**EPA STRIVE Programme 2007-2013**

# Soilscapes Extrapolation (Terra Incognita) Ireland

# (2007-S-CD-1-S1)

# ISIS Final Technical Report 4

Prepared for the Environmental Protection Agency
by
Teagasc and Cranfield University

**Authors:**

**Mayr, T.R., Zawadzka, J.E., Corstanje, R., Lo Papa, Fealy, R.**

**ENVIRONMENTAL PROTECTION AGENCY**
An Ghníomhaireacht um Chaomhnú Comhshaoil
PO Box 3000, Johnstown Castle, Co.Wexford, Ireland

Telephone: +353 53 916 0600 Fax: +353 53 916 0699
Email: info@epa.ie Website: www.epa.ie

## ACKNOWLEDGEMENTS

This report is published as part of the Science, Technology, Research and Innovation for the Environment (STRIVE) Programme 2007-2013. The programme is financed by the Irish Government under the National Development Plan 2007-2013. It is administered on behalf of the Department of the Environment, Heritage and Local Government by the Environmental Protection Agency which has the statutory function of co-ordinating and promoting environmental research.

This project involved many staff and students based at Teagasc Johnstown Castle, Teagasc Athenry, Cranfield University and University College Dublin. Many thanks go to all the staff, PhD students, placement students and FAS placements who contributed to this project. We also wish to thank the Steering Committee, Dr Alice Wemaere (EPA), Ms Niamh Connolly (EPA), Dr Mark Gibson (Teagasc), Prof Peter Loveland, Dr Erika Micheli and Dr Arwyn Jones.

## DISCLAIMER



The EPA STRIVE Programme addresses the need for research in Ireland to inform policymakers and other stakeholders on a range of questions in relation to environmental protection. These reports are intended as contributions to the necessary debate on the protection of the environment.

**EPA STRIVE PROGRAMME 2007-2013**

Published by the Environmental Protection Agency, Ireland

## Details of Project Partners

**Teagasc - Environment, Soils and Land-use Centre,**
Johnstown Castle, Wexford, County Wexford, Ireland.
Tel.: +353 (0)53 9171200
E-mail: rachel.creamer@teagasc.ie

**Teagasc – Spatial Analysis Group, Rural Economy and Development Programme.**
Ashtown, Dublin 15, Ireland.
Tel.: +353 (0)1 8059953
E-mail: reamonn.fealy@teagasc.ie

**Cranfield University - National Soil Resources Institute**
Cranfield Campus, Cranfield, Bedfordshire, MK43 0AL, UK
Tel.:+44 (0)1234 750111
Email: t.mayr@cranfield.ac.uk

**University College Dublin**
School of Biosystems Engineering, Agriculture and Food Science,
Belfield, Dublin 4, Ireland.
Tel.: +353 (0)1 7162808
Email: nick.holden@ucd.ie

# Executive Summary

The Irish Soil Information System (ISIS) project was established in 2008, following a comprehensive inventory of Irish soil data compiled by Daly and Fealy (2007) which highlighted that soil data coverage of Ireland was incomplete in both detail and extent. The ISIS project is funded under the Environmental Protection Agency STRIVE Research Programme 2007-2013 and co-funded by Teagasc. It was led by Teagasc with the participation of researchers from Cranfield University (UK) and University College Dublin. The overall objective of the ISIS project was to conduct a programme of structured research into the national distribution of soil types and construct a soil map, at 1:250,000 scale, which will identify and describe the soils according to a harmonised national legend. This map is now available in digital format and forms the basis of a new soil information system for Ireland (http://isis.teagasc.ie).

The ISIS project has utilised existing data and maps from the previous National Soil Survey (NSS) conducted by An Foras Talúntais (forerunner organisation to Teagasc). The NSS produced: mapping at 1:126,720 scale for 44% of the country; a General Soil Map of Ireland and a National Peatland map, both at 1:575,000 scale and other miscellaneous large scale mapping of experimental farms. In addition, more recent map products have been included such as the Indicative Soil and Subsoil mapping (Fealy and Green, 2009) with national coverage using GIS and remote sensing techniques.

Comparison of soil information at European scale has led to the requirement for the harmonisation and coordination of soil data across Europe, and, in light of the demands for soil protection on a regional basis within member states there is a growing need to support policy with a harmonised soil information system. The European Soil Bureau Network (ESBN) Technical Working Group dealing with Soil Monitoring and Harmonisation recommended a soil map of Europe at a scale of 1:250,000 as an economically feasible intermediate scale that can identify specific problems at regional scale (Montanarella and Jones, 1999).

The ISIS project adopted a combined methodology of utilising novel predicted mapping techniques in tandem with traditional soil survey applications. This unique combination at a national scale has resulted in the development of a new national soil map for Ireland. Building upon the detailed work carried out by the An Foras Talúntais (AFT) survey (known as *Terra Cognita*), the ISIS project generated soil-landscape models at a generalised scale of 1:250,000 for the counties of Carlow, Clare, Kildare, Laois, Leitrim, Limerick, Meath, Offaly, Tipperary South, Waterford, Westmeath, Wexford, West Cork, West Mayo and West Donegal. These soil-landscape models (also referred to as soilscapes) were used as the baseline data for statistical models (random forests, Bayesian belief networks and neural networks) to predict soil map units in counties where there was no map available (referred to as *Terra Incognita*). To validate the methodology, this work was supported by a 2.5 year field survey, in which 11,000 locations were evaluated for soil type, using an auger bore survey approach. These data were used to check the predicted soil mapping units (associations) for counties: Cavan, Dublin, East Cork, East Donegal, East Mayo, Galway, Kerry, Kilkenny, Louth, Monaghan, Roscommon, Sligo, Tipperary South and Wicklow, where a detailed soil survey map was not available. Where new soil information was generated, due to previously unknown combinations of soil-landscape units, profile pits were selected at representative locations across the country. These 225 pits were described and sampled in detail and were used to generate a new soil classification system for the country. The final product is a unique combination of new and traditional methodologies and soils data from both the AFT and the ISIS project. The final, soil association map of Ireland consists of 58 associations (excluding areas of alluvium, peat, urban, rock or marsh) that are made up from 213 soil series. Associated representative profile information is available in the online soil information system.

A key component of the ISIS project has been the development of a soil and land information system and associated public web site. This system has been designed to hold the complete set of information deriving both from the ISIS field programme and modelling activity, as well as the previously existing legacy soils information available for Ireland. Drawing on this information system, the web site is designed to hold and disseminate this information online both in cartographic and tabular form to stakeholders. Prior to this development, there has been no harmonised computerised system in place to hold and manipulate national Irish soils data. The information system therefore addresses the pressing need and requirement for a publicly-accessible, integrated IT framework based upon contemporary informatics standards to serve the many and varied stakeholders having an interest in soils information in Ireland.

# Technical Note on Soil Classification

Two Irish soil classification systems were developed during the ISIS project. An **Interim Soil Classification** was developed in the early stages of the project to enable the harmonisation and generalisation of the county soil maps published by An Foras Talúntais (AFT) and the rationalisation of the original AFT soil series. The **Interim Soil Classification** was used during the development of Work Packages (WP): WP1 and WP2, to produce the training data for the predictive mapping and for most of the field programme in WP3. In 2013/4, **the Interim Soil Classification** was modified following a World Reference Base style hierarchical approach that recognises Great Soil Groups and defines sub-groups by supplementary diagnostic horizons. The **Final Soil Classification** System was developed to provide a more user-friendly classification system that adopts the approach of a hierarchical key for recognition of Great Soil Groups and diagnostic horizons to define the sub-groups.

The **Final Soil Classification** System was subsequently implemented during the description of representative soil profiles, final map production and is included in the updated soil profile handbook, and national soil series list. This modified system is the **Final Soil Classification** system for Ireland that appears in the map and associated information system on the ISIS website.

This Final Technical Report was developed using the **Interim Soil Classification**, and describes a significant contribution to the production of the final New Soil Map of Ireland. Table B below details the differences between the **Interim** and the **Final Soil Classification** Systems.

The **Final Soil Classification** System for Ireland has 3 hierarchical levels:

### 1. Great Soil Groups:

The classification criteria for the Great Soil Groups (GSG) were based on recognisable features used by An Foras Talúntais (National Soil Survey of Ireland) to classify the soils of Ireland at Great Soil Group level. Table A provides an overview of the key criteria for recognizing the Great Soil Groups. The sequence follows World Reference Base (WRB) principles.

### 2. Soil Sub-groups:

The Irish Soil Classification of soil sub-groups (SSG) is based on the recognition of diagnostic horizons, properties and materials which, where possible, should be observed and measured in the field. The selection of diagnostic characteristics takes into account their relationship with soil forming processes. Diagnostic features are selected that are significant to soil management. Subgroups are named with a maximum of two diagnostic features that represent the most important processes occurring in the soil profile. Table B provides a look-up table between the interim and the modified classification systems, listing the Great Soil Groups and Sub-groups.

### 3. Soil Series

The classification of series is based on the same principles as the interim classification system. Within a sub-group a series is further defined by the nature of the soil texture and parent material.

## 4. Soil Associations

For mapping purposes, the soil series are combined to form soil associations that are identified by the most frequently occurring soil series and combinations of ancillary series. Each association is named after the key (lead) soil series, which is the most extensive soil in the association, e.g. Kilrush series is the dominant component in the Kilrush Association. To facilitate mapping, each soil association based on the Interim Classification is assigned an alphanumeric code that comprises the soil subgroup code (numeric) concatenated with a single alphabetic character, e.g. 711b for Kilrush Association. In the Final Soil Classification, the Kilrush Association is assigned the code 0700b in accordance with Tables A and B. With respect to classification terminology, the reports (3, 4, 5, 11 & 12) describing the predictive mapping programme refer only to soil association codes that relate to the Interim Soil Classification. However, the ISIS Soil Information System contains a translation table that links the interim soil association codes to the codes that relate to the Final Soil Classification. Thus the results of the predictive mapping can be linked to the final version of the New Soil Map of Ireland.

**Table A: Sequencing of the Great Soil Groups (GSG) in the Final Irish Soil Classification**

| Criteria | GSG code | Great Soil Group (GSG) |
|---|---|---|
| Soils with thick organic layers | 1 | OMBROTROPHIC PEAT |
| | 2 | MINEROTROPHIC PEAT |
| Shallow or extremely gravelly soils | 3 | RENDZINAS |
| | 4 | LITHOSOLS |
| Soils influenced by water | 5 | ALLUVIAL SOILS |
| | 6 | GROUNDWATER GLEYS |
| | 7 | SURFACE-WATER GLEYS |
| Soils affected by Fe/Al chemistry increase | 8 | PODZOLS |
| | 9 | BROWN PODZOLICS |
| Soils with clay enriched subsoil | 10 | LUVISOLS |
| Relatively young or soils with limited profile development | 11 | BROWN EARTHS |

For more details of the finalised Irish Soil Classification System please refer to the following documents:

*ISIS Final Technical Report 10: Simo et al. (2014). The Irish Field Handbook for Soil Profile Descriptions. Available from http://erc.epa.ie.safer/reports*

*ISIS Final Technical Report 13: Simo et al. (2014). The Irish Soil Information System Map and Legend. Available from http://erc.epa.ie.safer/reports*

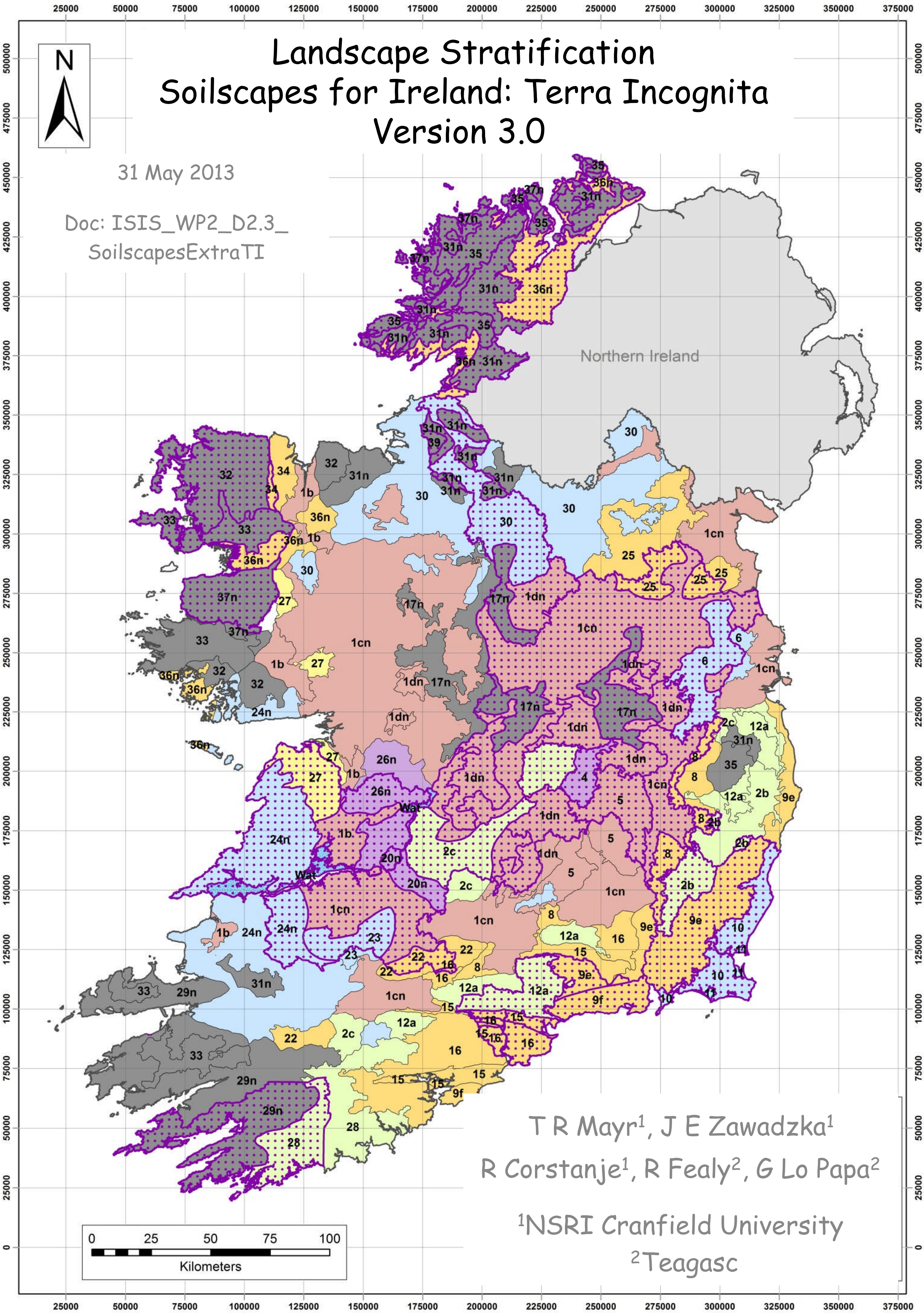
*ISIS Final Technical Report 9: Creamer et al. (2014). The Irish Soil Information System National Soil Series - Description and Classification of Representative Profiles. Available from http://erc.epa.ie.safer/reports*

| Table B Linkage between the Interim and Final Irish Soil Classifications for Soil Subgroups | | | |
|---|---|---|---|
| **Interim SSG_code** | **Interim Soil Subgroup (SSG)** | **SSG code** | **Soil Subgroup (SSG)** |
| 911 | Raw Ombrotrophic Peat Soils | 110 | Natural Ombrotrophic Peat Soils |
| 912 | Earthy Ombrotrophic Peat Soils | 170 | Drained Ombrotrophic Peat Soils |
| 913 | Cut-over Ombrotrophic Peat Soils | 180 | Cut-over Ombrotrophic Peat Soils |
| 914 | Industrial Ombrotrophic Peat Soils | 190 | Industrial Ombrotrophic Peat Soils |
| 921 | Raw Minerotrophic Peat Soils | 210 | Natural Minerotrophic Peat Soils |
| 922 | Earthy Minerotrophic Peat Soils | 270 | Drained Minerotrophic Peat Soils |
| | | 280 | Cut-over Minerotrophic Peat Soils |
| 211 | Typical Rendzinas | 300 | Typical Rendzinas |
| 215 | Histic Rendzinas | 310 | Histic Rendzinas |
| 213 | Humic Rendzinas | 360 | Humic Rendzinas |
| 214 | Stagnic Rendzinas | | |
| 212 | Gleyic Rendzinas | | |
| 111 | Typical Lithosols | 400 | Typical Lithosols |
| 113 | Histic Lithosols | 410 | Histic Lithosols |
| 112 | Humic Lithosols | 460 | Humic Lithosols |
| 821 | Typical Alluvial Gleys | 500 | Typical Alluvial Gley Soils |
| | | 510 | Histic Alluvial Gley Soils |
| 823 | Typical Calcareous Alluvial Gleys | 550 | Typical Calcareous Alluvial Gley Soils |
| | | 551 | Histic Calcareous Alluvial Gley Soils |
| 824 | Humic Calcareous Alluvial Gleys | 556 | Humic Calcareous Alluvial Gley Soils |
| 822 | Humic Alluvial Gleys | 560 | Humic Alluvial Gley Soils |
| 811 | Typical Brown Alluvial Soils | 570 | Typical Alluvial Soils |
| 812 | Gleyic Brown Alluvial Soils | 572 | Gleyic Alluvial Soils |
| 813 | Humic Brown Alluvial Soils | 576 | Humic Alluvial Soils |
| 721 | Typical Groundwater Gleys | 600 | Typical Groundwater Gleys |
| | | 610 | Histic Groundwater Gleys |
| 723 | Calcareous Groundwater Gleys | 650 | Calcareous Groundwater Gleys |
| | | 651 | Histic Calcareous Groundwater Gleys |
| 724 | Humic Calcareous Groundwater Gleys | 656 | Humic Calcareous Groundwater Gleys |
| 722 | Humic Groundwater Gleys | 660 | Humic Groundwater Gleys |
| | | 690 | Anthropic Groundwater Gleys |
| 711 | Typical Surface-water Gleys | 700 | Typical Surface-water Gleys |
| 712 | Humic Surface-water Gleys | 760 | Humic Surface-water Gleys |
| | | 790 | Anthropic Surface-water Gleys |
| 611 | Ferric Podzols | 800 | Typical Podzols |
| 621 | Typical Gley Podzols | 820 | Gleyic Podzols |
| 622 | Stagno-Gley Podzols | 830 | Stagnic Podzols |
| 632 | Iron-pan Stagno Podzols | 843 | Stagnic Iron-pan Podzols |
| 612 | HumoFerric Podzols | 860 | Humic Podzols |
| | | 890 | Anthropic Podzols |
| 631 | Ferric Stagno Podzols | | |
| 511 | Typical Brown Podzolics | 900 | Typical Brown Podzolics |
| 512 | Gleyic Brown Podzolics | 920 | Gleyic Brown Podzolics |
| 514 | Stagnic Brown Podzolics | 930 | Stagnic Brown Podzolics |
| | | 936 | Humi-Stagnic Brown Podzolics |
| 513 | Humic Brown Podzolics | 960 | Humic Brown Podzolics |
| | | 990 | Anthropic Brown Podzolics |
| 411 | Typical Luvisols | 1000 | Typical Luvisols |
| 412 | Gleyic Luvisols | 1020 | Gleyic Luvisols |
| | | 1026 | Humi-Gleyic Luvisols |
| 414 | Stagnic Luvisols | 1030 | Stagnic Luvisols |
| | | 1036 | Humi-Stagnic Luvisols |
| 413 | Humic Luvisols | 1060 | Humic Luvisols |
| 1020 | Technosols | 1090 | Anthropic Luvisols |
| 311 | Typical Brown Earths | 1100 | Typical Brown Earths |
| 312 | Gleyic Brown Earths | 1120 | Gleyic Brown Earths |
| | | 1126 | Humi-Gleyic Brown Earths |
| 314 | Stagnic Brown Earths | 1130 | Stagnic Brown Earths |
| 315 | Humi-stagnic Brown Earths | 1136 | Humi-Stagnic Brown Earths |
| 321 | Typical Calcareous Brown Earths | 1150 | Typical Calcareous Brown Earths |
| 322 | Gleyic Calcareous Brown Earths | 1152 | Gleyic Calcareous Brown Earths |
| 323 | Stagnic Calcareous Brown Earths | 1153 | Stagnic Calcareous Brown Earths |
| | | 1156 | Humic Calcareous Brown Earths |
| | | 1159 | Anthropic Calcareous Brown Earths |
| 313 | Humic Brown Earths | 1160 | Humic Brown Earths |
| | | 1190 | Anthropic Brown Earths |
| | | 1196 | Humi-Anthropic Brown Earths |

# Landscape Stratification
# Soilscapes for Ireland: Terra Incognita
## Version 3.0

31 May 2013

Doc: ISIS_WP2_D2.3_SoilscapesExtraTI





T R Mayr[1], J E Zawadzka[1]
R Corstanje[1], R Fealy[2], G Lo Papa[2]
[1]NSRI Cranfield University
[2]Teagasc

# IRISH SOIL INFORMATION SYSTEM (ISIS)

## Landscape Stratification

# Soilscapes for Ireland: Extrapolation to Terra Incognita

Version 3.0

T.R. Mayr[1], J.E. Zawadzka[1], R. Corstanje[1]

R. Fealy[2], G. Lo Papa[2]

[1]National Soil Resources Institute
School of Applied Sciences
Cranfield University
Cranfield MK43 0AL, UK

[2]Teagasc
Agriculture and Food Research Agency
Ashtown
Dublin, Ireland

Cranfield University
31 May 2013

[Document: ISIS_WP2_D2.3_SoilscapesExtraTI]

# Acknowledgement

This work is part of the Irish Soil Information System (ISIS) Project, managed by Teagasc (the Irish Agriculture and Food Development Authority) and co-funded by the Environmental Protection Agency (EPA) of Ireland through their Science, Technology, Research and Innovation for the Environment (STRIVE) Programme, as part of the National Development Plan 2007-2013.

Document: ISIS_WP2_D2.3_SoilscapesExtraTI

This document should be cited as:

Mayr T.R., Zawadzka J. E., Corstanje R., Lo Papa G., Fealy R. (2012). Soilscapes for Ireland: Extrapolation to Terra Incognita. ISIS Project Technical Report No. ISIS_WP2_D2.3, Cranfield University & Teagasc, 61pp.

## Tables



## Figures



## Appendices

# 1. Introduction

Soilscapes is a term introduced by Buol *et al*. (1973) and conceptually extended by Hole (1978) in the context of pedology, pointing to areas with a homogeneous composition of soils. The need to segment a landscape into soilscapes as the basis for digital soil mapping stems from the tremendous complexity of soil associations in some landscapes (McBratney *et al*., 1991; McSweeny *et al*., 1994; Kuhnt, 1994; Schmidt *et al*., 2010).

Soils in a landscape are associated spatially as well as taxonomically (Hole, 1978). However, spatially associated soils might not be associated taxonomically (Cambell and Edmonts, 1984). Thus, a spatial approach seems appropriate to derive homogeneous, non-fragmented soilscapes with smooth boundaries as a basis for subsequent digital soil-mapping (Schmidt *et al*., 2010)

In a supervised classification process one makes the implicit hypothesis that a soil-landscape is structured in such a way that it can be repeatedly predicted from a specific combination of soil-forming factors (Lagacherie, 2002). Therefore, a forthcoming question is to what extent this hypothesis is true? An attempt to digitally identify a relevant area for extrapolation was carried out by a measure of threshold distance to delineate a representative area where the model could be applied, based on elevation, slope and geology layers (Lagacherie *et al*., 2001).

Another approach involved the rule induction process to detect a relevant physiographic region using entire reference areas as single separate target classes (Bui and Moran, 2003; Grinand *et al*., 2007)

It is unlikely that a direct linear relationship exists between terrain attributes and soil property values; in fact, the relationships between soil property variation and underlying terrain variables can be very complex (Lark, 1999). The linearity and stationarity assumption and the data requirements of these techniques present stiff challenges to their application over large and diverse landscapes (Zhu, 2006).

# 2. Soilscape delineations

Predictive mapping of soil associations, using digital techniques, requires some stratification of the training data to develop models for soil-landscape units rather than treating the whole of the training area (Terra Cognita*)* as a single entity. Models developed within the soilscapes result in a different ranking of the importance of input environmental co-variates depending on the soil-landscape relationship. For example, in upland areas relief derivatives will have greater control over spatially differentiating soil types compared with undulating lowland drift-dominated terrain.

The concept of a soilscape is a broad soil-landscape unit that encompasses a number of soil associations. It groups soils formed primarily on similar substrate types linked to large scale landscape features. However, contrasting soils are included where they are restricted in extent but juxtaposed.

In addition to two manual approaches, two experimental approaches to soilscape delineation were investigated as part of this project, namely cluster analysis based on the LENZ approach and spatial segmentation to assess their capacity to automate the delineation procedure.

An initial soilscapes map (Figure 1) was produced by Brian Kerr based on geology maps, the Soil Association of Ireland and Land Use Potential Soils Bulletin 1980 with the accompanying 1:575,000 scale General Soils Map and consideration of the extent of previous glaciations. Basic regions were identified which were then sub-divided using soil and land-form information. For details see Appendix 1.

**Table 1: Environmental covariates used in cluster analysis**

| | |
|---|---|
| Subsoil - *p* | PET - *c* |
| Habitat - *o* | Prescott Index - *c* |
| DEM - *r* | Potential Soil Moisture Index - *c* |
| Mean annual precipitation - *c* | Mean annual radiation - *c* |
| Mean annual temperature - *c* | Vector Measure Ruggedness - *c* |
| Minimum temperature of coldest month - *c* | Slope - *r* |
| Maximum temperature of warmest month - *c* | Winter radiation - *c* |

In addition to two supervised approaches based on distance and rule induction as well as two unsupervised experimental approaches to soilscape delineation were investigated as part of this project, namely cluster analysis based on the LENZ approach and spatial segmentation to assess their capacity to automate the delineation procedure.

# 3. Unsupervised classification

## 3.1 Cluster analysis

The Land Environments of New Zealand – LENZ – (Leathwick *et al.*, 2003) was developed to describe an environmental classification of New Zealand in order to provide a framework for addressing a range of conservation and resource management issues. For further information on the LENZ classification see Appendix 2. Due to software limitations in PATN, LENZ was defined in two stages:

- In the *first stage*, a non-hierarchical classification technique (ALOC/ALOB) in which memory requirements rise only linear with the number of data points, was used to group together points located close to one another in environmental space. The average environmental values of the groups produced by this process can then be used as inputs to a conventional agglomerative classification process.
- In the second stage, the North and South Island output data set from ALOB was combined to form a matrix with 722 groups, and this was classified using a conventional agglomerative procedure. The PATN module GASO was used to calculate all inter-group. Once the inter-group distances had been calculated, the classification was defined using the flexible UPGMA sorting strategy as implemented in FUSE. The classification consists of 4 levels – Level with 20 classes (national), Level 2 with 100 classes (national/regional), Level 3 with 200 classes (regional) and Level 4 with 500 classes (regional/local).

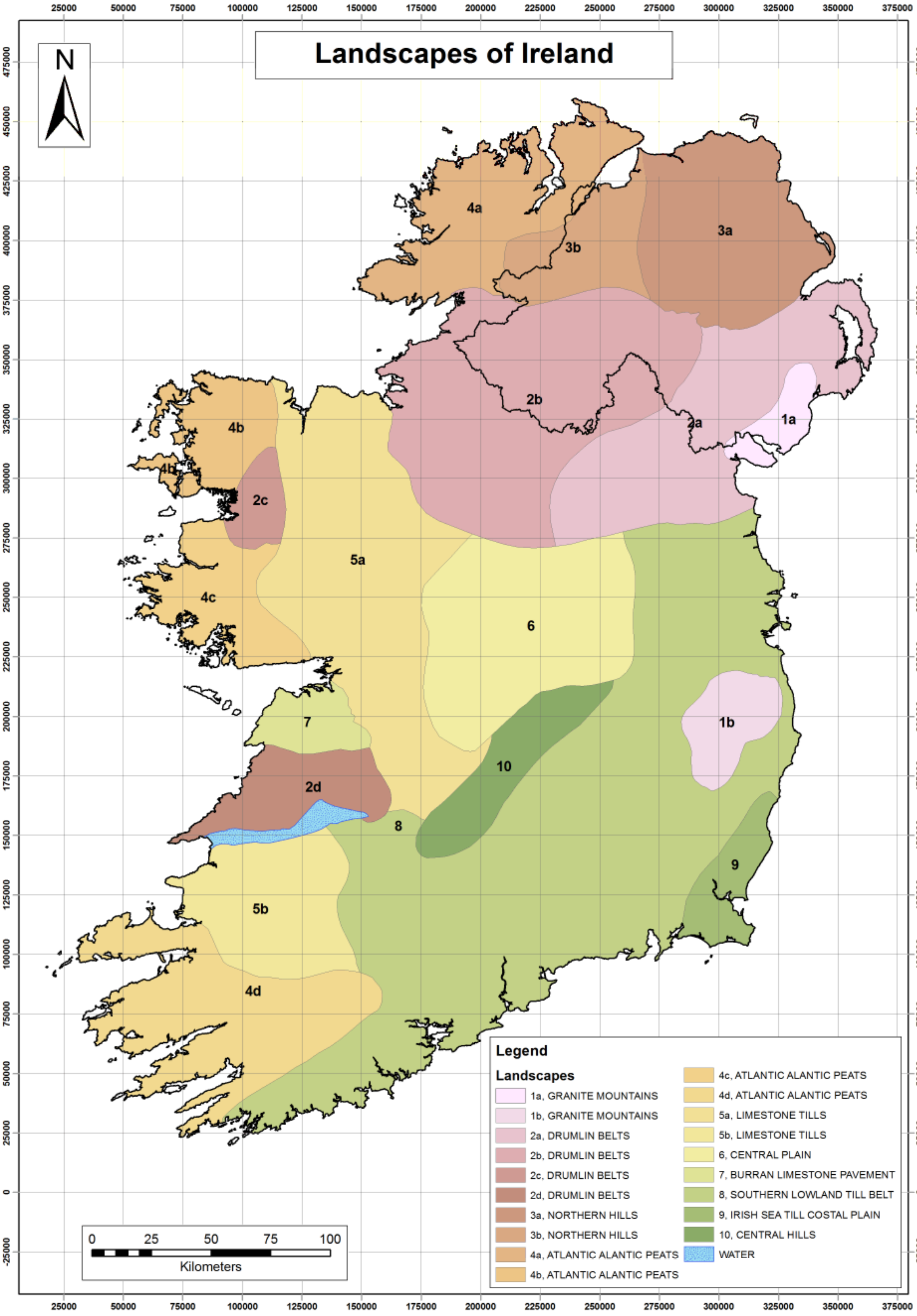



**Figure 1: Soilscapes (Brian Kerr)**

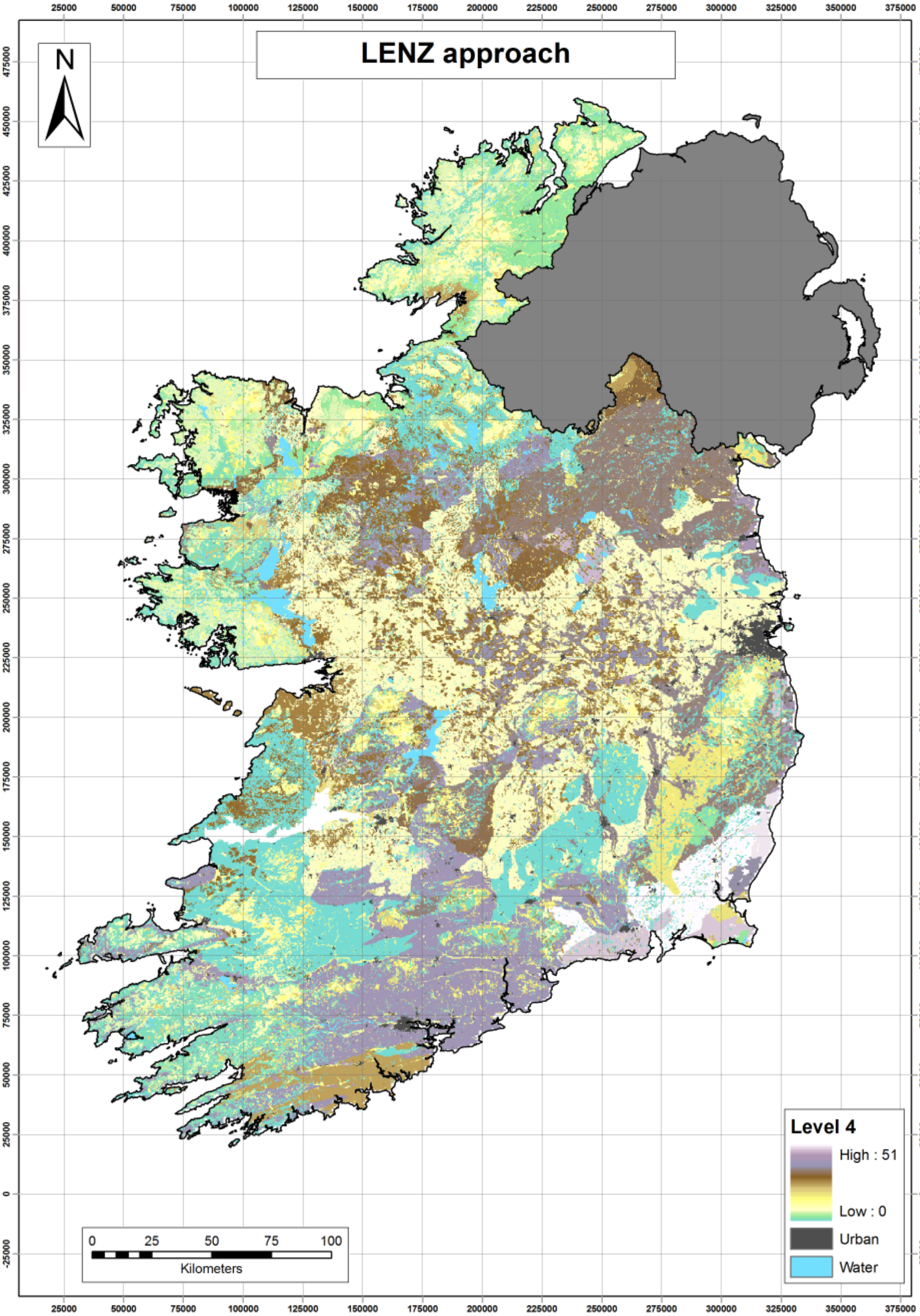



**Figure 2: LENZ Approach (Version_2-Level_4) using Cluster Analysis**

The approach was modified to evaluate the feasibility of determining soilscapes based on a range of environmental covariates covering the *scorpan* factors. In total, 14 environmental covariates were used in the clustering approach (Table 1). Due to software limitations (e.g. 300 clusters in Statistica, 400 clusters in SPSS, 255 clusters in Tanagra and 199 clusters in PATN), all environmental covariates were degraded to 320 m grid resolution resulting in 978'698 sampling points. K-mean-clustering was used for the first stage, followed by hierarchical agglomerate clustering. Four hierarchical levels were identified in the dendrogram with 3, 8, 12 and 50 classes respectively (Appendix 2, Version 2). Clustering results for 50 classes are shown in Figure 2.

## 3.2 Spatial segmentation

Soils in a landscape are associated spatially as well as taxonomically (Hole, 1978). However, spatially associated soils might not be associated taxonomically (Cambell & Edmonts, 1984). Thus, a spatial approach seems appropriate to derive homogenous, non-fragmented soilscapes with smooth boundaries as a basis for subsequent digital soil-sensing and soil-mapping purposes. The approach presented by Schmidt *et al.* (2010) is mainly based on a spatial moving-window analysis of the frequency distribution of the soils and a subsequent k-means cluster analysis on the frequencies. The window size, which determines the smoothness of the boundaries as well as he number of clusters, has to be optimised. Both influence the degree of spatial fragmentation.

The local frequency distribution (spatial arrangement/spatial association) of soil classes is the crucial element of the proposed landscape-segmentation approach. Therefore, a moving – window approach was implemented to compute the local frequency distribution of all soil classes of a map. Each soil-class frequency is then used as a feature in the subsequent cluster analysis.

Using small 'window' sizes apparently results in spatially more fragmented clusters then using larger window sizes. This is due to local noise or hot spots, which will be averaged out with larger window sizes, and therefore, more homogenous clusters will be returned. Thus fragmentation analysis is a key element of the approach introduced.

To determine the optimal window size thus the fragmentation, we plotted the overall perimeter against the cluster numbers. The optimal window size is found when the decrease in perimeter attenuates in relation to larger 'window' sizes.

A bespoke software program was written to extract the soil association composition using different window sizes and increments of moving the window across the landscape. The approach was tested with data from England and Wales (Figure 3).

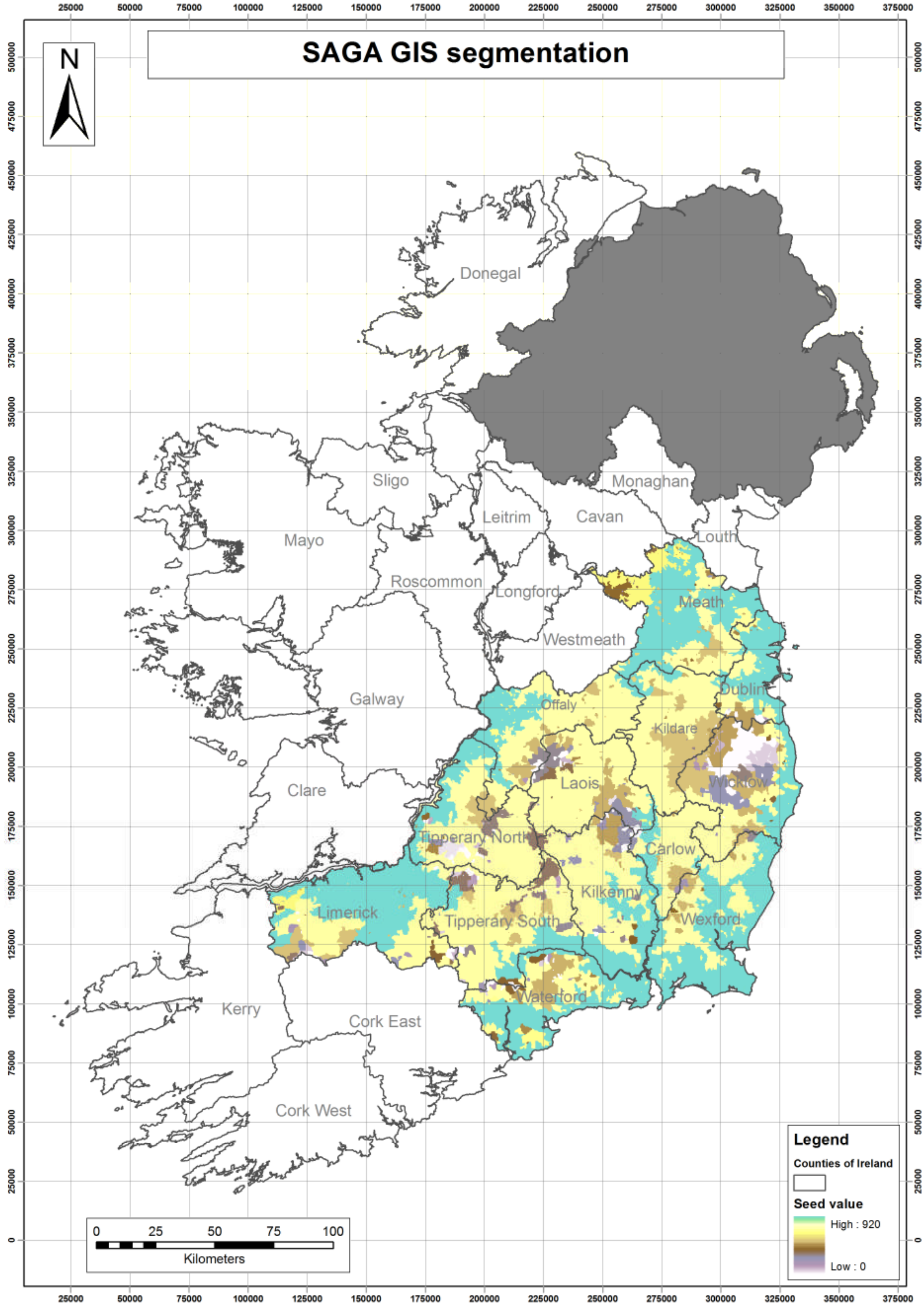



Figure 3: SAGA segmentation

# 4. Supervised classification

The initial maps were superseded once the national soil association map for Terra Cognita became available. The new map describes the national map units, which have been constructed from a generalisation process that harmonised and rationalised the soil series mapped at 1:126,720 scale (Jones *et al*., 2011, 2012). The new map has a nominal mapping scale of 1:250,000. Based on the new map of national soil associations, 35 soilscapes (Figure 4) were manually delineated for Terra Cognita based on expert knowledge according to Hannam *et al*. (2012).

Soilscapes have been predicted for Terra Incognita counties Cavan: Dublin, East Cork, East Mayo, Galway, Kerry, Kilkenny, Louth, Monaghan, Roscommon, Sligo, Tipperary South and Wicklow, based on environmental distance and rule inductions, as follows:.

- Environmental Distance
  - Feature Space Analysis
- Rule Induction
  - Bayesian Networks
  - Random forests
  - Neural Networks

Both the feature space analysis and the statistical modelling approaches establish associations between individual soilscapes and the environmental covariates. The three inference engines employed in the statistical modelling were selected for their different statistical concepts. This was achieved by generation of training data sets, which were attributed with the relevant covariates. Subsequently, predicted soilscapes were assigned to the closest combinations of environmental covariates found in Terra Incognita. The Bayesian Networks approach was run using Netica (v3.24). Random Forests and Neural Networks were run in Data Miner of Statistica 9.

## 4.1 Feature space analysis

Feature space analysis is assessing how closely a piece of land is related to a reference area in terms of environmental feature space. The feature space analysis was based on five covariates, namely the General soil Map, the Subsoil Map, the Geology Map, elevation (altitude) and slope. For each layer used in the analysis, the spatial extent of each class is determined. The individual class extents are converted into percentages which are used as weights. These weights are later used to assess how similar a particular piece of land is to a given reference unit.

The first step of soilscape extrapolation was the calculation of area percentages for each category within each layer for each soilscape found in Terra Cognita. The area percentages obtained were averaged for each soilscape to create the basis for extrapolation of soilscapes into Terra Incognita. In the second step, area percentages calculated in Terra Cognita were attributed to categories for each layer found in Terra Incognita. The methodology is detailed in Table 2. The first two steps were undertaken using the Spatial Analyst extension in ArcGIS™ (v9.3/10.0). The final step was to classify areas of Terra Incognita into soilscapes based on the highest area percentage score. This was achieved using bespoke FORTRAN programming.

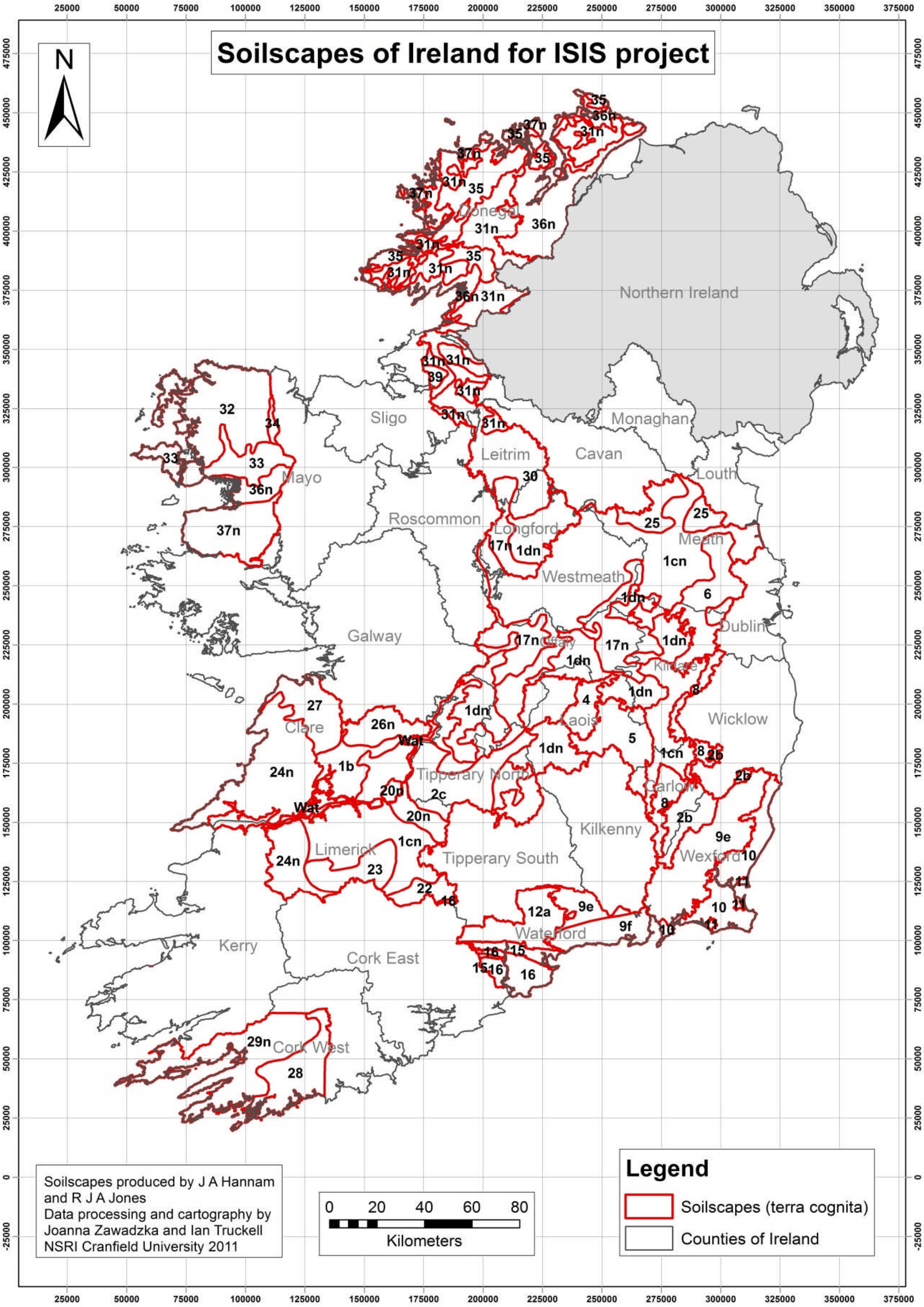



**Figure 4: Soilscapes in Terra Incognita** (after Hannam et al., 2012)

**Table 2: Feature-space analysis**

| **1) Prepare data layers (Terra Cognita)** | **3) Calculate scores** |
|---|---|
| Elevation – re-classify - categorical | Join data layers in Terra Incognita to LUTs |
| Slope – re-classify - categorical | Create composite layer |
| General Soil Map - categorical | Add scores |
| Subsoil Map - categorical | Divide by number of layers |
| Geology – categorical | |
| **2) Prepare LUT** | **4) Post processing** |
| Calculate % size of each class for each data layer | Remove areas less than 16 ha |
| **Prepare data layers (Terra Incognita)** | |
| Elevation – re-classify - categorical | |
| Slope – re-classify - categorical | |
| General Soil Map - categorical | |
| Subsoil Map - categorical | |
| Geology – categorical | |

## 4.2 Bayesian Belief Networks

Belief networks are a vital tool in probabilistic modeling and Bayesian methods. They are one class of probabilistic graphical model. In other words they are a marriage between two important fields: probability theory and graph theory. It is this combination that makes them a powerful methodology within machine learning and statistics.

While statistical models are able to rank variable importance, they do not define the relationship between soilscapes and the predictor variables, and are considered ‘black-box’ techniques. Bayesian belief networks represent a completely different approach to the statistical inference models used to date. BBNs are mathematical models that make predictions based on probabilities using a combination of measured data and expert opinion. They are of interest in that they can work well using limited data and offer a clear description of the relationships and interactions between predictor and target variables. Furthermore, they provide the means to formalise a soil surveyor’s knowledge into a set of rules and probabilities which can, in theory, be applied to regions beyond the study area.

The first stage of the modelling is to create a conceptual model diagram based on the previous modelling results and expert knowledge. The second stage is to allocate each node a set of values (these will be classes for both categorical data and the product of the discretization of continuous variables), Boolean where possible (Hough *et al*., 2010). This relies on the discretization of continuous variables, one of the major challenges involved in BBN modelling, which is usually a decision made using expert knowledge (Uusitalo, 2007).

The next task is to define the relationships between variables using conditional probability, which is the probability of an event A given B. For example the probability of a certain soil type at a given location will be influenced by the presence of a certain parent material and climatic conditions, which needs to be accounted in the model structure. Here the prior conditional probability will be calculated using Bayes rule. For variables which are not influenced by others, such as rainfall, prior unconditional probabilities that is usually a combination of expert knowledge and observed data. BBN also allows a phase of ‘structural learning' which shows, not only the variables that influence soilscapes most strongly, but also the effects variables have on predicting individual soilscapes. The models were generated using NETICA and NETICA API software packages from Norsys Software Corporation.

## 4.3 Random Forests

Random Forest (or random forests) is an ensemble classifier that consists of many decision trees and outputs the class that is the mode of the class's output by individual trees. The algorithm for inducing a random forest was developed by Leo Breiman and Adele Cutler.

Breiman (2001) introduces RF modelling as a method of improving predictions made using a classification and regression tree approach. The key features of this method are that trees are constructed using a bootstrap of the entire dataset and the splits at each node are not made by the best predictor from the entire suite of input variables, but from the best of a randomly selected subset, which prevents overfitting (Liaw & Wiener, 2002). The performance of the model is assessed by predicting the mean square error (MSE) of the 'out of bag' portion of the data at each tree, then averaging over the entire forest.

$$MSE_{OOB} = n^{-1}\sum_{i=1}^{n}(z_i - \hat{z}_i^{OOB})^2$$

where $z_i^{OOB}$ is the mean 'out of bag' prediction for the *i*th observation. RF also provides a measure of fit comparable to the $R^2$ values of the other models. This 'pseudo $R^2$' is labelled the 'percent variance explained' and is calculated using the following formula:

$$Var_{ex} = 1 - \frac{MSE_{OOB}}{\hat{\sigma}_y^2}$$

where $\hat{\sigma}_y^2$ is the total variance of the dependent variable calculated with $n$ as divisor (rather than $n - 1$) (Liaw & Wiener, 2002). The number of trees were set to 100 and the random test data proportion to 30 percent.

An interesting feature of RF is the ability to rank predictor variable in order of importance, which is done by measuring how much the error of the 'out of bag' estimates increases when data for a particular variable is 'removed' from the analysis and the other variables are left intact. This is done on a tree-by-tree basis for the entire forest. The models were generated using the STATISTICA 9 (StatSoft Inc., 2011).

## 4.4 Artificial Neural Networks

Artificial Neural Networks (ANN) constitute an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system.

The principles of ANN are well established (Bishop, 1995) with Maier & Dandy (2001) offering a practical guide for environmental modelling. The structure used for this modelling was a multilayer perceptron, a powerful predictive technique and the one most commonly applied in soil science (Agyare *et al*., 2007). Here, data are separated into a series of nodes, with similar nodes arranged into layers, typically, an input layer (containing the variables used for prediction), an output layer (where predictions are made) and in-between a single hidden layer that weights and transforms the data to extract meaningful relationships.

For each model the data were separated, with 70 per cent used for training, 15 per cent used for testing and 15 per cent used for independent validation. Splitting the data allowed the number of hidden nodes to be tested, which is important as the optimum number of nodes will differ depending on the problem at hand and the number of input variables. It is recommended that the number of hidden nodes should be half the number of input variables plus the number of output variables (Statsoft, Inc., 2011).

Generally, adding more nodes will increase the performance of the model; however, this may lead to over-fitting the data. To avoid this, the AAN use a back-propagation algorithm (Rumelhart *et al*., 1986) to test the performance of the network on both training and testing datasets. Training the network should reduce the 'error function' associated with predictions so when the error function of the testing dataset set plateaus or increases, it indicates the ANN has begun to overfit the data. The error function for regression is Sum of Squares error given below:

$$E_{SOS} = \sum_{i-1}^{N} (y_i - t_i)^2$$

where $N$ is the number of training cases, $y_i$ is the predicted value of the $i^{th}$ case and $t_i$ is the observed value. Ideally, when the differences in the error function are negligible, the network with the fewest nodes is chosen. As the test data set plays a role in developing the ANN infrastructure, a validation data set is used to independently test the predictive power of the models, the best performing were selected using $R^2$ values and RMSE.

ANNs can also rank variables in order of importance, although they use a different method from RFs. Here, data for each variable is replaced in turn by its mean value from the training data and the effect on the error function is recorded. The variables are then ranked by the amount their omission increases the overall error function (Lou & Nakai, 2001). The learning rate for the ANNs was set to 0.1 and the analysis was carried out using STATISTICA 9 (StatSoft Inc., 2011).

**Table 3: Stratified random sampling**

| | | | | | |
|---|---|---|---|---|---|
| s1b | 3,826 | s12a | 6,065 | s29n | 11,045 |
| s1cn | 62,864 | s15 | 1,611 | s30 | 12,711 |
| s1dn | 28,645 | s16 | 3,574 | s31n | 6,744 |
| s2b | 4,777 | s17n | 14,378 | s32 | 13,218 |
| s2c | 12,584 | s20n | 2,553 | s33 | 6,149 |
| s4 | 2,726 | s22 | 2,082 | s34 | 519 |
| s5 | 6,617 | s23 | 4,423 | s35 | 2,471 |
| s6 | 7,514 | s24n | 16,788 | s36n | 2,495 |
| s8 | 5,147 | s25 | 4,705 | s37n | 7,575 |
| s9e | 15,503 | s26 | 6,313 | s39 | 455 |
| s9f | 4,026 | s27 | 5,236 | | 298,466 |
| s10 | 7,069 | s28 | 6,058 | | |

**Table 4: Environmental Covariates (scorpan factors)**

| **s - Soil** | **r- Relief** |
|---|---|
| General Soil Map | LPIT2PEAK |
| **c - Climate** | PCTZ2TOP |
| Annual mean temperature | Protection Index (Reliefenergie) |
| Annual mean precipitation | SAGA Wetness Index (t=5, floating point)) |
| Annual mean solar radiation | Surface Curvature Index |
| Annual Mean evapotranspiration | ZCR2ST |
| Thornthwaite Global Humidity Index | ZPIT2PEAK |
| Potential Soil Moisture Deficit | ZTOP2PIT |
| **o - Organism** | **p – Parent material** |
| CORINE landcover map | Geology map (Version 2) |
| Habitat map | Subsoil map (Version 4) |
| **r- Relief** | Actual/potential drainage density ratio |
| DEM normalized by GSM soil units | **n – Landscape position** |
| Dissection Index (King, 1972) | Hammond (subclasses) |
| Downslope Flowpath Length | Iwahashi (16 classes) |
| DTM (Version 0) | SOTER |
| PMIN2MAX | Multi Resolution Valley Bottom Flatness Index |
| Elevation above pit | Catchment classification ( 7 classes) |

# 5.0 Inference models (Terra Cognita)

## 5.1 Training data

Training databases were developed linking the soilscapes with the environmental covariates.

***Reference area***
- Terra Cognita

***Target unit***
- Soilscapes

***Sample selection***
- Stratified Random Sampling (Table 3)

***Sample size***
- 10 observations per $km^2$
- 298,466 observations

***Environmental covariates***
- 32 environmental covariates (Table 4)

***Masking***
- Remove urban, industrial complexes, surface water, peat and rock
- Areas adjacent to rivers (2 pixels each side) were excluded to avoid artefacts from ‘burning’ the river network into the DTM

***Cleaning***
- Removing cases with ‘empty’ fields (i.e. geology)

***Re-coding***
- Re-coding categorical data layers as both subsoil maps and geology had comas in their fields

A bespoke Java program was used for sampling purposes, and spatial analyses were performed using ArcGIS™ (v9.3/10).

## 5.2 Model development

4 different modelling approaches were developed based on the inference engines discussed above as well as using feature space assessment.

### Feature Space Analysis

The 5 look-up-tables for each of the 37 soilscapes (Figure 3) were generated from the relevant environmental covariates. Elevation and slope were ‘binned’ into 12 and 8 classes respectively (**Table 5**).

- 5 environmental covariates (Table 2)
- Training accuracy not established
- No relative importance measure of environmental covariates
- Training accuracy of 30% (Table 6; Appendix 4)

**Table 5: Classification intervals for Feature Space analysis**

| Elevation [m] | | |
|---|---|---|
| Category | Min | Max |
| Class 1 | -5 | 25 |
| Class 2 | 25 | 50 |
| Class 3 | 50 | 75 |
| Class 4 | 75 | 100 |
| Class 5 | 100 | 125 |
| Class 6 | 125 | 150 |
| Class 7 | 150 | 175 |
| Class 8 | 175 | 200 |
| Class 9 | 200 | 300 |
| Class 10 | 300 | 400 |
| Class 11 | 400 | 500 |
| Class 12 | 500 | 1100 |

| Slope [%] | | |
|---|---|---|
| Category | Min | Max |
| Class 1 | 0 | 2 |
| Class 2 | 2 | 5 |
| Class 3 | 5 | 12 |
| Class 4 | 12 | 20 |
| Class 5 | 20 | 27 |
| Class 6 | 27 | 47 |
| Class 7 | 47 | 70 |
| Class 8 | 70 | max |

**Table 6: Accuracy assessment for FS, BN, NN and RF**

| Sample type | FSA | BN | NN1 | NN2 | RF |
|---|---|---|---|---|---|
| Train | 29.60 | **75.29** | 10.05 | 12.12 | 58.83 |
| Test | | | 10.79 | 12.56 | 58.59 |
| Validation | | | 10.94 | 12.81 | |

**Table 7: Accuracy assessment for BN networks**

| | error | accuracy | co-variates | | |
|---|---|---|---|---|---|
| with climate | 24.71 | 75.29 | 38 | | |
| without climate | 27.26 | 72.74 | 32 | | |
| scorpan | 19.12 | 80.88 | 12 | | |
| with climate | 17.99 | 82.01 | 8 | stepwise selection | |
| without climate | 19.46 | 80.54 | 8 | stepwise selection | |

## Bayesian Belief Networks

Initially five different Bayesian Belief Networks were constructed to assess different combinations of environmental covariates (Appendix 5). In all networks the numerical data were binned into 30 classes where possible.

**Network 1**

- 'with climate' layers
- **38** environmental covariates
- Training accuracy of **75%** (Table 7; Appendix 3)
- Relative o importance of environmental covariates are listed in Appendix 5
- User accuracies are listed in Appendix 7

**Network 2**

- 'without climate' layers
- ***32*** environmental covariates
- Training accuracy of **73%** (Tables 6 & 7, Appendix 3)
- Relative o importance of environmental covariates are listed in Appendix 5
- User accuracies are listed in Appendix 3

**Network 3**

- SCORPAN layers
- **12** environmental covariates
- Training accuracy of **81%** (Table 7 & Appendix 3)
- Relative importance of environmental covariates are also listed in Appendix 3
- User accuracies are listed in Appendix 3

**Network 4**

- 'with climate' layers – stepwise selection
- **8** environmental covariates
- Training accuracy of **82%** (Table 7 & Appendix 3)
- Relative o importance of environmental covariates are also listed in Appendix 3
- User accuracies are listed in Appendix 3

**Network 5**

- 'without climate' layers – stepwise selection
- **8** environmental covariates
- Training accuracy of **81%** (Table 7 & Appendix 3**)**
- Relative importance of environmental covariates are listed in Appendix 3
- User accuracies are listed in Appendix 3

## Neural Networks

Two Neural Networks were extracted based on 31 environmental covariates. In contrast to both Bayesian Networks and Random Forests, the relative importance of the environmental covariates is not provided.

**Network 1:**

- **30** environmental covariates
- Training accuracy of **10 %** (Table 6)
- No relative importance measure of environmental covariates
- Validation accuracy of **11%** (Table 6)
- User accuracies are listed in Appendix 4

**Network 2:**

- **30** environmental covariates
- Training accuracy of **12**%(Table 6)
- No relative importance measure of environmental covariates
- Validation accuracy of **13%** (Table 6)
- User accuracies are listed in Appendix 4

## Random Forests

A single Random Forest model was extracted.

- ***31*** environmental covariates
- Training accuracy of **59%** (Table 6)
- Relative importance of environmental covariates are listed in Appendix 6
- Validation accuracy of **59%** (Table 6)
- User accuracies are listed in Appendix 4

# 6. Predictive mapping of Soilscapes in Terra Incognita

Counties for which no detailed soil survey maps exist encompass Cavan, Dublin, East Cork, East Mayo, Galway, Kerry, Kilkenny, Louth, Monaghan, Roscommon, Sligo, Tipperary South and Wicklow.

## 6.1 Deployment data

Deployment data were compiled for Terra Incognita as follows

- 20 m grid sampling
- Urban, surface water, peat and rock and river buffer removed
- Areas adjacent to rivers (2 pixels each side) were excluded to avoid artefacts from “burning” the river network into the DTM
- Empty fields in the categorical data were marked with “no data”
- Data blocks of 500k points extracted from N to S
- Data blocks separated into individual soilscape files
- Soilscape files packaged into 250k blocks for processing

## 6.2 Deployment

The Feature Space Analysis was implemented using spatial modelling in the ArcGIS by applying the LUT’s for each of the soilscapes to Terra Incognita. The results were combined in a text file and processed using a dedicated FORTRAN programme. The results were mapped in ArcGIS. For the Bayesian Belief Networks, Neural Networks and Random Forests, the respective rules were applied to the deployment files and the results mapped in ArcGIS™.

Deployment resulted in the following maps:

- Feature Space analysis (Figure 5),
- Bayesian Belief Network (Figure 6),
- Neural Network 1 (Figure 7),
- Neural Network 2 (Figure 8)
- Random Forest (Figure 9).

## 6.3 Post-processing

A generalisation procedure was applied to up-scale the predicted soilscapes in *Terra Incognita* to a scale of 1:1,100,000 to achieve harmony with the soilscapes delineated for *Terra Cognita*. A bespoke ArcGIS™ tool was written by G. LoPapa (Teagasc) to undertake this process. Post-processing resulted in the following maps:

- Feature Space analysis (Figure 10),
- Bayesian Belief Network (Figure 11),
- Neural Network 1 (Figure 12),
- Neural Network 2 (Figure 13
- Random Forest (Figure 14).

**Table 8: Soilscapes mapped in Terra Incognita by the four different modelling approaches** (raw)

| Label | BN | FS | NN1 | RF1 |
|---|---|---|---|---|
| 1a | | 1a | | |
| 1b | 1b | 1b | 1b | |
| 1cn | 1cn | 1cn | 1cn | 1cn |
| 1dn | 1dn | 1dn | 1dn | 1dn |
| 2b | 2b | 2b | 2b | 2b |
| 2c | 2c | 2c | 2c | 2c |
| 4 | | 4 | 4 | |
| 5 | 5 | 5 | 5 | 5 |
| 6 | 6 | 6 | 6 | 6 |
| 8 | 8 | 8 | 8 | |
| 9e | 9e | 9e | 9e | 9e |
| 9f | 9f | 9f | 9f | |
| 10 | 10 | 10 | 10 | 10 |
| 11 | | | | |
| 12a | 12a | 12a | 12a | 12a |
| 15 | 15 | 15 | 15 | |
| 16 | 16 | 16 | 16 | |
| 17n | 17n | 17n | 17n | 17n |
| 20n | 20n | 20n | 20n | |
| 22 | 22 | 22 | 22 | |
| 23 | 23 | 23 | 23 | |
| 24n | 24n | 24n | 24n | 24n |
| 25 | 25 | 25 | 25 | 25 |
| 26n | 26n | 26n | 26n | |
| 27 | 27 | 27 | 27 | 27 |
| 28 | 28 | 28 | 28 | |
| 29n | 29n | 29n | 29n | 29n |
| 30 | 30 | 30 | 30 | 30 |
| 31n | 31n | 31n | 31n | 31n |
| 32 | 32 | 32 | 32 | 32 |
| 33 | 33 | 33 | 33 | 33 |
| 34 | 34 | 34 | 34 | |
| 35 | 35 | | 35 | |
| 36n | 36n | 36n | 36n | |
| 37n | 37n | 37n | 37n | 37n |
| 39 | 39 | | 39 | |
| | 33 | 33 | 34 | 19 |

# 7. Support assessment

For Terra Incognita, predictive maps, on 20 m resolution, were generated using Feature Space Analysis, Bayesian Networks, Neural Networks and Random Forests (Figures 5 to 9) and subsequently generalised to 1:1,100,000 scale (Figures 10 to 14). Table 8 lists the soilscapes mapped in Terra Incognita by the four methods.

## 7.1 Methodology

### Support assessment (generic)

The support assessment is expressed as the number of environmental covariates available in Terra Incognita compared with the relevant reference soilscape in Terra Cognita. The analysis is based on the number of classes in terms of categorical data and range in terms of numerical data. The generic support assessment is independent of the model used for predicting soilscapes in Terra Incognita. In this case, all categorical data were analysed and map units with less than 298 cases (0.1% of the training data) were removed. 99.9% confidence interval was used for all numerical data to determine the range of the data. A breakdown for both categorical and numerical data used in the analysis can be found in Appendix 5 for Stage 1 (5 covariates) and Appendix 6 for Stage 2 (31 covariates)

### Deployment assessment

Deployment assessment is expressed in terms of the environmental distance for the Feature Space analysis, beliefs for each state of the response node in the case of Bayesian Networks and posterior (final) prediction probabilities for each response category in the case of Random Forests. In both cases the relevant values for the dominant association are used in the analysis.

## 7.2 Support assessment (generic)

**Stage 1**

- Reference area: Terra Cognita
- Classes: 0-5 (Figure 15)

**Stage 2**

- Reference area: Terra Cognita
- Class: 0-31 (Figure 16)

## 7.3 Deployment assessment

**Stage 1**

- FS: environmental distance of dominant class (Figure 17)

**Stage 2**

- BN: probability of dominant class (Figure 18)
- RF: probability of dominant class not available
- NN: probabilities of dominant class not available

# 8. Thematic assessment

Table 8 lists the soilscapes mapped in Terra Incognita by the four methods. Statistical assessment of the 4 models can be found in Appendix 9.

## Feature Space Analysis

- Extrapolation (pre-processing) predicted in Terra Incognita 33of the 35 soilscapes in Terra Cognita (Table 8).
- Extrapolation (post-processing) predicted in Terra Incognita 24 of the 35 soilscapes inTerra Cognita.
- The spatial extent of different soilscapes in Terra Incognita as predicted by Feature Space Analysis is listed in Table 9.

## Bayesian Belief Networks

- Extrapolation (pre-processing) predicted in Terra Incognita 33 of the 35 soilscapes in Terra Cognita (Table 8**)**.
- Extrapolation (post-processing) predicted in Terra Incognita 30 of the 35 soilscapes in Terra Cognita.
- The spatial extent of different soilscapes in in Terra Incognita as predicted by Bayesian Belief Networks is listed in Table 9.

## Neural Networks (1)

- Extrapolation (pre-processing) predicted in Terra Incognita 34 of the 35 soilscapes in Terra Cognita (Table 8).
- Extrapolation (post-processing) predicted in Terra Incognita 27 of the 35 soilscapes in Terra Cognita.
- The spatial extent of different soilscapes in Terra Incognita as predicted by Neural Networks is listed in Table 9.

## Random Forests

- Extrapolation (pre-processing) predicted in Terra Incognita 19 of the 35 soilscapes in Terra Cognita (Table 8).
- Extrapolation (post-processing) predicted in Terra Incognita 12 of the 35 soilscapes in Terra Cognita.
- The spatial extent of different soilscapes in Terra Incognita as predicted by Random forests is listed in Table 9

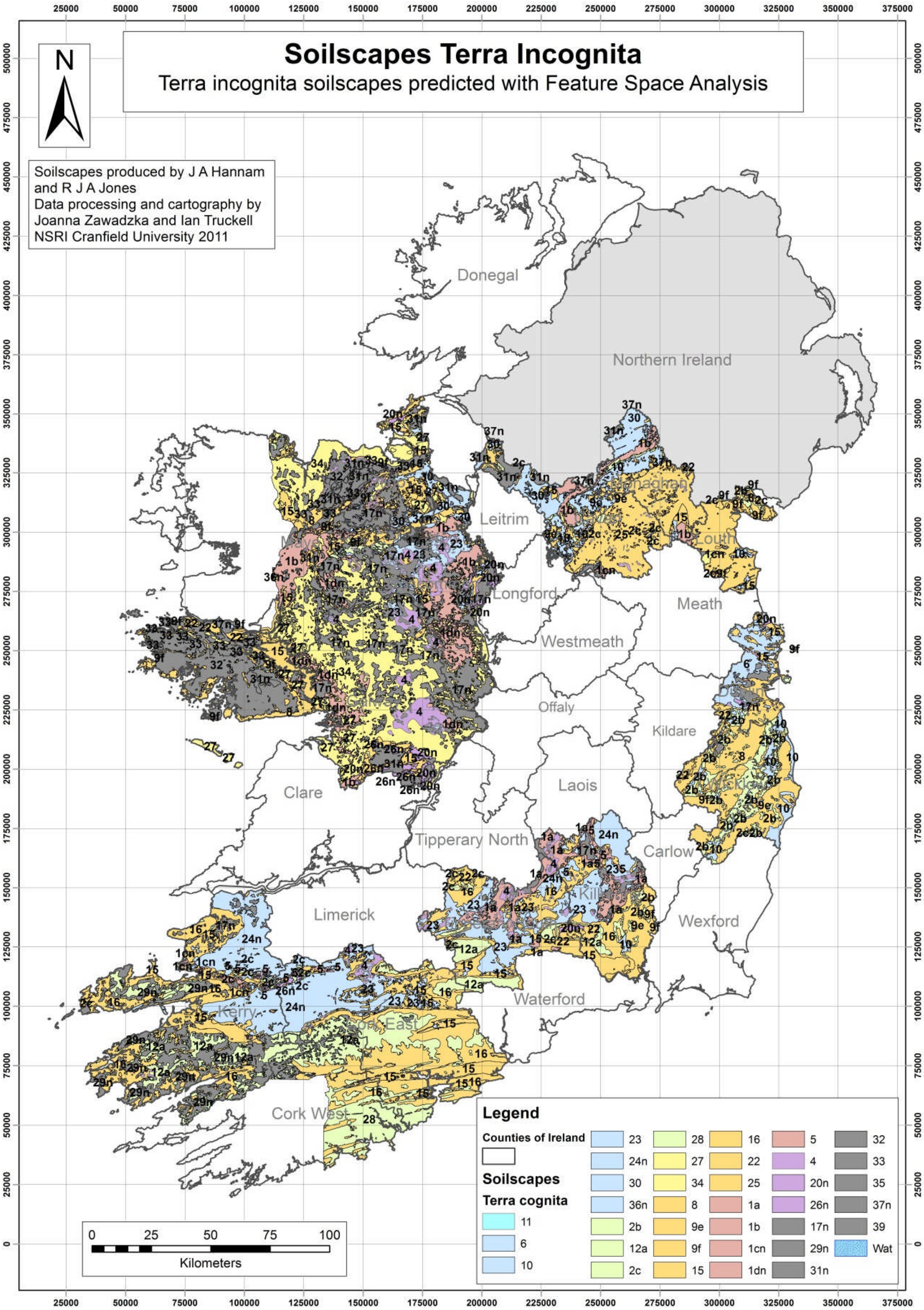



**Figure 5: Feature Space Analysis (raw predictions)**

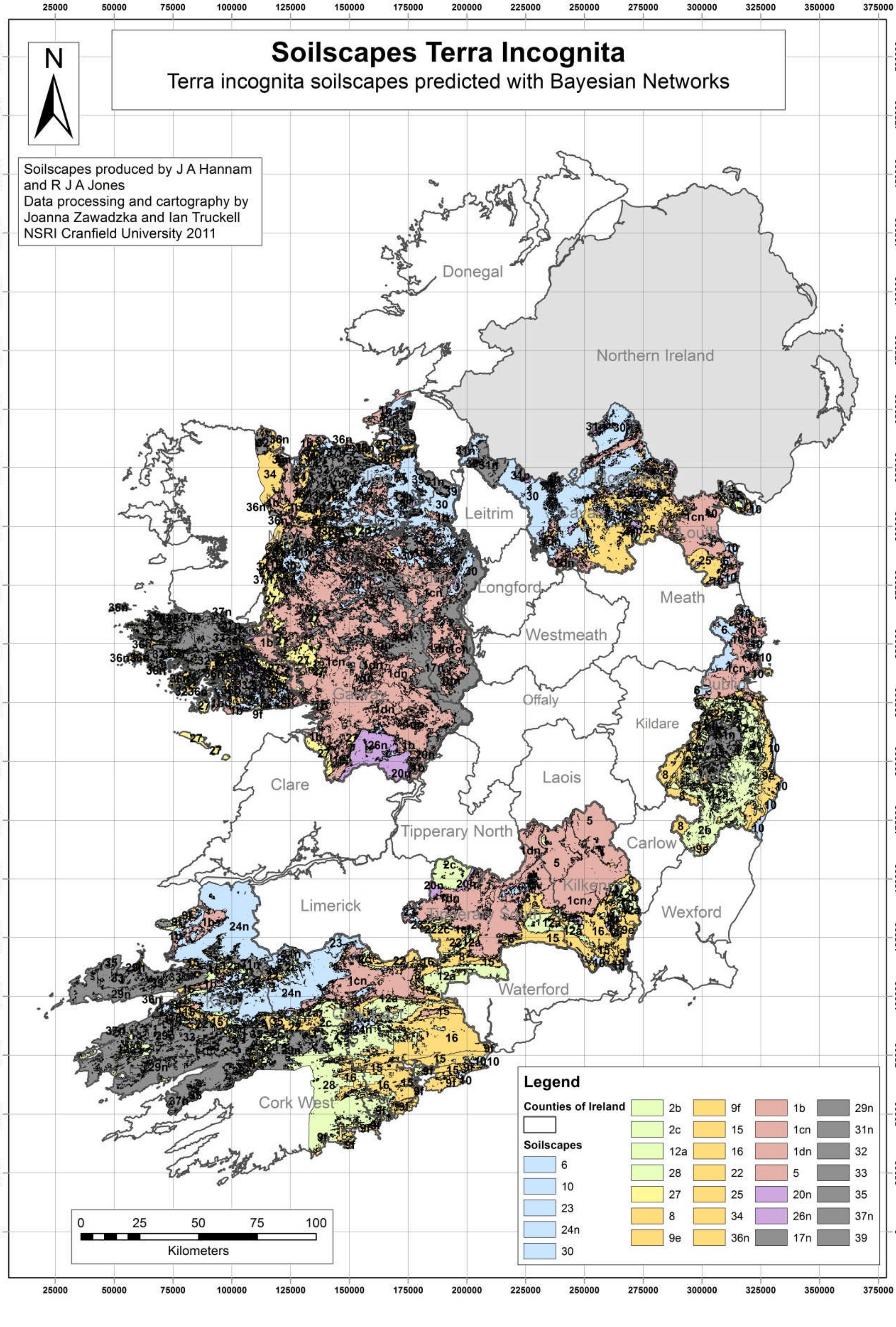



**Figure 6: Bayesian Belief Network (raw predictions)**

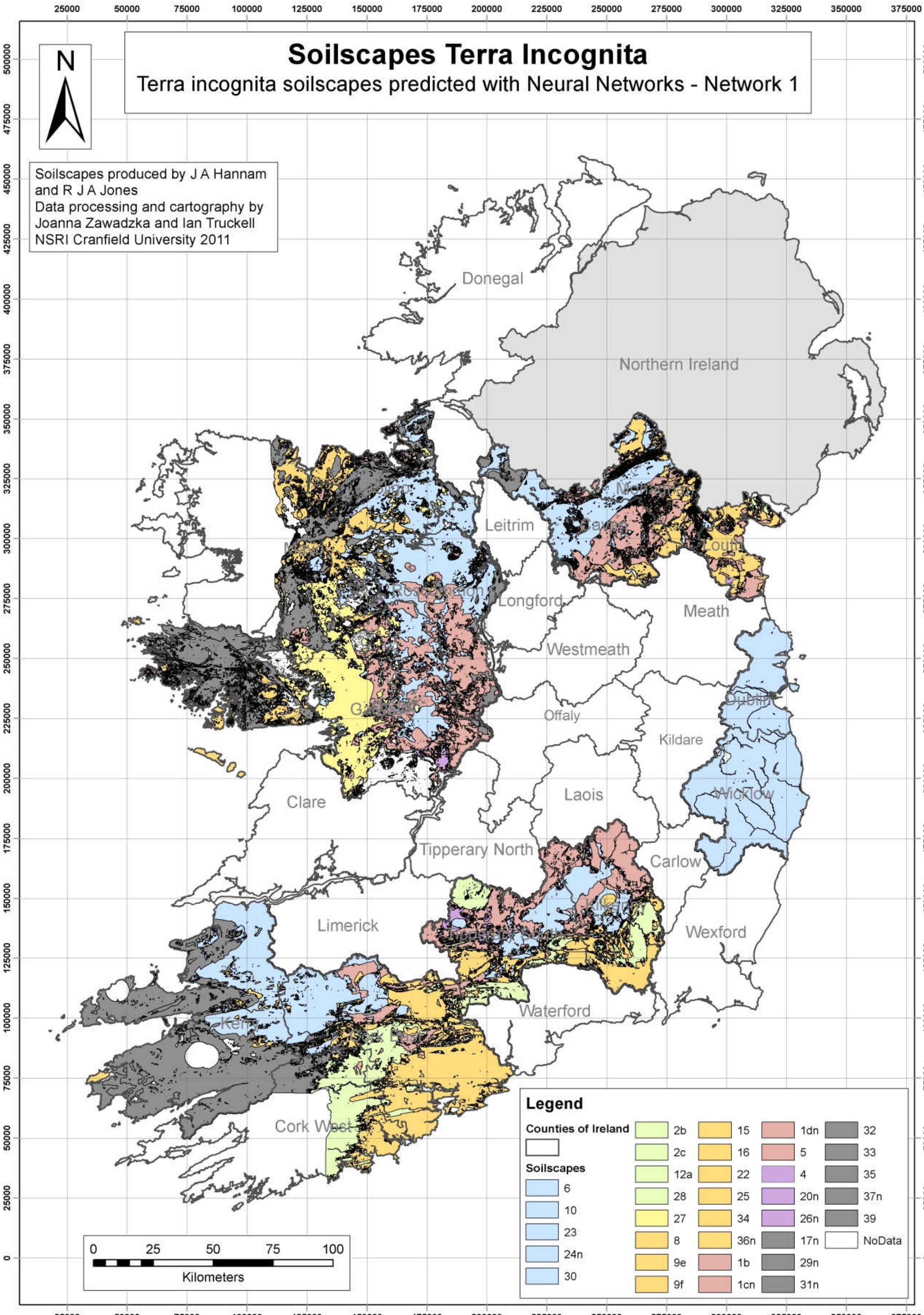



**Figure 7: Neural Networks (1) (raw predictions)**

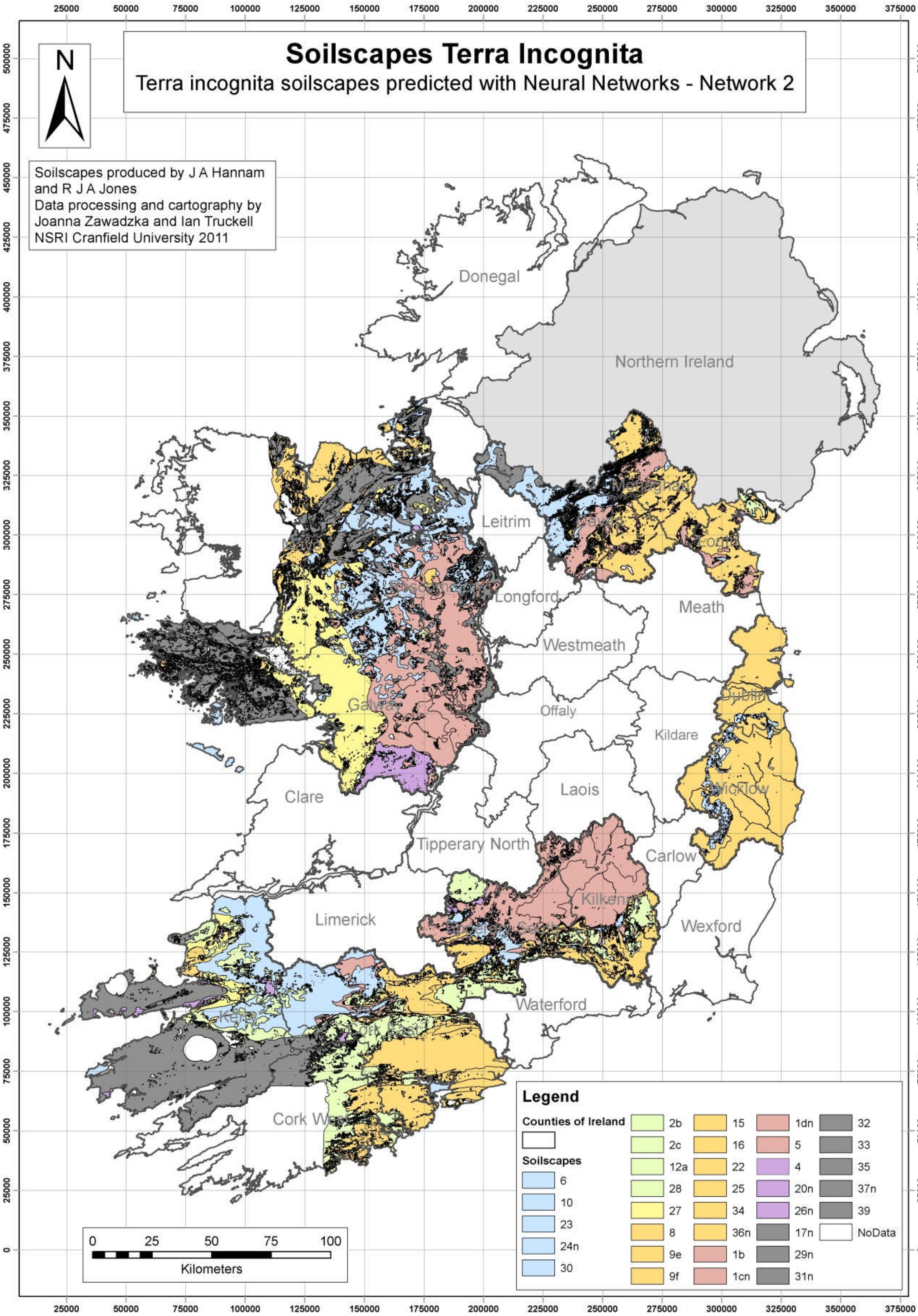



**Figure 8: Neural Networks (2) (raw predictions)**

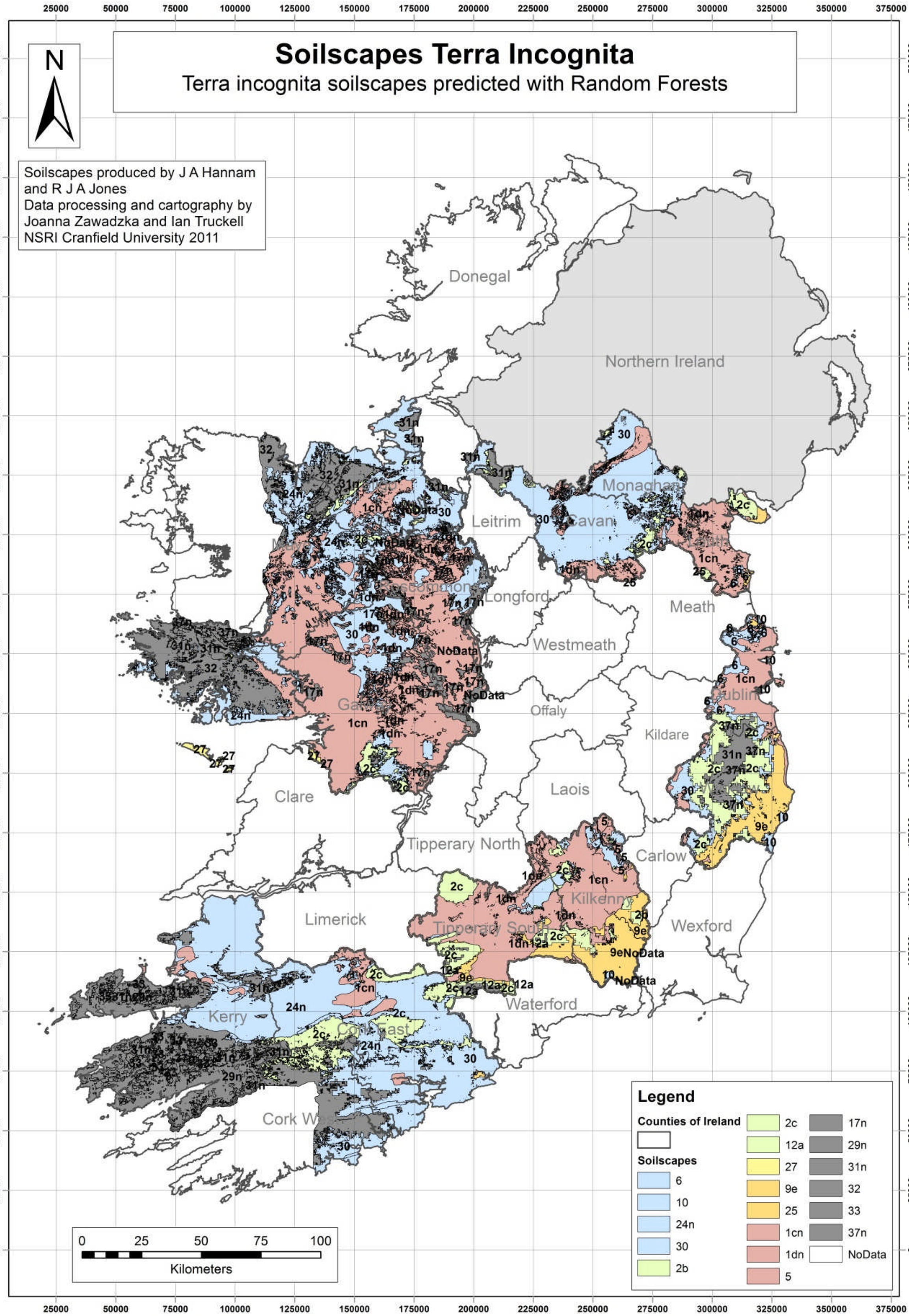



**Figure 9: Random Forests (raw predictions)**

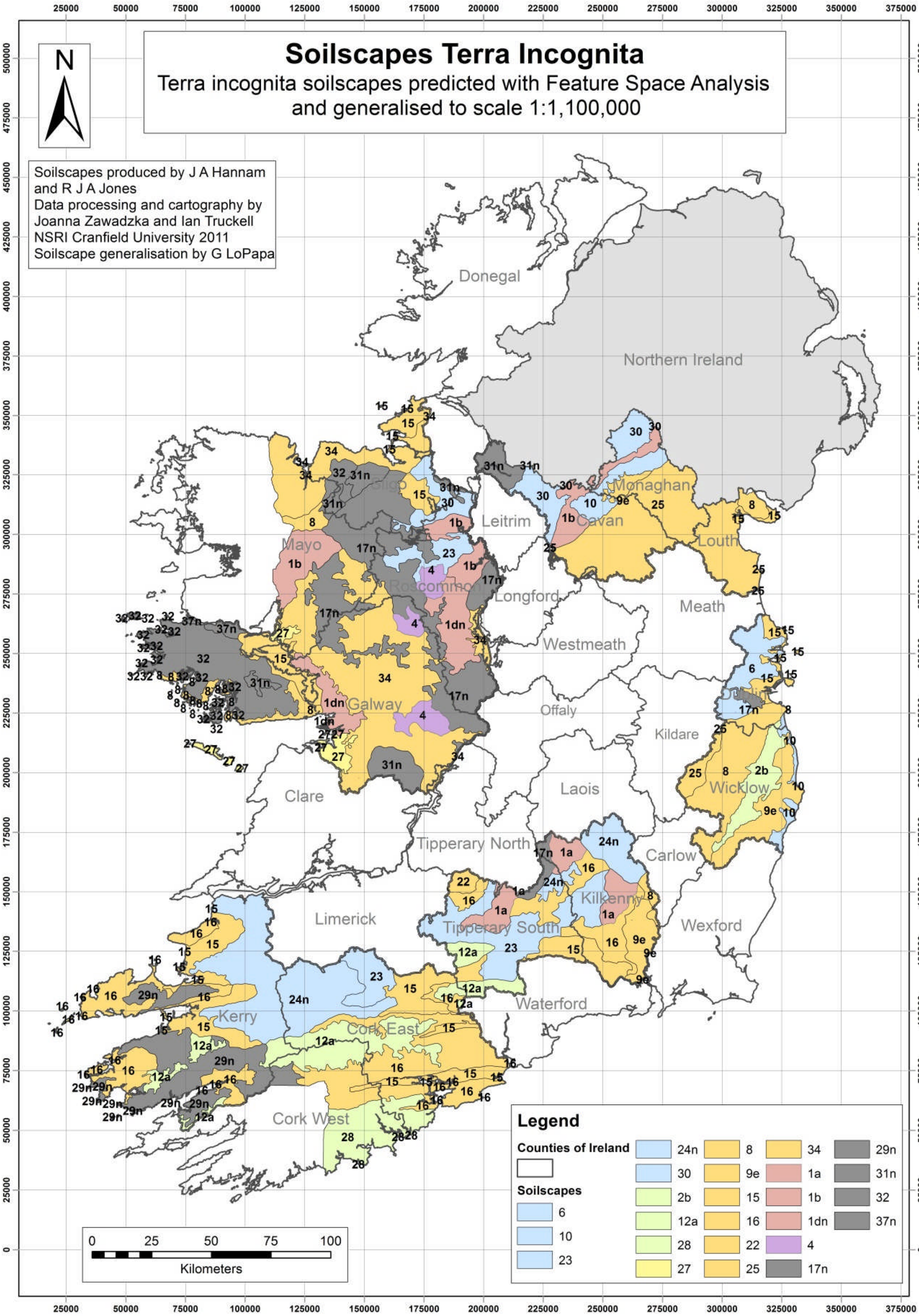



**Figure 10: Feature Space Analysis (generalised)**

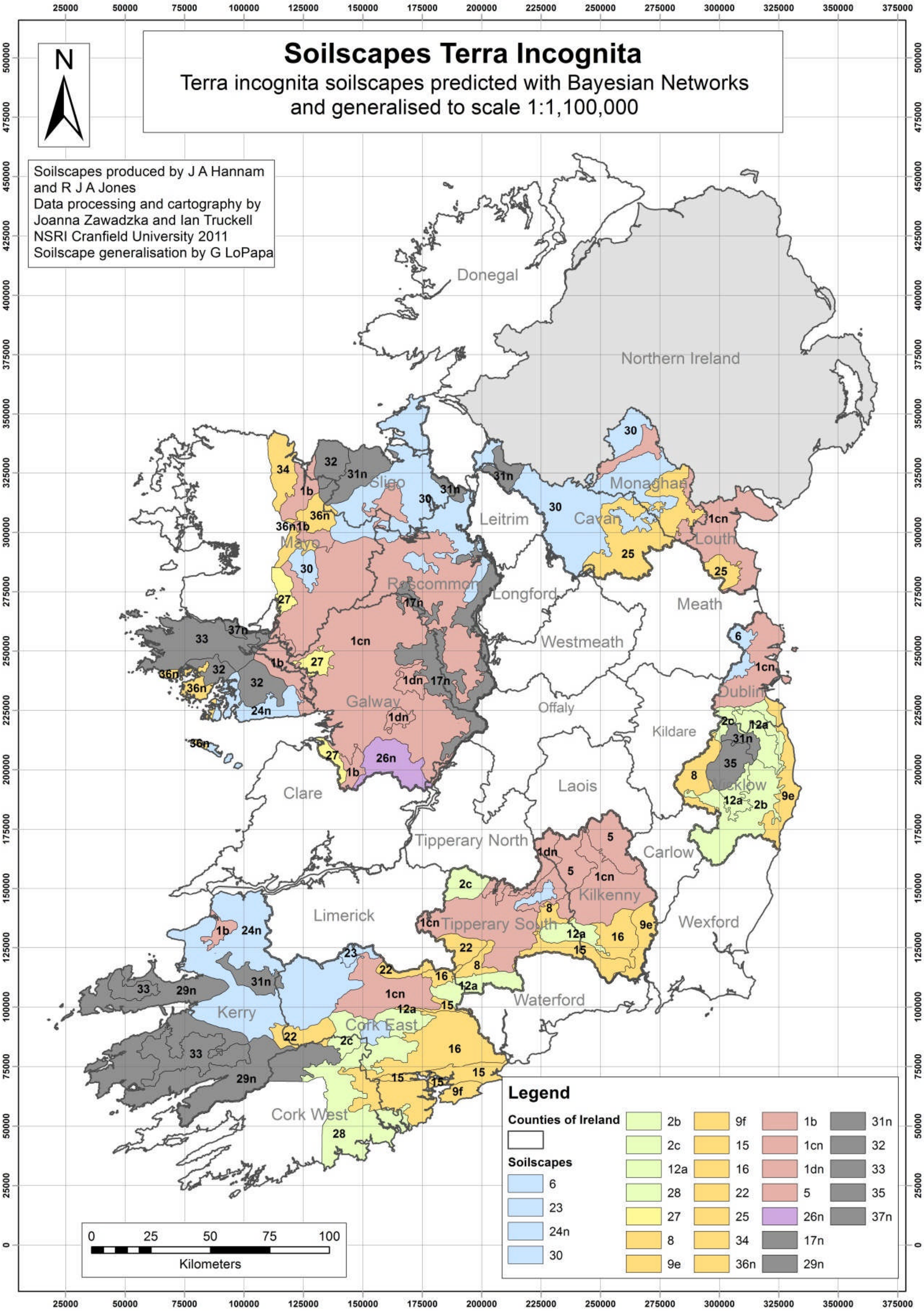



**Figure 11: Bayesian Belief Network (generalised)**

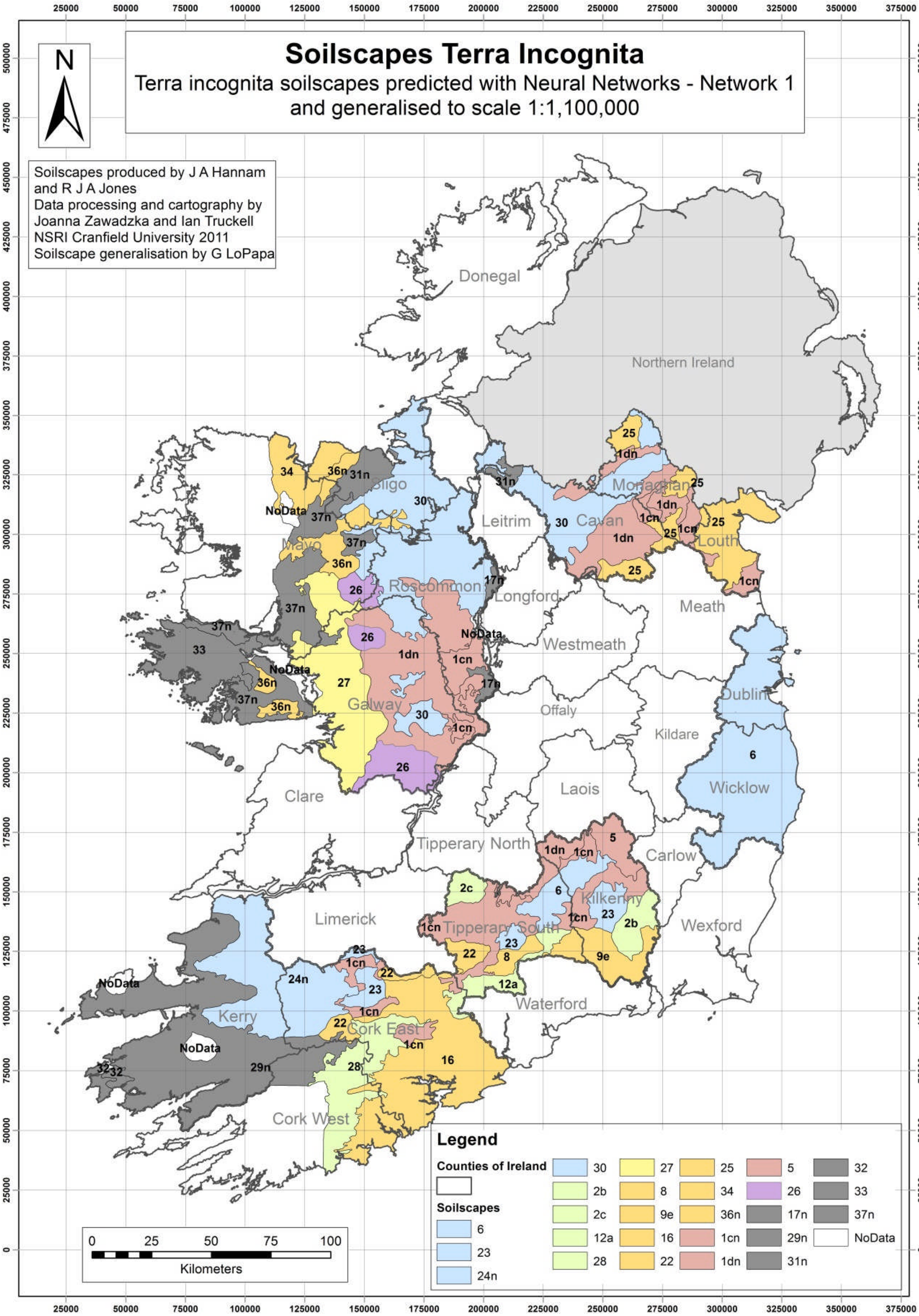



**Figure 12: Neural Network 1 (generalised)**

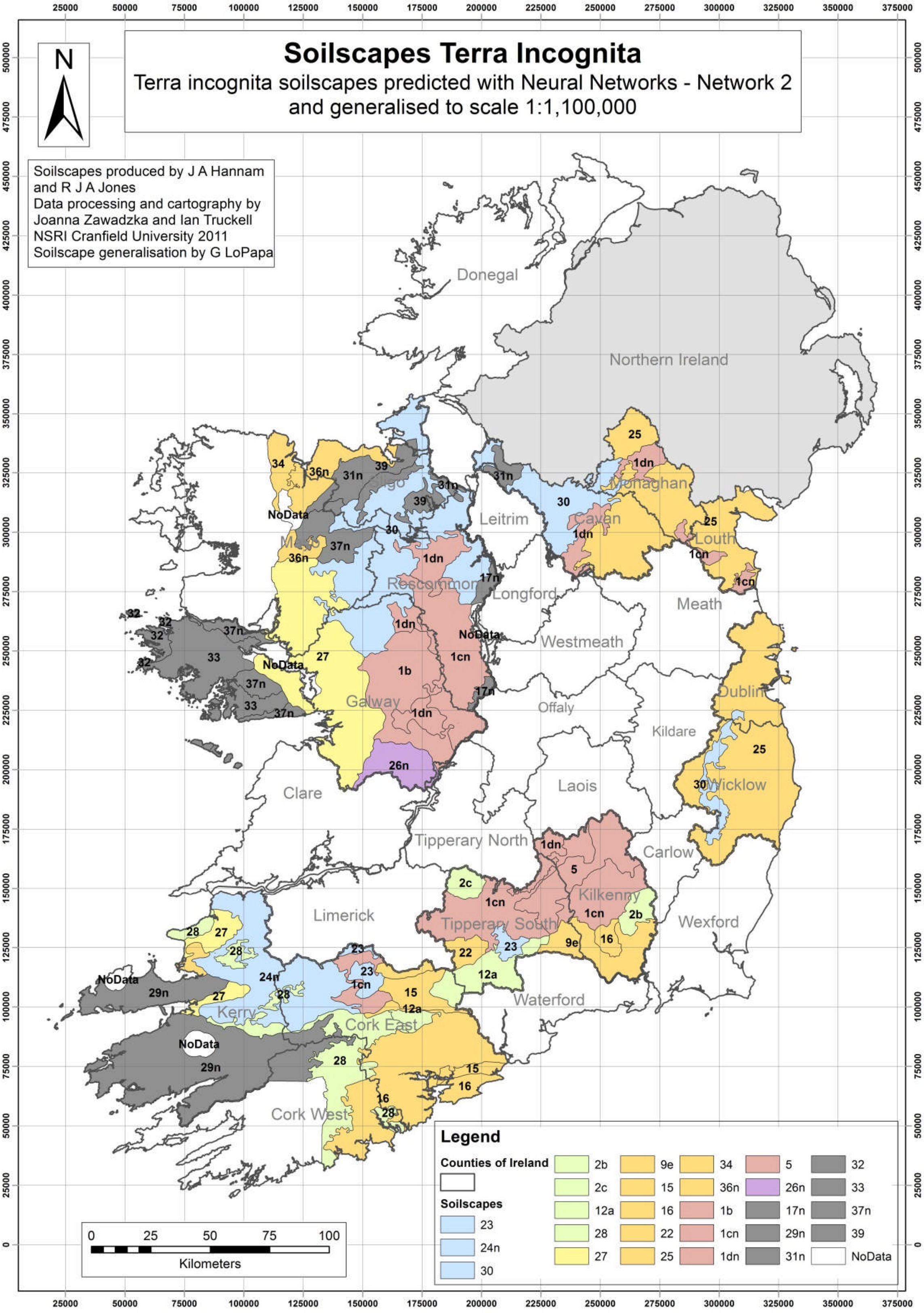



Figure 13: Neural Network 2 (generalised)

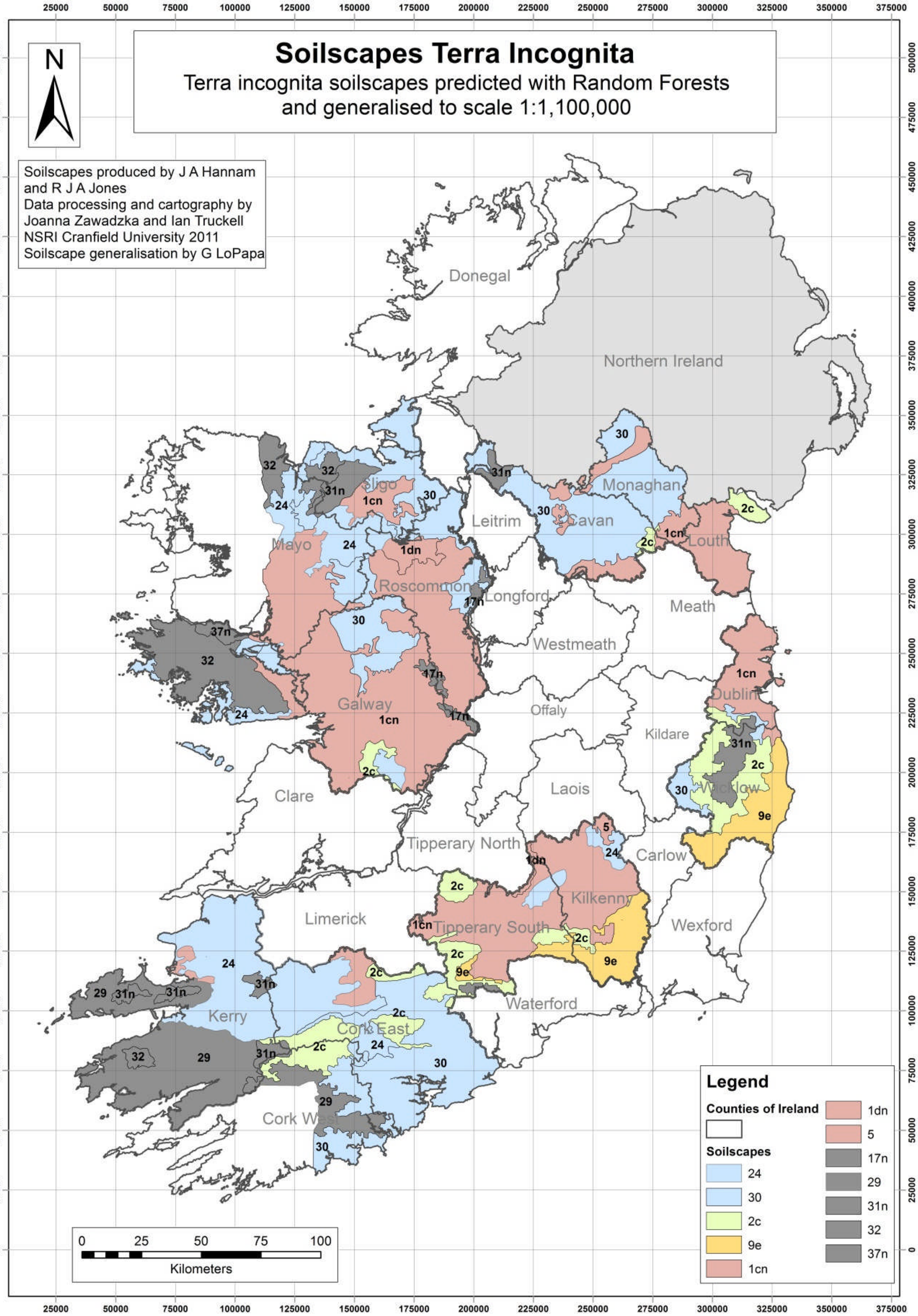



**Figure 14: Random Forest (generalised)**

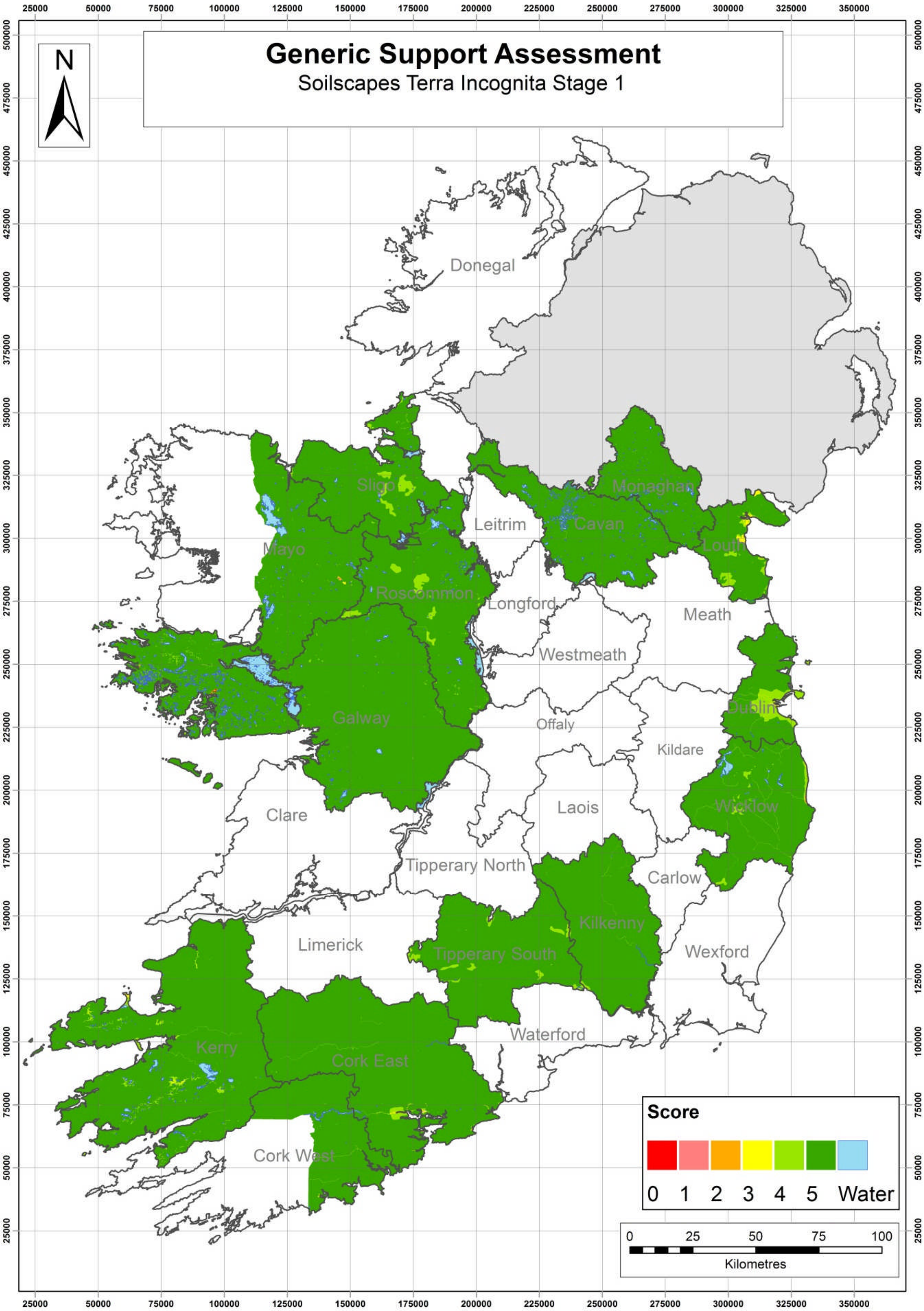



**Figure 15: Support Assessment Stage 1**

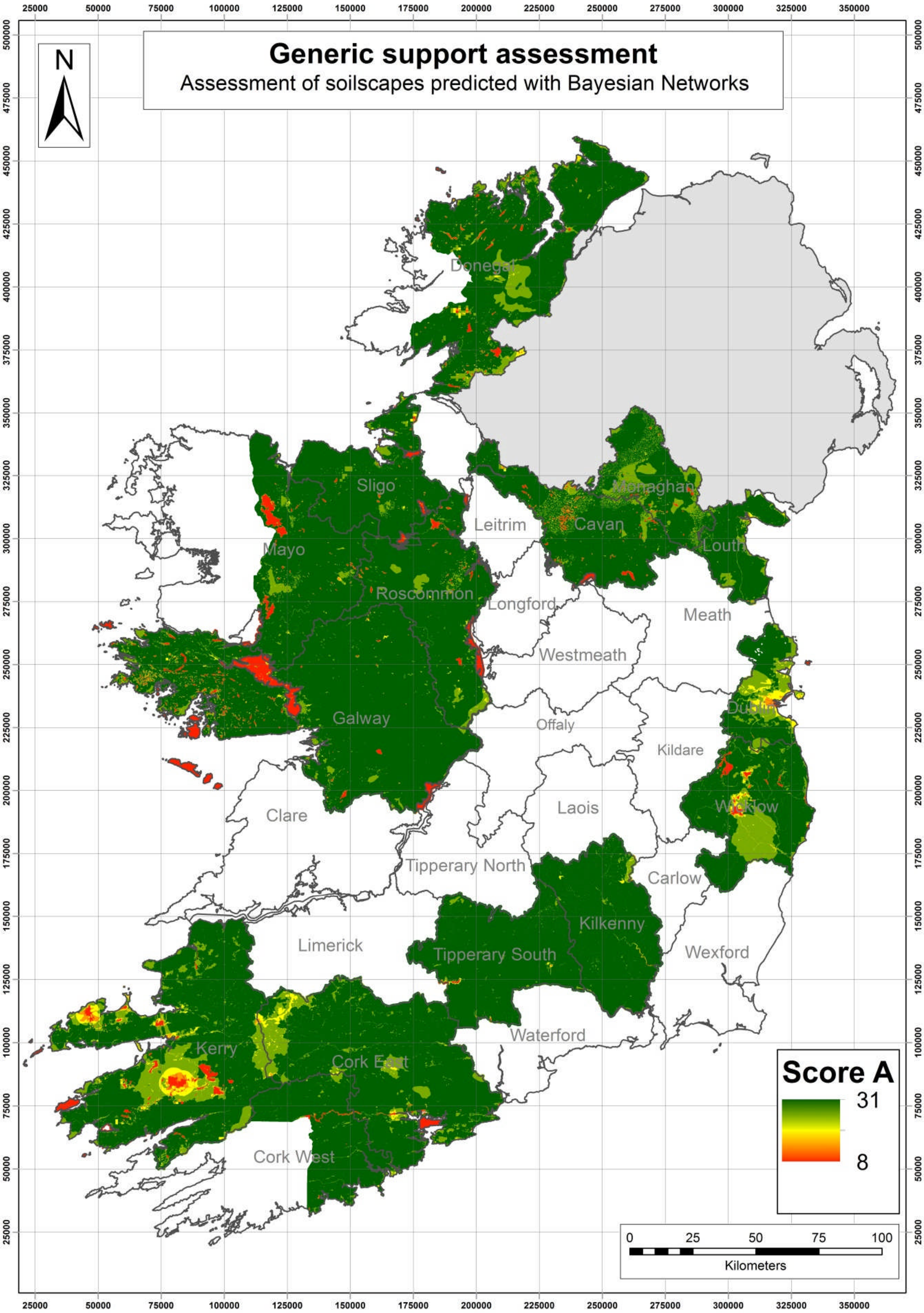



**Figure 16: Support Assessment Stage 2**

**Table 9: Relative extent of raw soilscape predictions Terra Incognita (km2)**

| Label | Area | BN | FS | NN1 | RF1 |
|---|---|---|---|---|---|
| 1a | | | 485.6 | | |
| 1b | 442.6 | 1150.4 | 758.1 | 243.9 | |
| 1cn | 6306.9 | 6848.9 | 20.0 | 2329.8 | 9770.4 |
| 1dn | 2864.4 | 1054.6 | 802.4 | 2674.8 | 620.4 |
| 2b | 480.1 | 911.1 | 646.4 | 251.4 | 0.6 |
| 2c | 1259.5 | 734.2 | 167.7 | 250.5 | 2700.4 |
| 4 | 273.6 | | 782.5 | 102.7 | |
| 5 | 663.8 | 815.7 | 107.0 | 297.6 | 106.0 |
| 6 | 757.3 | 181.5 | 462.1 | 3440.3 | 153.1 |
| 8 | 516.9 | 576.0 | 1108.8 | 310.3 | |
| 9e | 1561.7 | 585.4 | 1038.5 | 505.9 | 1431.0 |
| 9f | 431.3 | 288.1 | 85.2 | 70.6 | |
| 10 | 788.7 | 283.4 | 849.3 | 68.2 | 58.8 |
| 11 | 34.0 | | | | |
| 12a | 606.7 | 1267.9 | 2089.4 | 461.4 | 10.0 |
| 15 | 173.3 | 791.5 | 2854.1 | 77.6 | |
| 16 | 377.3 | 1389.8 | 3708.5 | 2091.2 | |
| 17n | 1437.8 | 1481.4 | 3530.1 | 239.5 | 733.3 |
| 20n | 291.5 | 65.6 | 179.5 | 187.6 | |
| 22 | 208.2 | 620.6 | 298.6 | 507.7 | |
| 23 | 442.3 | 174.3 | 1500.1 | 720.3 | |
| 24n | 1913.2 | 2223.9 | 2126.8 | 1882.9 | 4436.6 |
| 25 | 470.8 | 1067.0 | 1945.1 | 1350.5 | 0.1 |
| 26n | 723.1 | 632.4 | 90.9 | 751.4 | |
| 27 | 558.8 | 690.1 | 503.9 | 1763.9 | 59.5 |
| 28 | 707.9 | 1218.8 | 910.6 | 1112.5 | |
| 29n | 1338.2 | 1687.8 | 1261.8 | 3253.5 | 2622.0 |
| 30 | 1285.9 | 2964.9 | 675.5 | 3798.9 | 7775.6 |
| 31n | 2213.2 | 755.0 | 588.6 | 626.6 | 1528.1 |
| 32 | 1350.1 | 553.7 | 1254.6 | 386.2 | 2075.1 |
| 33 | 627.3 | 1185.1 | 313.0 | 720.5 | 12.5 |
| 34 | 51.9 | 404.6 | 3285.3 | 387.1 | |
| 35 | 1263.9 | 653.5 | | 5.1 | |
| 36n | 1645.0 | 730.0 | 0.2 | 856.4 | |
| 37n | 1124.2 | 421.4 | 61.0 | 1499.9 | 366.7 |
| 39 | 59.3 | 91.7 | | 18.2 | |
| Water | 312.8 | | | | |
| No Data | | | | 1255.6 | 40.0 |
| | 35563.6 | 34500.1 | 34491.3 | 34500.1 | 34500.1 |

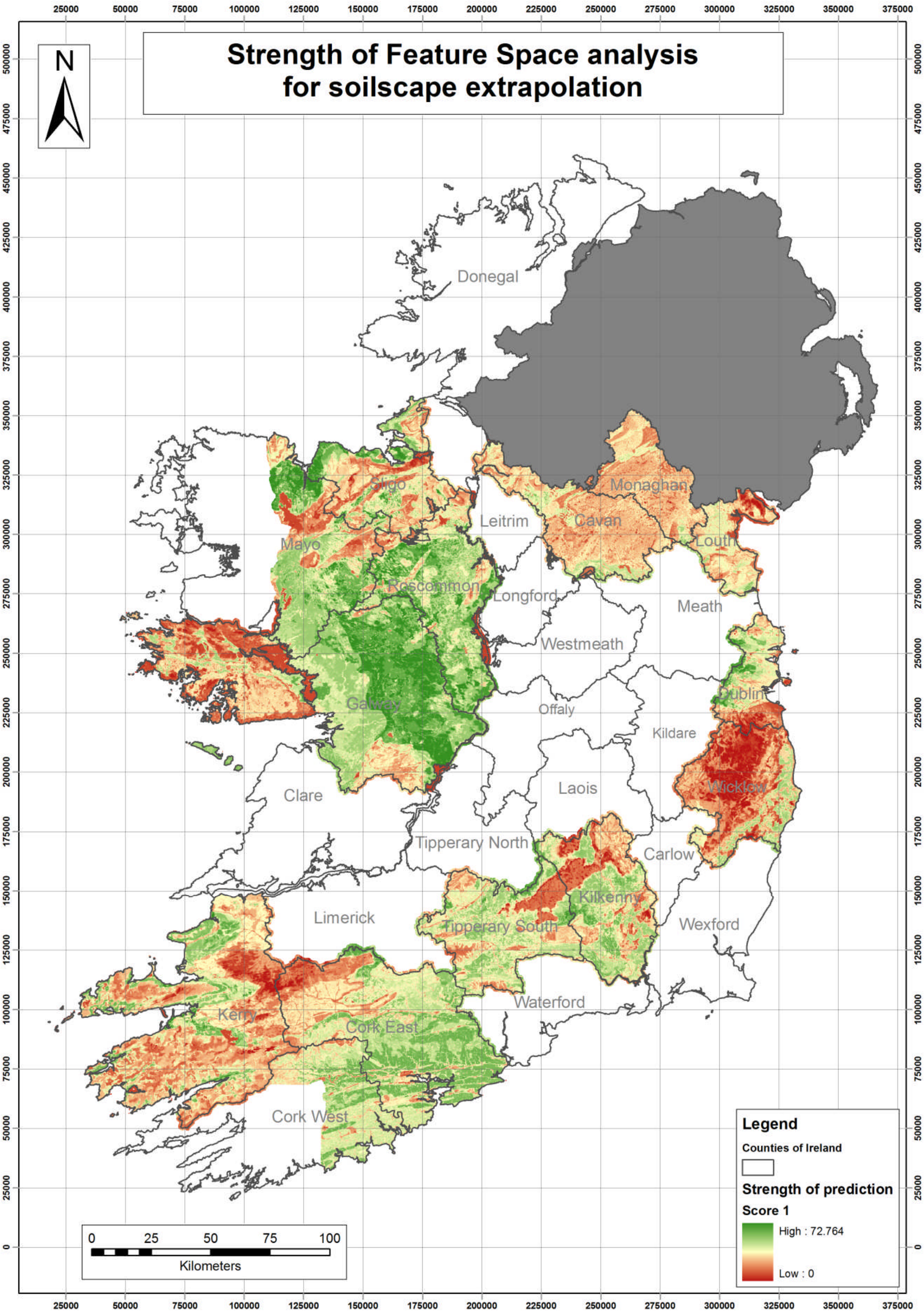



**Figure 17: Strength of Feature Space Analysis Predictions**

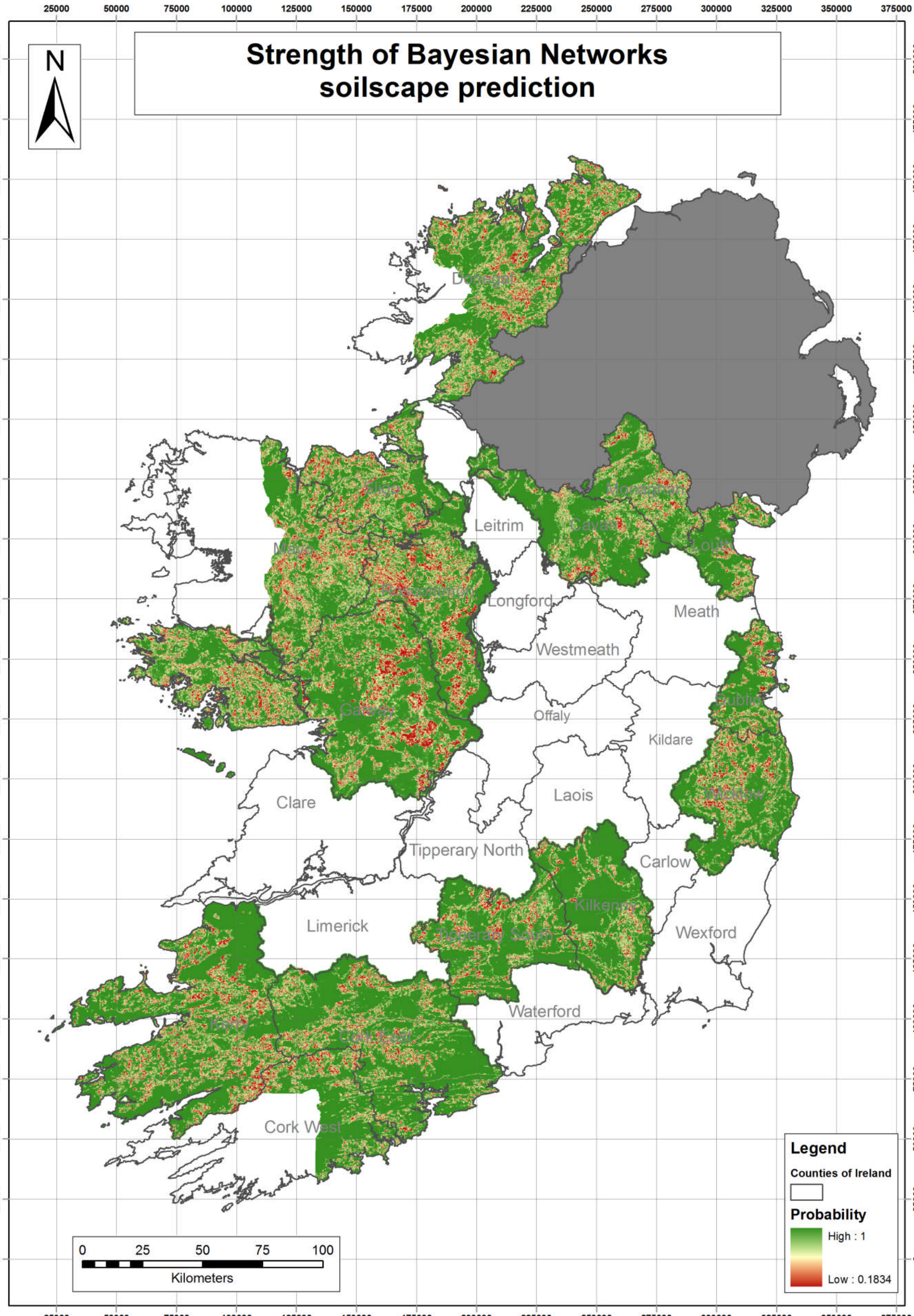



**Figure 18: Strength of Bayesian Belief Network predictions**

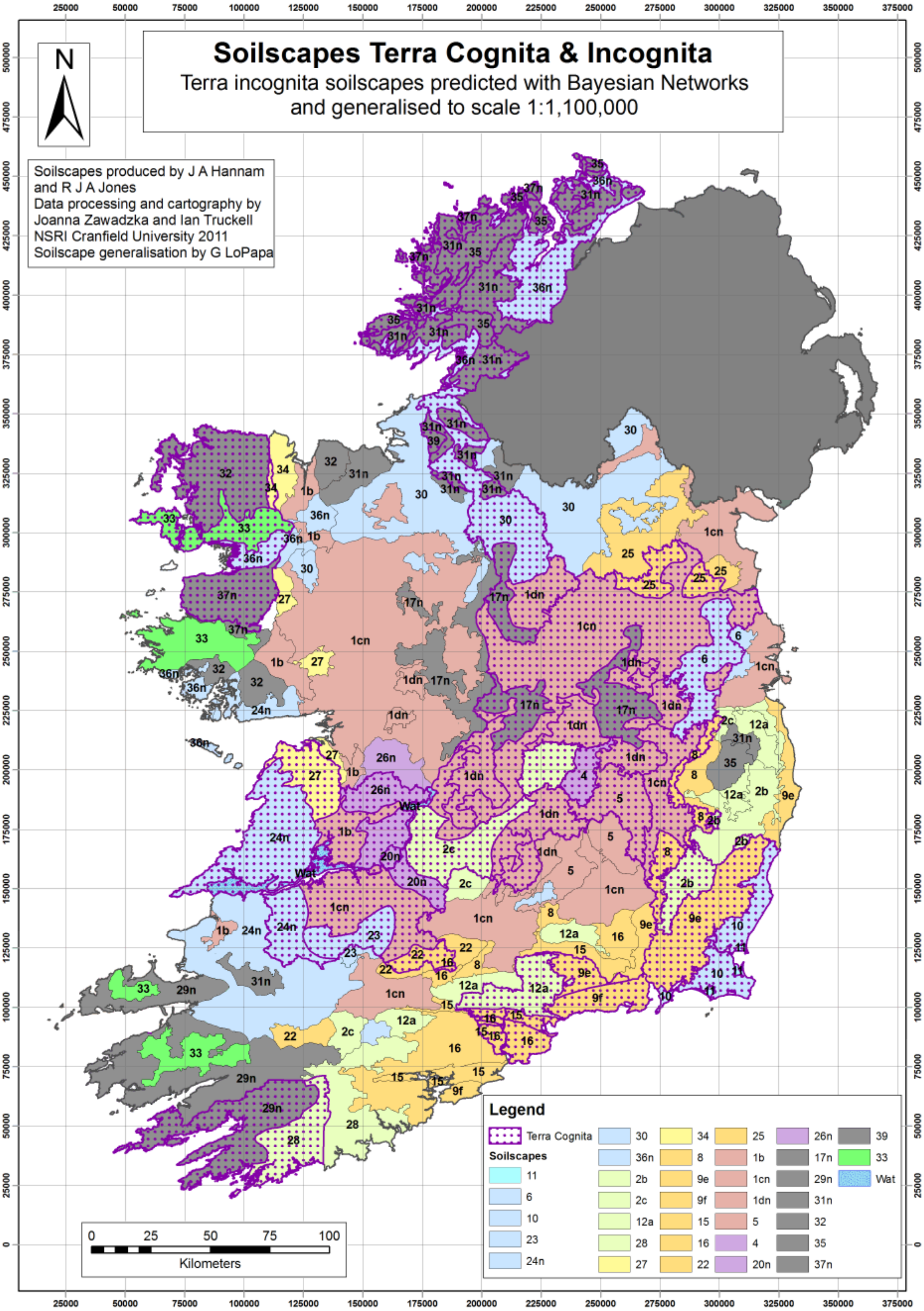



**Figure 19: Soilscapes for both Terra Cognita and Terra Incognita**

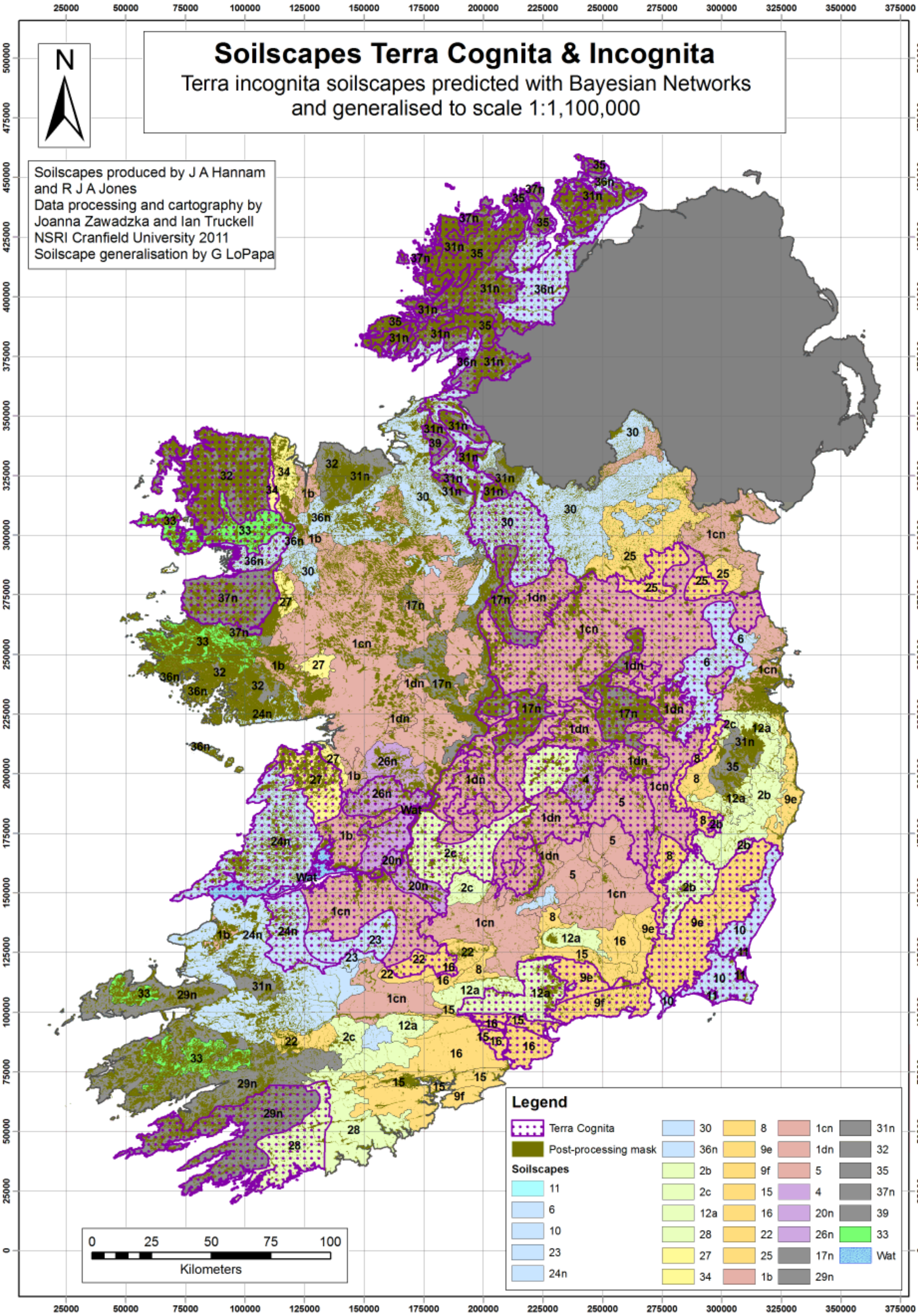



**Figure 20: Soilscapes for both Terra Cognita and Terra Incognita with Mask**

# 9. Conclusions

Both unsupervised and supervised approaches for soilscape delineations have been evaluated. Unsupervised approaches are interesting as they allow to delineate both Terra Cognita and Terra Incognita together, they are, however, there are both software limitations and it is difficult to generate coherent units. However, there is sufficient promise in those approaches to pursue them in the future.

Two different extrapolation techniques were evaluated based on environmental distance and induction rules. For the induction rule approach, three different inference engines were evaluated. For those, satisfactory accuracy levels (Table 6) were only achieved with Bayesian Belief Networks and Random Forests. In addition, Bayesian Belief Networks produced the highest number of soilscapes mapped in Terra Incognita. Furthermore, the predicted soilscapes in Terra Incognita fitted more closely with the soilscapes delineated in Terra Cognita than any other approach.

However, a recently completed MSc project (Smid, 2012) has established that the low number of mapped classes might be a consequence of low resolution climatic data, particularly the solar radiation layer. The influence of low resolution data on the performance of inference engines would need to be further investigated.

Although the result for Bayesian Belief Network 2 was slightly worse than for network 1, it did not contain the artefacts introduced by the low resolution of the climatic data. Hence, both the feature space analyses as well as Bayesian Belief Network 2 results were used as the basis of the predicative mapping of soil associations.

It has to be recognised that user accuracies for Bayesian Belief Networks 2 for soilscapes s35 (36.4%), s36n (46.0), s1b (48.8%) and s17n (49.2%) are below the 50% mark (Appendix 5), indicating a degree of misclassification compared with other soilscapes. The poor results for s39 (32.5%), can be explained by the small spatial extend, resulting in only 455 training cases.

Final soilscape maps based on Feature Space Analysis and Bayesian Belief Networks are illustrated in Figures 19 & 20.

## Acknowledgements

The authors would like to thank Khaled Taalab for his contributions on the descriptions of Belief Networks, Neural Networks and Random Forests.

## References

Agyare, W. A., Park, S. J. and Vlek, P. L. G. (2007) Artificial neural network estimation of saturated hydraulic conductivity, Vadose Zone Journal, 6, 423-431

Bishop, C. M. (1995) Neural networks: a principled perspective, Neural Networks - Producing Dependable Systems (ERA 95-0973)

Breiman, L. (2001) Random forests, Mach. Learning, 45, 5-32

Buol, S.W., Hole, F.D. amd McCracken, R.J. (1973) Soil genesis and classification. Univ. Press, Ames, IA, USA

Cambel, J.B. &Edmonts, W.J. (1984) The missing geographic dimension to soil taxonomy. Ann. Assoc. Am. Geogr. 74: 83-97

Hannam, J.A., Jones, R.J.A., Zawadzka, J.E., Palmer, R.C.. Truckell, I.G. & Mayr, T.R. (2011). Soilscapes for Terra Cognita Ireland. ISIS Project Technical Report No. ISIS_WP1_D2.2, 27pp. Cranfield University & Teagasc.

Hole, F.D. (1978) An approach to landscape analysis with emphasis on soils. Geoderma 21:1-23

Hough, R. L., Towers, W. and Aalders, I. (2010) The Risk of Peat Erosion from Climate Change: Land Management Combinations: An Assessment with Bayesian Belief Networks, Hum. Ecol. Risk Assess., 16, 962-976

Jones, R.J.A., Hannam, J.A., Palmer R.C., Truckell I.G., Creamer, R.E., McDonald, E. (2011). Classification, Harmonisation and Generalisation of County Soil Maps of Ireland. ISIS Project Technical Report No.ISIS_WP1_D1.1, 140pp. Cranfield University & Teagasc.

Jones, R.J.A., Hannam, J.A., Palmer R.C., Truckell I.G., Creamer, R.E., McDonald, E. (2012). Classification, Harmonisation and Generalisation of County Soil Maps of Ireland. ISIS Project Technical Report No. ISIS_WP1_D1.1(v2), 136pp. Cranfield University & Teagasc.

Mayr T., Bellamy P., Bradley I., Cooke H., Corstanje R., Burgess P., Palmer R., Pearn K., Rivas-Casado M., Truckell I. and Zawadzka J. (2009) Development of a pilot study for testing precise digital soil mapping techniques (at 50-100 m resolution) in a pilot area in England and Wales. Final Report for the Joint Research Centre, Ispa. pp. 217 excluding Appendices

Kuhnt, G. (1994) Regionale Repräsentanz. Beiträge zu einer Raumorientierten Messtheorie. Postdoctoral lecture qualification thesis, University Kiel, Germany

Lark, R.M. (1999) Soil-landform relationships at within-field scales: An investigation using continuous classification. Geoderma 92:141-165

Liaw, A., and Wiener, M. (2002) Classification and Regression by randomForest. R News: The Newsletter of the R Project, 2(3):18–22 (http://cran.r-project.org/doc/Rnews/)

Leathwick J., Morgan F., Wilson G., Rutledge D., McLoed M. and Johnston K. (2002) Land Environments of New Zealand. Technical Guide. Ministry of the Environment. 237 pp.

Lou, W. G. and Nakai, S. (2001) Application of artificial neural networks for predicting the thermal inactivation of bacteria: a combined effect of temperature, pH and water activity, Food Res. Int., 34, 573-579

Maier, H. R. and Dandy, G. C. (2001) Neural network based modelling of environmental variables: A systematic approach, Math. Comput. Model., 33, 669-682

McBratney, A.B., Hart, G.A. and McGarry, D. (1991) The use of region partitioning to improve the representation of geostatistically mapped soil attributes. J. Soil Sci. 42:513-532 et al., 1991

McSweeny, K., Slater, N.K., Hammer, R.D., Bell, J.C., Gessler, P.E. and Petersen, G.W. (1994) Towards a new framwwework for modeling the soil-landscape continiuum, In Amudson, R., Harden, J and Singer, J. (eds) Factors of Soil Formation: A 50th Anniversary Retrospective. SSSA Special Publication 33:127-145

Smid, M. (2012) Modelling and data fusion for soil mapping. MSc Thesis, Geographical Information Management, Cranfield University

Schmidt, K., Behrens, T., Friedrich, K. and Scholten, T. (2010) A method to generate soilscapes from soil maps. J. Plant Nutr. Soil Sci.

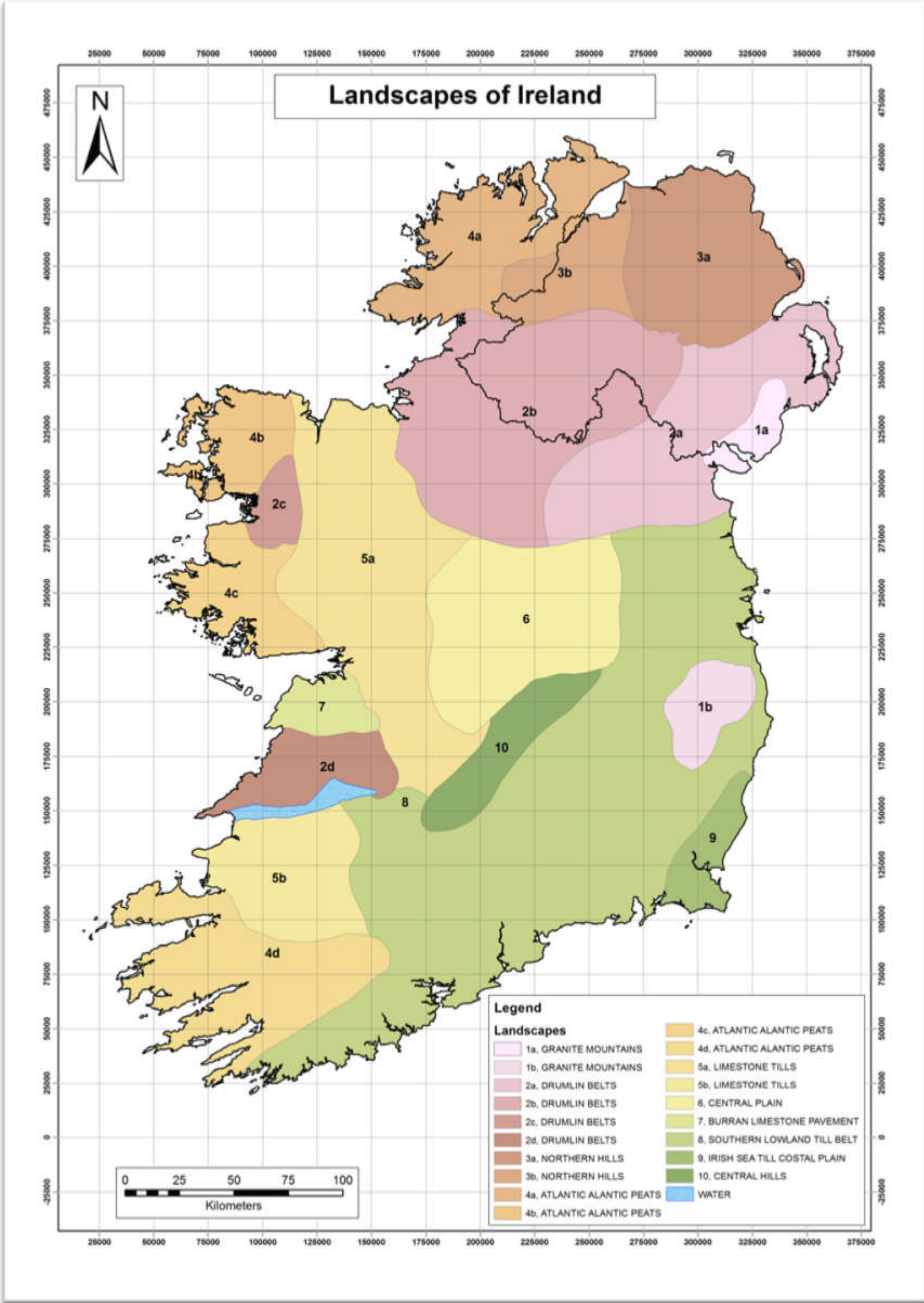
Uusitalo, L. (2007) Advantages and challenges of Bayesian networks in environmental modelling. Ecological Modelling 203 (3–4), 312–318.

Zhu, A.X., Moore, A. and Burt, J.E. (2006) Prediction of Soil Properties Using Fuzzy Membership. 2nd Global Workshop on Digital Soil Mapping, Rio de Janeiro, Brazil 4 to 7 July 2006

# Appendix 1 Ireland: Landscape-Soil Regions

**Version 1.2-03 -November 2009**

This is an interpretation of landscape and soil regions compiled by Brian Kerr.



| Code | Name | Landforms | Soils | Notes |
|---|---|---|---|---|
| 1 | *GRANITE MOUNTAINS* | | | |
| 1a | Northern Mourne Granites | High hills with abundant rock outcrops; shallow peat and some deeper bogs. Centre of local glaciation-some moraines | Peaty podzols; peat of various depths; recent moraines for the Loch Lomand readvance stage of the very late Pleistocene.Soils generally acid | Both the Mournes and the Wicklows are important for recreation and water catchments being on the doorstep of Belfast and Dublin urban conurbations. . |
| 1b | Wicklow Mountains | Incised by deep glens with important ancient woodland sites | As above | |
| 2 | *DRUMLIN BELTS* | | | |
| 2a | North east drumlin field over Ordovician greywackes and shales | -This is the classic drumlin belt of Ireland. There are numerous papers and arguments on the details of ice movements in this area. Inter-drumlin hollows are either gleyed with some peat, or open water | Acid brown earths; some rock outcrops; occasionally and gravel moraines. The parent material is a compacted brown or grey sandy or silty clay loam | Mineral soils on drumlins which are intensely farmed |
| 2b | North central and western drumlins over limestones | Generally wetter with larger inter-drumlin hollows and peat or water bodies | Less well farmed and more gleys present | The wetter climate begins to reduce the presence of brown soils and peaty gleys especially over limestone are common. Open water becomes frequent |
| 2c | Clew Bay drumlins | Limestone till classic drumlins with gley and peat in hollows | Gleys with interdrumlin peat. | |
| 2d | Clare and Limerick drumlins | As above | As above | |
| 3 | *NORTHERN HILLS* | | | |
| 3a | Antrim Plateau Basalt | Includes the blanket peat of the Antrim plateau | Blanket peat with deeper valley mires, otherwise gleys formed on basalt till, with brown mineral soils at lower elevations and on slopes. High base status soils and much reder in colour | The blanket peat thins towards the east coast. The basalt exposures on the north Antrim coast have produced the Giant's Causeway. All of this area is considered as a Severely disadvantaged farming zone |
| 3b | Over metamorphic rocks (schist, and mudstones) | Few drumlins-includes high hills of Sperrin Mountains. Gravel moraines which are leached are common | Brown podzolics with peat and some acid brown earths formed over hard metamorphic rocks. Acid and low fertility soils | All of this area is considered as a Severely disadvantaged farming zone |
| | | | | |
| 4 | *ATLANTIC ALANTIC PEATS* | | | |
| 4a | Donegal peat and pock complex | Low hills of ancient metamorphic rocks-mostly schists | Shallow peat and rock. Occassional brown mineral soils, at lower elevations | The soil pattern can only be mapped as complexes. Some drumlins may be large enough to isolate at larger scale maps |

| | | | | |
|---|---|---|---|---|
| 4b | Mayo | As above | | |
| 4c | Connemara | Complex of rock and peats or various depths,-high rainfall, wind exposed. | | Mapped as complexes |
| 4d | South west | Dingle rock and peat complex Mountains and high hills such as Boggeragh Mountains | | Mapped as complexes |
| | | | | |
| 5 | *LIMESTONE TILLS* | | | |
| 5a | Western Till Plain | Rolling lowland. Mostly grazing land with shallow till deposits over limestone. Peat in hollows | Often shallow mineral soils over till. Brown earths and brown podzolics are common | This is the classic grazing lands of the west and centre of Ireland |
| 5b | Kerry & Limerick till | This is the wetter western version of the above with gleys replacing the brown soils of further east | Mostly gleys | Blanket peat on higher hills |
| 6 | *CENTRAL PLAIN* | Basin peats with low limestone hills; esker and kame gravels and sands common | Deep peats-often cut over commercially exposing limestone till below | Sharp changes in topography, soils and land use in short distance. Classic eskers at Trim |
| 7 | *BURRAN LIMESTONE PAVEMENT* | Classic limestone pavement with solution features | Rendzina soils and rock | Clint and grike landscape-may be a World heritage site in places-need to check this |
| 8 | *SOUTHERN LOWLAND TILL BELT* | Grazing lands of Cork and Waterford and into Midland counties of Kilkenny and Carlow. Assumes these glacial deposits are older and more weathered than all the boulder clays further north | Brown earths and brown podzolics | This is some of the best land in Ireland with brown mineral soils which are better drained especially in the east. Heartland of the Irish dairy industry |
| | | | | |
| 9 | *IRISH SEA TILL COSTAL PLAIN* | Narrow coastal plain with heavy soils deposited by Irish Sea ice | Gleys- | Flat lowland coastal plain |
| 10 | *CENTRAL HILLS ( Slive Broom)* | Older shale and sandstone which form low hills in the central plain | Peats and rock at higher levels with peaty podzols | Only major units of relief in the central plain |

Notes to the map Legend

1. This is a very approximate classification of landscape based on geology maps, the Soil Association of Ireland and Land Use Potential Soils Bulletin 1980 with the accompanying 1:575,000 scale General Soils Map, and my university notes on glaciation, c 1970.
2. The soils information used is the standard older terminology and makes no attempt to use more recent classifications
3. There is an accompanying map1, which sketches in TEN basic Regions based on geology which are then sub-divided using what soils and land form information I have.
4. All of this is in outline form only but it does match with my knowledge of the landscape from field work in the past.
5. Points to remember are:
    a. Except for soils in the southern fringe, soils in Ireland are younger than the last glaciations and some have formed since the late glacial re-advances which are responsible for moraines which are still very fresh and sharp.
    b. In general soils are shallow over hard rock of compacted till which often has a fragipan
    c. The climatic gradient plays a great part in soil formation. Atlantic facing western areas have high rainfall, and strong winds which limit vegetation growth, and therefore soil development
    d. The understanding of the soils is very closely related to the understanding of the glacial history. Hence the kame and kettle topography of the central plain; the northern drumlin fields; and the heavy gleys of the Wexford area formed on Irish sea deposits pushed on land by late Pleistocene Irish Sea Ice.
    e. In the west the hard metamorphic rocks of the Dingle Peninsula ,Connemara, Mayo, and Donegal form high hills and in the SW considerable mountains , and combine with the unfavourable climate to reduce soil formation to blanket peat and rock with occasional mineral soils. These can only be mapped sensibly as complexes.
    f. There has been a long history of peat extraction in all areas of Ireland (there is no large scale coal deposits and forest clearance began in Neolithic times). Therefore cut –over peat is common. There are pictures on your photo data base of large scale peat extraction for power stations in the central plain

# Appendix 2 LENZ Approach

## Introduction

The Land Environments of New Zealand (LENZ) (Leathwick *et al*., 2003) was developed to describe an environmental classification of New Zealand in order to provide a framework for addressing a range of conservation and resource management issues.

Approximately 2.7 million data points are contained in a 100 m resolution raster layer covering New Zealand. Due to software restrictions, a 25% subset of the occupied data cells was used to define the classification, which these selected by taking every second cell of every second row.

Due to software limitations in PATN, LENZ was defined in two stages. Initially, a non-hierarchical classification technique (ALOC/ALOB) in which memory requirements rise only linear with the number of data points, was used to group together points located close to one another in environmental space. The average environmental values of the groups produced by this process can then be used as inputs to a conventional agglomerative classification process.

In a second stage of classification, the North and South Island output data set from ALOB were combined to form a matrix with 722 groups, and this was classified using a conventional agglomerative procedure. The PATN module GASO was used to calculate all inter-group. Once the inter-group distances had been calculated, the classification was defined using the flexible UPGMA sorting strategy as implemented in FUSE. The classification consists of 4 levels – Level with 20 classes (national), Level 2 with 100 classes (national/regional), Level 3 with 200 classes (regional) and Level 4 with 500 classes (regional/local).

Once the classification had been defined, there remained the challenge of mapping the geographic distribution of the environmental groups defined by the non-hierarchical classification. This was archived using purpose written code that calculated environmental distances between each 100 m grid across New Zealand and the centroids for the 722 environmental groups used in Phase 2.

### Version 1 (November 2009)- Giuseppe

#### *Environmental variables (13)*

- Mean annual temperature - c
- Mean annual precipitation - c
- Mean minimum temperature of the coldest month - c
- Mean annual temperature of the warmest month - c
- Mean annual solar radiation - c
- Winter solar radiation - c
- Annual potential evapo-tranpiration - c
- Potential soil moisture deficit - c
- Precott Index - c
- Subsoil - p
- DEM - r
- Slope - r
- Ruggedness (VRM) - r

***Input files***

- Grid to Point
- Attach all relevant layers

***Categorical data:***

Binarized categorical data

***Sample size:***

320 m spatial resolution – 978’698 points

***Clustering approach:***

2 stage approach, first k-means, and then hierarchical agglomerate clustering

***Classification***

3 levels with 6, 44 and 100 environments respectively [Figures 1 to 3]

## Version 2 (March 2010) - Giuseppe

***Environmental variables (14)***

- Subsoil - p
- Habitat - o
- DEM - r
- Mean annual precipitation - c
- Mean annual temperature - c
- Minimum temperature of coldest month - c
- Maximum temperature of warmest month - c
- PET - c
- Prescott Index - c
- Potential Soil Moisture Index - c
- Mean annual radiation - c
- Vector Measure Ruggedness - c
- Slope - r
- Winter radiation - c

***Input files***

- Grid to Point
- Attach all relevant layers

***Categorical data:***

Binarized categorical data

***Sample size:***

320 m spatial resolution – 978’698 points

***Clustering approach:***

2 stage approach, first k-means, and then hierarchical agglomerate clustering

***Classification***

4 hierarchical levels – I (3 classes), II (8 classes), III (12 classes) and IV (50 classes) [ISISData\dsm_new\data_from_Athenry\LENZ]

## Version 3

### *Study area:*

Phase 1 including Carlow, Laois, Limerick, Meath, Kildare, Offaly, Tipperary North, Waterford, Wexford, Dublin, Wicklow, Kilkenny, and Tipperary South.

### *Input layers:*

- Dtm_v0_200 m - r
- Slope (based on dtm_v0_200 m) - r
- Aa122 - n
- Aa162 - n
- Prsc - c
- Hab - -o
- Rann - c
- Rockunitgroups_0604.shp (ROCKUNIT field) - p
- CORINE2000.shp (Lev3_Class field) - o
- Subsoils_v4.shp (SUB_TYP field) - p

### *Dataset preparation:*

- Resampling of isis_dtm_v0 into200 m resolution with nearest neighbour method and Resample tool.
- DTM of 200 m resolution was extracted for Phase 1 area with 'Extract By Mask' tool.
- Points 200 m apart were created with Raster to Point tool based on the extracted DTM.
- IDnum, X, Y fields were added and calculated accordingly.
- Spatial join tool was used to attach vector datasets.
- Sample tool was used to extract raster values at point locations.
- Table results from Sample tool were attached to point files based on FID and MASK fields from shapefile and table respectively.
- Fields that were not needed were removed from attribute tables of point shapefiles.
- Attribute tables were exported to text files.

### *Statistical Analysis*

- Lenz dataset was opened in Statistica.
- Numerical data were standardised (Data| Standardize)
- Variable specifications were modified for those categorical variables that had numerical entries → data type was set to 'text'.
- Data Mining | Generalize EM& K-means cluster analysis
- V-fold cross validation with default settings was used to generate optimal number of clusters. Output saved as lenz_ph1_cl002_v1.
- Classification into 3 clusters returned one cluster taking up over 99% of data point. The remaining ones should be considered as outliers and removed from the analysis. Thus the result was exported to text file with IDnum and XY coordinates and all cases that were classified into class2 or 3 deleted. Subsequently table holiding IDnum of non-outliers was merged with table holding all data based on IDnum field. Missing cases were deleted during the merge. Save new input dataset as lenz_data_200 m_stand_v2.sta.
- Perform v-fold cross validation again with Chebychev distances as a measure of distances between clusters. Again, 2 clusters were obtained with cluster 1 taking up over 99% of the dataset. Output saved as lenz_ph1_cl002_v2.
- New dataset exluding new outliers was generated and saved as new input dataset as lenz_data_200 m_stand_v3.sta. V-fold validation was performed again, with similar results.

- It has been found that the problem with persistence of outliers occurred only when Chebychev distances were used.  Euclidean distances produce meaningful results.
- Use lenz_data_200 m_stand.sta and Euclidean distances.
- Export results to text file along with IDnum, X, Y fields.

***Post-processing***

In ArcCatalog:

- Right click on each txt file and select Create feature class from X Y table
- Create rasters from point files using Point to Raster tool
- Convert rasters to Integer using Int (sa) tool
- Display outputs in ArcMap.
- Use focal statistics tool 5x5 rectangle majority filter to smooth out the map outputs.

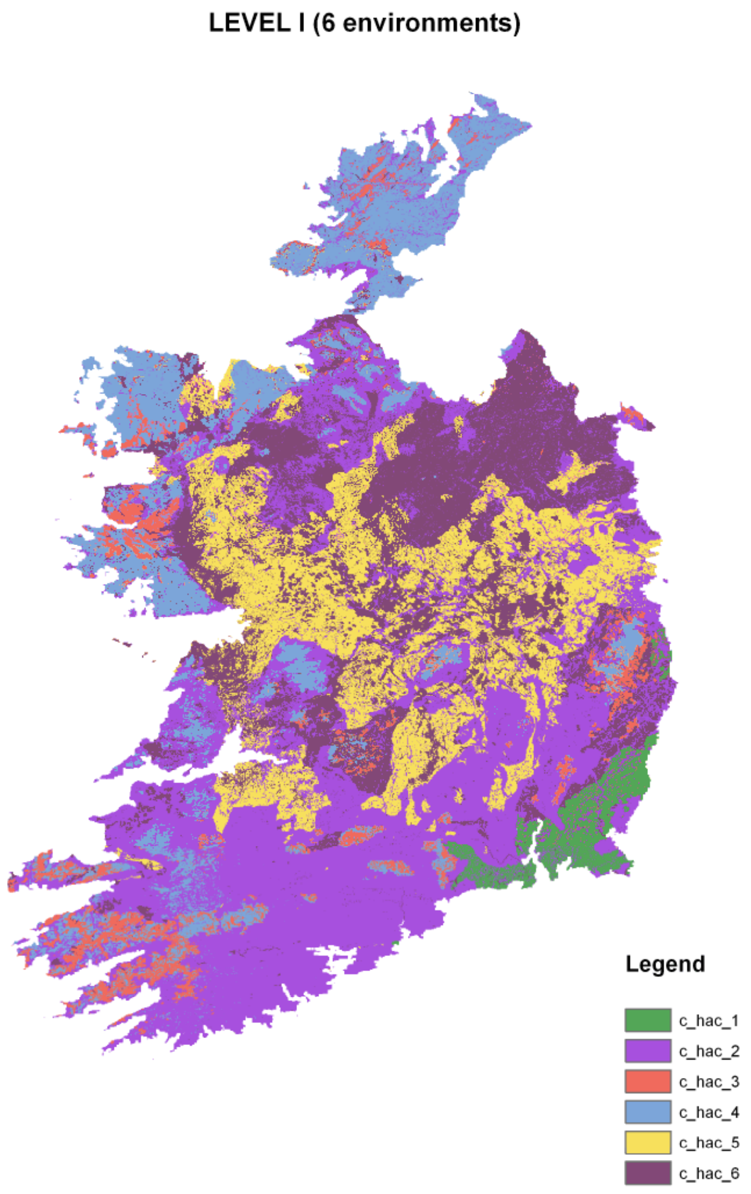



**Figure 1: Version 1**

## LEVEL II (44 environments)

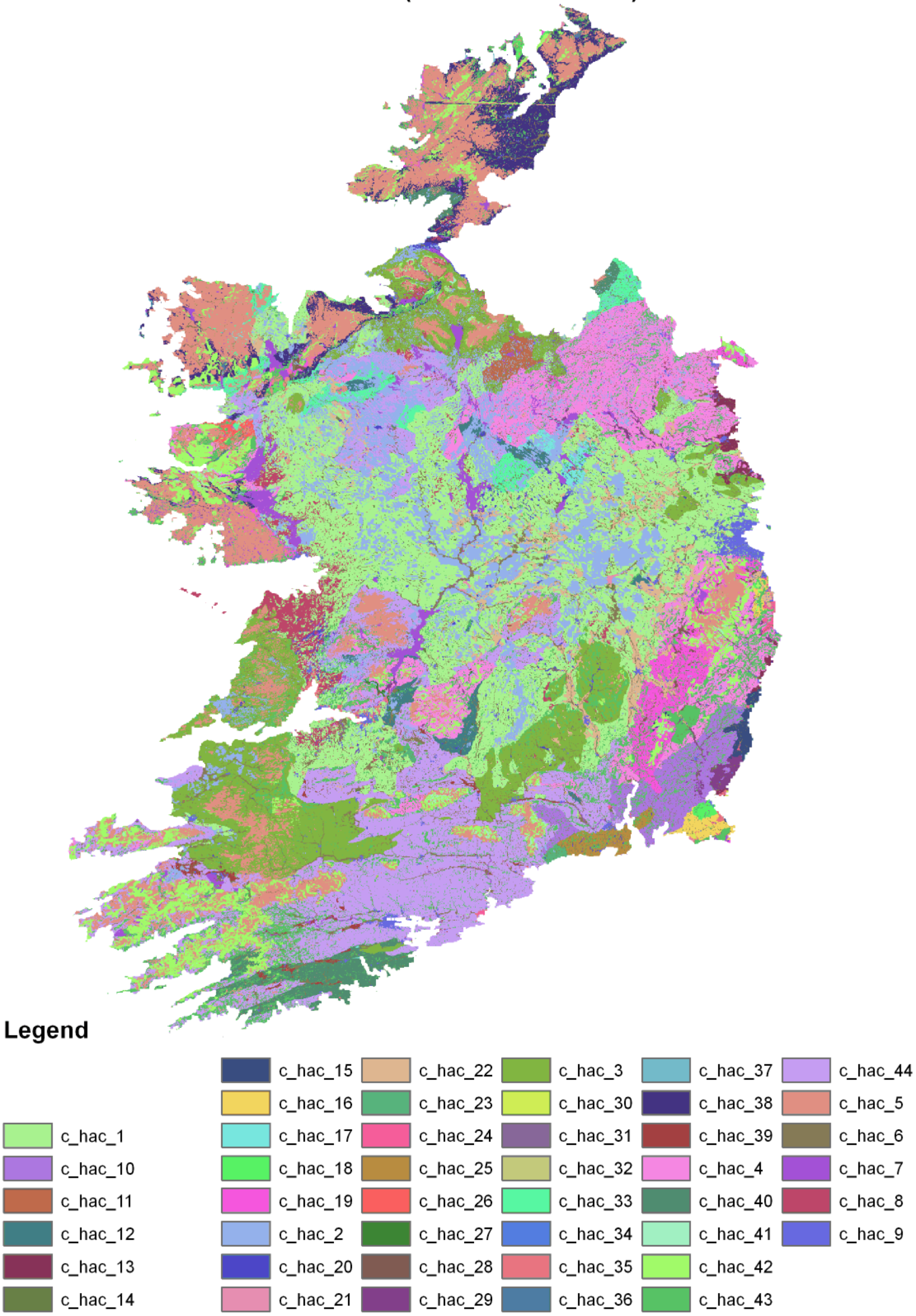



**Figure 2: Version 1**

## LEVEL III (100 environments)

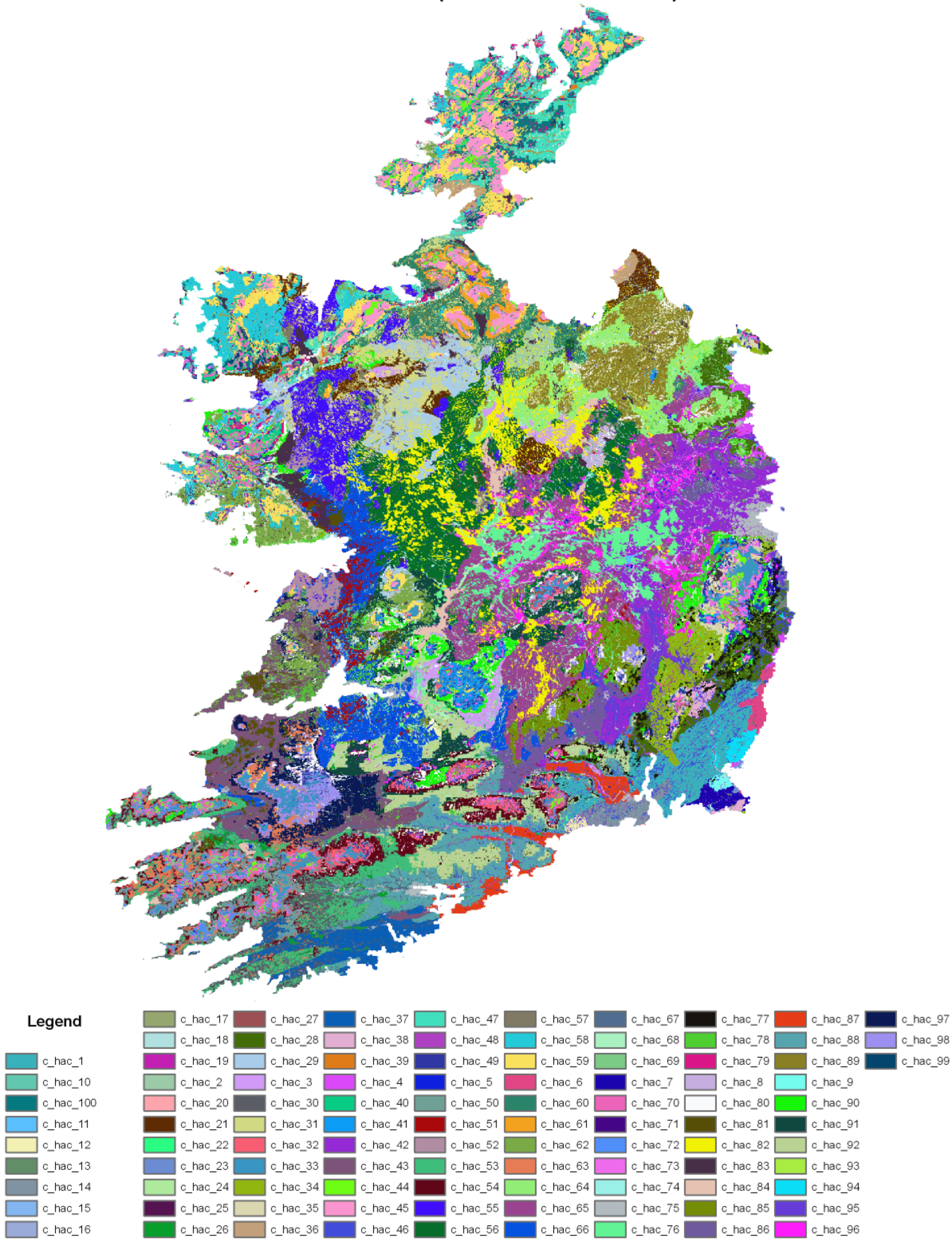
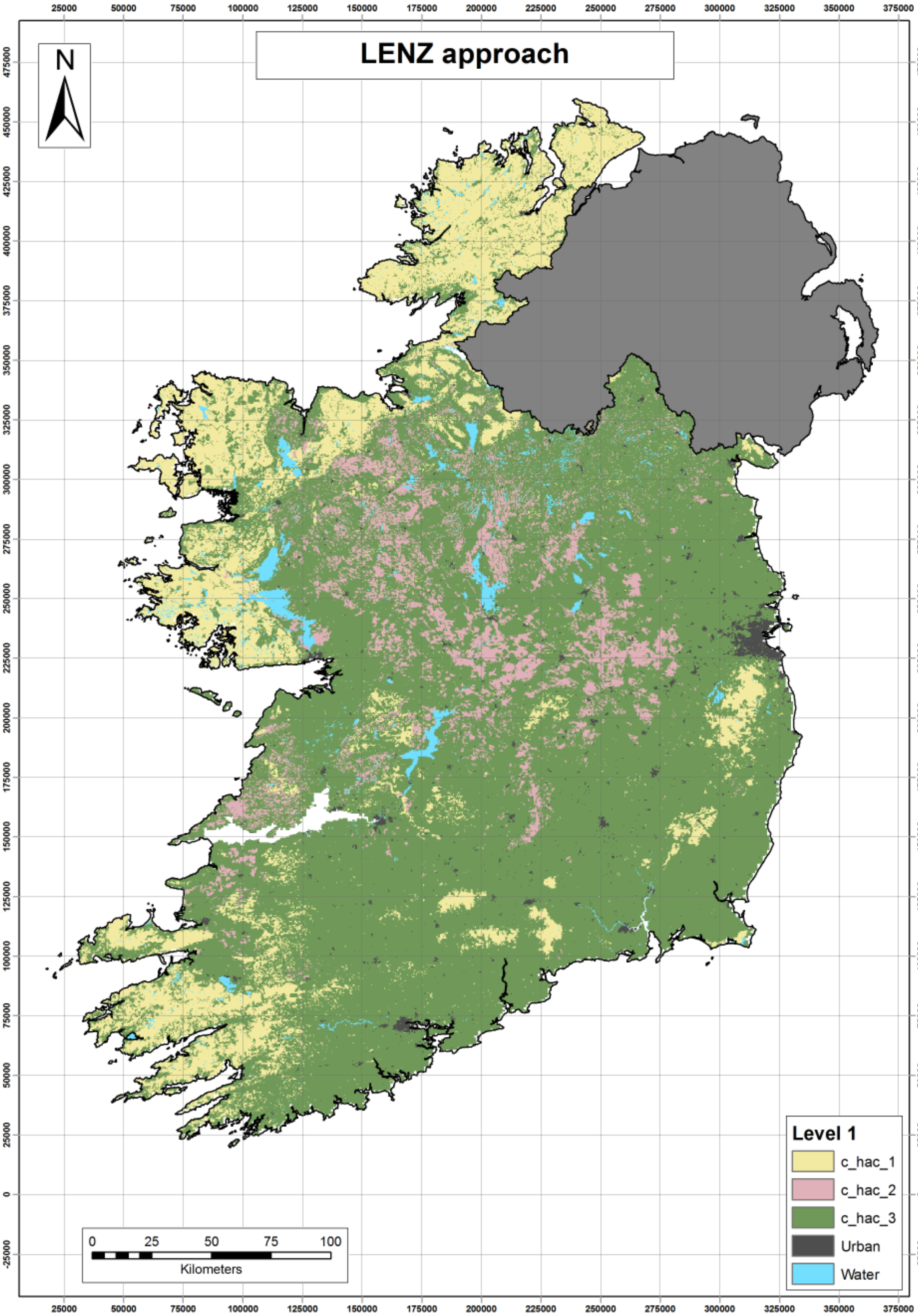
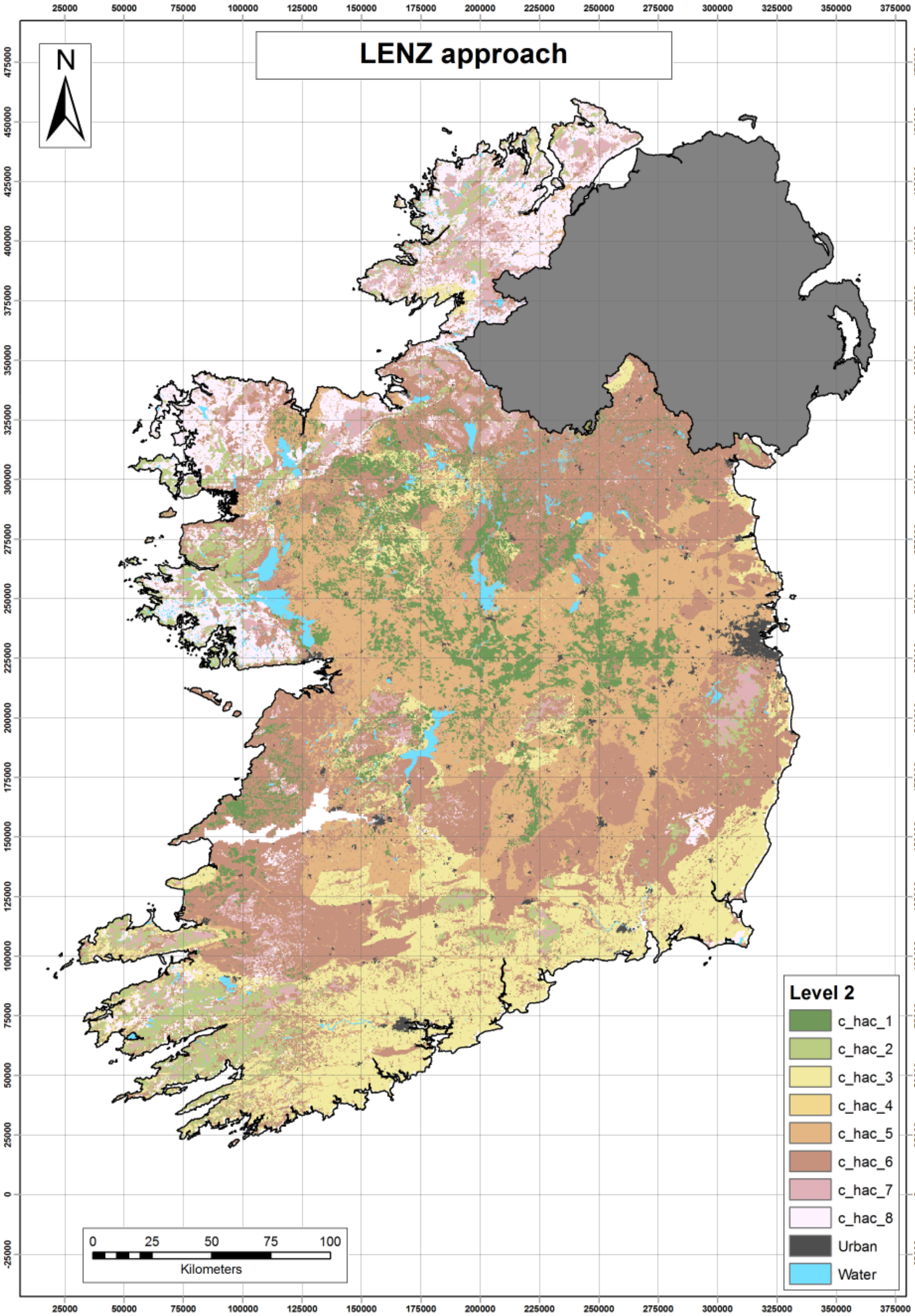


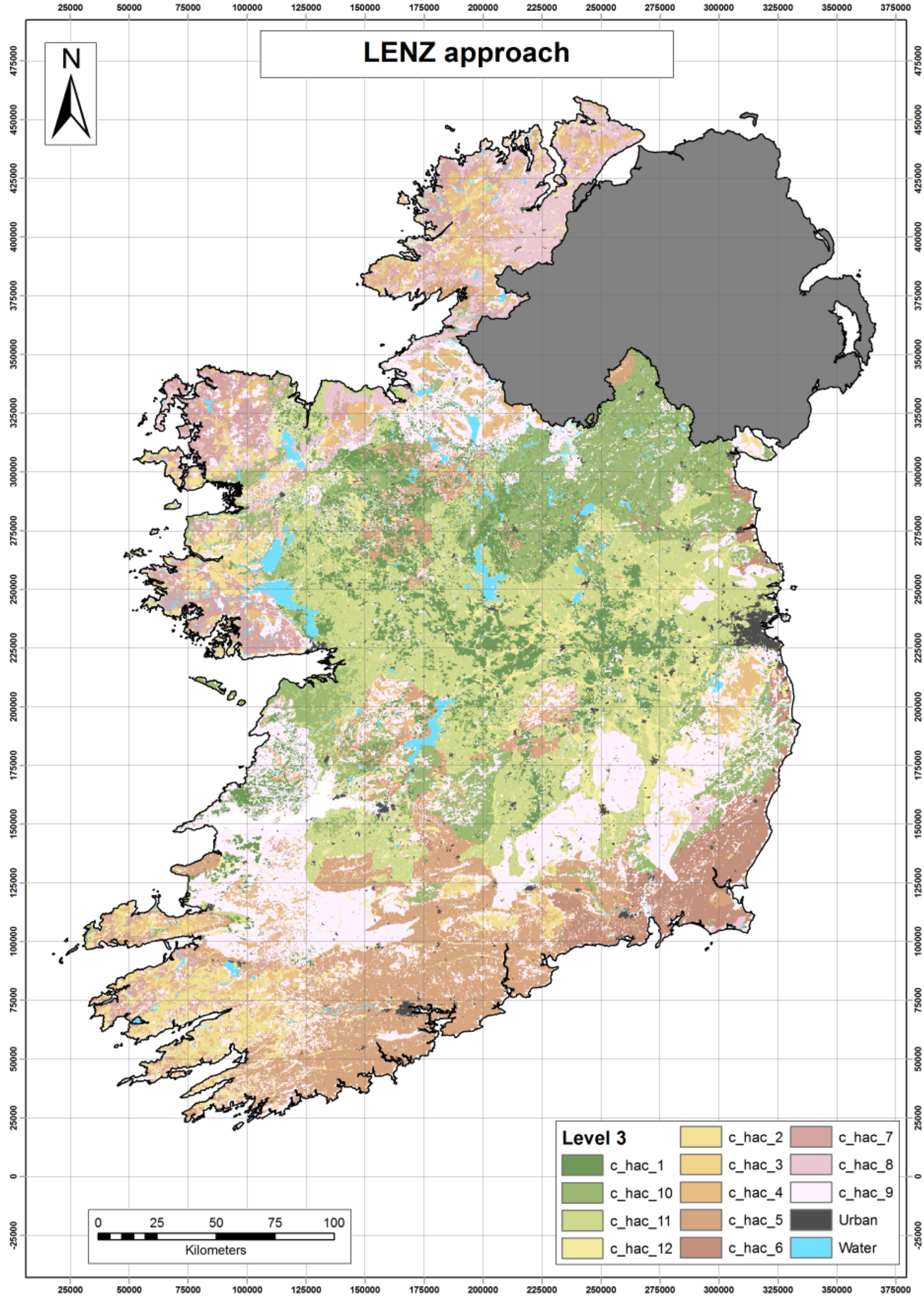

**Figure 3: Version 1**



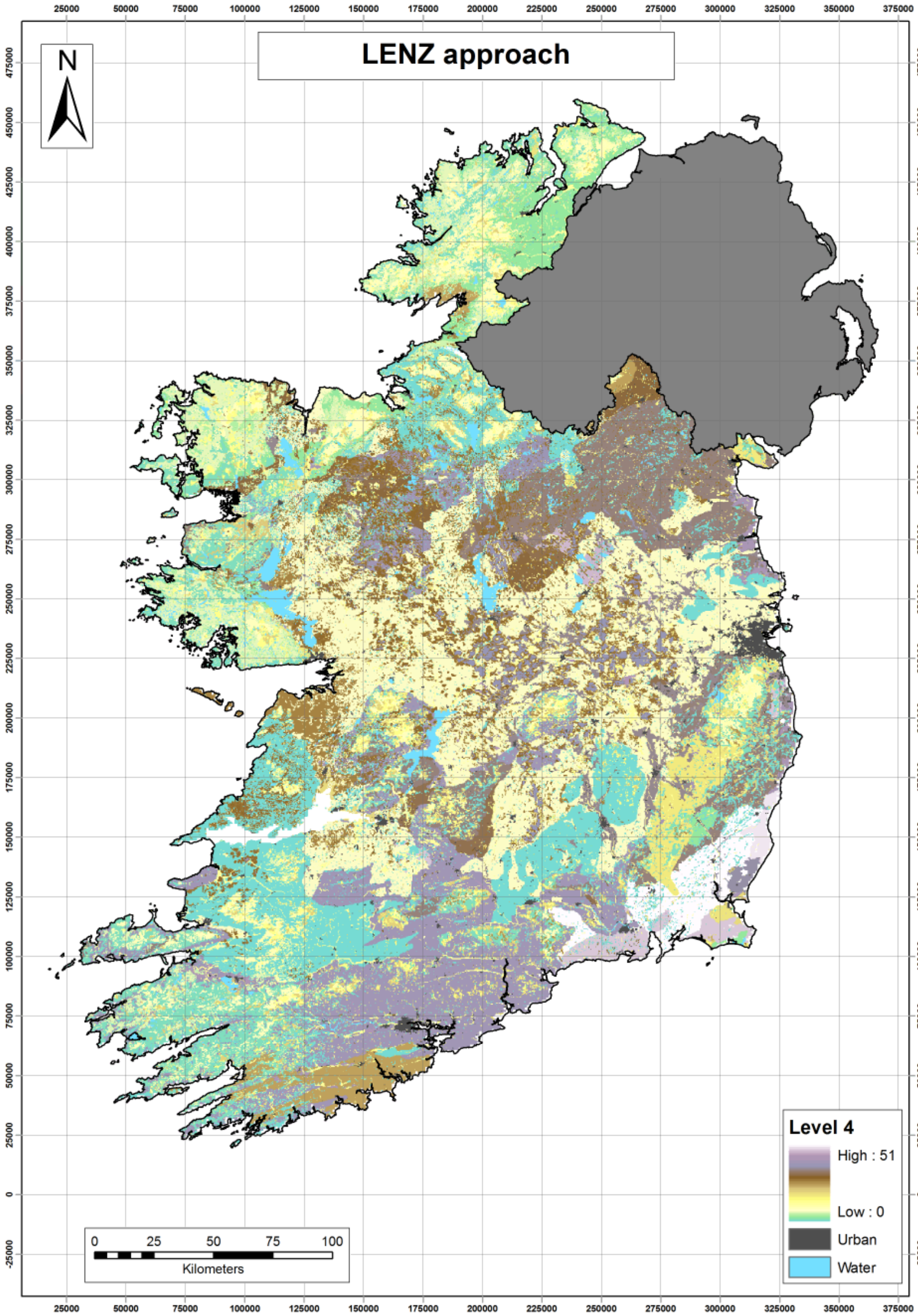

**Version 2 – Level 1**




**Version 2 – Level 2**



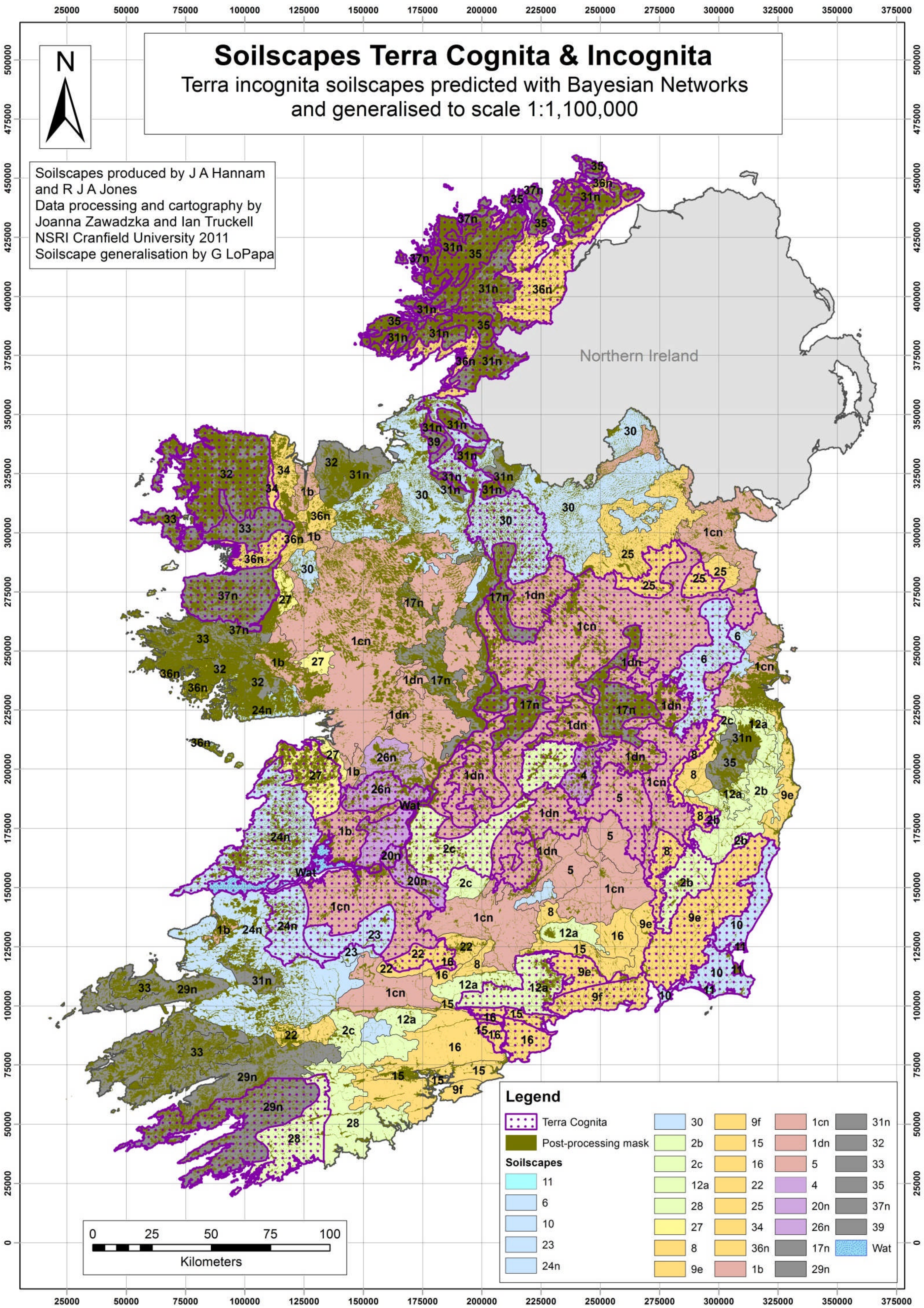

**Version 2 – Level 3**




**Version 2 – Level 4**

# Appendices 3-6 are available only in electronic form

| | |
|---|---|
| Appendix_3_BN_Evaluation | .xls |
| Appendix_4_Inference_Evaluation | .xls |
| Appendix_5_Support_Assessment_Stage1 | .xls |
| Appendix_6_Support_Assessment_Stage2 | .xls |


Soilscapes Terra Cognita & Incognita
Terra incognita soilscapes predicted with Bayesian Networks
and generalised to scale 1:1,100,000
N
Soilscapes produced by J A Hannam
and R J A Jones
Data processing and cartography by
Joanna Zawadzka and Ian Truckell
NSRI Cranfield University 2011
Soilscape generalisation by G LoPapa
Northern Ireland
Legend
Terra Cognita
Post-processing mask
Soilscapes
11
6
10
23
24n
30
2b
2c
12a
28
27
8
9e
9f
15
16
22
25
34
36n
1b
1cn
1dn
5
4
20n
26n
17n
29n
31n
32
33
35
37n
39
Wat
0 25 50 75 100
Kilometers