Data Mining for Biodiversity Prediction in Forests

Barry O'Sullivan[1], Steven Keady[2], Enda Keane[3], Sandra Irwin[2], and John O'Halloran[2]

[1] Cork Constraint Computation Centre
Department of Computer Science, University College Cork, Ireland
b.osullivan@cs.ucc.ie
[2] Dept. of Zoology, Ecology and Plant Science, University College Cork, Ireland
{s.keady|s.irwin|j.ohalloran}@ucc.ie
[3] Treemetrics Ltd., Cork, Ireland
ekeane@treemetrics.com

Abstract. There is international consensus on the key elements of sustainable forest management. Biological diversity has been recognised as one of them. This paper investigates the usefulness of terrestrial laser scanning technology in forest biodiversity assessment. Laser scanning is a rapidly emerging technology that captures high-resolution, 3-D structural information about forests and presently has applications in standing timber measurement. Forest biodiversity is influenced by structural complexity in the forest although precise repeatable measures are difficult to achieve using traditional methods. The aim of the research presented here is to apply laser scanning technology to the assessment of forest structure and deadwood, and relate this information to the diversity of plants, invertebrates and birds in a range of forest types including native woodlands and commercial plantations. Procedures for forest biodiversity assessment are known to be expensive due to their reliance on labour-intensive field visits. We describe our progress on the application of terrestrial laser scanning in an automated approach to biodiversity assessment. We apply regression techniques from the field of data mining to predict several biodiversity measures using physical attributes of the forest with very promising results.

1 INTRODUCTION

The Convention on Biological Diversity[4] defines biodiversity as the variability among living organisms from all sources including among other things, terrestrial, marine, and other aquatic ecosystems, and the ecological complexes of which they are a part; this includes diversity within species, between species and of ecosystems. At least 40% of the world's economy and 80% of the needs of the poor are derived from biological resources. The richer the diversity of life, the greater the opportunity for medical discoveries, economic development, and adaptive responses to new environmental challenges such as climate change and disease outbreaks.

Ground flora diversity is one of the most important elements of biodiversity in forest ecosystems. It plays an important role in forest ecosystem functioning, both directly and indirectly. It can account for a high proportion of annual litterfall and allow for rapid return of nutrients to the soil, thereby contributing to forest productivity. Ground flora diversity also contributes to the diversity of heterotrophic organisms, stabilising the biogeochemical cycle by balancing production and mineralisation. It can also influence bird diversity and mammal community composition, and can provide habitats for insects and other invertebrates important to ecosystem functioning. The presence of ground flora can also promote percolation of water and minimise erosion. In plantation forests, the presence of ground flora can also enhance the aesthetic environment.

The *objective* of the research presented in this paper is to investigate the usefulness of terrestrial laser scanning technology in forest biodiversity assessment. Terrestrial Laser Scanners are ground based devices that use laser to digitise the geometry of real environments with high precision by capturing large point clouds (millions of single points) of structural data in a very short time. The result is a very large volume of detailed, three-dimensional information about an object's dimensions, spatial positioning, texture and colour that can have a number of applications in industry, architecture, science and engineering. In recent years this technology has gone through an intense development phase and has recently been exploited for its potential application in vegetation mapping in the forest environment, where it can be used in forest inventory sampling as an alternative to traditional manual survey techniques.

[4] http://www.cbd.int

(a) A photograph of a native woodland.

(b) A photograph of a conifer forest.

(c) A laser scan 'point cloud' of a native forest.

(d) A laser scan 'point cloud' of a conifer forest.

Fig. 1. Some example photographs and laser scans of the forest types considered in this study.

The *contribution* of this paper is to show that data mining techniques can be successfully used in conjunction with terrestrial laser scanning technology to accurately predict biodiversity measures in forests. Our empirical evaluation is based on a detailed field study that involved the scanning of several Irish woodlands. Based on this field study we were able to generate a range of biodiversity predictors that exhibited strong correlation coefficients. Therefore, our results confirm the conventional wisdom that the physical structure of a forest is a major determinant of biodiversity.

The remainder of the paper is structured as follows. In Section 2 we present a detailed discussion of the importance of physical structure in biodiversity prediction in forests. Section 3 presents a detailed overview of the process of terrestrial scanning of forests. We present our empirical results in Section 4, showing that biodiversity can be accurately predicted by reasoning about physical forest structure. We present our conclusions and plans for future work in Section 5.

2 THE IMPORTANCE OF STRUCTURE

We are concerned with the application of terrestrial laser scanning as a tool in the management of forests for biodiversity conservation. Terrestrial laser scanning has been developed to capture detailed, three-dimensional information about an object's dimensions, spatial positioning, texture and colour. This technology is widely used in architectural, engineering and industrial measurement. In recent years it has been adapted for use in the forest industry where it is now used for taking measurements from standing timber in a non-destructive manner in order to inform optimal harvest decision-making and reduce waste (Figure 1). These measurements would previously have been obtained through labour intensive manual field surveys but can be gathered much more efficiently and at much higher levels of accuracy and precision using laser scanning [1–3]. An Irish company, TreeMetrics Ltd., have developed a fully automated laser scanning system for pre-harvest timber measurement, which we employ in this study.

The data collected using laser scanning has a wide range of potential applications in forests outside those concerned with timber production including monitoring of carbon sequestration, and the measurement of structural characteristics of forest stands related to biodiversity [4–6]. Biodiversity is integral to sustainable forest management, both as a means of ensuring efficacy and stability of ecosystem functions that are vital to forest health and commercial viability, and as a goal of forest managers seeking to provide non-timber services.

Because traditional methods of structural complexity and deadwood assessment require many field visits our approach will develop competence in new emerging technologies for both forest and broader biodiversity assessment. Our innovative and novel approach gives precise measures which will provide baseline data and support national infrastructure for future monitoring. Our approach can also be calibrated to provide assessment across Europe independent of the observer (i.e. fauna or flora person) that can be quality assured. This provides a potentially innovative, novel technology for biodiversity assessment and monitoring.

Over the last half of the 20th century, forest cover in Ireland increased from less than 1% to almost 10%, largely through the planting of non-native trees such as Sitka spruce. Although this is one of the lowest proportional forest covers in Europe [7] a strategic government aim is to increase forest cover to 17% by 2030 [8]. This intense afforestation constitutes a major ecological change in the landscape, and there is an urgent need to assess its effect on native flora and fauna. Furthermore Ireland is required to halt biodiversity loss by 2010 and to attempt to meet this challenge one of the requirements is to assess biodiversity. In particular, scientists and managers need to improve their understanding of the impact of silviculture regimes on the biota of Ireland's new forests, the ecological differences between native woodlands and plantations, and the management required to optimise the biodiversity of commercial plantations.

Forest structure is both a product and a driver of ecosystem processes and biological diversity, shaping the spaces in which forest animals and plants live, determining the extent to which organisms can move through and between forest patches, and regulating the penetration of light from canopy to forest floor [9–11], and has traditionally been measured using time consuming and expensive manual surveys [12, 13]. The advancement of remote sensing technology presents the opportunity to automate and improve forest survey methods [3, 5, 14, 15]. The methodology used here relies on terrestrial laser scanning to measure structure and relate this to biodiversity with a view to using the outputs from scanning to assess biodiversity.

3 TERRESTRIAL SCANNING

Terrestrial laser scanning is a rapidly developing technology that has huge potential for yielding data on forest structure at a previously unattainable level of accuracy and resolution of detail. The scanning hardware is used to generate a point cloud description of an area of forest – see Figures 1(c) and 1(d). A point cloud description is simply a representation of the image of a forest such that every point is modelled in terms of its (x, y, z) coordinates, along with some representation of the intensity associated with each point.

The terrestrial scanner used in this research is shown in Figure 2. In Figure 2(a) we show the full scanner on its tripod in-situ in a forest. The head of the scanner (Figure 2(b)) revolves in a full circle recording the intensity of the reflection from the laser emitted from the scanner's head.

By applying algorithms to identify the individual trees it is then possible to automatically determine traditional forest measurement parameters such as diameter at breast height (DBH), tree height and tree count. In addition, various steps of forest reconstruction are automatically undertaken with a tree being realised through a set of circles fitted to represent the shape up along its stem. This profile has been processed using robust filter routines to exclude effects caused by branches, occlusions, weather etc. The result is shown in Figure 3, which presents a screenshot of the AutoStem[5] software used to process the scans we obtained. One specific tree is highlighted in the right-hand portion of the window. On the left-hand side we are presented with the "depth at breast height" (DBH) of each tree found in the point cloud. The DBH

[5] `http://www.treemetrics.com/technology/index.html`

(a) A Faro laser scanner.

(b) The head of the scanner.

Fig. 2. A typical terrestrial laser scanner. This one is made by Faro and was used to gather the data analysed here.

is the diameter of the tree taken at a height of 1.3m above the ground. Currently, a typical scan of a forest takes around 5 minutes, but technology advances continue to improve both the speed and mobility of the hardware.

The requirement for collaboration between remote sensing and biodiversity research communities to fully exploit the potential of remote sensing in biodiversity studies is well recognised [4] but collaborative work between these two fields in the quest for biodiversity conservation is in its infancy. Although this technology exists in Ireland, its huge potential as a tool for biodiversity research and management has not yet been exploited. While ongoing progress is being made in the improvement of scanning technology, the principal constraint on the use of laser-scanning data for biodiversity research is the development of algorithms for extracting useful ecological parameters from the very large sets of X-Y data points. An objective of this project is to derive algorithms for quantifying canopy structure, canopy openness and deadwood volumes.

Exploitation of existing, but hitherto unused, technologies enable detailed structural assessment of the habitat of many species of canopy invertebrates, epiphytes and birds, which until now has only been possible on a relatively small scale, has required specialist skills, and has been extremely time-consuming [12, 13].

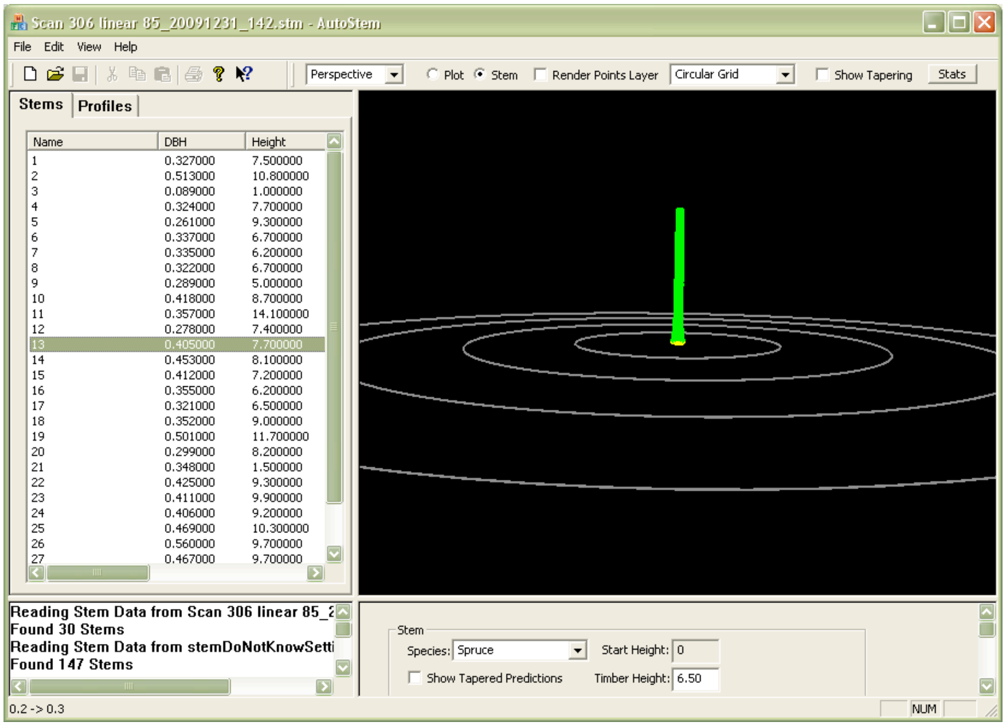


Fig. 3. An example visualisation of a tree geometry extracted from a point cloud scan from a forest.

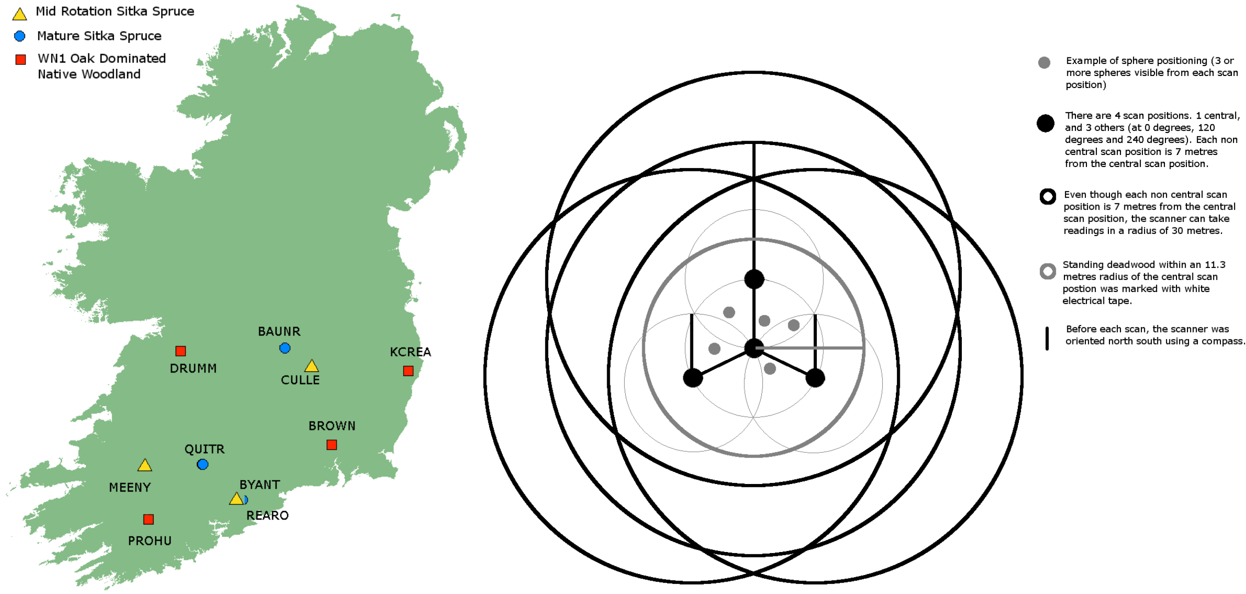
Terrestrial laser-scanning will also enable rapid and accurate quantification of deadwood volume, arguably the most important determinant and indicator of biodiversity value in temperate native forests [16]. By improving our understanding of the relationship between forest structure and biodiversity, these data will improve the ecological relevance of scientific advice concerning forest management. In particular, laser scanning will enable commercial management procedures such as thinning, stocking and harvesting techniques to be more explicitly linked to outcomes for forest biodiversity. Ultimately such research will enable Ireland to comply with the convention on biological diversity and to enhance opportunity for carbon sinking.

4 EXPERIMENTAL EVALUATION

We evaluated the utility of terrestrial laser scanning as a basis for biodiversity prediction in a variety of forest settings. Terrestrial laser scanning was conducted at four native woodland sites during summer and winter (leaf-on and leaf-off condition), and at six conifer plantation forest sites during summer (3 mid-rotation and 3 mature). One mature conifer plantation was also visited during winter. The locations of the sites are shown in Figure 4(a): those marked in yellow are mid-rotation Sitka spruce sites, in blue are mature Sitka spruce sites, and in red are WN2 Oak dominated native woodland sites. The *objective* of the experiment was to determine whether a number of specific biodiversity measures could be predicted based on the structural information extracted from terrestrial laser scans. The specific biodiversity measures we considered were: population abundance and species richness for birds, beetles and spiders based on measurements of these values taken at the sites that were surveyed using the laser scanner.

4.1 Settings in the Field

Laser scanning point clouds were obtained in each forest site, at one or two plots in each, for which biodiversity surveys had been taken previously. Scan positions were marked at each plot according to Figure 4(b). At each plot terrestrial laser scanning was conducted at four points using a FARO laser scanner. The centre of the plot was marked with a white topped cane. A rangefinder and compass were used to mark non-central scan positions with canes. A compass was used to ensure that the scanner was oriented north-south



(a) Forest sites used in our experimental evaluation.

(b) Plot layout for terrestrial laser scanning. Even though each non-central scan position is 7m from the central scan, readings could be taken in a radius of 30m. Standing deadwood within 11.3m radius of the central scan was marked with white tape.

Fig. 4. The site location and protocol used in our experimental evaluation.

before the commencement of each scan. For each scan the number of the scan, time and wind strength were recorded. Once all scans were completed photographs were also taken of the site to record non-scanner specific information on the site at the time of scanning. Hemispherical photographs were also taken at the centre of each plot using a camera with a fisheye lens. At sites where repeat visits were necessary a number of metrics relating to scan position were taken to ensure that repeat scans were conducted at the same site.

The scans for each site were processed by AutoStem and hand-written code to extract the structural description of the forest. Specifically, we extracted the number and position of all trees in the scans. In addition, we computed the minimum, average and maximum tree stem diameters at 10cm height intervals. The minimum, average and maximum stem height was also computed. A number of auxillary features were computed based on these values, such as variance in stem diameters at different heights, as well as density measures for the forest as a whole. In addition to scan data, we recorded the age and type of the forests at each site.

4.2 Prediction Experiment

We framed the problem of predicting each of the six biodiversity measures based on the physical descriptions we obtained of the forests as a data mining task. We used the Weka, a standard open-source data mining system[6], for this purpose. We used a variety of regression techniques in Weka, specifically:

- least median squared linear regression,
- linear regression,
- multi-layer perceptron,
- pace regression linear models, and
- regression trees.

The results of this experiment are presented in Table 1 using 10-fold cross-validation. In this table we present the correlation coefficients between the true biodiversity measure and the predictions one gets from the

[6] `http://www.cs.waikato.ac.nz/ml/weka/`

concepts generated by each technique. Our data-set contained measurements for 37 different forest locations, so correlation coefficients in excess of 0.325 (0.418) are statistically significant at the 95% (99%) level; this was determined using R, the statistical package.[7] Therefore, these are very strong results, most significant at the 99% level.

Table 1. Correlation coefficients between actual and predicted biodiversity.

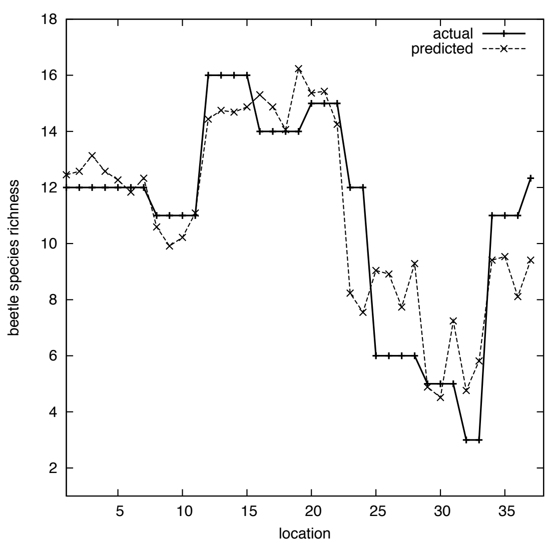
Technique	Birds species	Birds pop	Beetles species	Beetles pop	Spiders species	Spiders pop
Least Med. Sq.	0.41	0.04	0.57	–	0.41	0.47
Linear Reg.	0.43	0.57	0.78	0.71	0.66	0.76
MLP	0.46	0.55	0.66	0.68	0.67	0.73
Pace Reg.	0.23	0.68	0.86	0.73	0.70	0.68
RepTree	0.10	0.51	0.69	0.70	0.42	0.21

Note that both beetles and spiders are easier to predict than birds; the correlations we get for birds are lower. This makes intuitive sense, since bird populations in Ireland do not tend to be distributed on the basis of forest type.

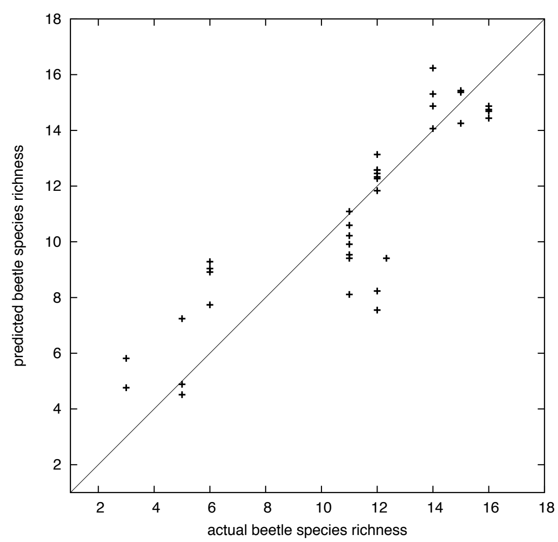
We present an example of the predictor we obtained for the specific case of beetle species richness (BSR) using the concept generated by pace regression in Figure 5. The concept generated in this case was as follows, where the age of the forests in the data-sets is either mature or mid-rotation, which were numerically interpreted as 0 and 1, respectively:

$$BSR = 34.8974 - 13.0242 \times Age - 0.062 \times AvgStemDiameter.$$

Figure 5(a) presents the actual and predicted values at each forest location (37 in all). Figure 5(b) presents the linear relationship between the values; perfect prediction would lie on the diagonal. The very strong prediction is encouraging.



(a) Comparison by forest location.



(b) Actual versus Prediction.

Fig. 5. An example biodiversity predictor for beetle species richness obtained using pace regression.

[7] `http://www.r-project.org`

5 CONCLUSION AND FUTURE WORK

In this paper we presented a novel approach to predicting biodiversity in forests by reasoning about their physical structure. Our approach is based on terrestrial scans, from which a rich physical description of a forest can be obtained. Based on such a description, we have shown how several standard data mining techniques can accurately predict six biodiversity measures of the species richness and abundance of birds, spiders and beetles.

Our work has the potential to automate the development of a world inventory of forests rich with environmental concerns. Specifically, our solution can be used to measure the environmental impact of harvesting trees in a particular forest.

In our future work we plan to significantly expand the set of features that we extract from the laser scans. Of greatest importance is to be able to factor in measurements of standing deadwood in a clear and user-friendly manner.

Acknowledgements. Barry O'Sullivan is supported by Science Foundation Ireland (Grant 05/IN/I886). The contribution of the other authors is co-funded by the EPA and COFORD under the National Development Plan 2007-2013.

References

1. Bienert, A. et al., 2007. Tree detection and diameter estimations by analysis of forest terrestrial laserscanner point clouds. *ISPRS Workshop on Laser Scanning 2007*, Finland.
2. Nsset, E. et al., 2004. 'Laser scanning of forest resources: the Nordic experience'. *Scandinavian Journal of Forest Research*, 19:482-499
3. Weltz, M. A. et al., 1994. 'Comparison of laser and field measurements of vegetation height and canopy cover'. *Water Resources Research*, 30: 1311-1319.
4. Innes, J. L. & Koch, B. 1998. 'Forest biodiversity and its assessment by remote sensing'. *Global Ecology and Biogeography Letters*, 7: 397-419.
5. Maltamo, M. et al., 2005. 'Identifying and quantifying structural characteristics of heterogeneous boreal forests using laser scanner data'. *Forest Ecology and Management*, 216: 41-50.
6. Patenaude, G. et al., 2004. 'Quantifying forest above ground carbon content using LiDAR remote sensing'. *Remote Sensing of Environment*, 93: 368-380
7. Httl, R. F. et al., 2000, 'Forests of the temperate region: gaps in knowledge and research needs'. *Forest Ecology and Management*, 132:83-96.
8. Department of Agriculture Food and Forestry, 1996. *Growing for the future. A strategic plan for the development of the forestry sector in Ireland*, The Stationery Office, Dublin.
9. Hunter, M. L. 1999. *Maintaining biodiversity in forest ecosystems.* Cambridge University Press. 714pp.
10. Smith, G. F. et al., 2006, *Biodiversity assessment of afforestation sites.* Report prepared for COFORD and EPA.
11. Spies, T.A. 1998. *Forest Structure: A Key to the Ecosystem.* Northwest Science, 72:34-36.
12. Gil-Tena, A. et al., 2007. 'Effects of forest composition and structure on bird species richness in a Mediterranean context: Implications for forest ecosystem management'. *Forest Ecology and Management*, 242: 470-476.
13. Nordn, B. et al., 2004. 'Dead wood in semi-natural temperate broadleaved woodland: contribution of coarse and fine dead wood, attached dead wood and stumps'. *Forest Ecology and Management*, 194: 235-248.
14. Parker, G. G. et al., 2001, 'Light transmittance in forest canopies determine using airborne laser altimetry and in-canopy quantum measurement'. *Remote Sensing of the Environment*, 76:298-309.
15. Pesonen, A. et al., 2008. 'Airborne laser scanning-based prediction of coarse woody debris volumes in a conservation area'. *Forest Ecology and Management*, In Press.
16. Travaglini, D. et al., 2007, 'Forest inventory for supporting plant biodiversity assessment - ForestBIOTA data on deadwood monitoring in Europe'. *Plant Biosystems*, 141:222-230.