NRC Research Press

ARTICLE

Enhancing the Irish NFI using *k*-nearest neighbors and a genetic algorithm

Daniel McInerney, Frank Barrett, Ronald E. McRoberts, and Erkki Tomppo

Abstract: This paper presents a nationwide application of *k*-nearest neighbors (*k*-NN) to estimate growing stock volume per hectare for the Irish National Forest Estate using optical satellite imagery and field inventory data from the second National Forest Inventory (NFI). Two approaches are tested: an unweighted *k*-NN and an improved version (i*k*-NN) that is optimised using a genetic algorithm. The performance of the models is assessed in terms of the root mean square error (RMSE) and prediction error. From the simulations, it was found that the optimal value of *k* was 3, and the smallest pixel-level RMSE for growing stock was 126 $m^3 \cdot ha^{-1}$ when i*k*-NN was used. Comparisons with estimates from the NFI show that the i*k*-NN technique can enhance the Irish NFI. These improvements include a total estimate of growing stock volume of 102 million m^3 with a confidence interval of ±3%, which is smaller than the NFI-reported confidence interval of ±5%. In addition, while total county-level estimates of growing volume estimated using i*k*-NN were consistent with those published from the NFI, their corresponding confidence intervals were much narrower, in the range of a two- to four-fold reduction in the width of the confidence interval.

Key words: forest inventory, remote sensing, *k*-NN, nearest neighbors, genetic algorithm.

Résumé : Cet article présente une application à l'échelle nationale de la méthode des *k* plus proches voisins (*k*-NN) pour estimer le volume sur pied par hectare pour le domaine forestier national irlandais à l'aide d'imagerie par satellite optique et des données d'inventaire sur le terrain du second Inventaire forestier national (IFN). Deux approches sont testées : une méthode *k*-NN non pondérée et une version améliorée (i*k*-NN) optimisée à l'aide d'un algorithme génétique. La performance des modèles est évaluée en termes d'erreur quadratique moyenne (EQM) et d'erreur de prédiction. Les simulations ont montré que la valeur optimale de *k* était de 3 et, qu'à l'échelle du pixel, l'EQM la plus faible pour le volume sur pied était de 126 $m^3 \cdot ha^{-1}$ et était obtenue avec la méthode i*k*-NN. Les comparaisons avec les estimations de l'IFN montrent que la technique i*k*-NN peut améliorer l'IFN irlandais. Ces améliorations comprennent une estimation totale du volume sur pied de 102 millions de m^3 avec un intervalle de confiance de ±3 %, ce qui est inférieur à l'intervalle de confiance de ±5 % associé à l'IFN. De plus, bien que les estimations totales du volume sur pied faites avec l'i*k*-NN à l'échelle des comtés soient cohérentes avec celles publiées par l'IFN, leurs intervalles de confiance respectifs étaient beaucoup plus étroits, avec une réduction de la largeur de l'intervalle de confiance de l'ordre de deux à quatre fois. [Traduit par la Rédaction]

Mots-clés : inventaire forestier, télédétection, *k*-NN, plus proches voisins, algorithme génétique.

1. Introduction

National forest statistics are primarily estimated from field-based National Forest Inventory (NFI) programmes that collect information on a range of forest-related parameters (Lawrence et al. 2010). However, they are typically based on a percentage sampling, with sample intensities commonly determined by forest area. These methods are employed as complete enumeration of all forests is not feasible due to cost and time. The sampling intensities employed in operational inventories are too small to provide accurate representations of the forest resources at local scales (Bottai et al. 2002; Tomppo et al. 2014). Furthermore, there are specific limitations associated with large-scale field inventories as they cannot identify within-stand variability caused by differences in site productivity or active forest management operations such as thinning or selective felling (Gjertsen 2007). Given that the intention of these inventories is to serve strategic purposes and national reporting, they are not designed to produce statistically robust estimates of forest resources at local scales despite the increasing need for such information.

Research carried out in Finland during the early 1990s sought to overcome this issue and to leverage multispectral satellite imagery, field inventory data, and a variety of geospatial data in conjunction with a statistical estimator to produce small-area estimates (several thousands of hectares) of forest parameters (Tomppo 1991). This research was primarily driven by the need to produce forest resource information for small administrative areas (i.e., municipalities) in a more cost-effective way than would be possible from the sole use of field inventory data. Since then, the approach has been routinely applied in Finland and Sweden to update forestry databases pertaining to national and regional forest inventories (Tomppo et al. 2008; Barrett et al. 2016). McRoberts (2012) provides a comprehensive overview of where *k*-nearest neighbors (*k*-NN) has been applied in the context of forest inventory and identifies four principal areas where *k*-NN has been used,

Received 11 January 2018. Accepted 18 September 2018.

D. McInerney. Coillte Forest, Castletroy, Limerick, Ireland.
F. Barrett. Department of Agriculture, Food and the Marine, Johnstown Castle Estate, Co. Wexford, Ireland.
R.E. McRoberts. Northern Research Station, U.S. Forest Service, 1992 Folwell Avenue, Saint Paul, MN 55038, USA.
E. Tomppo. University of Helsinki, Department of Forest Sciences, P.O. Box 27, 00014 University of Helsinki, Finland; Aalto University, Department of Electronics and Nanoengineering, P.O. Box 11000, 00076, AALTO, Finland.
Corresponding author: Daniel McInerney (email: daniel.mcinerney@coillte.ie).

Copyright remains with the author(s) or their institution(s). Permission for reuse (free in most cases) can be obtained from RightsLink.

Published at www.nrcresearchpress.com/cjfr on 25 September 2018.

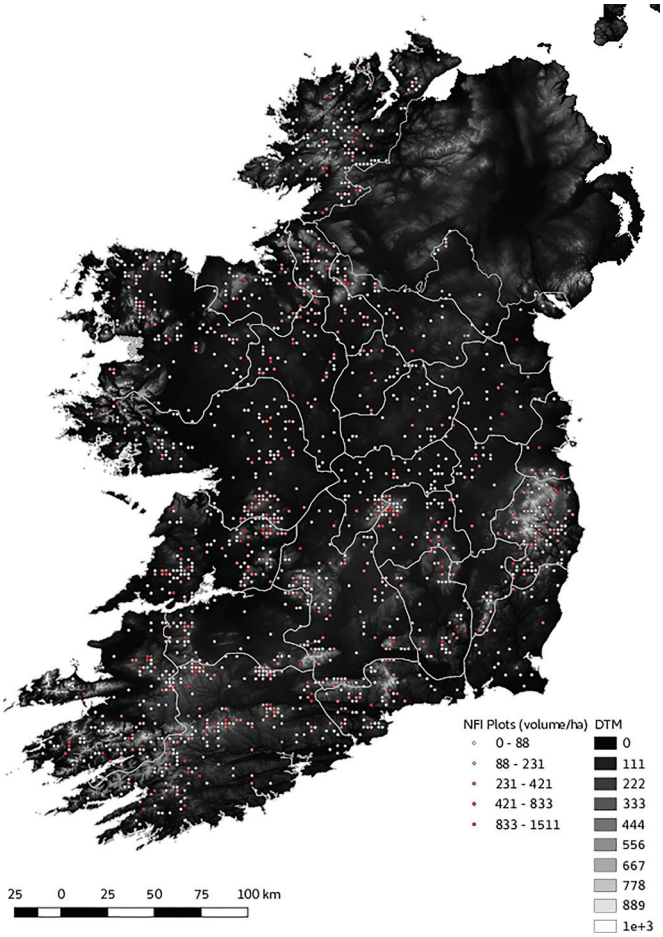
which include (*i*) imputation of missing forest inventory variables, (*ii*) forest mapping, (*iii*) small-area estimation, and (*iv*) probability and model-based inference. While *k*-NN remains the technique most widely applied in forestry, other variations of this nonparametric technique have also been used in a forestry context, e.g., most similar neighbor (MSN) (Moeur and Stage 1995), gradient nearest neighbor (GNN) (Ohmann and Gregory 2002), and the random forest algorithm (McInerney and Nieuwenhuis 2009). A detailed meta-analysis and literature review of these techniques is covered in Chirici et al. (2016).

The success of *k*-NN and the fact that it has been widely used to produce current estimates of forest resource information can be largely attributed to the following advantages. Firstly, *k*-NN is practical and statistically oriented in that it can simultaneously predict a wide range of forest variables (Mäkelä and Pekkarinen 2004). Secondly, it is nonparametric and does not require any a priori information regarding the distribution of the data. Thirdly, it is cost-effective in that it can generate more detailed statistics for small areas and map products than can be achieved using field data alone. Fourthly, it is a versatile technique that can be used with different datasets and in different geographic locations (Chirici et al. 2016).

With a forest cover of slightly more than 10.5%, reference has often been made to Ireland being one of the least forested countries in Europe. At the beginning of the 20th century, forest cover was less than 1.4%, but since the 1950s, Ireland's rate of afforestation has been one of the greatest in the EU, and its current government policy is to increase national forest cover to 18% by 2046 (Department of Agriculture, Food and the Marine (DAFM) 2014). Planting of privately owned lands has been driven by EU and State grant payments (Zanchi et al. 2007). Under such programmes, the area of privately owned forests has increased from 81 958 ha in 1973 to 335 900 ha in 2012, over a four-fold increase. During the same period, the state-owned forest area has also increased substantially from 242 056 ha to 395 760 ha (DAFM 2018*a*). The average size of private afforestation sites since 1980 to 2016 was 8.8 ha, resulting in the development of a fragmented forest landscape with patterns of spatial development related to existing agricultural activities, availability of grant aid, and environmental constraints precluding forest development (Gillmore 1988). These developments, coupled with the current structure of the Irish Forest Estate, are almost unique and present challenges in terms of how the Forest Estate is measured and how robust estimates are produced using i*k*-NN, a variation of *k*-NN that is optimized using genetic algorithms (sections 2.6 and 2.7).

There are many examples of where *k*-NN has been successfully implemented as part of an NFI programme, for instance in Minnesota (Franco-Lopez et al. 2001), Sweden (Reese et al. 2005), and Italy (Chirici et al. 2008), while additional examples are provided in Chirici et al. (2016). However, one common feature of many of the published examples of *k*-NN is that they have largely been carried out in countries characterised by vast areas of large percentage forest cover consisting of a uniform forest structure in terms of age class distribution. In this study, we investigate the use of i*k*-NN to increase the accuracy and precision of Irish NFI estimates in a landscape dominated by forests at a juvenile development stage and characterised by a very sparse and fragmented forest cover, which are largely managed on short (35–45 year) rotations. In addition, there is no standard set of configuration parameters that can be applied to different forest environments, and each implementation of *k*-NN requires optimization to local conditions and selection of different explanatory variables and estimation parameters (distance metric, value of *k*, selection of weights). Furthermore, this study provides an additional example to a small set of studies that have applied the *k*-NN approach but optimised it through the use of a genetic algorithm and coarse-scale variables. Specific examples of this approach include the Finnish NFI (Tomppo and Halme 2004) and for the prediction of categorical forest variables (Tomppo et al. 2009). The particular challenge in Ireland is the small forest percentage and scattered forest patches, which in turn implies that the coarse-scale variables as explanatory variables play an even more important role. Furthermore, there are few operational examples of *k*-NN that have applied a detailed error estimation approach, which is also presented in this article.

Fig. 1. Spatial distribution of the NFI field plots overlaid on the digital terrain model (DTM). [Colour version available online.]



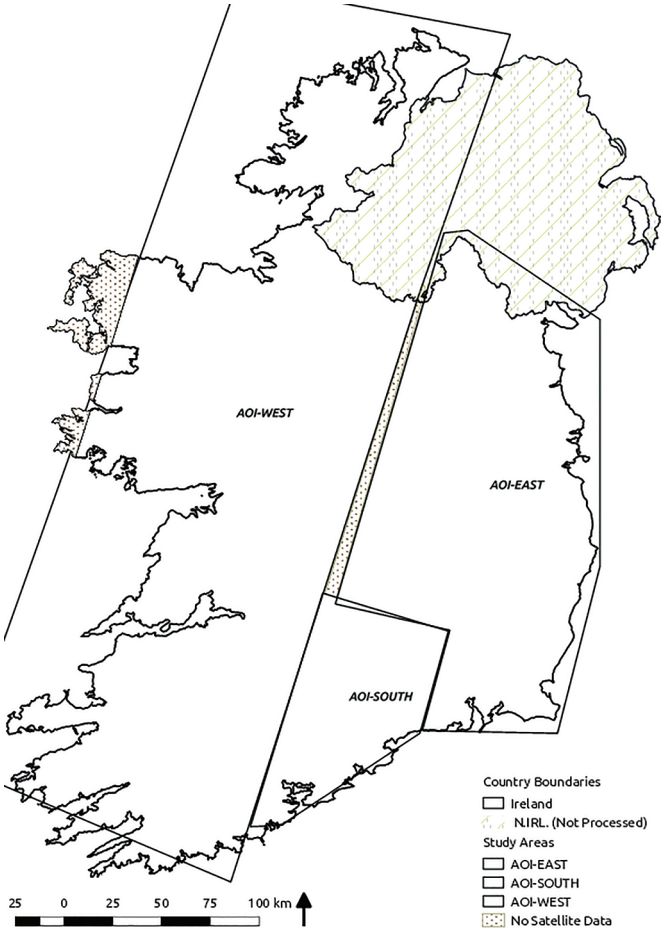
The objective of this paper is to present an operational use of i*k*-NN, which is optimised using a genetic algorithm and coarse-scale variables and a detailed error estimation technique. A related objective is to clearly describe an operational technique applied in a novel way to a unique national Forest Estate. The paper builds on previous studies carried out in Ireland (McInerney et al. 2005; McInerney and Nieuwenhuis 2009). However, the paper extends these studies, as well as many others, by presenting a detailed description of the methods and data required to carry out this type of analysis on an operational scale across a large area.

2. Materials

2.1. Study area

The analysis focuses on the Irish Forest Estate, including both public and privately owned stands, with specific reference to the country's counties (Fig. 1). For the purposes of the processing and availability of satellite imagery, the country has been divided into three areas of interest (AOI) (Fig. 2). The total forest area in 2013 has been estimated as 731 652 ha, which represents 10.5% of the total land area. The Irish Forest Estate consists largely of plantation forests managed on commercial rotations ranging in length

Fig. 2. Map detailing the extents of the three AOIs analysed as part of this research. In addition, areas unprocessed due to lack of satellite data are also mapped. [Colour version available online.]



from 35 to 45 years. The majority (more than 60%) of them are plantations consisting of either monocultures or mixtures of two coniferous species, with Sitka spruce (*Picea sitchensis* (Bong.) Carr.) and lodgepole pine (*Pinus contorta* Douglas ex Loudon) being the dominant species. The Forest Estate is very young, with two-thirds of the forests less than 20 years of age. Approximately 31% of this juvenile-stage forest is held in private ownership, which is spatially fragmented with the private plantation average size being 8.8 ha (DAFM 2018*a*). Despite these characteristics, there is a substantial quantity of growing stock present in older broadleaf forests, and increasingly, a more diverse range of species is being planted by Irish forest owners (Forest Service 2013*b*).

A small portion of the Forest Estate has been excluded from the analysis due to a lack of satellite imagery or cloud cover; the corresponding areas are 6234 ha and 9256 ha, respectively, which represents 2.1% of the total Forest Estate. While these areas are not processed, it is considered that their overall contribution to the estimates are minimal given that more than 60% of this area is less than 20 years of age.

2.2. Applying best practices in the use of plot-based NFI data and Earth observation data

Based on questionnaire responses from NFIs representing 65% of global forest area, Barrett et al. (2016) provided specific guidelines and identified best practices for countries initiating, expanding, or revising the use of remotely sensed data in their NFIs, including the following recommendations.

1. In some instances, it may be required to remove non-applicable field plot data caused by differences between the date of field data collection and the acquisition of satellite imagery. Plot-level disturbances caused by biotic or abiotic factors between the two dates can cause large errors in estimates and map products. Plots with these issues should be either removed or updated.
2. Adequate handling of positional error in regard to georectified imagery is of critical importance and appropriate postprocessing of GPS data should be carried out to reduce these errors.
3. Range of variation in both response and remotely sensed variables in the sample or reference data should cover ranges in the same variable in the entire inventory area of interest.
4. Map predictions should be done simultaneously and in a consistent manner using a multivariate method such as the *k*-NN technique.
5. From a technical perspective, the input data must provide the required accuracy for estimates using the selected analytical methods.
6. Uncertainty in forest attribute maps and map-assisted estimates should be estimated using statistically rigorous methods.
7. Map-based estimates should be adjusted to compensate for prediction errors (typically by comparing map predictions and ground data for a probability sample using techniques such as the model-assisted, generalized regression estimator).

These recommendations for best practices were applied in the use of NFI plot data, satellite imagery, auxiliary geospatial data, model diagnostics, a genetic algorithm, and the i*k*-NN method as discussed further below.

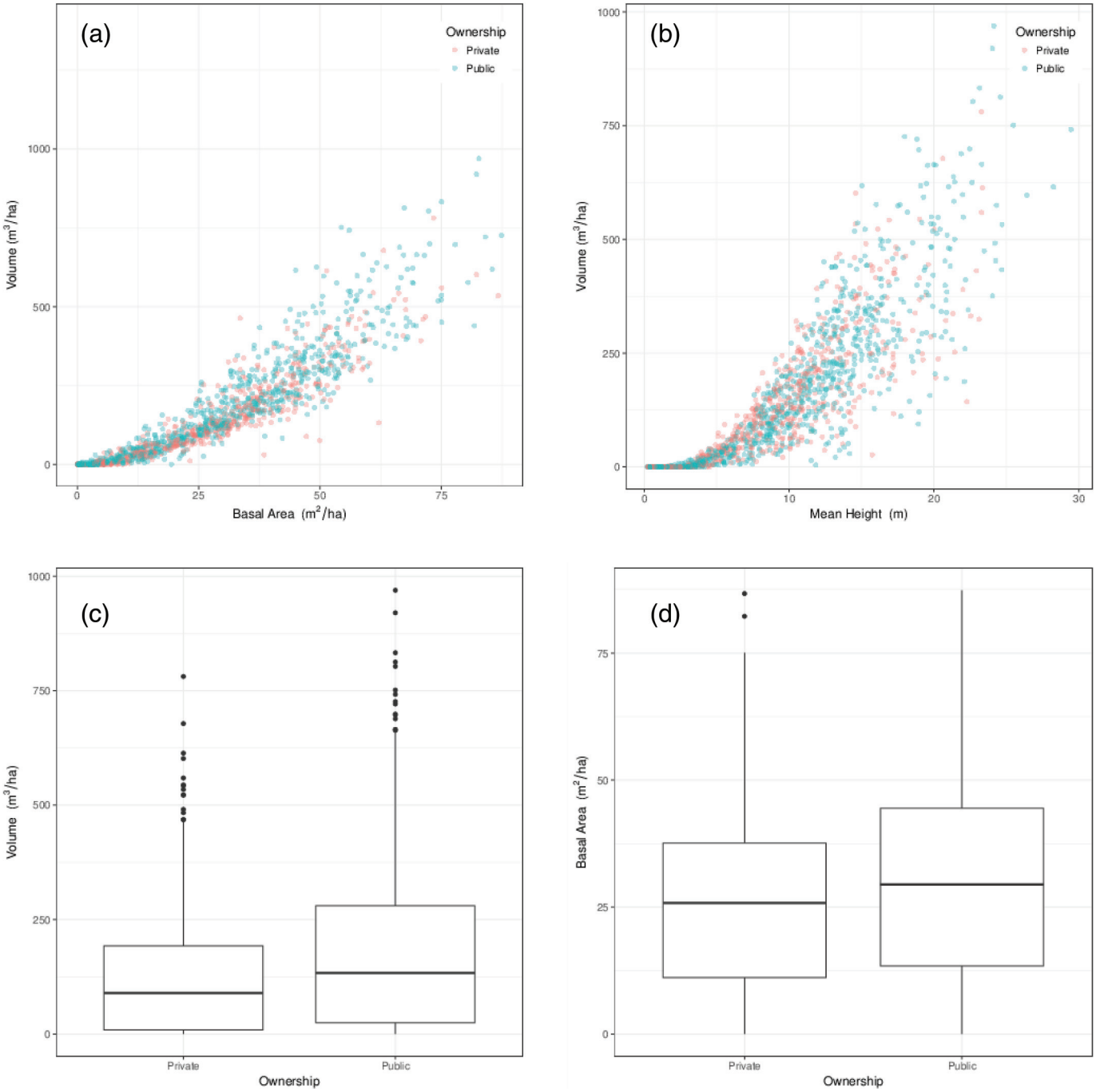
2.3. National Forest Inventory data

Field inventory data from the 2nd NFI were used as a reference dataset in the *k*-NN estimation. The NFI is based on a sample of all forests using a 2 × 2 km systematic sampling design across the entire country with the inventory designed for a 5-year rotation. It consists of 1747 permanent sample plots located on forest lands (Forest Service 2013*a*). The field data were collected between 2010 and 2012. The plot design is a set of concentric circular plots with a maximum radius of 12.62 m centred on the plot coordinate encompassing a total area of 500 m^2. The plot centres are located using a differential GPS, which also used real-time RINEX correction of the signals. It was estimated that the average location error was less than 5 m. In situations where there was insufficient GPS signal, laser range finders were used to navigate from known GPS coordinates to the plot centre. A total of 105 timber and non-timber variables is assessed on each plot. The primary focus of our analysis is on plot-level estimates of growing stock volume per hectare ($m^3 \cdot ha^{-1}$).

For the purposes of this study, estimates were only produced for stocked forests. As a result, NFI plots that were temporarily unstocked (i.e., clearfelled) were re-labelled as non-forest plots and not used as part of the analysis. This was done to ensure that estimates were only produced for areas that were stocked with forests; clearfelled and temporarily unstocked forest lands are still considered forest from a land-use perspective, but they will clearly be identified as non-forest from a landcover or remote sensing context. Furthermore, field inventory plots that had a volume of more than 50 m^3 but a normalised difference vegetation index (NDVI, an indicator of green vegetation) of less than 0.25 were considered anomalies and were removed from the reference dataset. This represented a total of 22 plots and was likely caused by a clearfell intervention during the intervening period between the field inventory survey and the acquisition of the satellite data. The spatial distribution of the NFI plots is presented in Fig. 1.

Estimates of growing stock volume per hectare and basal area per hectare were estimated using the Irish single tree volume

Fig. 3. NFI-2 scatterplots of (*a*) basal area by volume and (*b*) mean height on plot by volume. Boxplot distributions of (*c*) growing stock volume per hectare and (*d*) basal area per hectare by ownership.



models that were developed in conjunction with the NFI (Forest Service 2013*a*). Uncertainty relating to the allometric model predictions is acknowledged but has not been actively considered as part of this study as they are deemed negligible relative to the effects of sampling variability, which has been demonstrated in similar types of studies (McRoberts and Westfall 2014). However, there is a growing amount of literature on this topic of attempting to quantify this difference (Breidenbach et al. 2014), but it is considered out of scope for this paper. Figure 3 presents the scatterplots of growing stock volume per hectare, basal area per hectare, and mean plot height for both public and private plots. In addition, boxplot figures of volume and basal area per hectare are presented and show that the respective means for the public Forest Estate are greater than for the private estate, which is expected as the publicly owned Forest Estate is older.

2.4. Satellite imagery

Optical satellite imagery from SPOT-4 and Indian Resource Satellite (IRS LISS 3) acquired as part of the IMAGE2012 programme was used as the source of the explanatory variables in this analysis. The images were all acquired during the spring–summer of 2011 or 2012, which corresponded with the acquisition of the NFI field data. Each satellite image scene covered a specific part of the country and therefore the images were processed separately. In the case of the IRS data acquired in the western half of the country, three scenes from the same path were acquired on the day and

Table 1. Spectral bands, ground sampling distance (GSD), and the swath of satellite sensors.

Data source	Spectral bands	GSD (m)	Swath (km)
IRS-P6 LISS-3	Four spectral bands		141
	Band 1 (green) (0.52–0.59 μm)	23.5	
	Band 2 (red) (0.62–0.68 μm)	23.5	
	Band 3 (NIR) (0.77–0.86 μm)	23.5	
	Band 4 (mid-IR) (1.55–1.70 μm)	23.5	
SPOT-4	Four spectral bands		60
	Band 1 (green) (0.5–0.59 μm)	20	
	Band 2 (red) (0.61–0.68 μm)	20	
	Band 3 (NIR) (0.78–0.89 μm)	20	
	Band 4 (SWIR) (1.58–1.75 μm)	20	

Note: IR, infrared; NIR, near infrared; SWIR, short-wave infrared.

were mosaiced together and treated as one scene. Therefore, only NFI plots located within a scene were used for the i*k*-NN analysis on that scene. Table 1 outlines the specific properties of each satellite sensor with respect to its spectral bands and spatial resolution.

The paper refers to three distinct areas of interest, which correspond to the three separate image datasets (Fig. 4):

1. AOI West — three IRS scenes from the same path and acquired on 29 April 2011;
2. AOI South — one SPOT scene acquired on 20 June 2012;
3. AOI East — two IRS scenes acquired on 27 May 2012.

A mask of cloud and cloud shadow was manually digitised to remove cloud-affected forest areas from the analysis; the cumulative area of the cloud mask affecting forest land was 9256 ha.

In addition to the spectral bands, band ratios were calculated as explanatory variables. IRS LISS-III and SPOT-4 both have four spectral bands as outlined in Table 1. Practical tests have shown that the use of the band ratios, in addition to the original bands, assists in distinguishing among forest types and tree species. Band ratios are similar to vegetation indices, and for the purposes of this study, the following sets of band ratios were calculated: 2/1, 3/1, 4/1, 3/2, 4/2, and 4/3.

2.5. Auxillary data

A number of auxillary datasets were also generated and used as explanatory variables for the weighted *k*-NN estimation. Details regarding these datasets are outlined below.

- Digital terrain model (DTM) with a spatial resolution of 50 m from which the slope, aspect, and angle were calculated. This DTM was produced by the Ordnance Survey Ireland (OSi) using photogrammetry and their stereo orthophotography campaigns from 2004.
- The Irish Forest Service's National forest map from 2012 was used as a forest mask as part of this analysis. This vector-based dataset was estimated from a combination of an automated classification of satellite imagery and the on-screen interpretation of Landsat TM imagery (1993–1997), panchromatic orthophotos (1995), and the Ordnance Survey topographic maps. Forest boundaries were digitised using orthophotos (to within 2 m accuracy), and where appropriate, the boundaries observed on the Ordnance Survey topographic maps (map series referred to as the OSi 25 in. series). This national forest map has since been updated for newly planted areas on an annual basis using digital orthophotos available to the Irish Forest Service. The resulting inventory database includes all forest areas of more than 0.2 ha with a classification accuracy of 88% for the forest species and age classes comprising the database (Gallagher et al. 2000; DAFM 2018*b*).
- An indicative soils map of Ireland was produced and published in 2006 at a working scale of 1:150 000 by the Spatial Analysis Unit (SAU) within Teagasc as part of the Subsoils, Land Cover, Habitat and Indicative Soil Mapping/Modelling Project funded initially by the Irish Forest Service and subsequently by the Department of Environment, Heritage and Local Government (Fealy et al. 2009).
- Accumulated temperature across the country (Black et al. 2010). Baseline and future climate data for Ireland have been calculated from simulations of future climate scenarios (Intergovernmental Panel on Climate Change (IPCC) scenarios A2 and B1). The simulations are from a regional climate model (RCM) developed by the Rossby Centre in Sweden (McGrath et al. 2005). The data have been calculated using a dynamic downscaling method, published by the Community Climate Change Consortium for Ireland (C4I) and validated using back-casting techniques (McGrath et al. 2005). The simulated daily mean temperatures, daily total rainfall, and daily total evaporation have been compiled into mean monthly values, while the accumulated temperature grid at 1 km resolution as used in this study was calculated for the growing season of March to October inclusive (Ray 2002).
- Coarse-scale variables of growing stock volume per hectare were generated for the three main forest type groups, which included conifer-dominated forests, broadleaf-dominated forests, and mixed forests. The datasets were generated by interpolating the NFI field plot estimates of growing stock using

Fig. 4. Extent of satellite image scenes used for the nationwide i*k*-NN estimation: IRS LISS-III (AOIs West and South) and SPOT-4 (AOI East). Areas in white within the Irish land boundary are excluded from the analysis as no imagery was readily available for these areas. [Colour version available online.]

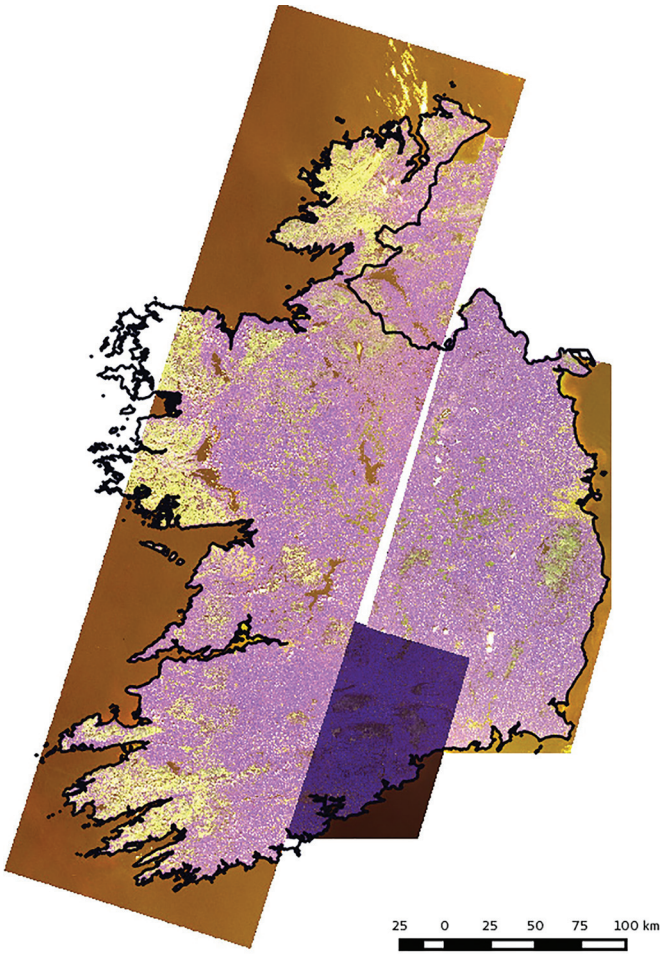
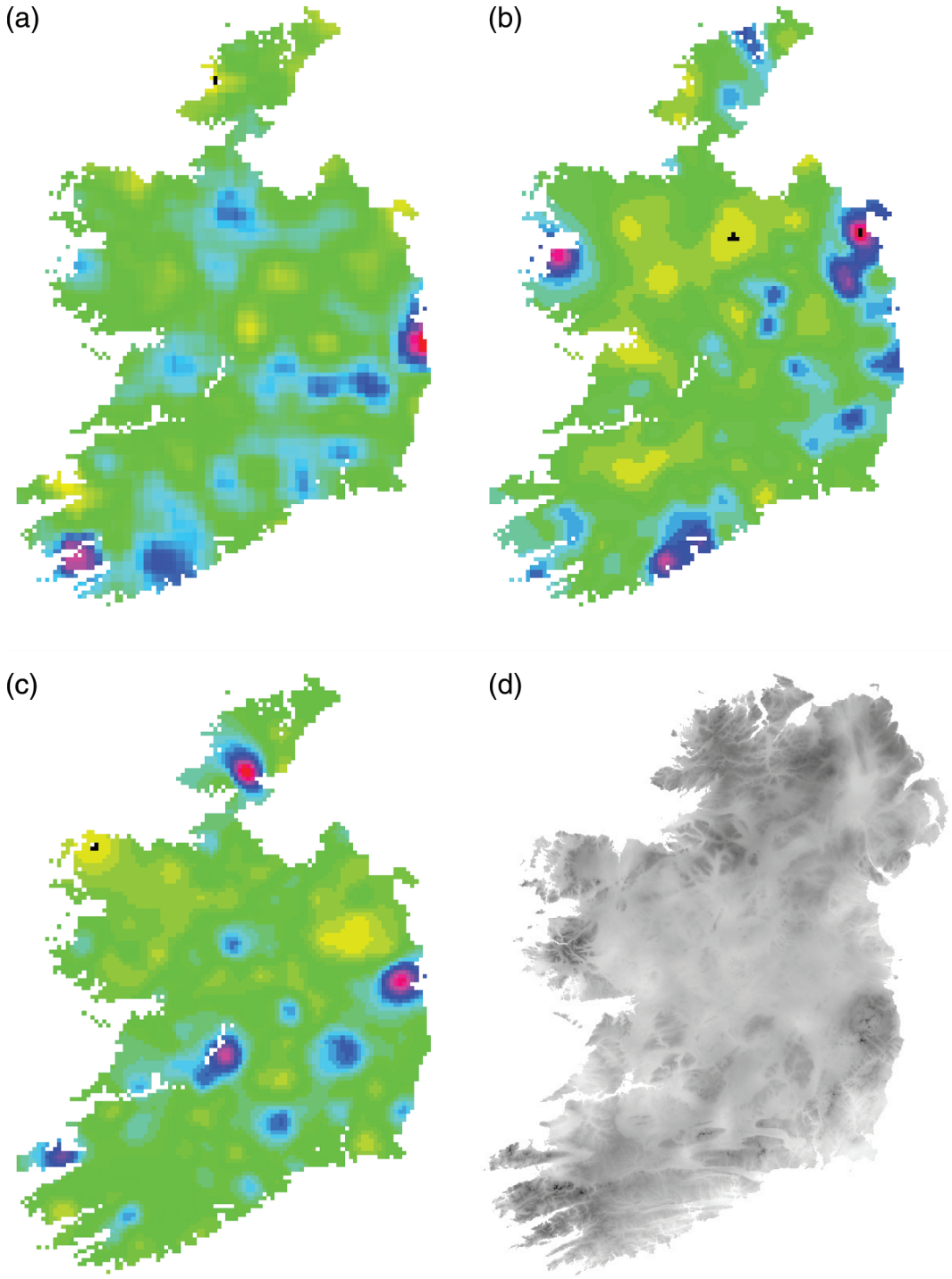


Fig. 5. Summary of coarse-scale variables used by the genetic algorithm for the optimisation of explanatory variable weights: (*a*) conifer forest volume; (*b*) broadleaf forest volume; (*c*) mixed forest volume; and (*d*) accumulated temperature day degrees.



inverse distance weighting and kernel filtering. Due to the sparseness of the field plots, the coarse-scale variables represent the variation of growing stock at the scale of tens or hundreds of kilometres. These coarse-scale variables assist in directing the search of the nearest neighbors in the i*k*-NN estimation.

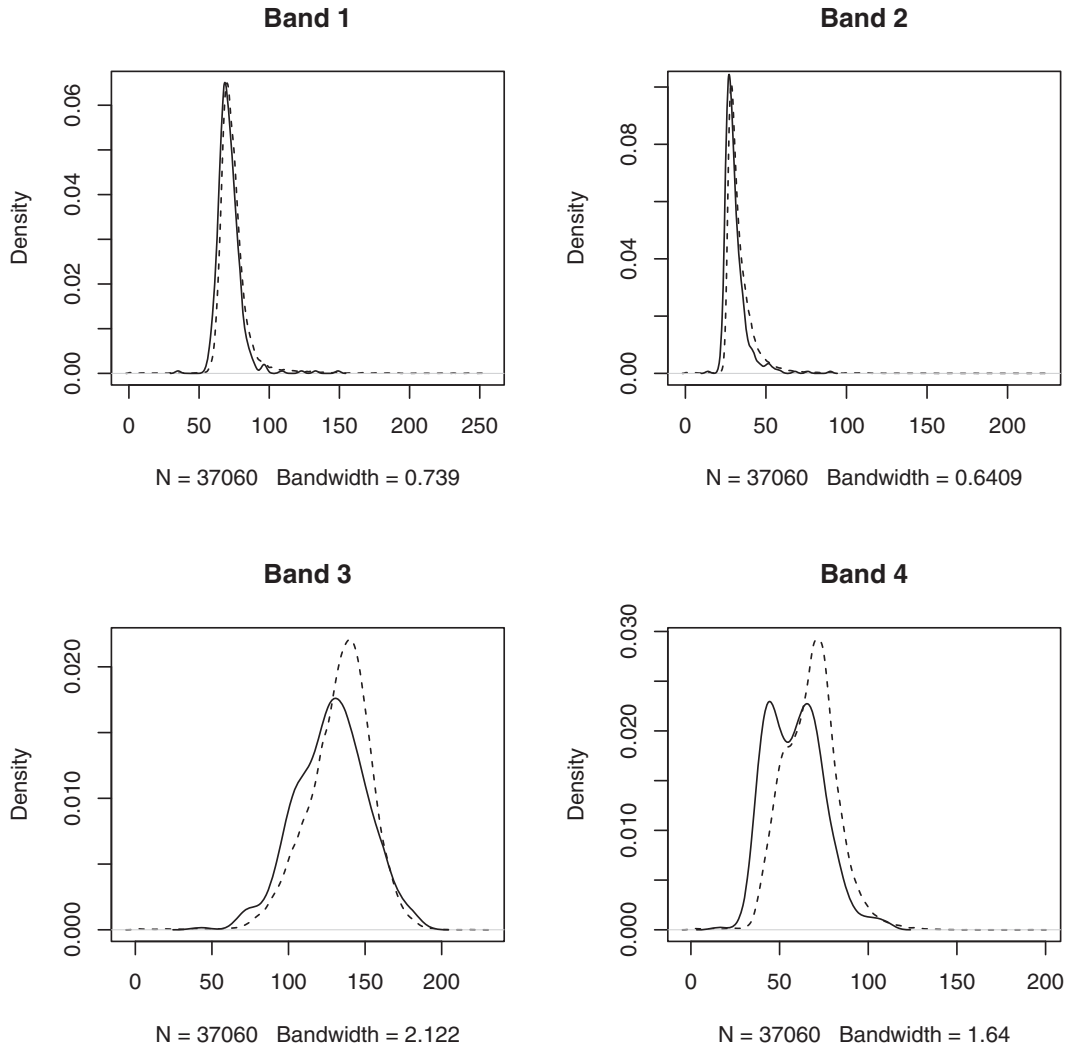
- Administrative county boundaries for which forest variable estimates were made.

Examples of the coarse-scale variables are included in Fig. 5.

The above auxillary datasets were combined in a common file based on the Easting and Northing grid coordinates of the NFI plot location, as well as the pixel information from associated bands of the IRS and SPOT satellite images.

Using pktools, a collection of computer based utilities written in C++ for image processing (McInerney and Kempeneers 2015), an ASCII file containing the Easting and Northing coordinates of the field inventory plots was created by extracting the pixel values for all NFI plots from the feature variables (satellite imagery and aux-

Fig. 6. Density distribution of the reference data (dashed line) and the target dataset (solid line) for the four bands from the LISS-III data. Note that the *y* axis is the density and the *x* axis is the digital number of the satellite image band.



illary spatial data layers) and formatted to the required specifications for i*k*-NN estimation. Weighted mean pixel values were extracted based on a 3 × 3 window of pixels around the target pixel to reduce uncertainty relating to the location of the field plots.

2.6. Model diagnostics

McRoberts (2009) detailed diagnostic tools for evaluating and enhancing nearest neighbours prediction for continuous, univariate response variables. One specific diagnostic that is used prior to *k*-NN estimation is that the reference dataset adequately covers the distribution of the target dataset. This test ensures that the estimation process will not extrapolate predictions for all target variables. Figure 6 presents the density distributions for the reference and target datasets. This diagnostic test demonstrates that the distribution of the satellite image bands for the reference plots fully covers the range of values in the target dataset and would therefore not extrapolate variable predictions.

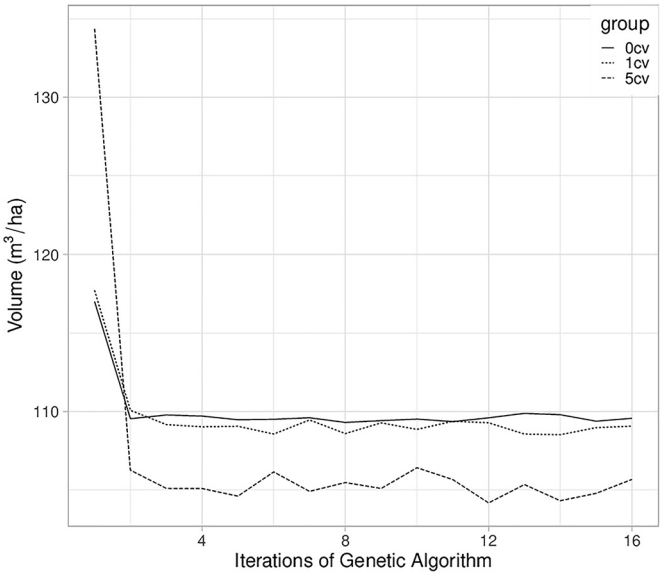
2.7. Explanatory variable weight determination using a genetic algorithm

Genetic algorithms (GAs) are iterative search algorithms that are based on the principles of evolutionary processes to solve optimisation problems. In the context of i*k*-NN, GAs begin with the random selection of a set of values for the weights for the explanatory variables. The search uses genetic operators such as inheritance, selection, mutation, and crossover to ultimately find the values that optimise the search criterion (Tomppo and Halme 2004; Tomppo et al. 2009; McRoberts et al. 2016). Three simulations were constructed to identify the optimal selection of weights for the explanatory bands and to evaluate the effectiveness of the coarse-scale variables (CV). These included

- zero CV (OCV): spectral bands, band ratios but no additional coarse-scale variables were used;
- one CV (1CV): spectral bands and band ratios were used with one additional coarse-scale variable used (in this case DTM);
- Five CV (5CV): spectral bands and band ratio were used in conjunction with five additional coarse-scale variables, which included DTM, accumulated temperature, and interpolated growing stock volume for conifer, broadleaf, and mixed forest dominated stands.

The genetic algorithm was parameterised to run 16 times as this was considered sufficient to estimate the appropriate weights; Figure 7 presents the impact on the root mean square error (RMSE) for growing stock volume per hectare for the three constructed simulations. The graph demonstrates that the smallest RMSE is achieved after 12 iterations for model 5CV, which uses the spectral bands, band ratios, and five coarse-scale variables. It is also noteworthy that this model shows a decrease in RMSE of

Fig. 7. Three models used in the genetic algorithm: 0CV, zero coarse-scale variables; 1CV, one coarse-scale variable; 5CV, five coarse-scale variables.



approximately 5 $m^3 \cdot ha^{-1}$ after the third iteration compared with the two other simulations.

2.8. i*k*-NN method

k-NN is a supervised learning technique that uses a reference dataset consisting of sample population units that contain observations of both response and feature variables. The feature variables are auxiliary variables, which in this case are multispectral satellite image bands and auxillary spatial layers. Within this study, two *k*-NN approaches are used for the purposes of comparison, namely an unweighted *k*-NN method and the i*k*-NN technique. The unweighted *k*-NN approach used only the satellite image bands and the corresponding band ratios as explanatory variables and applied equal weights to each.

The i*k*-NN technique was developed in Finland and is underpinned by a genetic algorithm, which is used to find optimal weights for explanatory variables. These weights correspond to the diagonal values for a weighted Euclidean distance matrix. i*k*-NN extends the conventional *k*-NN technique, which has been employed in the Finnish Multi-Source NFI to combine field plot data, satellite images, and digital map data since 1990 (Tomppo 1991) and its improved version since early 2000 (Tomppo and Halme 2004; Tomppo et al. 2014).

The core of the i*k*-NN estimation method is the new weights (eq. 3) for each field plot by spatially discrete computation units, e.g., by municipalities (Tomppo 1996). The weights are computed for each field sample plot $i \in F$, where F is the set of field plots on forestry land. These plot weights are sums of the weights that are computed for the field plots over all satellite image pixels on the forest land mask of the computation unit. The plot weights corresponding to a single pixel (eq. 1) are, in turn, computed by a nonparametric *k*-NN estimation method (Tomppo 1991; Tomppo et al. 2008). The method utilises the distance metric d, defined in the feature space of the satellite image data and coarse-scale forest variables.

Following Tomppo et al. (2008, 2014), let us denote the k nearest feasible field plots by $i_1(p), \ldots, i_k(p)$. The weight $w_{i,p}$ of field plot i to pixel p is defined as

$$(1)\qquad w_{i,p} = \frac{1}{d^t_{p_i,p}} \Big/ \sum_{j \in \{i_1(p),\ldots,i_k(p)\}} \frac{1}{d^t_{p_j,p}}, \quad \text{if and only if } i \in \{i_1(p), \ldots, i_k(p)\}$$
$$= 0, \quad \text{otherwise}$$

The distance weighting power t is a real number, usually $t \in [0,2]$. Leave-one-out cross-validation tests were carried out using the field plot data and explanatory variables to find the values of t and k that minimize the leave-one-out RMSEs. A small quantity, greater than zero, is added to d when $d = 0$ and $i \in \{i_1(p), \ldots, i_k(p)\}$. The distance metric d employed was

$$(2)\qquad d^2_{p_j,p} = \sum_{l=1}^{n_f} \omega^2_{l,f}(f_{l,p_j} - f_{l,p})^2 + \sum_{l=1}^{n_g} \omega^2_{l,g}(g_{l,p_j} - g_{l,p})^2$$

where $f_{l,p}$ is the lth normalised intensity value of the spectral band image variable, normalising done on the basis of digital elevation model, $f_{l,p_j} = f^0_{l,p_j} / \cos^r(\alpha)$, where α is the angle between sun illumination and terrain normal, r is the user given power for the cosine correction, $g_{l,p}$ is the large-area prediction of the lth applied forest variable, n_f is the number of image variables (or features), n_g is the number of coarse-scale forest variables, and ω_f and ω_g are the weight vectors for image features and coarse-scale forest variables, respectively.

The values of the elements of the weight vectors ω_f and ω_g are computed by means of a genetic algorithm (Tomppo and Halme 2004). Further details on this methodology are provided in Tomppo and Halme (2004).

The first phase of the improved version of *k*-NN (i*k*-NN) is to run the optimization algorithm by strata. The estimation after that is similar to the basic *k*-NN estimation.

For computing forest parameter estimates for computation units, sums of field-plot weights to pixels, $w_{i,p}$, are calculated by computation units and by map stratum h over the pixels belonging to the unit u. The weight of the plot i in map stratum h to computation unit u is denoted

$$(3)\qquad c_{i,h,u} = a \sum_{p \in u_h} w_{i,p}$$

where u_h is the set of the pixels in the map stratum h, and a is the pixel size. Ratio estimators with these plot weights, that is, with inclusion probabilities, are used for parameter estimation.

Predictions of some forest variables are expressed in the form of a digital map during the procedure (Tomppo 1991, 1996). A pixel-level prediction of variable Y for pixel p on forest land is defined as

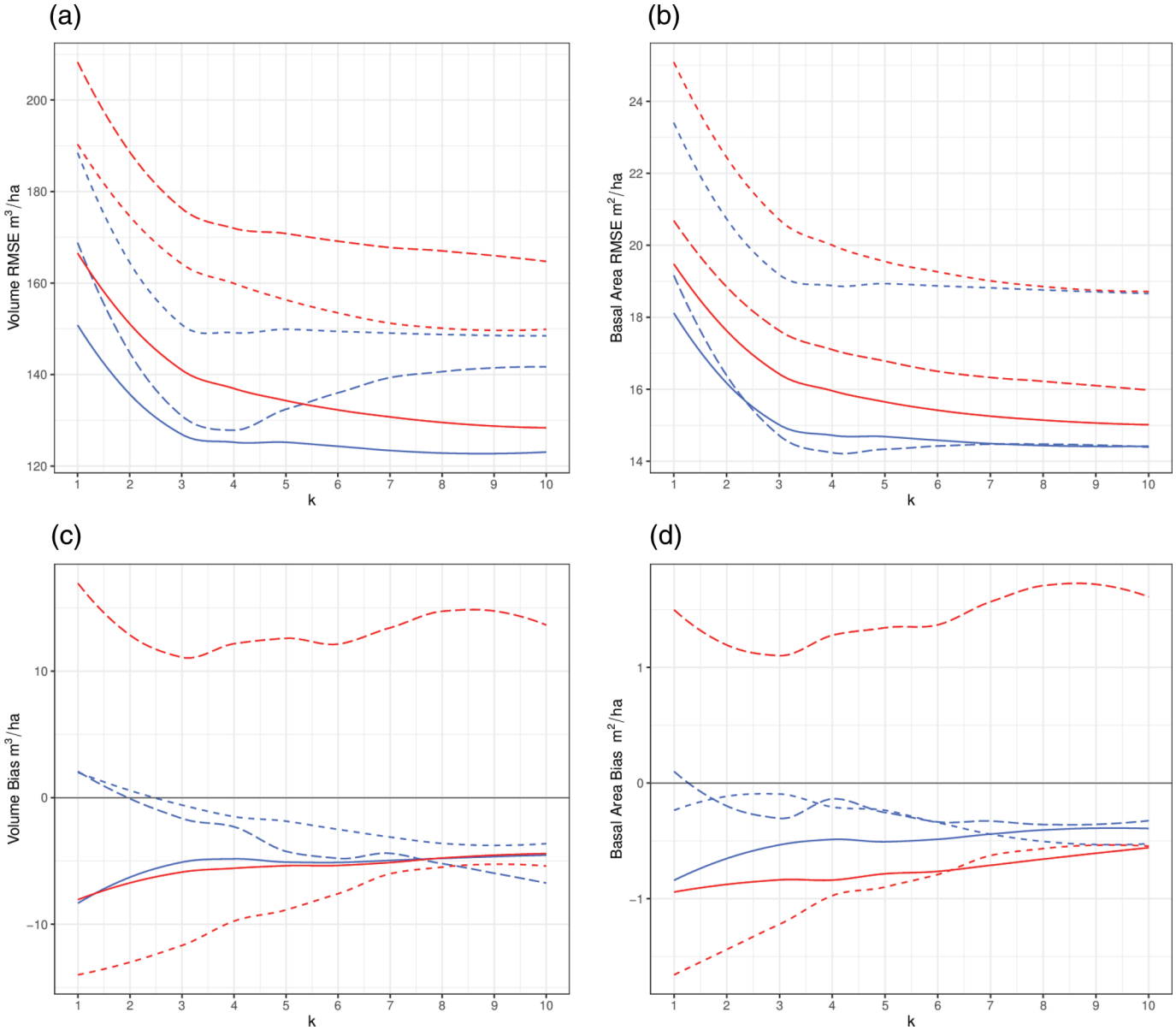
$$(4)\qquad \widetilde{y_p} = \sum_{i \in I_h} w_{i,p} y_i$$

where y_i is the value of the forest variable Y on plot i, I_h is the set of the field plots belonging to map stratum h, and F is the set of the plots i belonging to forest land. The mode or median value can be used instead of the weighted average for categorical variables. Mode has turned out to work better than median in the practical tests (Tomppo et al. 2009).

The Euclidean distance metric is widely used in many *k*-NN applications, but a distance metric that is based on the weighted sum of spectral components and coarse-scale variables was developed by Tomppo and Halme (2004). This distance metric underpins the i*k*-NN methodology and also uses a genetic algorithm for optimisation of the feature variable weights. In the current work, we focus on the basal area and growing stock volume estimates.

To compensate for systematic model prediction errors, we used model-assisted estimators for the final estimates (Särndal et al.

Fig. 8. Plots of root mean square error (RMSE) and bias for unweighted *k*-NN (red) and i*k*-NN (blue) from the optimal run of the genetic algorithm for *k* = 1 to 10: (*a*) RMSE volume per hectare; (*b*) RMSE basal area per hectare; (*c*) bias volume per hectare; and (*d*) bias basal area per hectare. AOI West, solid line; AOI South, long dash; AOI East, short dash.



1992; Baffetta et al. 2009). For the totals, calculated with the inclusion probabilities (eq. 6), "the correction term", C, is

$$C = \hat{A} \times \frac{1}{n}\sum_{i}^{n} \epsilon_i \tag{5}$$

where ϵ_i, $i = 1, \ldots, n$, are the prediction errors in the reference dataset set of a size of n, obtained using, e.g., leave-one-out cross validation, and $\hat{A}$ is the estimate for the area in question, e.g., for forest area. The estimates of the means are corrected in similar manner, that is, just leaving out the estimate $\hat{A}$ in eq. 5.

A model-assisted estimator was used to calculate the confidence intervals of the *k*-NN and i*k*-NN estimates. The estimates are corrected using the estimated bias resulting from the systematic prediction error by comparing the reference data (field inventory plots) and the map data (*k*-NN and i*k*-NN estimates). The model-assisted general regression estimators provided by Särndal et al. (1992) are as follows:

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} \hat{y}_i - \frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i) \tag{6}$$

and

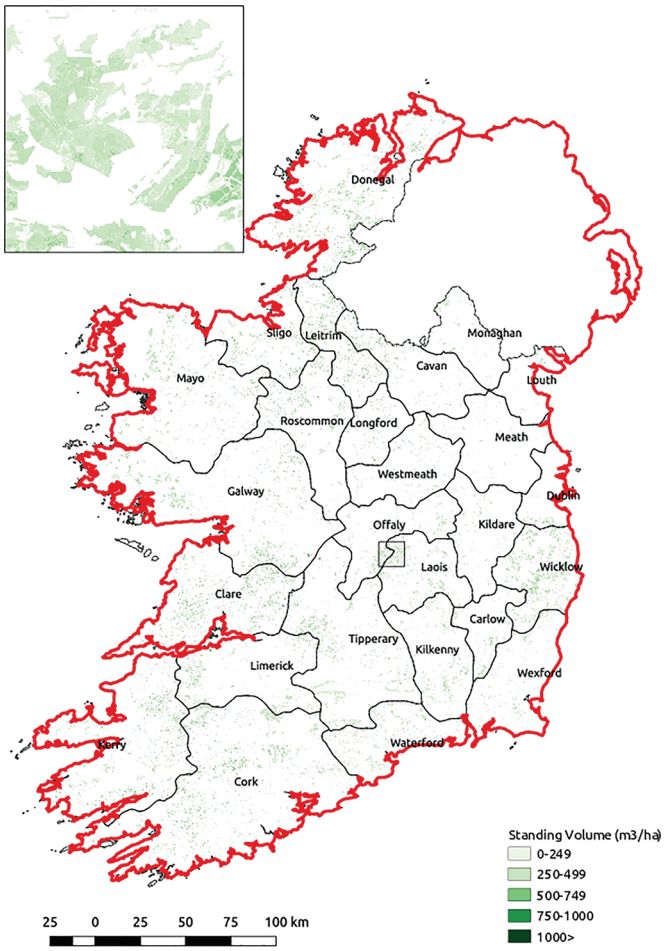
$$\widehat{\mathrm{Var}}(\hat{\mu}) = \frac{1}{n(n-1)}\sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \tag{7}$$

3. Results

3.1. Model prediction errors

The root mean square error (RMSE) for growing stock volume and the associated prediction error (mean deviations) are pre-

Fig. 9. Nationwide map of growing stock volume per hectare at the national scale estimated using i*k*-NN and the genetic algorithm with county boundaries included. The inset map shows a subset of the Slieve Blooms forest, located in Laois. [Colour version available online.]



sented in Fig. 8 for values of *k* from 1 to 10 from i*k*-NN using the optimal run of the genetic algorithm heuristic search (blue lines) and the unweighted *k*-NN (red lines). This graph demonstrates that the optimised i*k*-NN yields smaller RMSEs across all values of *k* in the three AOIs when compared with the unweighted *k*-NN. It also shows that $k = 3$ provides the smallest pixel-level RMSE for volume predictions or represented the value of *k* at which the RMSEs levelled off. These values were 126, 128, and 148 $m^3 \cdot ha^{-1}$ for AOIs West, South, and East, respectively (Fig. 8). Practical tests showed that the effect of the value of *k* on the regional-level estimates played a smaller role than on the pixel-level predictions. For consistency and due to the fact that some of our estimation regions were fairly small with a relatively small number of sample plots, we selected the smallest feasible value of *k*, i.e., 3. As a result, this value is used for the estimation process for the entire study area. The RMSEs expressed as a percentage of the mean of the NFI field plots are 55%, 42%, and 60% for each of the respective AOIs.

The estimation process was used to estimate the parameters across the entire study area excluding areas that were covered by cloud. The estimates were corrected based on the prediction error using eq. 5 as defined previously. The map and corresponding table of estimates by county are provided in Fig. 9 and Table 2. The map is an indicative presentation of the distribution of growing stock volume across the entire country. The table of estimates by county includes those from NFI, *k*-NN, and i*k*-NN. At the national level, the NFI estimated the total growing stock for the entire country to be 97 million m^3 (±4.4%), whereas the growing stock estimated using i*k*-NN was 102 million m^3 (±3%), which is within the upper confidence interval of the NFI estimate. The total estimate of gross standing volume calculated using *k*-NN was 126 million m^3 (±4%), which is considerably larger than the NFI and i*k*-NN estimates. These estimates are the cumulative total and therefore correspond to all tree species within the Forest Estate. At the county level, the county with the largest forest area, county Cork, has a total growing stock of 12 million m^3 ± 12% as estimated from the NFI, while i*k*-NN estimated that the total growing stock was 11.8 million m^3 ± 6%. County Louth, in the northeast of the country, has the smallest forest area and correspondingly the smallest total volume in both the NFI and i*k*-NN estimates with the largest associated confidence intervals, 100% and 76% for the NFI and i*k*-NN, respectively. Despite this anomaly in county Louth, which is due to the small forest cover (and small number of NFI plots, six in total), the estimates from i*k*-NN are consistent with the NFI estimates, but typically with a narrower confidence interval, whereas the county-level estimates produced using the unweighted *k*-NN are all larger than either the NFI or i*k*-NN despite having similar confidence intervals.

4. Discussion

This paper has presented a study that used i*k*-NN estimation optimised with a genetic algorithm to estimate growing stock volume per hectare for the entire country and compared the results with the Irish NFI results (Forest Service 2013*b*). The research builds on the initial pilot projects that were carried out in counties Clare and Wicklow using data from the first NFI (McInerney and Nieuwenhuis 2009) and provides a robust approach to processing the entire national Forest Estate using optical satellite imagery. There are multiple reasons that make i*k*-NN estimation an appropriate technique that can be easily integrated into forest inventories. Firstly, it is cost effective and produces spatially contiguous estimates that could not be obtained when field data are used alone. Secondly, it uses a synthetic estimator, which utilises data outside the area of interest. Thirdly, it is nonparametric and statistically oriented, meaning that there are no assumptions made regarding the underlying distribution of the data and the estimation process preserves the natural dependence between the forest variables.

In terms of comparisons with other studies, the meta-analysis carried out by Chirici et al. (2016) provides a good benchmark of the most important *k*-NN applications carried out operationally around the world. In terms of the selection of *k*, the majority, more than 60% of the *k*-NN applications, used a value of $k \leq 5$. This demonstrates that the selection of a value of $k = 3$ used in the study is consistent with the majority of applications. In terms of the reported values of relative RMSE, it was found that almost 70% of the studies reported a relative RMSE between 20% and 40%, which is smaller than that reported in this analysis. However, if we consider the RMSEs obtained in studies focusing on temperate oceanic forests and (or) those based on medium-resolution satellite data (e.g., SPOT), the median relative RMSEs were 38% and 44%, respectively. These figures are therefore consistent with the findings and results from this study.

More specifically, a study carried out in Japan to estimate stand volume using multispectral satellite data achieved relative RMSEs between 58% and 60% (Tanaka et al. 2015). These findings are broadly similar to the relative RMSEs achieved in this study across the three AOIs (between 42% and 59%). Furthermore, Fazakas et al. (1999) achieved relative RMSEs of 66.6% in Sweden to estimate wood volume using satellite imagery. Unlike many studies dealing with *k*-NN applications, this paper presents a detailed error

Table 2. Gross standing volume by county estimated from NFI and i*k*-NN estimation including forest area, with 95% confidence intervals in parentheses.

County	Total forest (000s ha)	NFI total volume (000s m³)	i*k*-NN total volume (000s m³)	*k*-NN total volume (000s m³)
Carlow	8400	1272 (±47%)	1273 (±33%)	1936 (±23%)
Cavan	16 830	1931 (±41%)	2011 (±15%)	1548 (±22%)
Clare	53 290	7071 (±17%)	6981 (±10%)	9911 (±9%)
Cork	83 620	12 513 (±12%)	11 790 (±6%)	15 135 (±7%)
Donegal	55 380	6202 (±21%)	8584 (±6%)	8888 (±7%)
Dublin	5220	1197 (±52%)	809 (±20%)	1216 (±20%)
Galway	59 410	8003 (±13%)	8793 (±5%)	9654 (±6%)
Kerry	53 180	6921 (±22%)	8296 (±6%)	10 955 (±8%)
Kildare	10 400	1393 (±35%)	1045 (±21%)	1684 (±18%)
Kilkenny	19 010	3209 (±24%)	3365 (±15%)	4134 (±12%)
Laois	25 260	4632 (±21%)	4875 (±17%)	6024 (±11%)
Leitrim	26 500	3071 (±26%)	3577 (±5%)	3723 (±7%)
Limerick	26 340	3295 (±25%)	3253 (±8%)	4293 (±8%)
Longford	8360	1059 (±33%)	1438 (±9%)	1199 (±17%)
Louth	2420	455 (±100%)	12 (±76%)	186 (±25%)
Mayo	51 730	5583 (±17%)	5949 (±7%)	5483 (±14%)
Meath	12 490	1566 (±40%)	1399 (±17%)	1867 (±17%)
Monaghan	5600	482 (±44%)	537 (±14%)	982 (±16%)
Offaly	24 520	2486 (±28%)	2354 (±14%)	3665 (±13%)
Roscommon	25 880	4060 (±27%)	4555 (±11%)	5590 (±10%)
Sligo	20 590	3108 (±27%)	4036 (±7%)	4179 (±10%)
Tipperary	47 850	5865 (±17%)	6125 (±8%)	6507 (±9%)
Waterford	26 550	3267 (±24%)	3266 (±15%)	4566 (±12%)
Westmeath	13 180	1084 (±41%)	890 (±11%)	1634 (±15%)
Wexford	13 840	1705 (±26%)	1799 (±14%)	2816 (±13%)
Wicklow	35 860	6037 (±24%)	5665 (±8%)	8391 (±8%)
Total	731 710	97 476 (±5%)	102 677 (±3%)	126 166 (±4%)

assessment, which uses a model-assisted approach to calculate the confidence intervals of the estimates.

It is important to note that while the integration of Earth observation data into field inventory campaigns can contribute positively to the improvements in estimating forest parameters, there will continue to be a requirement for field surveys for the purposes of creating or maintaining auxiliary reference layers (e.g., the forest mask maintained by the Irish Forest Service as part of their forestry vector database), as well as generating the coarse-scale variables used as part of the i*k*-NN analysis. It is also necessary to note that the products generated from this approach can be used for indicative forest management at local scales, but there is too much uncertainty of the estimates at the individual stand level. The i*k*-NN method applied in this study reduces the prediction errors and standard errors for standing volume when compared with the NFI results (Table 2). Consequently, techniques such as i*k*-NN and map products are intermediate products that allow for the production of confidence intervals.

One of the distinct benefits of this approach is that NFI data can be combined with pre-existing and largely free or low-cost satellite imagery to extend the applications and use of the NFI data. As a result, it can produce information that has the potential to inform policy and act as an aid to sustainable forest management by providing data on the national resource that is timely, spatially complete, and verifiable and can be used for retrospective analysis. From an operational perspective, it is regarded that the NFI provides country-level estimates pertaining to Ireland's forests, but that it is not designed, nor applicable, at the regional level. Consequently, the data products generated using i*k*-NN can be used to assist in the development of national and regional forest policies. From an operational planning perspective, many of the current models rely on either out-of-date forest inventory data and (or) are based on assumptions regarding forest growth and developments. As a result, the data products generated using i*k*-NN provide an independent data source that can be used to validate and complement existing timber supply forecasts. This will enable national forest authorities to ascertain whether the assumptions underlying the forecast have been implemented correctly and whether they are biased.

With the advent of sensors such as Landsat-8 and Sentinel-2, there are more opportunities to avail of multi-temporal cloud-free observations of the Forest Estate, which can be combined with NFI data for these types of applications and which will allow for i*k*-NN analyses to be re-run at shorter time intervals to update map products and statistics.

Barrett et al. (2016) provided specific guidelines and identified best practices for countries initiating, expanding, or revising the use of remotely sensed data in their NFIs. With the exception of adjustments to map-based estimates for prediction errors, this study can be viewed as the best practice in the use of plot-based NFI data and Earth observation data. Future work will attempt to specifically focus on time-series analysis of i*k*-NN estimates with a view to quantifying changes in regional timber volume stocks.

5. Conclusions

The following four primary conclusions can be drawn from this study.

1. The i*k*-NN method can add value to Irish NFI field plot data by using Earth observation data to enhance estimates for the most meaningful and commonly reported forest resource parameters and provides more accurate estimates than a simple, unweighted *k*-NN method.
2. The precision of the i*k*-NN estimates are increased when the field data are combined with remote sensing data and a model-assisted estimator.

3. The i*k*-NN method developed by Tomppo and Halme (2004) is flexible and, when adapted to local conditions, is suited to the specific Irish Forest Estate.
4. The operational use of Earth observation data with Irish NFI field plot data can promote more efficient use of limited financial resources and increase the accuracy and precision of Forest Estate estimates at geographic scales beyond what was originally intended. This is becoming increasingly relevant with the widespread availability of free, multispectral satellite imagery such as Landsat-8 OLI and Sentinel-2. Furthermore, the derived products from this type of analysis are needed for the preparation of up-to-date timber supply forecasts and regional management plans as they provide robust information concerning the spatial distribution of timber volume.

Vegetation height data such as those provided by airborne laser scanning (ALS) are being used more and more as auxiliary variables for parameter estimation in operational forest inventories, but the main obstacle associated within their operational use on a wall-to-wall basis is the cost and availability of the data. For this reason and given the relatively small cost and wider availability of multispectral satellite imagery, methodologies that employ these latter data sources will remain more attractive for bodies charged with carrying out NFIs at least in the short term. Furthermore, while frequent and extensive cloud cover has historically been a barrier to national-scale remote sensing surveys using optical sensors in Ireland, multiple or dual constellation optical sensors satellite such as RapidEye and Sentinel 2 (McInerney et al. 2016) can now potentially provide the required image data to facilitate i*k*-NN parameter estimation in a flexible, timely, and reliable way.

Acknowledgements

This article is a product of Working Group 2 of COST Action FP1001 (USEWOOD). COST (European Cooperation in Scientific and Technical Research) financially supported a Short Term Scientific Mission to Finland to initiate the analysis and research related to the results presented in this article. Additional support was provided by the Irish Forest Service and the Finnish Research Institute (now referred to as the Natural Resources Institute Finland). The authors acknowledge the two anonymous reviewers for their detailed comments and suggestions on the manuscript.

References

Baffetta, F., Fattorini, L., Franceschi, S., and Corona, P. 2009. Design-based approach to *k*-nearest neighbours technique for coupling field and remotely sensed data in forest surveys. Remote Sens. Environ. **113**(3): 463–475. doi:10.1016/j.rse.2008.06.014.

Barrett, F., McRoberts, R., Tomppo, E., Cienciala, E., and Waser, L. 2016. A questionnaire-based review of the operational use of remotely sensed data by national forest inventories. Remote Sens. Environ. **174**: 279–289. doi:10.1016/j.rse.2015.08.029.

Black, K., Xenakis, G., and Ray, D. 2010. Climate change impacts and adaptive strategies. Ir. For. **67**(1–2): 66–85.

Bottai, L., Chirici, G., Corona, P., Marchetti, M., Maselli, F., and Ricotta, C. 2002. Stand basal area extension over forest land by different classification algorithms applied to Landsat ETM+ imagery. *In* Proceedings of ForestSAT 2002 Symposium, Heriot-Watt University, Edinburgh, Scotland, August 2002.

Breidenbach, J., Antón-Fernández, C., Petersson, H., McRoberts, R., and Astrup, R. 2014. Quantifying the model-related variability of biomass stock and change estimates in the Norwegian National Forest Inventory. For. Sci. **60**(1): 25–33. doi:10.5849/forsci.12-137.

Chirici, G., Barbati, A., Corona, P., Marchetti, M., Travaglini, D., Maselli, F., and Bertini, R. 2008. Non-parametric and parametric methods using satellite images for estimating growing stock volume in alpine and Mediterranean forest ecosystems. Remote Sens. Environ. **112**(5): 2686–2700. doi:10.1016/j.rse.2008.01.002.

Chirici, G., Mura, M., McInerney, D., Py, N., Tomppo, E., Waser, L., Travaglini, D., and McRoberts, R. 2016. A meta-analysis and review of the literature on the *k*-nearest neighbors technique for forestry applications that use remotely sensed data. Remote Sens. Environ. **176**: 282–294. doi:10.1016/j.rse.2016.02.001.

Department of Agriculture, Food and the Marine (DAFM). 2014. Forests, products and people. Ireland's forest policy — a renewed vision. Department of Agriculture Food and the Marine, Dublin, Ireland.

Department of Agriculture, Food and the Marine (DAFM). 2018*a*. Forest statistics — Ireland 2017 [online]. Department of Agriculture, Food and the Marine, Johnstown Castle Estate, Co. Wexford. Available from https://www.agriculture.gov.ie/media/migration/forestry/forestservicegeneralinformation/ForestStatisticsIreland2017090318.pdf [accessed 30 August 2018].

Department of Agriculture, Food and the Marine (DAFM). 2018*b*. Forest cover datasets [online]. Available from https://www.agriculture.gov.ie/forestservice/forestservicegeneralinformation/foreststatisticsandmapping/forestcoverdatasets/ [accessed 30 August 2018].

Fazakas, Z., Nilsson, M., and Olsson, H. 1999. Regional forest biomass and wood volume estimation using satellite data and ancillary data. Agric. For. Meteorol. **98–99**: 417–425. doi:10.1016/S0168-1923(99)00112-4.

Fealy, R.M., Green, S., Loftus, M., Meehan, R., Radford, T., Cronin, C., and Bulfin, M. 2009. Teagasc EPA soil and subsoils mapping — Final Report. Volume I. Teagasc, Dublin.

Forest Service. 2013*a*. The Second National Forest Inventory — Republic of Ireland — field procedures and methodology covering the National Forest Inventory, 2009 to 2012. Department of Agriculture, Fisheries and Food, Johnstown Caste Estate, Co. Wexford, Ireland.

Forest Service. 2013*b*. The Second National Forest Inventory, Republic of Ireland, results. Department of Agriculture, Fisheries and Food, Johnstown Castle Estate, Co. Wexford, Ireland.

Franco-Lopez, H., Ek, A., and Bauer, M. 2001. Estimation and mapping of forest stand density, volume, and cover type using the *k*-nearest neighbors method. Remote Sens. Environ. **77**(3): 251–274. doi:10.1016/S0034-4257(01)00209-7.

Gallagher, G., Dunne, S., Jordan, P., and Stanley, B. 2000. Ireland's forest inventory and planning system. *In* Proceedings of IUFRO Conference on Remote Sensing and Forest Monitoring, Rogow, Poland, 1–3 June 1999. Warsaw Agricultural University Faculty of Forestry. pp 182–194.

Gillmore, D.A. 1988. Trends and spatial patterns in private afforestation in the Republic of Ireland. Ir. For. **55**(1): 10–25.

Gjertsen, A.K. 2007. Accuracy of forest mapping based on Landsat TM data and a kNN-based method. Remote Sens. Environ. **110**(4): 420–430. doi:10.1016/j.rse.2006.08.018.

Lawrence, M., McRoberts, R., Tomppo, E., Gschwantner, T., and Gabler, K. 2010. Comparisons of National Forest Inventories. *In* National Forest Inventories: pathways for common reporting. *Edited by* E. Tomppo, T. Gschwantner, M. Lawrence, and R. McRoberts. Springer, Dordrecht, Netherlands. pp. 19–32. doi:10.1007/978-90-481-3233-1_2.

Mäkelä, H., and Pekkarinen, A. 2004. Estimation of forest stand volumes by Landsat TM imagery and stand-level field-inventory data. For. Ecol. Manage. **196**(2–3): 245–255. doi:10.1016/j.foreco.2004.02.049.

McGrath, R., Nishimura, E., Nolan, P., Semmler, T., Sweeney, C., and Wang, S. 2005. Climate change: regional climate model predictions for Ireland. Environmental Protection Agency, Dublin, Ireland.

McInerney, D., and Kempeneers, P. 2015. Open source geospatial tools: applications in Earth observation. Springer, Cham, Switzerland. doi:10.1007/978-3-319-01824-9.

McInerney, D., and Nieuwenhuis, M. 2009. A comparative analysis of kNN and decision tree methods for the Irish National Forest Inventory. Int. J. Remote Sens. **30**(19): 4937–4955. doi:10.1080/01431160903022936.

McInerney, D., Pekkarinen, A., and Haakana, M. 2005. Combining Landsat ETM+ and field inventory data for the national forest inventory — a pilot study for Co. Clare. *In* Proceedings of ForestSAT 2005, Boras, Sweden. Available from http://shop.skogsstyrelsen.se/sv/publikationer/rapporter/proceedings-of-forestsat-2005-in-boras-may-31-june-3-rapport-200.html.

McInerney, D., Barrett, F., Landy, J., and McDonagh, M. 2016. Remote sensing assisted rapid damage assessment of windblow in Irish forests following storm Darwin. Ir. For. **73**(1–2): 161–180.

McRoberts, R. 2009. Diagnostic tools for nearest neighbors techniques when used with satellite imagery. Remote Sens. Environ. **113**(3): 489–499. doi:10.1016/j.rse.2008.06.015.

McRoberts, R. 2012. Estimating forest attribute parameters for small areas using nearest neighbors techniques. For. Ecol. Manage. **272**: 3–12. doi:10.1016/j.foreco.2011.06.039.

McRoberts, R., and Westfall, J. 2014. Effects of uncertainty in model predictions of individual tree volume on large area volume estimates. For. Sci. **60**(1): 34–42. doi:10.5849/forsci.12-141.

McRoberts, R., Domke, G., Chen, Q., Naesset, E., and Gobakken, T. 2016. Using genetic algorithms to optimize *k*-Nearest Neighbors configurations for use with airborne laser scanning data. Remote Sens. Environ. **184**: 387–395. doi:10.1016/j.rse.2016.07.007.

Moeur, M., and Stage, A. 1995. Most similar neighbour: an improved sampling inference procedure for natural resource planning. For. Sci. **41**(2): 337–359.

Ohmann, J., and Gregory, M. 2002. Predictive mapping of forest composition and structure with direct gradient analysis and nearest-neighbor imputation in coastal Oregon, USA. Can. J. For. Res. **32**(4): 725–741. doi:10.1139/x02-011.

Ray, D. 2002. Modelling the future climatic suitability of plantation forest tree species. *In* Climate change: impacts on UK forests. *Edited by* M. Broadmeadow. Forestry Commission, Edinburgh, Forestry Commission Bulletin 125.

Reese, H., Granqvist-Pahlen, T., Edberth, M., Nilsson, M., and Olsson, H. 2005. Automated estimation of forest parameters for Sweden using Landsat data and the kNN algorithm. *In* Proceedings of the 31st International Symposium on Remote Sensing, St. Petersburg, Russia, 20–24 June 2005. Available from https://catalog.hathitrust.org/Record/004197342.

Särndal, C., Swensson, B., and Wretman, J. 1992. Model assisted survey sampling. Springer-Verlag, New York.

Tanaka, S., Takahashi, T., Nishizono, T., Kitahara, F., Saito, H., Iehara, T., Kodani, E., and Awaya, Y. 2015. Stand volume estimation using the *k*-NN technique combined with forest inventory data, satellite image data and additional feature variables. Remote Sens. **7**(1): 378–394. doi:10.3390/rs70100378.

Tomppo, E. 1991. Satellite image-based National Forest Inventory of Finland. *In* Proceedings of the Symposium on Global and Environmental Monitoring, Techniques and Impacts, Victoria, B.C., 17–21 September 1990. International Archives of Photogrammetry and Remote Sensing, Victoria, B.C. **28**(1–7): 419–424.

Tomppo, E. 1996. Multi-source national forest inventory of Finland. *In* New Thrusts in Forest Inventory: Proceedings of the Subject Group S4.02-00 — Forest Resource Inventory and Monitoring and Subject Group S4.12-00. *Edited by* R. Pivinen, J. Vanclay, and S. Miina. IUFRO XX World Congress, Tampere, Finland, 6–12 August 1995. Remote Sensing Technology, Vol. 1. pp. 27–41.

Tomppo, E., and Halme, M. 2004. Using coarse scale forest variables as ancillary information and weighting of variables in k-NN estimation: a genetic algorithm approach. Remote Sens. Environ. **92**(1): 1–20. doi:10.1016/j.rse.2004.04.003.

Tomppo, E., Katila, M., Makisara, K., and Peräsaari, J. 2008. Multi-source National Forest Inventory: methods and applications. Springer, Dordrecht, Netherlands. doi:10.1007/978-1-4020-8713-4.

Tomppo, E., Gagliano, C., De Natale, F., Katila, M., and McRoberts, R. 2009. Predicting categorical forest variables using an improved k-nearest neighbour estimator and Landsat imagery. Remote Sens. Environ. **113**(3): 500–517. doi:10.1016/j.rse.2008.05.021.

Tomppo, E., Katila, M., Makisara, K., and Perasaari, J. 2014. Multi Source National Forest Inventory of Finland — methods and results 2011 [online]. Working Papers of the Finnish Forest Research Institute. Vol. 319. Available from http://www.metla.fi/julkaisut/workingpapers/2014/mwp319.pdf [accessed 30 August 2018].

Zanchi, G., Theiel, D., Green, T., and Lindner, M. 2007. Forest area change and afforestation in Europe: critical analysis of available data and the relevance for international environmental policies. European Forest Institute, Joensuu, Finland, EFI Technical Report 24.