipitation trends. However, they are difficult to determine or predict due to the complex interactions between the various factors that influence the soil moisture content (e.g. soil texture, topography and vegetation cover) (Tromp-van Meerveld & McDonnell, 2006). The intensity and frequency of precipitation
events play an important role in determining soil water movement in terms of infiltration and percolation processes. Changes in the moisture content of the soil can result in large changes in radar backscatter. The Essential Climate Variable Soil Moisture (ECV SM) product (Liu et al., 2012) (Wagner et al., 2012) was used in this study to provide an estimate of the soil moisture conditions of the two study areas on the SAR acquisition dates. For both counties, soil moisture estimates are averaged over four ECV pixels (0.25 degree spatial resolution) that encompass each county. Figure 8 displays the time series of daily ECV soil moisture values with in situ daily accumulated precipitation from nearby Met Éireann synoptic meteorological stations. The Mount Dillon and Markree station data are available only from summer 2008 while observations for the entire year are available from both the Ballyhaise and Knock Airport stations. The uncertainty range of the soil moisture values increases dramatically during the summer months. Soil moisture variability increases strongly during the vegetative growth period due to increased evapotranspiration and water uptake by plants (Hupet & Vanclooster, 2002; Illston et al., 2004). No clear relationship between the median C-band backscatter measurements (Fig.3) and soil moisture on different acquisition dates was observed. This is most likely due to the C-band signal interacting primarily with the uppermost layer of the grass canopy, increasing volume scattering and minimising influence from the underlying soil. Loew et al. (2006) found considerable (C-band VV) backscatter differences between fields with the same soil moisture content and attributed these to varying biomass of the studied grasslands. Similarly, Schieche et al. (1999) found the soil water content to have little influence on ERS-1/2 $\sigma^0$ values in the presence of vegetation, which may help explain why no consistent distinction can be observed between GSw and GAd that would be as a result of soil moisture differences. For L-band, the GSw class displays a higher median backscatter value than the other grassland types for both the HH and HV polarisations. This may be a manifestation of the soil moisture influence on the longer wavelength measurements. It could also possibly be due to enhanced volume scattering from Juncus spp. and Carex spp., although previous studies have noted an almost translucent nature of the grass blade-like structures at L-band (Dobson et al., 1996; Costa & Telmer, 2006).
5.1.2 Backscatter changes with grassland phenology

The degree to which the backscatter response can be attributed to vegetation or underlying soil conditions varies throughout the year as the grassland species are in different growth stages (e.g. flowering, senescence). Composite NDVI values for every 16-day interval between January 2007 and December 2009 were used to qualitatively analyse the influence of the grassland growth stage on the radar backscatter return (see Fig. 2). There are marked seasonal variations both within and between years, mainly due to meteorological factors and management (e.g. fertiliser application, grazing intensity) (Hurtado-Uria et al., 2013). The typical pattern is low or no growth over the winter months, due to low temperatures and low levels of solar radiation, with significant growth commencing in February or March and continuing to peak growth in June. The structure of the sward also depends on whether it is managed by grazing or cutting, with grazing usually leading to a more varied sward structure (e.g. due to trampling, dung patches). Improved grasslands generally display a higher NDVI throughout the year, apart from a period during the summer of 2008 in Longford when the semi-improved grasslands display higher NDVI values. Semi-improved grasslands tend to have a sward at a range of heights with greater botanical diversity than improved grasslands (see Table 3), with each species exhibiting a characteristic growth pattern during its seasonal development. Usually, longest heights are during summer when plants are flowering and setting seed and shortest during autumn or spring, when most species are germinating. GSDc are typically shorter than GSdh grasslands, but dominated by similar fine-leaved bent and fescue grasses. Either these grassland types are usually left fallow or only lightly grazed during the summer as livestock are moved to more productive pastures. It is possible that there may be a relationship between the L-band backscatter and grass height, as GAr generally has the lowest median backscatter values of all classes and the smallest grass heights. Similar results were also observed by Hill et al. (1999). Considering the low number of L-band acquisitions, it is difficult to investigate thoroughly the temporal variation throughout the growing season. A general decrease in backscatter strength is observed for the L-band HH acquisitions over the three dates. An increased vegetation biomass attenuates the backscatter signal and similar results have been observed by Dubois et al. (1995) and Rombach & Mauser (1997). To explore comprehensively the effect of grassland growth stage on the radar signal return, further investigations
are needed which focus at the field-scale, with study fields grouped by growth stage and complemented with *in situ* biophysical measurements of the grass canopy.

### 5.2 Feature Importance

The varying importance of the ancillary data on the different classes is interesting to note. For example, in Longford the soils and sub-soils have a dominant influence in the classification of dry calcareous semi-improved grasslands (given their usual confinement to limestone areas and alkaline soils). Similarly, the high importance of elevation is expected as these grasslands are largely confined to the slopes of esker ridges and moraines in the midlands (Fossitt, 2000). As found by Rodriguez-Galiano et al. (2012), it can be observed that elevation is most important for those classes whose spatial distribution is conditioned by relief (e.g. dry improved grasslands are mostly located in lowland areas and dry semi-improved humic grassland mainly occurs in upland areas). It is not as important for wet semi-improved grasslands as these are usually found on flat terrain in both upland and lowland areas. The sensitivity of the C-band intensity measurements to phenological differences in terms of the importance scores can be observed for improved grasslands (GAd) in both Longford and Sligo where the spring acquisitions have a higher importance than non-spring acquisitions.

### 5.3 Accuracy Assessment

The observed increases in classification accuracy when both frequencies are combined are consistent with the results of Lardeux et al. (2009) and Turkar et al. (2012). In all three sets of classifications (C-band, L-band and combined C and L-band), the Longford results outperform the Sligo results. This may be due to several factors; Sligo has a much larger area and a landscape with increased topographical variation when compared to Longford. In addition, there are some areas with bare rock outcrops and coastal areas which cause some confusion, although these are almost exclusively within the non-grassland classes. In future studies, it may be worthwhile including an additional ‘other land’ class to take these areas into account.

Not surprisingly, the cover types with high intra-class variability (i.e. grasslands) were the most difficult to reliably classify as grasslands form a continuum of types and there is confusion between class boundaries. A similar observation was made by Waske & Braun (2009) using multi-temporal C-
band SAR data with a RF classifier. These misclassifications were expected given their similar
descatter profiles and the heterogeneous nature of the Longford and Sligo landscapes. Clear-cut
boundaries between the different classes are not readily apparent and thereby contribute to the
confusion between classes. For the Sligo dataset, the majority of misclassifications occur between the
GSdc and GSdh classes, while in Longford, the GAr class is the most difficult to reliably classify. It is
observed that the accuracies also vary considerably depending on the classifier used. The increased
performance of the ERT classifier is observed across all datasets for both study areas. The OA
accuracies for the ERT classifier increase by 2% and 3% for the Longford and Sligo datasets
respectively when the multitemporal texture measures are excluded from the classification. The ERT
class-specific accuracies are less variable (lower standard deviations) for the final classifications
compared to the other classifiers. There is also a more discernible increase in SVM accuracy after the
variable exclusion when compared to the improvement in accuracy of the RF and ERT classifiers.

5.4 Comparison of different classification results with and without ancillary datasets

To quantitatively assess the influence of the ancillary datasets on the classification accuracies, a
number of classification permutations were carried out (see Table 5). Classifications were performed
using all radar and ancillary data (a), all data without soils (b), all data without elevation (c), and radar
intensities only (d). When the C-and L-band datasets are analysed separately, the differences in
accuracies (OA and κ) after the ancillary data are excluded from the classifications are considerably
larger than when the frequencies are combined. For Longford, there are small differences (3.6 - 5%) between the complete dataset and the radar only dataset for the three classifiers. The differences for
the same datasets for Sligo are larger at between 7.9 and 9.3%. A much higher decrease is observed
for the L-band than the C-band classifications when the ancillary data are excluded. This may be
explained by the fewer radar acquisitions that make up the L-band dataset compared to the C-band.
For all classification scenarios, the ERT classifier outperforms SVM and RF. In addition to the above,
the classifiers were run using only the ancillary variables (soils and elevation) as input. The ancillary
data results (see Table 5) outperform the separate C- and L-band radar only (d) results for both
Longford and Sligo. The combined C-and L-band radar intensities produce higher accuracies than using the ancillary data alone.

These findings show that combining radar datasets with ancillary data significantly improves the accuracy of distinguishing grasslands. Similar findings were found in the case of wetlands (Corcoran et al., 2013; Marti-Cardona et al., 2013). To further improve grassland classifications, the combination of optical with SAR and ancillary datasets may result in increased accuracies (Hill et al., 2005; Bagan et al., 2012; Smith & Buckley, 2011) although some studies (Price et al., 2002; Dusseux et al., 2012) have found a combined approach unsuccessful in yielding more accurate results. This would have obvious limitations; especially from an operational context as consistent and systematic cloud-free optical imagery may not be available for specific areas on a yearly basis. In addition, the extra effort (in terms of additional optical image processing and analysis) and cost may not be worthwhile in practice, as the findings from this study have shown that both single-frequency and multi-frequency multi-temporal SAR and ancillary data are capable of providing high classification accuracies in the absence of optical data.

6. Conclusion

SAR data are less frequently used in land-cover classification studies than optical data, yet they can be an important alternative or complementary data source for areas with persistent cloud cover. In Earth Observation and with SAR data in particular, we are usually faced with high-dimensional datasets. More sophisticated classifiers are needed to deal with such data, and machine learning algorithms such as SVM and RF are among the most effective methods currently available. In this study we have presented one of the first known applications of the Extremely Randomised Trees (ERT) algorithm for grassland discrimination using SAR data. The results provide for the first time fine spatial resolution land cover classifications for two counties in Ireland, showing the spatial distribution of different grassland classes based on the integration of multi-temporal, multi-sensor C- and L-band SAR and ancillary soils and elevation data. All three algorithms produce high classification accuracies for both study areas. The best results are achieved when both frequencies are used in the classifications, agreeing with previous studies which have highlighted the limitations of
using single polarisation and frequency data (Ferrazzoli et al., 1999; Blaes et al., 2005). An almost consistent, although at times moderate, superiority of ERT over RF and SVMs was observed for all datasets. Consistent with the results of Loosvelt et al. (2012b), a decrease in the number of variables led to a strong reduction in the data dimensionality and a more parsimonious dataset with increased overall accuracies. Overall, the high accuracies are very encouraging and the presented approach has demonstrated comparable results for two different large and heterogeneous areas. This is an important aspect in terms of the operational viability of the approach in being applied on a national scale across all counties in Ireland. If carried out nationally on an annual basis, the classifications could contribute to future assessments of Ireland’s greenhouse gas (GHG) inventory for the (extended) Kyoto protocol (2013-2020), EU reporting and other national assessment requirements. This will be critical for monitoring the impacts of achieving the productivity and environmental sustainability targets as set out in *Food Harvest 2020* and the *Green Low-Carbon Agri-Environmental Scheme (GLAS)* (DAFM, 2014) (introduced as part of the Rural Development Plan 2014-2020 that aims to work within the framework of key EU Directives and national and international targets for preserving grassland habitats and low input pastures) respectively.

Radar (and optical) EO technology will undoubtedly become an integral part of grassland resource management within the next decade. The availability of multi-temporal and multi-configuration C- and L-band data will increase with Sentinel-1A/B, and future planned SAR missions such as the Radarsat Constellation and ALOS PALSAR-2. The forthcoming availability of S-band data (NovaSAR-S) will present further opportunities, in addition to those already presented by X-band sensors such as TerraSAR-X and COSMO-SkyMed. These datasets will be invaluable for future studies where further research on the effect of specific management practices (e.g. cutting and grazing, drainage) on the backscatter response is required. Ideally, dedicated field studies coincident with the time of image acquisition are needed in order to fully understand and profile the scattering mechanisms caused by different grassland conditions (e.g. density, height, moisture content) under observation.
Acknowledgements

The authors would like to acknowledge the Irish Environmental Protection Agency (EPA) for funding this research under the Science, Technology, Research & Innovation for the Environment (STRIVE) Programme and the European Space Agency (ESA) for providing all SAR data through Cat-1 proposal ID 11768. The authors would also like to acknowledge the anonymous reviewers for their constructive comments and suggestions which helped to improve the manuscript. The assistance of Peter Hallahan of the Ordnance Survey Ireland and Kevin Black of Forest, Environmental Research and Services (FERS) is also appreciated.
References


Canadian Journal of Remote Sensing, 38(01), 60-68. doi: 10.5589/m12-012


List of Figure Captions

**Figure 1** Location of the study counties of Longford and Sligo in the Republic of Ireland (shaded). The blue markers indicate the locations of the Irish Meteorological Service (Met Éireann) synoptic stations, where MK=Markree, KA=Knock Airport, MD=Mt Dillon, and BH=Ballyhaise.

**Figure 2** MODIS NDVI time series for improved (green) and semi-improved (red) grasslands in a) Longford, and b) Sligo over a three year period from January 2007 to December 2009. There are 23 MODIS datasets per year and error bars represent 1 standard deviation. The data has been smoothed using the Hodrick-Prescott-(HP) filter.

**Figure 3** Temporal C- and L-band backscatter profiles of the grassland classes for Longford and Sligo. The boxes represent the lower (25th percentile) and upper (75th percentile) quartile values of the data with a black line at the median. The whiskers extend to the minimum and maximum values of the data (excluding outliers).

**Figure 4** Feature importance scores of the backscatter intensities, SAR-derived multi-temporal features, and ancillary data sources for C-band (a-b), L-band (c-d), and combined C- & L-band (e-f) datasets for Longford and Sligo.

**Figure 5** Feature importance scores of the individual grassland classes for the combined C- and L-band datasets for Longford (left) and Sligo (right). For Longford, the first 12 features are the C-band backscatter intensities (shaded), next three the L-band HH backscatter intensities followed by the two L-band HV backscatter intensities (dotted) and the four ancillary variables (diagonal hatch); soils, subsoils, elevation and slope. For Sligo, the first 15 features are the C-band backscatter intensities (shaded), next three the L-band HH backscatter intensities followed by the two L-band HV backscatter intensities (dotted) and the four ancillary variables (diagonal hatch); soils, subsoils, elevation and slope.
Figure 6 Classification results of the ERT classifier as applied to the C- and L-band SAR and ancillary dataset for Sligo (left) and Longford (right).

Figure 7 ERT Classification probabilities of dry improved and wet semi-improved grassland for Sligo and Longford. Low probabilities are shown in white with high probabilities in black.

Figure 8 Time series of daily ECV soil moisture values (solid lines) and error range (shaded) covering the two counties (Longford (a-b); Sligo (c-d)) with in situ daily accumulated precipitation from nearby Met Éireann synoptic meteorological stations. All SAR acquisition dates are marked with blue crosses.
### Table 1 SAR data characteristics

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Date</th>
<th>Width</th>
<th>Azimuth</th>
<th>Pol</th>
<th>Pass</th>
<th>Track</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERS-2</td>
<td>2008-03-23</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>D</td>
<td>80</td>
<td>2525</td>
</tr>
<tr>
<td>ERS-2</td>
<td>2008-03-27</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>A</td>
<td>144</td>
<td>1076</td>
</tr>
<tr>
<td>ASAR</td>
<td>2008-04-08</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>D</td>
<td>309</td>
<td>2526</td>
</tr>
<tr>
<td>ERS-2</td>
<td>2008-05-01</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>A</td>
<td>144</td>
<td>1076</td>
</tr>
<tr>
<td>ASAR</td>
<td>2008-06-17</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>D</td>
<td>309</td>
<td>2526</td>
</tr>
<tr>
<td>ERS-2</td>
<td>2008-07-10</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>A</td>
<td>144</td>
<td>1076</td>
</tr>
<tr>
<td>ASAR</td>
<td>2008-07-22</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>D</td>
<td>309</td>
<td>2526</td>
</tr>
<tr>
<td>ERS-2</td>
<td>2008-08-10</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>D</td>
<td>80</td>
<td>2525</td>
</tr>
<tr>
<td>ERS-2</td>
<td>2008-08-14</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>A</td>
<td>144</td>
<td>1076</td>
</tr>
<tr>
<td>ERS-2</td>
<td>2008-08-26</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>D</td>
<td>309</td>
<td>2522</td>
</tr>
<tr>
<td>ASAR</td>
<td>2008-09-14</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>D</td>
<td>80</td>
<td>2524</td>
</tr>
<tr>
<td>ERS-2</td>
<td>2008-11-23</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>D</td>
<td>80</td>
<td>2525</td>
</tr>
<tr>
<td>PALSAR FBS</td>
<td>2008-01-27</td>
<td>23.6cm</td>
<td>38°</td>
<td>HH</td>
<td>A</td>
<td>1</td>
<td>1060/1070</td>
</tr>
<tr>
<td>PALSAR FBD</td>
<td>2008-04-28</td>
<td>23.6cm</td>
<td>38°</td>
<td>HH/HV</td>
<td>A</td>
<td>1</td>
<td>1060/1070</td>
</tr>
<tr>
<td>PALSAR FBD</td>
<td>2008-06-13</td>
<td>23.6cm</td>
<td>38°</td>
<td>HH/HV</td>
<td>A</td>
<td>1</td>
<td>1060/1070</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sligo</th>
<th>Date</th>
<th>Width</th>
<th>Azimuth</th>
<th>Pol</th>
<th>Pass</th>
<th>Track</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERS-2</td>
<td>2008-02-01</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>D</td>
<td>352</td>
<td>2512</td>
</tr>
<tr>
<td>ERS-2</td>
<td>2008-03-23</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>D</td>
<td>80</td>
<td>2514</td>
</tr>
<tr>
<td>ASAR</td>
<td>2008-04-11</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>D</td>
<td>352</td>
<td>2511</td>
</tr>
<tr>
<td>ASAR</td>
<td>2008-07-09</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>D</td>
<td>123</td>
<td>2509</td>
</tr>
<tr>
<td>ASAR</td>
<td>2008-07-25</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>D</td>
<td>352</td>
<td>2511</td>
</tr>
<tr>
<td>ERS-2</td>
<td>2008-07-29</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>A</td>
<td>416</td>
<td>1086</td>
</tr>
<tr>
<td>ERS-2</td>
<td>2008-08-10</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>D</td>
<td>80</td>
<td>2514</td>
</tr>
<tr>
<td>ERS-2</td>
<td>2008-08-17</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>A</td>
<td>187</td>
<td>1084</td>
</tr>
<tr>
<td>ERS-2</td>
<td>2008-08-29</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>D</td>
<td>352</td>
<td>2512</td>
</tr>
<tr>
<td>ASAR</td>
<td>2008-09-14</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>D</td>
<td>80</td>
<td>2513</td>
</tr>
<tr>
<td>ASAR</td>
<td>2008-09-17</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>D</td>
<td>123</td>
<td>2509</td>
</tr>
<tr>
<td>ASAR</td>
<td>2008-10-03</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>D</td>
<td>352</td>
<td>2511</td>
</tr>
<tr>
<td>ASAR</td>
<td>2008-10-19</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>D</td>
<td>80</td>
<td>2506</td>
</tr>
<tr>
<td>ERS-2</td>
<td>2008-11-23</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>D</td>
<td>80</td>
<td>2514</td>
</tr>
<tr>
<td>ASAR</td>
<td>2008-12-31</td>
<td>5.6cm</td>
<td>23°</td>
<td>VV</td>
<td>D</td>
<td>123</td>
<td>2509</td>
</tr>
<tr>
<td>PALSAR FBS</td>
<td>2008-03-01</td>
<td>23.6cm</td>
<td>38°</td>
<td>HH</td>
<td>A</td>
<td>3</td>
<td>1070/1080</td>
</tr>
<tr>
<td>PALSAR FBD</td>
<td>2008-06-01</td>
<td>23.6cm</td>
<td>38°</td>
<td>HH/HV</td>
<td>A</td>
<td>3</td>
<td>1070/1080</td>
</tr>
<tr>
<td>PALSAR FBD</td>
<td>2008-07-17</td>
<td>23.6cm</td>
<td>38°</td>
<td>HH/HV</td>
<td>A</td>
<td>3</td>
<td>1070/1080</td>
</tr>
<tr>
<td>Level 0</td>
<td>Level 1</td>
<td>Level 2</td>
<td>Level 3</td>
<td>Description</td>
<td>#Longford</td>
<td>#Sligo</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reclaimed</td>
<td></td>
<td>Highly managed pasture over peats or heavy gleys</td>
<td>303</td>
<td>539</td>
<td></td>
</tr>
<tr>
<td>Semi-improved</td>
<td>Wet [GSw]</td>
<td>Humid</td>
<td></td>
<td>Grassland on poorly drained soils, low management, rushes and/or sedges</td>
<td>320</td>
<td>597</td>
<td></td>
</tr>
<tr>
<td>Grassland</td>
<td></td>
<td></td>
<td></td>
<td>often present</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dry [GSd]</td>
<td></td>
<td>Semi-improved dry grassland over acid soils</td>
<td>229</td>
<td>191</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Calcareous</td>
<td></td>
<td>Semi-improved dry grassland over basic soils/limestone karst</td>
<td>242</td>
<td>393</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Forest</td>
<td></td>
<td></td>
<td>Areas dominated by trees and woody vegetation (minimum 20% canopy closure</td>
<td>843</td>
<td>1023</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>and area of 0.5ha)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Settlement</td>
<td></td>
<td></td>
<td>All developed land, including transportation infrastructure and human</td>
<td>379</td>
<td>631</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>settlements</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Water</td>
<td></td>
<td></td>
<td>Includes all bodies of permanent fresh and saltwater.</td>
<td>785</td>
<td>612</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Peatland</td>
<td></td>
<td></td>
<td>Raised bogs, blanket bogs and cutover bog</td>
<td>1280</td>
<td>1111</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cropland</td>
<td></td>
<td></td>
<td>Arable and tillage land</td>
<td>459</td>
<td>165</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total</td>
<td>5202</td>
<td>6442</td>
<td></td>
</tr>
</tbody>
</table>
**Table 3 Grassland Descriptions.**

<table>
<thead>
<tr>
<th>Grassland Class</th>
<th>Species Composition (common name – botanical name)</th>
<th>Management Regime</th>
<th>Approximate Height Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA Dry</td>
<td>Perennial ryegrass (<em>Lolium perenne</em>) White Clover (<em>Trifolium repens</em>) Timothy (<em>Phleum pratense</em>)</td>
<td>Intensively managed – typically reseeded frequently, heavily fertilised and grazed or cut for silage. Generally high stocking densities.</td>
<td>3 - 20cm (post grazing - ) pre-grazing</td>
</tr>
<tr>
<td>GA Reclaimed</td>
<td>Italian ryegrass (<em>Lolium multiflorum</em>) Meadow Fescue (<em>Festuca pratensis</em>) Meadow foxtail (<em>Alopecurus pratensis</em>) Dandelions (<em>Taraxacum spp.</em>) Nettle (<em>Urtica dioica</em>) Thistle (<em>Cirsium arvense</em>)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSD Humid</td>
<td>Bents (<em>Agrostis spp.</em>) Fescues (<em>Festuca spp.</em>) Sweet vernal-grass (<em>Anthoxanthum odoratum</em>) Wavy hair grass (<em>Deschampsia flexuosa</em>) Mat grass (<em>Nardus stricta</em>)</td>
<td>Extensively managed</td>
<td>10 – 80cm</td>
</tr>
<tr>
<td>GSD Calcareous</td>
<td>Rigwort plantain (<em>Plantago lanceolata</em>) Cockfoot (<em>Dactylis glomerata</em>) Bents (<em>Agrostis spp.</em>) Meadow grasses (<em>Poa spp.</em>) Timothy (<em>Phleum pratense</em>) Fescues (<em>Festuca spp.</em>) Red Clover (<em>Trifolium pratense</em>) Blue moor grass (<em>Sesleria caerulea</em>) Downy Oat-grass (<em>Avenula pubescens</em>) Yellow oat-grass (<em>Trisetum flavescens</em>) Quaking grass (<em>Briza media</em>)</td>
<td>Extensively managed</td>
<td>10 – 60cm</td>
</tr>
</tbody>
</table>
Table 4 C-, L-, and combined C/L-band classification results for Longford and Sligo (RF=Random Forest, SVM=Support Vector Machines, ERT=Extremely Randomised Trees, PA=producer’s accuracy, UA=user’s accuracy). v0 indicates the initial classification where all input variables were included, v1 indicates the classification after the least important variables were excluded.

<table>
<thead>
<tr>
<th>Class</th>
<th>Longford</th>
<th>Sligo</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>v0_ld</td>
<td>v1_ld</td>
</tr>
<tr>
<td></td>
<td>RF PA UA</td>
<td>SVM</td>
</tr>
<tr>
<td>C-band</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSdh</td>
<td>0.92 0.96</td>
<td>1 0.88</td>
</tr>
<tr>
<td>GSdc</td>
<td>0.97 0.97</td>
<td>1 0.93</td>
</tr>
<tr>
<td>GDa</td>
<td>0.92 0.95</td>
<td>0.89 0.86</td>
</tr>
<tr>
<td>GAr</td>
<td>0.90 0.83</td>
<td>0.81 0.77</td>
</tr>
<tr>
<td>GSsw</td>
<td>0.92 0.82</td>
<td>0.85 0.75</td>
</tr>
<tr>
<td>OA</td>
<td>95.7% 1</td>
<td>93.7%</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.95</td>
<td>0.93</td>
</tr>
<tr>
<td>L-band</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSdh</td>
<td>0.87 0.74</td>
<td>0.74 0.72</td>
</tr>
<tr>
<td>GSdc</td>
<td>0.89 0.95</td>
<td>0.83 0.79</td>
</tr>
<tr>
<td>GDa</td>
<td>0.91 0.85</td>
<td>0.79 0.77</td>
</tr>
<tr>
<td>GAr</td>
<td>0.86 0.65</td>
<td>0.66 0.50</td>
</tr>
<tr>
<td>GSsw</td>
<td>0.81 0.70</td>
<td>0.69 0.55</td>
</tr>
<tr>
<td>OA</td>
<td>92.4% 1</td>
<td>87.7%</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.91</td>
<td>0.86</td>
</tr>
<tr>
<td>C/L-band</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSdh</td>
<td>0.96 0.94</td>
<td>0.90 0.89</td>
</tr>
<tr>
<td>GSdc</td>
<td>0.96 0.98</td>
<td>0.91 0.95</td>
</tr>
<tr>
<td>GDa</td>
<td>0.95 0.95</td>
<td>0.92 0.84</td>
</tr>
<tr>
<td>GAr</td>
<td>0.91 0.86</td>
<td>0.81 0.80</td>
</tr>
<tr>
<td>GSsw</td>
<td>0.91 0.87</td>
<td>0.87 0.83</td>
</tr>
<tr>
<td>OA</td>
<td>97.2% 1</td>
<td>95.3%</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.97</td>
<td>0.94</td>
</tr>
</tbody>
</table>
Table 5 Comparison of different classification results with and without ancillary datasets

<table>
<thead>
<tr>
<th></th>
<th>C-band</th>
<th>L-band</th>
<th>C/L-band</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Longford</td>
<td>Sligo</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RF</td>
<td>SVM</td>
<td>ERT</td>
</tr>
<tr>
<td></td>
<td>OA</td>
<td>k</td>
<td>OA</td>
</tr>
<tr>
<td>a) All Data</td>
<td>96.9% 0.96</td>
<td>97.1% 0.97</td>
<td>97.9% 0.98</td>
</tr>
<tr>
<td>b) No Soils</td>
<td>94.8% 0.94</td>
<td>91.5% 0.90</td>
<td>95.8% 0.95</td>
</tr>
<tr>
<td>c) No Elevation</td>
<td>93.5% 0.92</td>
<td>93.9% 0.93</td>
<td>93.9% 0.93</td>
</tr>
<tr>
<td>d) Radar only</td>
<td>87.2% 0.85</td>
<td>83.3% 0.80</td>
<td>88.5% 0.87</td>
</tr>
<tr>
<td>a) All Data</td>
<td>95.3% 0.94</td>
<td>93.5% 0.92</td>
<td>95.8% 0.95</td>
</tr>
<tr>
<td>b) No Soils</td>
<td>89.4% 0.87</td>
<td>80.2% 0.77</td>
<td>89.8% 0.88</td>
</tr>
<tr>
<td>c) No Elevation</td>
<td>88.3% 0.86</td>
<td>86.1% 0.83</td>
<td>89.0% 0.87</td>
</tr>
<tr>
<td>d) Radar only</td>
<td>76.2% 0.72</td>
<td>57.3% 0.49</td>
<td>76.9% 0.73</td>
</tr>
<tr>
<td>a) All Data</td>
<td>98.0% 0.98</td>
<td>97.9% 0.98</td>
<td>98.7% 0.98</td>
</tr>
<tr>
<td>b) No Soils</td>
<td>96.9% 0.96</td>
<td>96.9% 0.96</td>
<td>97.5% 0.97</td>
</tr>
<tr>
<td>c) No Elevation</td>
<td>96.7% 0.96</td>
<td>97.0% 0.96</td>
<td>97.4% 0.97</td>
</tr>
<tr>
<td>d) Radar only</td>
<td>93.7% 0.93</td>
<td>92.9% 0.92</td>
<td>95.1% 0.94</td>
</tr>
<tr>
<td>Ancillary Data</td>
<td>88.5% 0.86</td>
<td>84.4% 0.81</td>
<td>87.3% 0.85</td>
</tr>
</tbody>
</table>
Figure 2

(a) Longford

(b) Sligo
Figure 3

(a) C-band - Longford

(b) C-band - Sligo

(c) L-band HH - Longford

(d) L-band HV - Longford

(e) L-band HH - Sligo

(f) L-band HV - Sligo

Legend:
- GSdh
- GSdc
- GSsw
- GAd
- GAr
Figure 4

(a) C-band Longford
\(n=46\)

(b) C-band Sligo
\(n=59\)

(c) L-band Longford
\(n=29\)

(d) L-band Sligo
\(n=29\)

(e) C/L-band Longford
\(n=71\)

(f) C/L-band Sligo
\(n=84\)
Figure 6
Figure 8