PROJECT INFORMATION

Project title: Distribution maps of forest tree species

Project ID: 63

Contact person: Jesus San-Miguel (jesus.san-miguel@jrc.ec.europa.eu)

PROJECT DESCRIPTION

The European Commission, Joint Research Centre, hosts the European Forest Data Centre (EFDAC at http://efdac.jrc.ec.europa.eu) of the Forest Information System for Europe (FISE). The EFDAC-FISE platform is envisioned to transparently integrate information referring to forest resources in Europe, including taxa-specific information. The exercise involves the integration of:

1. a harmonized collection of reference maps describing the European-wide distribution of forest tree species along with their habitat suitability (both current and under varying climate change scenarios);
2. a detailed analysis of the implications that the uncertainties – induced by the many available datasets and the required computational modelling strategies – impose for the correct interpretation and use of each reference map, also for science-based regional and continental-scale environmental policies;
3. an updated review of taxa-specific research highlights and major topics in the corresponding science-policy interface at regional, continental and global scale.

The Atlas of European Forest Tree Species is a project within EFDAC-FISE that will cover these taxa-specific aspects. It is proposed to be a semantically enhanced, systematic online collection of taxa-specific articles analyzing the state-of-the-art information of forest tree species in Europe. The overall approach is set in the context of a continuous integration of multiple sources of available information, from field observations (e.g. national forest inventories and other field data), bio-geo-climatic analysis and remote sensing. The final maps of taxa distribution and habitat suitability (derived by the iterative integration of all the available sources of information) will be made freely available to the public (see http://ec.europa.eu/geninfo/legal_notices_en.htm). All the sources of information will be managed according to their specific distribution policies and acknowledged.

Methods

The exploited methods follow three main steps:

1. Data collection, pre-processing and setting, e.g. of species presence/absence, climate data, solar radiation, etc.
2. State-of-art and modelling strategy design.
3. Modelling: model fitting, accuracy assessment and validation of generated spatially distributed maps.

Steps 1) and 2) are developed in parallel with a continuous feedback due to the high degree of dependence between the modelling design and both analysis and semantic interpretation of available datasets.
In this study taxa maps are modelled integrating one or more of the following main components:

- a geo-referenced sample of observed tree species in Europe (E-Forest EFDAC and BioSoil datasets);
- a set of biophysical factors classified in three groups: bioclimatic, terrain and topography; and
- a model describing the relationship of the presence/absence of species with the biophysical factors.

The distribution maps generated for the Atlas may be categorized as referring to two main analysis methods.

The first method has the objective of reducing as much as possible the number of data-transformation and modelling steps. This way, the biases and modelling errors generated due to assumptions and hypotheses required by intensive data-processing are reduced to a minimum degree. On the other hand, it should be underlined that the heterogeneity of the available datasets at European scale is remarkable. Therefore, the local density of reported observations varies highly from Country to Country and sometimes from region to region. Despite all efforts for generating harmonised datasets, systematic biases may affect specific datasets and national inventories so that differences may be perceived in the local reported frequency of taxa from Country to Country. The concept of “model-free” distribution, meaning an ideally “undisturbed” spatial representation of field observations which would not be biased by modelling processing steps, is unfortunately not realistic in a complex and highly heterogeneous spatial extent as the European one.
For example, the influence of sudden changes in the spatial density and distribution of observations is often evident along the boundaries between different countries or even smaller administrative units with autonomous responsibility for the local data collection. This kind of artifacts may be present even in the spatial zoning of other aspects related to forest resources (e.g. de Rigo et al., 2013a) and the cumulated impact of these phenomena may be mitigated with the help of integrated statistical modelling (de Rigo et al., 2014).
A lightweight approach is followed, trying to reduce the modelling steps to the required ones in order for interesting information to emerge and be easily visualised. Following this lightweight approach, a spatial statistics aggregates the field observations at different spatial scales.

These modelling-derived family of visualizations is based on the Geospatial application (de Rigo et al., 2013b) of the Semantic Array Programming paradigm (de Rigo, 2012a,b). In particular, dynamic aggregation tools have been required in order for the arrays of field observations to be processed. The observations refer to four harmonised datasets in the EDFAC, FISE. The observations are aggregated by considering both the frequency of observed taxa and the spatial density of available observations (including the ones in which the analysed taxa are not reported). This second aspect of the distribution of field observations is here considered to qualitatively estimate the level of accuracy of the aggregated frequency. In the frequency-accuracy maps, the frequency represents the proportion1 of field-observations in a given spatial block where at least one of the taxa has been reported. The accuracy represents a nonnegative2 spatially-explicit index summarising how many field observations are available in each block.

The second methodology for estimating taxa distribution is the constrained spatial multi-frequency analysis (C-SMFA). The maps have been prepared using a multi-scale spatial analysis approach (with a modelling terminology, multiple spatial frequencies). The available presence/absence records are divided into training and validation subsets.
The training set is iteratively split into subsets and processed to estimate the spatial probability of presence of each tree taxon using kernel density methods. Multiple kernels are exploited with different diameters to smooth the estimated probability of presence. The goal of the iterative splitting procedure is to allow cross-validation and to select the optimal kernel diameters. The outcome of all iterations is finally merged to obtain the aggregated maps, and constrained to be coherent with available remote-sensing/land-cover maps.
The habitat suitability estimates refer to all mentioned steps linking taxa-specific field observations with bio-geo-climatic information. The algorithms exploit robust ensemble methods with the relative distance similarity (RDS) technique as provided within the Semantic Array Programming paradigm (e.g. see other applications in de Rigo et al., 2013a, Bosco et al. 2011, 2013).

References

1. de Rigo, D., Barredo, J. I., Busetto, L., Caudullo, G., San-Miguel-Ayanz, J., 2013a. Continental-scale living forest biomass and carbon stock: a robust fuzzy ensemble of ipcc tier 1 maps for Europe. IFIP Advances in Information and Communication Technology 413, 271-284. http://dx.doi.org/10.1007/978-3-642-41151-9_26
2. de Rigo, D., Caudullo, G., Busetto, L., San-Miguel-Ayanz, J., 2014. Supporting EFSA assessment of the EU environmental suitability for exotic forestry pests: Final report. EFSA Supporting Publications 2014 (EN-434), 61pp. http://www.efsa.europa.eu/en/supporting/doc/434e.pdf
3. de Rigo, D., Corti, P., Caudullo, G., McInerney, D., Di Leo, M., San-Miguel-Ayanz, J., 2013b. Toward open science at the European scale: Geospatial semantic array programming for integrated environmental modelling. Geophysical Research Abstracts 15, 13245+. http://dx.doi.org/10.6084/m9.figshare.155703
4. de Rigo, D., 2012a. Semantic array programming for environmental modelling: Application of the Mastrave library. In: Seppelt, R., Voinov, A. A., Lange, S., Bankamp, D. (Eds.), International Environmental Modelling and Software Society (iEMSs) 2012

International Congress on Environmental Modelling and Software. Managing Resources of a Limited Planet: Pathways and Visions under Uncertainty, Sixth Biennial Meeting. pp. 1167-1176. http://www.iemss.org/iemss2012/proceedings/D3_1_0715_deRigo.pdf
5. de Rigo, D., 2012b. Semantic Array Programming with Mastrave - Introduction to Semantic Computational Modelling. http://mastrave.org/doc/MTV-1.012-1/
6. Bosco, C., de Rigo, D., Dijkstra, T. A., Sander, G., Wasowski, J., 2013. Multi-Scale robust modelling of landslide susceptibility: Regional rapid assessment and catchment robust fuzzy ensemble. IFIP Advances in Information and Communication Technology 413, 321-335. http://dx.doi.org/10.1007/978-3-642-41151-9_31
7. Bosco, C., de Rigo, D., Dewitte, O., Montanarella, L., 2011. Towards the reproducibility in soil erosion modelling: a new pan-European soil erosion map. In: Keesstra, S. D., Mol, G. (Eds.), Wageningen conference on applied soil science : 'Soil Science in a Changing World'. Wageningen University. http://arxiv.org/abs/1402.3847