European Commission

JRC TECHNICAL REPORTS

EFDM-geo: A spatially-explicit pan-European application of the European Forestry Dynamics Model

Development of a modelling tool for the forest-based bioeconomy

Sarah Mubareka, Maarten Hilferink, Andrea Camia

2016

EUR xxxxx xx

Joint Research Centre

Contact information
Name: Sarah Mubareka
Address:
E-mail:sarah.mubareka@jrc.ec.europa.eu
Tel.:+39 0332 78 6741

JRC Science Hub
https://ec.europa.eu/jrc

JRCxxxxx

EUR xxxxx xx

Print	ISBN xxx-xx-xx-xxxxx-x	ISSN xxxx-xxxx	doi:xx.xxxx/xxxxxx
PDF	ISBN xxx-xx-xx-xxxxx-x	ISSN xxxx-xxxx	doi:xx.xxxxx/xxxxxx

Luxembourg: Publications Office of the European Union, 2016

How to cite: Author(s); title; EUR; doi

Printed in *xxx (country name)*

Table of contents

Acknowledgements

We recognize the excellent groundwork set by the first developers of EFDM: Ola Sallnäs, Tuula Packalen, Seija Sirkiä, Maria Eronen, Minna Räty, Ambros Berger & Renats Trubins. We also acknowledge the indispensable help of Roberto Pilli for support in UBald data provision; Lorenzo Busetto, Daniel Mc Inerney and Roberto Pilli who compiled the Forestats database; Daniele De Rigo, Jose Barredo, who provided support and advice in using their satellite imagery-based maps; and Jesus San Miguel Ayanz for initiating the EFDM "project" in the first place. Thank you to the opinions of external experts who have improved the quality of the report: Mart-Jan Schelhaas and Hans Verkerk.

The development of the EFDM-geo modelling environment was possible thanks to the following individuals who tested the EFDM model in their own countries:

Ambros Berger, Klemens Shadauer (Austrian Federal Research and Training Centre for Forests, Natural Hazards, and Landscape, BFW); Adrian Lanz, Ester Thürig (Swiss Federal Institute for Forest, Snow and Landscape Research WSL); Lukáš Kratěna, Jiři Závodsky, Radim Adolt (Czech Forest Management Institute, UHUL); Susann Klatt, Joachim Rock, Heino Polley (Johann Heinrich von Thünen Institute, Federal Research Institute for Rural Areas, Forestry and Fisheries, Germany); Iciar Alberdi, Maria Pasalodos, Laura Hernández, Joan Josep Ibañez, Isabel Cañellas (Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, INIA - Centro de Investigación Forestal, CIFOR); Roberto Vallejo (Ministerio de Agricultura, Alimentación y Medio Ambiente); Olli Salminen, Tuula Packalen, Kari Korhonen (Natural Resources Institute Finland, LUKE) ; Antoine Colin, Thierry Belouard, Marion Pilate, Pierre Merian, Nicolas Robert, Dominique Leclerc (Institut national de l'information géographique et forestière, IGN) ; Gintaras Kulbokas (Aleksandras Stulginskis university, ASU); Jan Oldenburger, Peter Schnitger (PROBOS); Mart Jan Schelhaas (ALTERRA); Francisco Castro Rego, Graça Louro, Leónia Nunes and Susana Dias (Centre for Applied Ecology "Prof. Baeta Neves", CEABN - Research Network in Biodiversity and Evolutionary Biology, InBIO; School of Agriculture ,University of Lisbon, ISA-UL); Margarida Tomé, Susana Barreiro (Centro de Estudos Florestais, Instituto Superior de Agronomia, Universidade de Lisboa); Ola Sallnäs, Renats Trubins (Swedish University of Agricultural Sciences, SLU); Boštjan Mali, Mitja Skudnik (Slovenian Forestry Institute).

Abstract

A spatially-explicit forest growth modelling system, using the concepts of the European Forestry Dynamics Model (EFDM) as a core engine for forest growth, was developed to support the assessment of woody biomass availability in Europe for the forest-based sector of the bioeconomy. This version of EFDM is named "EFDM-geo". This document describes EFDM-geo as it was implemented for seamless spatially-explicit pan-European modelling, allowing for continental-wide scenario runs and the integration of demand for harvest levels of timber assortments.

The model contains a forest growth component that simulates natural growth for different forest types in Europe. It also simulates management activities that affect the natural growth process. Three activities: harvesting, thinning and no management, are configured to respond according to the demand for wood for material and energy uses. Demand for industrial solid wood products is considered, broken down into four commodities: coniferous sawlogs, non-coniferous sawlogs, coniferous pulpwood and non-coniferous pulpwood.

The national-level data on market demand is downscaled to NUTS 1 regional level. At this level, bio-geo-climatic variables such as the forest type (species, age and volume classes), topography and road accessibility are included in the computation of which forest stands are more likely to be subject to some management activity. As the simulation progresses in time, the forest type evolves according to the predefined rules governing growth for given management activities. As a recursive model, the forest composition can be estimated at any point in time during the simulation for any individual NUTS 1 region in Europe or for the whole EU territory.

The spatially-explicit component of this model is useful for computation of environmental indicators. When overlain with data such as waterways, soil type and areas prone to natural disaster, the impact of forest growth and activities in the forest can be studied. Furthermore, the impacts of spatially-explicit policy such as zoning and protection of certain areas; or the simulation of new biomass processing plants can be assessed through scenarios.

The implementation of the model and the modelling environment are available as open source software.

Highlights

- EFDM-geo is a working software package that is able to handle NUTS 1 regional –level and 1-ha raster level pan-European modelling of forest growth. It is compete with protocol for file structure and data integrity checks; delivering meaningful error messages when data input is incomplete, and compiling raw datasets to matrices on the fly
- EFDM-geo is a quasi-dynamic recursive forest growth model composed of a top-down NUTS 1 regional-level sub-model, and a bottom-up raster-level sub-model
- The regional sub-model computes the movement of forest areas through an age/volume class matrix within regional and species class bounds, and according to regional and species-specific parameters
- The raster sub-model takes local landscape and anthropogenic characteristics into consideration to determine the likelihood of management treatments to take place based on cost
- The amount and quality of wood harvested is governed by an exogenous forest sector model
- Management activities are scaled to meet harvest demand
- The location of the management activities is the result of an optimisation of resources available and location of those resources
- EFDM-geo is equipped with the capability to model net changes in wood-producing areas; this will facilitate the integration of short rotation forestry
- EFDM-geo is implemented in a free and open source software and has been released under the EUPL license here: https://webgate.ec.europa.eu/CITnet/stash/projects/FISE
- EFDM-geo only applies to even-aged forests at the time of writing

1 Introduction

The assessment of biomass availability and potentials is of growing importance in Europe and globally. Woody biomass is at the core of the bio-based economy aiming at replacing fossil-based materials with renewable sources. The GHG reduction targets and the related affected policies are pushing a stronger demand for bioenergy, which is mostly based on woody biomass. Given the strategic importance of woody biomass as a raw material for the European bioeconomy, the assessment of its current and future availability under different policy and climate scenarios is crucial for policy makers and for the sustainable use of renewable resources. What is described in this report results from the recognized need for the further development and refinement of modelling tools to model European forests and the interactions with the forest-based sector at continental scale.

The European Forestry Dynamics Model (EFDM) incorporates different aspects of forest dynamics and forest management. The reimplementation of parts of EFDM in a spatial-explicit environment, as described in this document, is called "EFDM-geo". EFDM-geo facilitates both linkages with other forest-based sector models accounting for supply-demand interactions in a broader bioeconomy perspective and the future integration with pan-European land use models. Furthermore, it respects the objectives outlined in Mubareka et al, 2014 by using a software that offers a form of array programming with internal integrity checks for data consistency, and is developed and available as free and open source[1]. The purpose of such a system is to support the assessment of potential impacts of different policy and climate scenarios on the EU's forests and the forest-based sector. The allocation of management activities to specific forest types is guided by

[1] available as open source by Object Vision BV, see: http://www.objectvision.nl/geodms

information at two levels: a bottom-up model based on local characteristics of each site (raster cell); and an overarching, top-down, national-level demand for wood products.

This document describes a functioning modelling framework. It demonstrates possible model output, but does not report any real results, thus reported numbers are to be taken as examples.

1.1 Background

1.1.1 EFDM

EFDM simulates the development of the forest and estimates volume of wood harvested, integrating management schemes as part of its input. The model was jointly developed by the Natural Resources Institute Finland and the Swedish University of Agricultural Sciences for the European Commission (Packalen et al, 2014). EFDM was conceived as a flexible system for harmonized forestry modelling for all European countries.

EFDM belongs to the forest model family of matrix models, born in the 1940's to model future plant and animal population structures (Liang and Picard, 2013). It is built around a basic matrix structure, defined by a set of fixed states, between which "units" of forest move over time. The core of the matrix structure is the dynamic state-space, which can be defined by, for example, volume and age. Matrix models rely on a series of transition matrices. When applied in the forestry sector, the transition matrices express the probability of a forested area leaving its current position within a matrix to join a different, usually higher, position within the matrix, thus acquiring the characteristics, and therefore the probabilities for transitions, of this new category. More details on the origins of EFDM can be found here: Sallnäs (1990).

1.1.2 EFDM-geo

In order to compute pan-European, spatially-explicit model runs, some components of the EFDM model (originally coded in R, with details of that implementation on the JRC Science Hub[2]), were re-coded in GeoDMS, while many new components were added. The package of recoded EFDM plus additional features is referred to as the *EFDM-geo*. New components include the capability of spatially disaggregating forest inventory data; creating initial state matrices on the fly from this data; reading demand files for commodities; re-scaling activities probabilities; and spatially disaggregating activities.

When modelling forestry dynamics in an area with complex land use such as Europe, it is useful to maintain a spatial disaggregation of forest resources for the model run. In EFDM-geo, areas of forest are the sum of 1-ha cell to regional or national level. It produces output on a grid-unit basis, while taking into consideration the spatial location of each cell by changing how the model behaves based on spatial criteria such as the profile of neighbouring cells; proximity to elements such as roads or processing plants; spatially explicit climate and geographical parameters; and differing management practices within, for example, protected areas. The spatially-explicit approach used in EFDM-geo can furthermore account for land-use changes, thus including changes in forest area.

It is not the objective of the spatially-explicit modelling tool to "predict" future maps of forest types in Europe. Indeed the forest growth component of the model aggregates data at regional level within the forest area matrix, just as it does for EFDM. Forest type totals within the matrix are indeed consistent with EFDM if the same data sources are used. The added value of the spatial dimension comes into play during the scaling of activities, as we will explain in the following sections.

---

[2] https://ec.europa.eu/jrc/en/european-forestry-dynamics-model

2 Technical implementation

EFDM-geo operates at two levels: an overarching, top-down, national-level demand for wood products called the regional model; and a bottom-up model called the raster model, which takes local characteristics of each site into consideration at pixel-level. In the following sections, we describe these two levels in detail. Here we give a brief overview to describe how they come together in EFDM-geo.

The regional model refers to the model runs at NUTS 1 level. In the regional model, the demand for wood products governs the amount of wood that is to be removed from the forest in a given country. This demand is given at a national level by an exogenous source. The requested wood harvested in the forests is computed at every time step, given the potential for management activities to produce the right kind of wood to suit the demand. At the start of the simulation, each forest type is assigned a probability for a given management activity. What happens in the regional model, is that this probability is rescaled to match the demand. Thus, the management activities that could potentially take place given the species class, age class and volume class, are rescaled so that only the requested amount of wood, in the optimal configuration of assortments, is actually harvested. If the scaling factors for the activities is equal to 1, the result at the regional level is equivalent to EFDM.

In the regional model, resources are sought in the forests in order to immediately satisfy the demand at every time step. The choices made by the model as to which management activities are applied and where, are made in the "Raster Model". At every simulation step, the distribution of the forest area within the volume-age class matrix, changes. The new distribution of the forest area depends on whether or not any silvicultural treatments took place. If none took place, the shift in forest area is governed by transition probabilities (described in further detail in Section 3.2). Transitions may furthermore take place based on the type of treatment that was applied. In order to assess whether the forest area will undergo treatment, and if so, which treatment will take place, a probability of treatment (called "activity") is assigned. The spatial context within which the resources are located is associated with the lowest possible associated cost for treatment in EFDM-geo. This is accounted for by introducing spatial layers relative to forest management. Elements such as roads and topography; as well as neighbourhood influences, history of management and, of course, current forest type, play a role in the likelihood of any given management treatment. This likelihood, or the inverse of the cost, is re-evaluated at every time step.

Table 1 summarizes the indices used to describe both the raster and the regional models.

Table 1. Indices used to describe the raster and regional models

Variable	Definition	Description
i	Raster cell	1 cell = 1 hectare (or 1km in section 3.1.1)
f	Forest type	Collectively: age, volume, species
v	Volume classes	Class widths and number of classes can vary between NUTS 1 regions
g	Age classes	Class widths and number of classes can vary between NUTS 1 regions
s	Species	In EFDM-geo: broadleaves and conifers
r	Region	NUTS 1 in EFDM-geo
a	Activity	No Management, thinning, harvesting, recruitment at volume and age classes = 0.
t	Time	Each time step corresponds to the width of the age class
c	Commodity	Coniferous (Cf) sawlogs, non-coniferous (nCf) sawlogs, Cf pulpwood, nCf pulpwood

Table 2 summarises the input data for the model. This data includes cost maps, the administrative regions map (NUTS 0 and NUTS 1 shapefiles[3]), maps required for generating the initial state on the fly (age, volume and tree species probability of presence); transition probabilities matrices; initial estimate of activities probabilities; yield matrices; and demand for commodities.

The notation conforms to the following rules: the lower left indicates the main simulation loop; data for different times are found at different sequenced instantiations. If different run-chains were made for different scenarios, a scenario index would also be in the lower left corner. An index for country (for which demands are given, containing multiple regions) is hidden for simplicity but should be in the lower left as well if it were to be notated explicitly. Upper left indicates different attributes (arrays) close to each other, often generated by functions within a metascript[4], for example to be applied to *all* regions (*r*), or to *all* species (*s*). Lower right indicates the indices that extend across the domain of the arrays that are used to represent the input data of state variable. Upper right indicates the value range of data (as in incidence maps) or the column range (in probability matrices where each row represents a "from" state and each column represents a "to" state).

[3] http://ec.europa.eu/eurostat/web/nuts/overview
[4] Using http://objectvision.nl/geodms/operators-a-functions/metascript/for_each

Table 2. The input data or the model.

${}^{a}S_{i}$	Cost map	Cost in cell (i) per management activity (a)
R_i^r	Region incidence map	A region within a range (r) is attributed to each cell (i)
${}_{0}Y_{i}^{s,v,g}$	Initial state incidence map, t=0	Initial state of cell describing the forest type within a range of species (s), age (g) and volume (v) classes for each cell (i) at time=0
${}^{r,a,s}P_{v',g'}^{v,g}$	Transition probability matrices	This is implemented as a matrix for each species (s), activity (a) and region (r). v and g refer to the transition probability for each age and volume class; v’ and a’ refer to volume and age for the following time step within that species, activity and region.
${}^{r,s}A_{v,g}^{a}$	Activity fraction prior matrix	Initial activities’ probabilities within a range of activities (a) for each age (g) and volume (v) class. These are generated for each species (s) and region (r)
${}^{r,s}H_{v,g}^{c,a}$	Harvest matrices	Harvest within an age and volume class as the result of a given activity (a) to satisfy the request of an assortment (c) for a given species group (s) and region (r)
${}_{t}D^{c}$	Demand for assortment	Demand per assortment (c) is given at every time step (t)

Table 3 summarises the variables associated to the endogenous calculations or steps within the model.

Table 3. State variables.

${}^{r,s}_{0}X_{vg}$	Initial state vector, t=0	Sum of hectares associated to each volume (v) and age (g) class combination at time = 0. The class ranges will vary for each species (s) in each region (r)
${}^{r,s}_{t}X_{vg}$	Initial state vector, t<0	Same as above, but for time steps during the simulation
${}_{t}Y_{i}^{s,v,g}$	Initial state incidence map, t<0	Initial state of cell at successive time steps for each cell within the range of species (s), age (g) and volume (v) classes
${}_{t}M_{i}^{a}$	Activity incidence map, t<0	Activities incidence map, within a range of given activities (a), for a period during the simulation
${}^{r,s}_{t}B_{v,g}^{a}$	Activity fraction posterior matrices	Rescaled activities in a region and per species, given the exogenous demand for wood
${}^{s}_{t}\sigma^{a,b}$	Balancing factors	Result of iterative proportional fitting to optimize allocation of activities to meet demand

2.1 Regional Model

In EFDM-geo's regional model, the main driver for the model is the harvest requested to satisfy the demand for the following commodities: non-coniferous pulpwood, non-coniferous sawlogs, coniferous pulpwood, and coniferous sawlogs. This demand is given at a national level, as an exogenous piece of information, and is downscaled to NUTS 1 regional level endogenously. Each activity, depending on what forest type it is applied to, harvests a specific quantity of wood to satisfy the demand. The sum of this harvest estimate is region-specific, meaning it is constrained within the region without any spill-over between regions.

Each forest type, given the region in which it is found (f_r), has a probability of succumbing to a management activity. For each region, there is an activity fraction prior per forest type. This means that the activity probabilities will vary from forest type to forest type within a region, and this variance is also inter-regional. We can therefore express different probabilities for final felling for a given forest type for different regions. This allows us to implicitly take the growing conditions into account, as well as the regional management practices. This initial probability for activities is fixed over time; what changes in time is the actual activity that takes place.

The model is able to optimize the assignment of assortments to forest types through iterative proportional fitting if region-specific coefficients that map forest types to assortments are available. In such an approach, balancing factors need to be computed so that the demand is met by the potential harvest from the activities given the state of the forest at any given time step. These balancing factors are used to solve the problem of which activities will actually need to take place in order to meet the demand for each assortment (eqs. 1,2):

$$\forall c\colon \sum_{r,a,f} {}^{r,s}H_{v,g}^{c,a} \cdot {}^{r,s}_{t}B_{v,g}^{a} \cdot {}^{r,s}_{0}X_{vg} = {}_{t}D^{c} \qquad \text{eq. 1}$$

where the activity fraction posterior matrix is computed with the balancing factors (eq. 2)

$$ {}^{r,s}_{t}B^{\,a}_{\,v,g} \coloneqq \frac{{}^{s}_{t}\sigma^{\,a} \cdot {}^{r,s}A^{\,a}_{\,v,g}}{\sum_b {}^{s}_{t}\sigma^{\,b} \cdot {}^{r,s}A^{\,b}_{\,v,g}} \qquad \text{eq. 2} $$

Each assortment is linked to a set of combinations of activities and forest types that could contribute most to it. Although the split between species is clear, the relative contribution of different age and volume class combinations to pulpwood and sawlogs respectively, is not. Hence, for cases where no information on splitting between pulpwood and sawlogs is available, a simplified approach is implemented whereby the total demand for coniferous and non-coniferous products respectively, is split evenly over all forest types. This results in a linear effect, requiring only one iteration to solve the problem.

The posterior activities fractions and the relative associated transitions are then applied to the state to compute the state for the next time step (eq. 3):

$$ {}^{r,s}_{t+1}X_{\,v'g'} \coloneqq \sum_{a,v,g} {}^{r,a,s}P^{\,v,g}_{\,v',g'} \cdot {}^{r,s}_{t}B^{\,a}_{\,v,g} \cdot {}^{r,s}_{t}X_{\,vg} \qquad \text{eq. 3} $$

This newly computed level of activity is referred to as the "activity fraction posterior matrix". The demand is therefore met after considering the initial state, the activities that will take place and the harvest given by the activities. This is repeated at every time step.

In terms of scenario configurations, it is possible to influence the activity in a cell according to the region in which the cell is found. A forest type might be more or less likely to be thinned given the region in which it is found due to subsidies, taxes or other political reasons. We call this inter-regional differentiation "shadow price". It is analogous to the difference in the cost of an apartment in a major city, and the cost of an apartment of similar characteristics in the countryside. The shadow price for each activity therefore influences all cells in the region, accounting for inter-regional differences despite raster cell similarities. Despite the shadow prices, a situation whereby several cells may have the same associated management costs, the discrete allocation will make an arbitrary choice, as described in section 2.2.2.

2.1.1 Claims for management activities

The calculation of the claims to certain management activities is the bridge from the demand for forest assortments, to the actual forest resources. The claims refer to the area of forest that should undergo a treatment for each age/volume class in order to satisfy the demand for forest commodities. This demand is derived from the Global Forest Trade Model (GFTM, Jonsson et al 2015[5]). The claims are recalculated at each time step given the demand, which in this prototype are the same for every time step to facilitate analysis, and the resources available to suit the demand. The resources, as we have said before, are divided by conifers (CF) and broadleaves (BL), thus this first division leaves little doubt. What needs to be implicitly calculated is the division between pulpwood and sawlogs. Figure 1 shows an example of the claims on forest areas for the first three time steps for conifers, for the first age/volume classes in a region in Europe:

---

[5] http://publications.jrc.ec.europa.eu/repository/bitstream/JRC96814/lb-na-27360-en-n%20.pdf

	2015 noMg	2015 ff	2015 th	2020 noMg	2020 ff	2020 th	2030 noMg	2030 ff	2030 th
id	nomgmt	finalfell	thin	yr2025/raste	yr2025/raste	yr2025/raste	yr2030/raste	yr2030/raste	yr2030/raste
0 (0..5[yr] 0..36.3333[m3/ha];)	69361	0	3682	581494	0	27401	508259	0	22495
1 (0..5[yr] 36.3333..72.6667[m3/ha];)	14632	0	776	0	0	0	0	0	0
2 (0..5[yr] 72.6667..109[m3/ha];)	0	0	0	0	0	0	0	0	0
3 (0..5[yr] 109..145.333[m3/ha];)	0	0	0	0	0	0	0	0	0
4 (0..5[yr] 145.333..181.667[m3/ha];)	0	0	0	0	0	0	0	0	0
5 (0..5[yr] 181.667..INF[m3/ha];)	0	0	0	0	0	0	0	0	0
6 (5..10[yr] 0..36.3333[m3/ha];)	0	0	0	38579	28	1501	326969	229	11960
7 (5..10[yr] 36.3333..72.6667[m3/ha];)	0	0	0	39190	48	1418	234476	270	7975
8 (5..10[yr] 72.6667..109[m3/ha];)	0	0	0	3121	5	103	2554	4	79
9 (5..10[yr] 109..145.333[m3/ha];)	0	0	0	0	0	0	0	0	0
10 (5..10[yr] 145.333..181.667[m3/ha];)	0	0	0	0	0	0	0	0	0
11 (5..10[yr] 181.667..INF[m3/ha];)	0	0	0	0	0	0	0	0	0
12 (10..15[yr] 0..36.3333[m3/ha];)	0	0	0	0	0	0	14736	22	463
13 (10..15[yr] 36.3333..72.6667[m3/ha];)	0	0	0	0	0	0	45163	112	1253
14 (10..15[yr] 72.6667..109[m3/ha];)	0	0	0	0	0	0	17025	60	407
15 (10..15[yr] 109..145.333[m3/ha];)	0	0	0	0	0	0	1560	7	31
16 (10..15[yr] 145.333..181.667[m3/ha];)	0	0	0	0	0	0	51	0	0
17 (10..15[yr] 181.667..INF[m3/ha];)	0	0	0	0	0	0	0	0	0
18 (15..20[yr] 0..36.3333[m3/ha];)	0	0	0	0	0	0	0	0	0
19 (15..20[yr] 36.3333..72.6667[m3/ha];)	0	0	0	0	0	0	0	0	0
20 (15..20[yr] 72.6667..109[m3/ha];)	0	0	0	0	0	0	0	0	0
21 (15..20[yr] 109..145.333[m3/ha];)	0	0	0	0	0	0	0	0	0
22 (15..20[yr] 145.333..181.667[m3/ha];)	0	0	0	0	0	0	0	0	0
23 (15..20[yr] 181.667..INF[m3/ha];)	0	0	0	0	0	0	0	0	0
24 (20..25[yr] 0..36.3333[m3/ha];)	0	0	0	0	0	0	0	0	0
25 (20..25[yr] 36.3333..72.6667[m3/ha];)	0	0	0	0	0	0	0	0	0
26 (20..25[yr] 72.6667..109[m3/ha];)	0	0	0	0	0	0	0	0	0
27 (20..25[yr] 109..145.333[m3/ha];)	0	0	0	0	0	0	0	0	0
28 (20..25[yr] 145.333..181.667[m3/ha];)	0	0	0	0	0	0	0	0	0
29 (20..25[yr] 181.667..INF[m3/ha];)	0	0	0	0	0	0	0	0	0
30 (25..30[yr] 0..36.3333[m3/ha];)	0	0	0	0	0	0	0	0	0
31 (25..30[yr] 36.3333..72.6667[m3/ha];)	0	0	0	0	0	0	0	0	0
32 (25..30[yr] 72.6667..109[m3/ha];)	0	0	0	0	0	0	0	0	0
33 (25..30[yr] 109..145.333[m3/ha];)	0	0	0	0	0	0	0	0	0
34 (25..30[yr] 145.333..181.667[m3/ha];)	0	0	0	0	0	0	0	0	0

Figure 1. Claims for management activities to occur in different age/volume groups for three time steps: 2020, 2015, 2020. Units are in forest area affected (ha).

In summary, the demand for wood, which is in m^3, per time-step, per assortment, is confronted with the production potential of the forests at regional level for the same time step. The activities are then rescaled to make them meet. At the start of any given time step, the initial factor at which the rescaling starts is taken into account. This number is inherited from the last iteration from the previous period, which means that for the first time step of the simulation, a rescaling factor of 1.0 because there was no previous demand upon which to rescale. The previously-used factors are then adjusted for any mismatch between demand and production.

2.2 Raster model

In EFDM-geo’s raster model, an overall cost for each management activity type (final felling, thinning or no management) is associated to each raster cell. Only one type of activity will be assigned to each one-hectare cell. The total cost for an activity for a cell is determined based on a combination of several layers affecting the total cost to manage forests in any given cell; and the probability of each activity being applied to the forest type found in that raster cell. Naturally, the potential for each cell to actually succumb to final felling or thinning, despite its relative probability within the activities choices, will depend on the demand for wood products, solved in the regional model. This means that although a raster cell may contain a forest type that is ripe for final felling, it may not be felled at all if the demand is insufficient to request the wood contained in that cell.

2.2.1 Allocation of management activities

At the heart of the simulation is the discrete allocation of activities to 1-ha parcels of forested land. The regional amount of activities is disaggregated among the cells based on their characteristics (location, forest type) while minimizing cost.

Each raster cell (i), has an associated forest type (f) (refer to Table 1 for notations). This means that the cell is either labelled as coniferous or broadleaved at time=0 (this remains intact for the duration of the simulation), and has a given age and volume class (these vary throughout the simulation). Furthermore, each raster cell belongs to a region (r). The description of each cell at time=0 (T0), or start of the simulation, is what is called the “initial state” and is denoted as ${}^{rf}_{0}Y_{vg}$. The sum of the initial states of all cells must equal the sum of the initial state vector (used in the regional model) of the region in which the cells are found (eq 4):

$$\forall r, f\colon {}^{r,s}_{0}X_{vg} = \textstyle\sum_i R_i^r \cdot {}_{0}Y_i^{s,v,g} \qquad \text{eq. 4}$$

The initial state in a matrix model is not usually spatially-explicit and is indeed presented as a vector in the regional model. In the raster model however, we transform the initial state into a spatially-explicit input into the model with the help of satellite data. We derive three layers using satellite data: age, volume and species (see Section 3.1 for details). Each raster cell also has an activity potential based on the forest type present. Furthermore, given its location in space, each cell will have an associated cost, influencing the likelihood that the activity will actually take place or not. The minimum total cost (presented as the inverse of cost in the annotation ${}^{a}S_{i}$), is sought during the process of allocating the activities to the individual cells (eq. 5):

$$\textstyle\sum_{i,a} {}^{a}S_{i} \cdot {}_{t}M_{i}^{a} \qquad \text{eq.5}$$

Equation 5 is resolved while taking into account the important following constraint, which is the sum of the activities for the entire region (eq. 6):

$$\forall f, a\colon \textstyle\sum_i R_i^r \cdot {}_{t}Y_i^{s,v,g} \cdot {}_{t}M_i^{a} = {}^{r,s}_{t}B^{a}_{v,g} \cdot {}^{r,s}_{t}X_{vg} \qquad \text{eq. 6}$$

Decisions are taken locally based on a cost that is offset by a shadow price per region per activity to meet these constraints. A positive shadow price reflects a 'subsidy' that might be required to get sufficient locations for an activity whereas a negative shadow price reflects a 'taxation' to prevent an excessive allocation. This is a handy mechanism for scenario configuration as well. Regional or local-level subsidies to harvest a certain forest type can be represented with this mechanism. When too many different locations have the same relative cost for one activity, the discrete allocation will only allocate the required amount of cells by taking arbitrary choices. In these cases, costs are artificially perturbated[6].

The allocated activities and associated transitions are then used to calculate the probability distribution (which is equal to 1) for the new state of each cell for the following time step (eq. 7):

$$Prob({}_{t+1}Y_i^{s,v',g'} = 1) \coloneqq \sum_{avg} {}^{a,s}P_{v',g'}^{v,g} \cdot {}_{t}M_i^{a} \cdot {}_{t}Y_i^{s,v,g} \qquad \text{eq. 7}$$

The expected occurrence of the new state will balance with the new regional totals (eq. 8a, or eq. 8b):

$$\sum_{avg} \left[{}^{a,s}P_{v',g'}^{v,g} \cdot \sum_i R_i^r \cdot {}_{t}Y_i^{s,v,g} \cdot {}_{t}M_i^{a} \right] = {}_{t+1}^{r,s}X_{v'g'} \qquad \text{eq. 8a}$$

$$\sum_{avg} {}^{a,s}P_{v',g'}^{v,g} \cdot {}_{t}^{r,s}B_{v,g}^{a} \cdot {}_{0}^{r,s}X_{vg} = {}_{t+1}^{r,s}X_{v'g'} \qquad \text{eq. 8b}$$

2.2.2 Cost maps

Cost maps are crucial to the simulation. In this context, forests situated in the least expensive areas for management activities are more likely to be managed. The cost maps can be specific to different management practices as well as regions due to different harvesting technologies for different species. Figure 2 shows the example map for final felling. The colour gradient from reds to greens represents a decreasing cost in felling the forests. A generic map combining riparian areas, slope, topography and roads is used in this configuration, although we know that this issue is site and species specific, and depends heavily on the technology available in the regions. This is therefore subject to further improvement.

[6] More details here: http://wiki.objectvision.nl/index.php/Virtual_perturbation

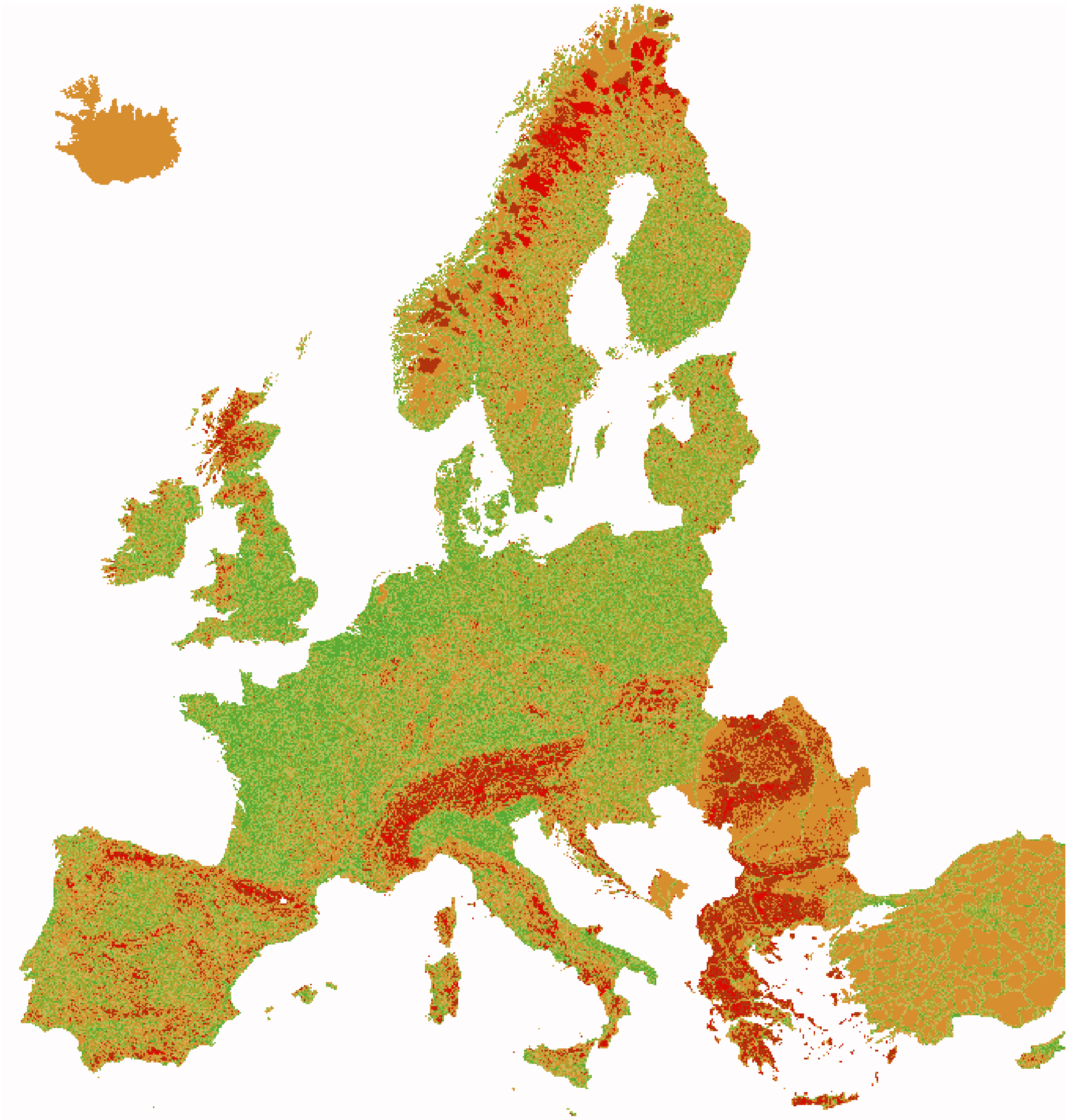


Figure 2. Costs for final felling, based on a combination of factors described in 2.2.2

(red->green=high cost->low cost)

2.3 Summary

Figure 2 summarises the processes in EFDM-geo, including the exogenous data that is generated in different modelling environments. First, EFDM-geo generates the initial state on the fly. It does so by reading text files where volume and age class divisions are stipulated. The volume and age class divisions and thresholds are usually different for different NUTS 1 regions within large countries. It then reads the portion of the pan-European maps of volume and age that is geographically relevant for the user. This is usually one NUTS 1 region, but can be several regions or countries if a batch is launched. Following this, the optimisation process to allocate species to 1-ha cells is initiated and a species distribution map is generated on the fly[7]. Once the species distribution map is in place, the software can then generate a vector of the initial state(s) on the fly. There will be one vector per NUTs 1 region per species. EFDM-geo then reads the demand file for assortments computed by the GFTM model, and translates the assortments to categories of forest types that could satisfy the requested harvest level. The management activities probabilities are then read by EFDM-geo. These probabilities can be generated in different ways. In this implementation, the datasets were generated using an R script, which generically assigns higher probabilities for final felling to larger age and volume classes; and higher probabilities of thinning in medium age and volume classes[8]. Once the demand and management potential are known to the model, it proceeds to optimise the allocation of management activities to meet the demand, based on the stock available in the forest. This is done at a NUTS 1-level of computation. The model then applies the management activities to where it thinks are the most logical locations, using cost maps to guide the choices. Using the transition probabilities matrices, which are usually generated in the EFDM R-code, the forests transit through the areas matrix according to the rules governing the management

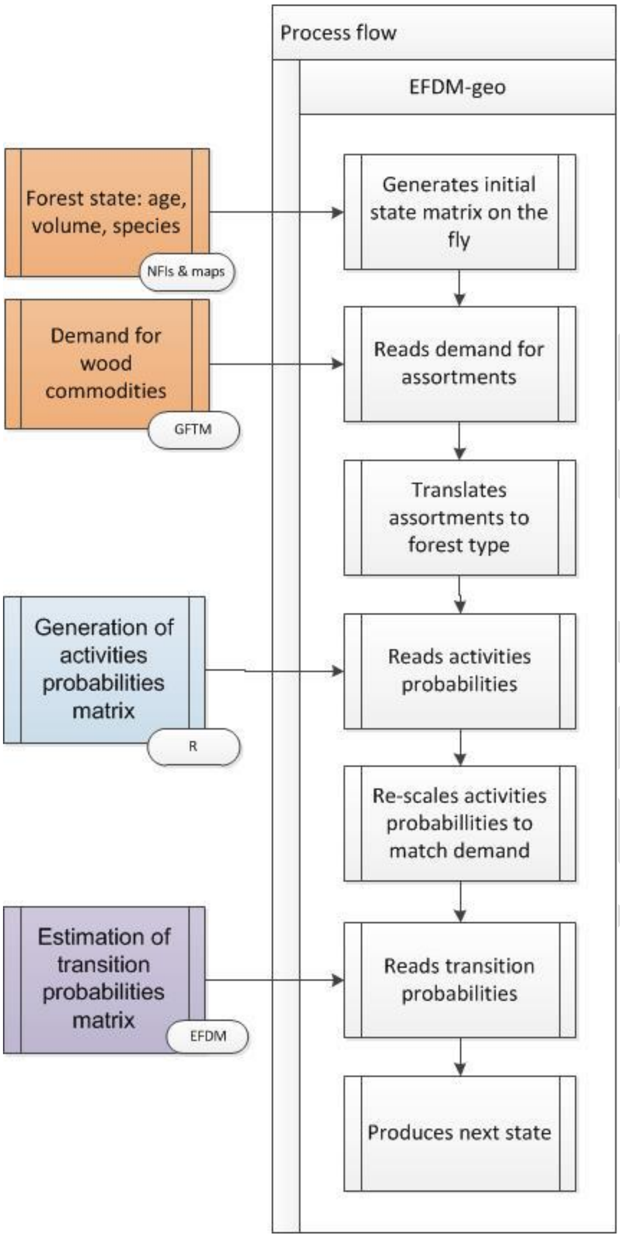


Figure 3. Overview of EFDM-geo processing steps, including the exogenous data and origin.

[7] It is foreseen to generate the volume and age maps in a similar way. At the time of writing, these two layers are generated outside of EFDM-geo.

[8] This estimation should be replaced by local experts before the model becomes operational.

activities – including the 'no management' option. Thus, the state of the forest for the following time step is computed.

3 Data Preparation

A major inhibitor to spatial modelling of forest resources is the availability of seamless datasets. These datasets should be sufficiently descriptive to assign a profile that is detailed enough to distinguish between forest types. Furthermore, they should be accurate enough to produce realistic results. The data behind the initial state matrix would ideally be provided by national experts. It is indeed one of the objectives of this publication to peak interest of national experts so they may be prompted to join this initiative in order to improve the tools described here. In the absence of detailed national-level data, publically-available datasets are used. This section outlines a possible methodology for producing seamless spatially-explicit input data at the best spatial resolution available from public sources of data.

3.1 Initial State

The initial state is a description of the forest at the beginning of the simulation, which is the year 2015. The area of forest in each volume and age class category, for each species group (broadleaves or conifers) is quantified. The initial state is a matrix of forest types, with each "box" of the matrix containing their associated area in hectares. In EFDM-geo, each species has its own initial state file. This is to facilitate control on a per species basis, with the foresight of detailing species groups beyond what is applied here (broadleaves and conifers as two main groups). Furthermore it allows the user the freedom to configure species by age and volume class breakdown, in a completely independent manner. In the original configuration of the model in R, the initial state was analogous to the "state space", referred to when describing Markov models (further detailed in Section 3.2.2), which implied that the initial state files had to contain all possible states regardless of whether or not they can be verified as actually being realistic with respect to the field data. Since the state space had to contain an element for each potential combination of factor levels, the total number of possible combinations was always a multiple of the factors used to describe the forest type. However in EFDM-geo, the initial state files do not have to contain all possible states, although the unnamed states will be generated and filled in a following time step, should any forest area occupy that space in the matrix.

The following datasets are used in the current configuration, for estimating forest area per volume and age class (Table 4):

- Estimation of the relative probability of presence for 78 taxa and genus on a 1km x 1km grid (de Rigo et al., 2016). Each layer considers only the probability of presence of the relevant taxa, and not information from other taxa. The maps are generated based on a frequency-of-observations analysis, whereby presence and absence of species given by field observations within each square kilometer is factored into the algorithm, as is the Corine Land Cover 2006 classification of broadleaves and conifers (method detailed in de Rigo et al. 2014)
- Maps of the growing stock derived from satellite imagery and expert knowledge. There are different datasets available, including products for conifers and broadleaves by Gallaun et al (2010) or from the Biomasar (Santoro et al., 2011) for all tree species groups combined
- Forest map 2006 at 25 m resolution (Kempeneers et al, 2013)
- Corine Land Cover[9]
- UBald dataset of general equations to calculate average volume per species per age (Pilli et al 2013)

[9] http://www.eea.europa.eu/publications/COR0-landcover

- “ForeStats” in-house JRC database: the most up-to-date, publicly available data for forest parameters of European countries (EU27 + Norway, Switzerland and Balkan countries), at the highest possible level of spatial disaggregation (i.e., at the smallest available NUTS level). ForeStats includes forest area, growing stock volume, annual increment of growing stock volume and forest biomass.
- State of Europe’s Forests 2015

Table 4. Summary of input data

Data source	Authors	Resolution	Units
Probability of presence of tree species	De Rigo et al (2016)	1km x 1km	0-100%
Growing stock volume	Gallaun et al (2013) or Santoro et al (2011)	1km x 1km	m^3 / ha
Forest non-forest	Kempeneers et al. (2013) and Pekkarinen et al. (2009)	25m x 25m	boolean
Corine Land Cover	European Environment Agency http://www.eea.europa.eu/publications/COR0-landcover	250m x 250m (rescaled to 1-ha)	classes
UBald database of growth functions	Pilli et al (2013)	Site-specific	coefficients
ForeStats	Busetto, Pilli and Mc Inerney (JRC internal documentation)	Best available, regional	Different for different parameters
State of Europe’s Forests 2015	Ministerial Conference on the Protection of Forests in Europe	Country	Different for different parameters

Based on the available data, mapped estimates of forestry resources needed for spatial modelling are produced. The maps are generated with 1km x 1km grid cells (*i*). A description of our approach in generating the following maps is given in this chapter in the respective sub-sections[10]:

1. Estimated area occupied per species group per square kilometre [${}^{s}S_i$]
2. Estimated total growing stock volume per species per square kilometre [${}^{s}V_i$]
3. Estimated average age per species per square kilometre [${}^{s}G_i$]

3.1.1 Species distribution

Species distribution is the result of a balanced allocation of cells to broadleaves (BL), conifers (CF) and other (not forest). Allocation, a process also used elsewhere in the model

[10] The notation used in these sections follow those described in section 2; simulation loop is always 0 for this section and is therefore not explicitly notated.

to allocate management activities, is the optimized assignment of a specific category from a set within the category (such as species classes) to a raster cell, based on several criteria. In this case, the allocation is "discrete" or whole (no partial allocation of species classes to fractions of cells) because each cell is allocated to only one out of a limited set of species classes.

There are three main required data sets for discrete allocation: 1. A set of numbers describing how many hectares should be assigned to each class; 2. Suitability criteria, analogous to the cost mapping approach we described earlier in section 2.2.2, only this time for species distribution; 3. A base map from which to start the "seeding" of the allocation process.

3.1.1.1 Land claims for broadleaves and coniferous forests

The areas of land to be covered in either broadleaves or conifers are given by National Forest Inventory Data in the ForeStats database at the best available NUTS level (usually NUTS 2), normalised by the SOEF 2015 "Forest available for wood supply" (FAWS) areas at national-level. In this way, the forest areas always sum to those reported in SOEF for 2015 for any given country.

3.1.1.2 Suitability maps

The main dataset used for estimating suitability for broadleaves and conifers is the forest map 2006 at 25 m resolution (Kempeneers et al, 2013). In many cells, according to the high-resolution raw data, cells can be shared between broadleaves and conifers, for a total of full coverage, and the sum of cover of CF + BL can be less than 100%. (Figure 4):

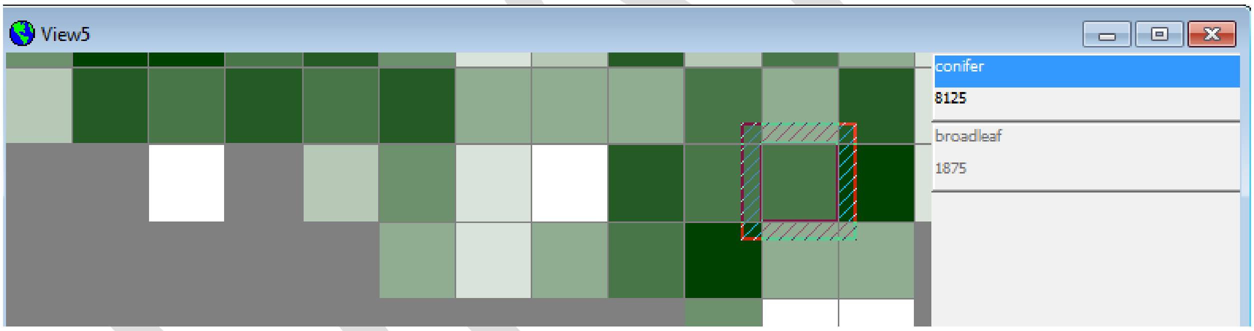


Figure 4. Example of 1-ha cell partially covered by broadleaves (1,875 m^2) and conifers (8125 m^2).

Some cells with the lowest fractions are considered as "other land use" while other cells with high fractions are considered as 100% CF or 100% BL. Figure 5 shows an example of the cell density in a region in central Finland.

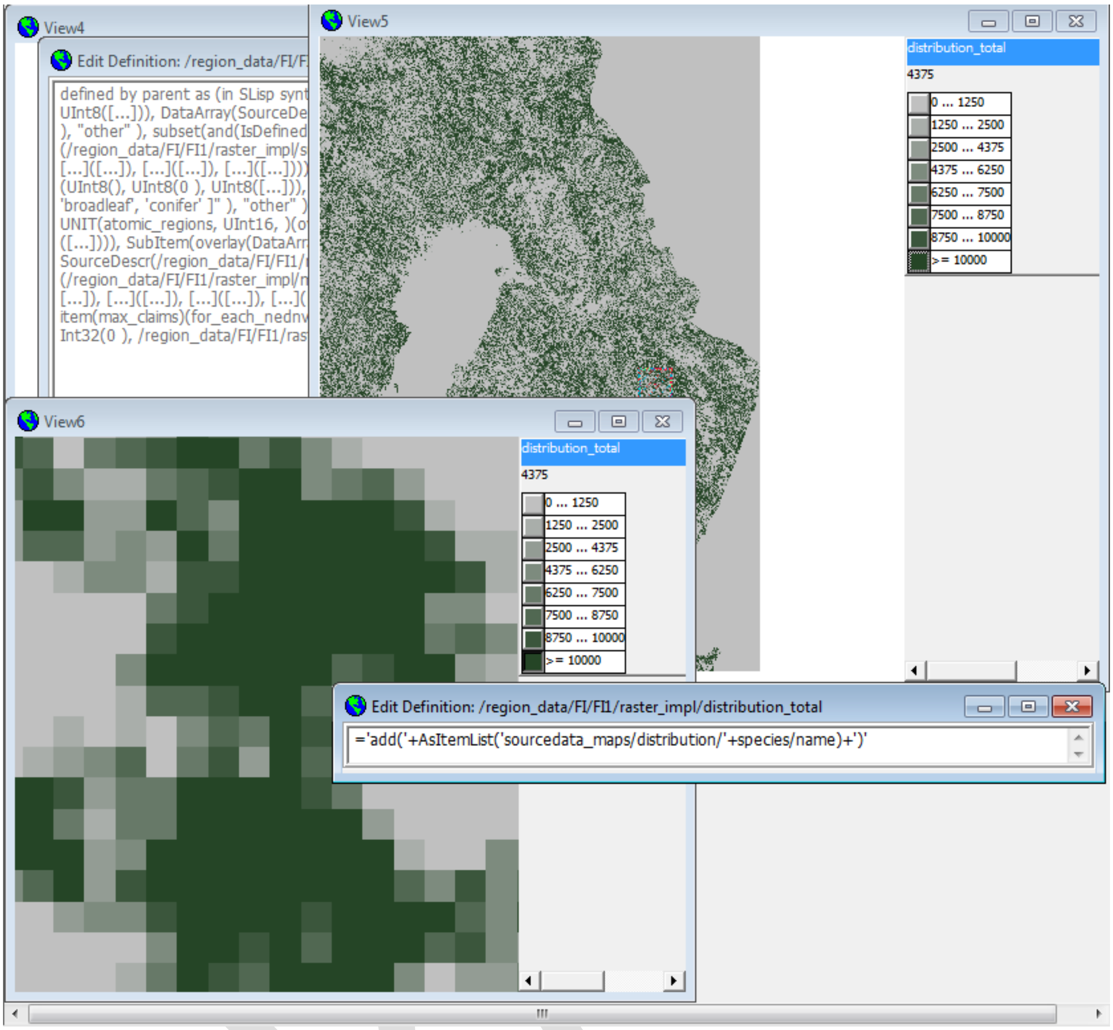


Figure 5. Cells can be fully forested (100%=10,000 m^2), partially forested (1250-9,999 m^2), or considered to not be forested (<1250 m^2).

The discrete allocation algorithm is used to assign areas of forest per species to the most suitable cells, whereby a single species class (broadleaf or conifer) is assigned to each cell, given the probability of presence maps (e.g. Figure 5).

3.1.1.3 Base map

A base map is required for the model, in order to give a guidance of where coniferous and broadleaved forests are roughly located. The base map (Figure 6) was generated using the forest-non-forest map, the probability of presence map (aggregated by species types) and Corine Land Cover (CLC) forest classes as the base map for the allocation. The forest/not-forest map is used to determine whether or not a cell is dominantly forested. If so, it is assigned to either the conifer or the broadleaf class according to the probability of presence map. There tended to be an underestimation of forest area with the 2006 forest map, causing an overflow of demand because there were too few allocatable cells. This was corrected by adding cells from the third data source, the CLC map, whose classes corresponded to forests. These were mosaicked in the background. If the CLC classes were of mixed forests, burnt areas or transitional woodland/shrubland, they were labelled as "mixed". Mixed forest cells are used as passive land use to absorb the overflow of either broadleaves or coniferous areas, as reported in SOEF 2015. A fourth class, labelled as "other" is used as a placeholder for an eventual expansion of the model to handle afforestation, deforestation, and the diversification of non-forested land to short-rotation forestry land.

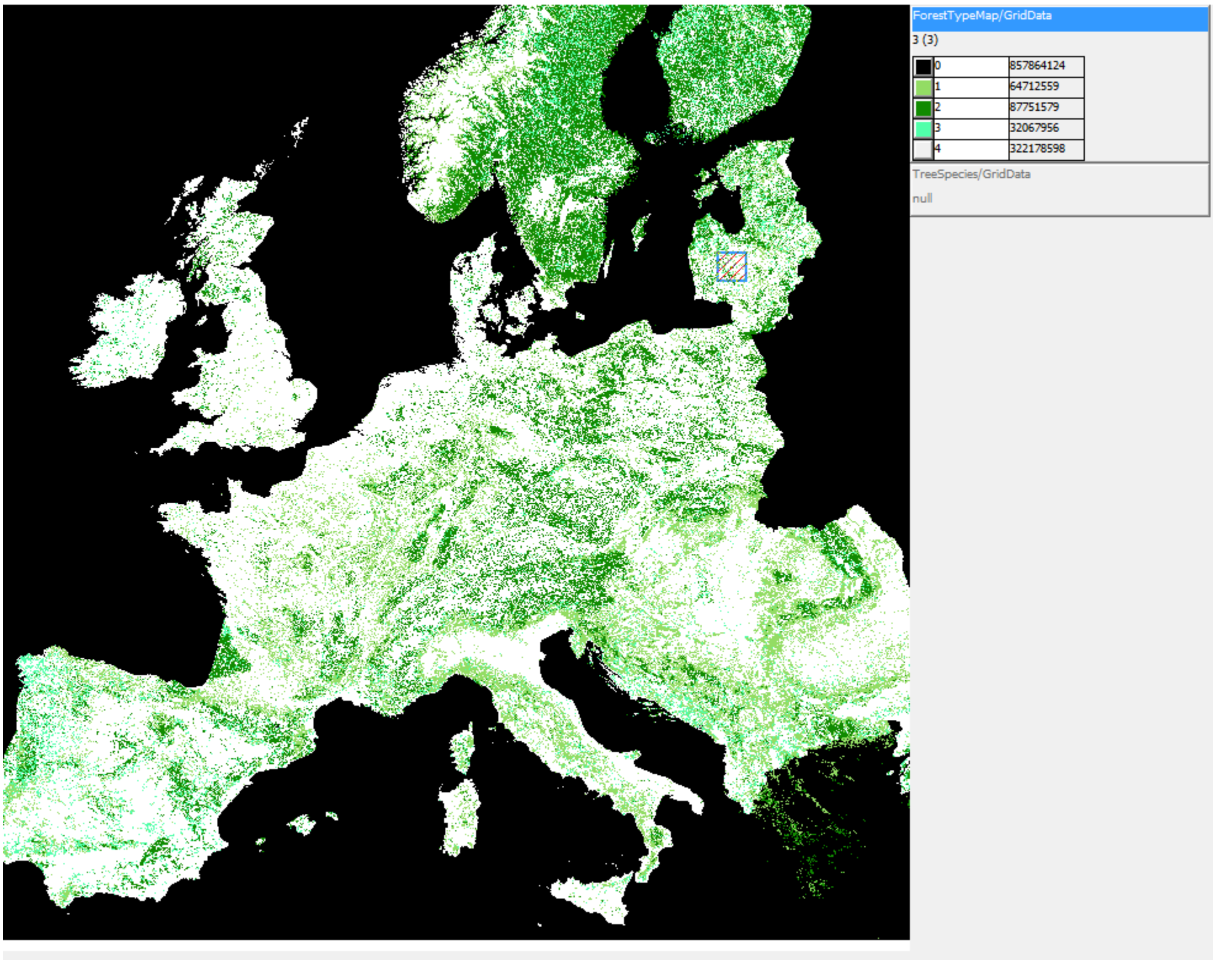


Figure 6. Base map for allocating broadleaves and conifers to 1-ha cells. Class 1=”broadleaves”; class 2 = “conifers”; class 3 = “mixed forest”; class 4 = ”other”.

3.1.2 Volume distribution

A volume, in m^3 per species per region, is given by the NFI data (vNFI). vNFI is disaggregated using the volume given by a remote sensing-derived product (vRS). Remotely-sensed volume estimates are corrected by NFI figures to give absolute values of volume of species per cell (eq. 9):

$$ {}^{r,s}V_i = \left(\frac{{}^{r,s}vRS_i}{\sum {}^{r,s}vRS_i}\right) * {}^{r,s}vNFI \qquad \text{eq .9} $$

3.1.3 Age distribution

The model requires the designation of a time-step. Regardless of the length of this interval, which could be considered analogous to a growing period, the individual trees in each class may either remain in the same class with a given probability; increase one size class with a given probability; increase more than one size class; be harvested; or be thinned. Estimating age per species is therefore a crucial step. The ForeStats database does not always contain information on the breakdown of age by species and volume classes; it may only report the proportion of forest area per age group. The challenge therefore lies in spatially allocating this forest area per age group based on some proxy.

Species and site-specific yield curves can be used as a proxy to determine the likely age given volume and increment data per species in ForeStats. The increment is the derivative measuring the sensitivity of volume changes (dependent variable), to changes in age (independent variable). In order to do this, a formula describing the shape of the curve relating age to volume is identified. Pilli et al (2013) use the Chapman-Richards function (eq. 10), describing the cumulative volume of the forest.

$$y = a(1 - e^{-bt})^c \qquad \text{eq. 10}$$

The authors then fitted data from yield tables to the function, deriving the parameters for the function: a,b,c per species per site class in different countries. Figure 7 shows an example for accumulating volume plotted against age and the differential form of eq. 10 (eq. 11), describing the increment plotted against age for Pinus sylvestris (Figure 8). The differential form of the equation contains both the expansion and decline:

$$y' = a * b * c * e^{-bt} * (1 - e^{-bt})^{c-1} \qquad \text{eq. 11}$$

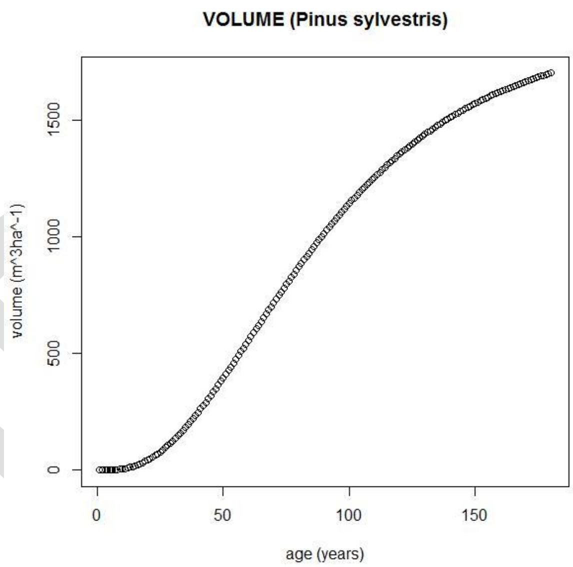


Figure 7. Volume curve for Pinus sylvestris using equation 10 and coefficients estimated by Pilli et al (2013).

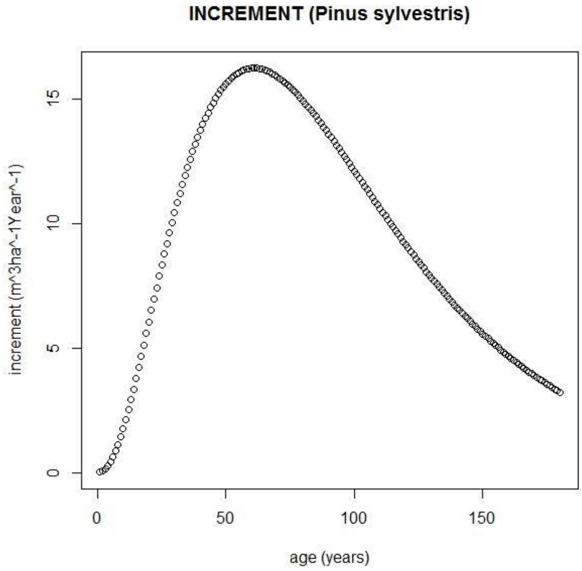
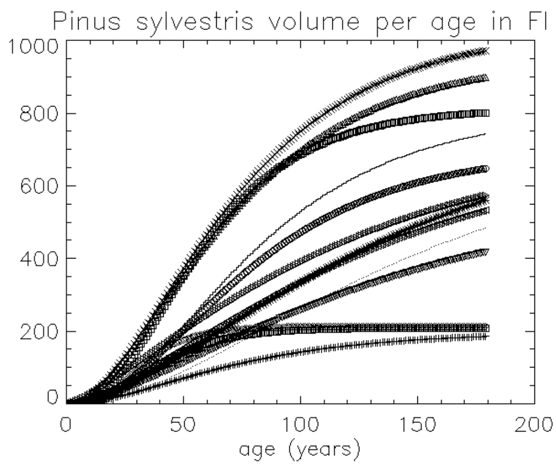
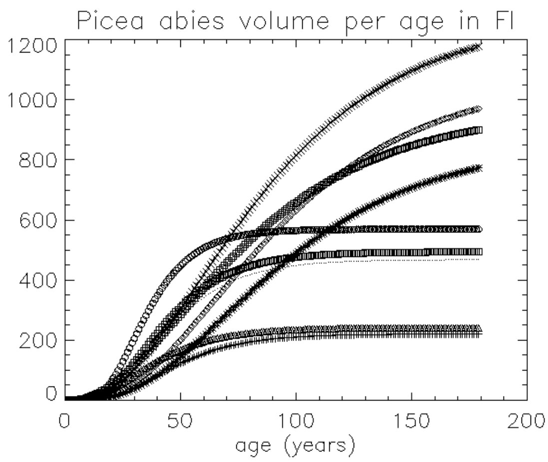


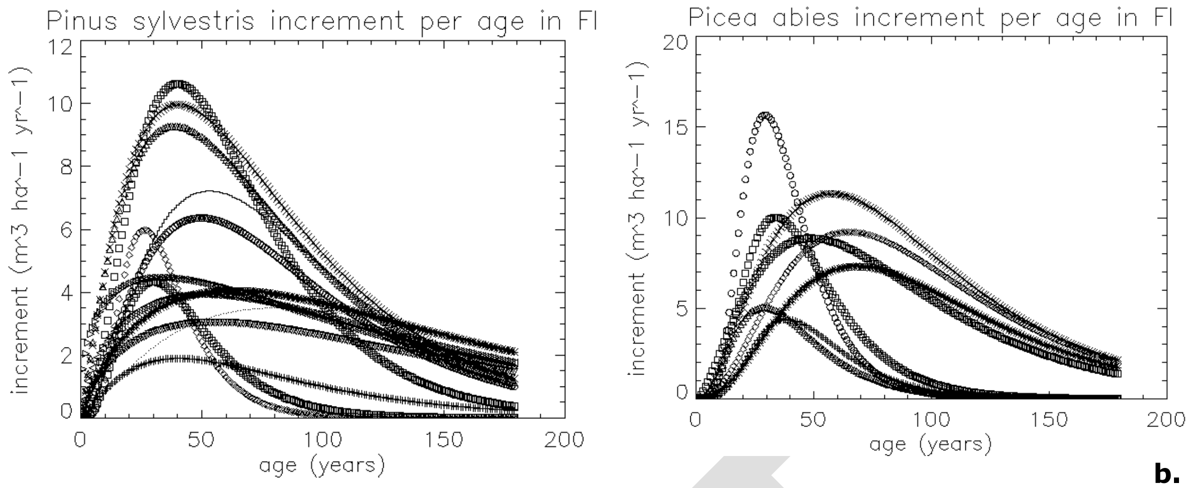
Figure 8. Increment curve for Pinus sylvestris using equation 11 and coefficients estimated by Pilli et al (2013).

Knowing the increment and volume, but not knowing the age (*G*), we run equations 10 and 11 for age classes between 0-180 years. There are usually several sites per dominant species within a single country. There are therefore several curves for a species type within a country (Figure 9, a & b).





a.



b.

Figure 9. Example of different curves for Finland expressing a) Cumulative volume over time for several sites for pine and spruce; b) Increment over time for several sites for pine and spruce.

As several sets of coefficients exist for each species, we compute several curves for both volume and increment. The sets of coefficients depend on the site quality, and we do not have a source for site quality for the mapped area. Since we are interested in estimating the age of given species, we use the increment and volumes given in the curves to match the ForeStats (NFI) data, where volume and increment is given by species group. The age cannot be determined from one parameter or the other. The age will therefore be estimated using both parameters: cumulative volume and increment per species class.

The differences between the calculated increment and volume can be used to estimate the average age of the region's forest by assessing the point on the graph where the percent difference between NFI data and the growth function (GF) data, for both volume and increment, are minimal. This can be done at either regional or cell level. The compound difference can be minimal at several points throughout the age series, or at a single point that could correspond to a low error in one parameter, and a high value in the other (Figure 10).

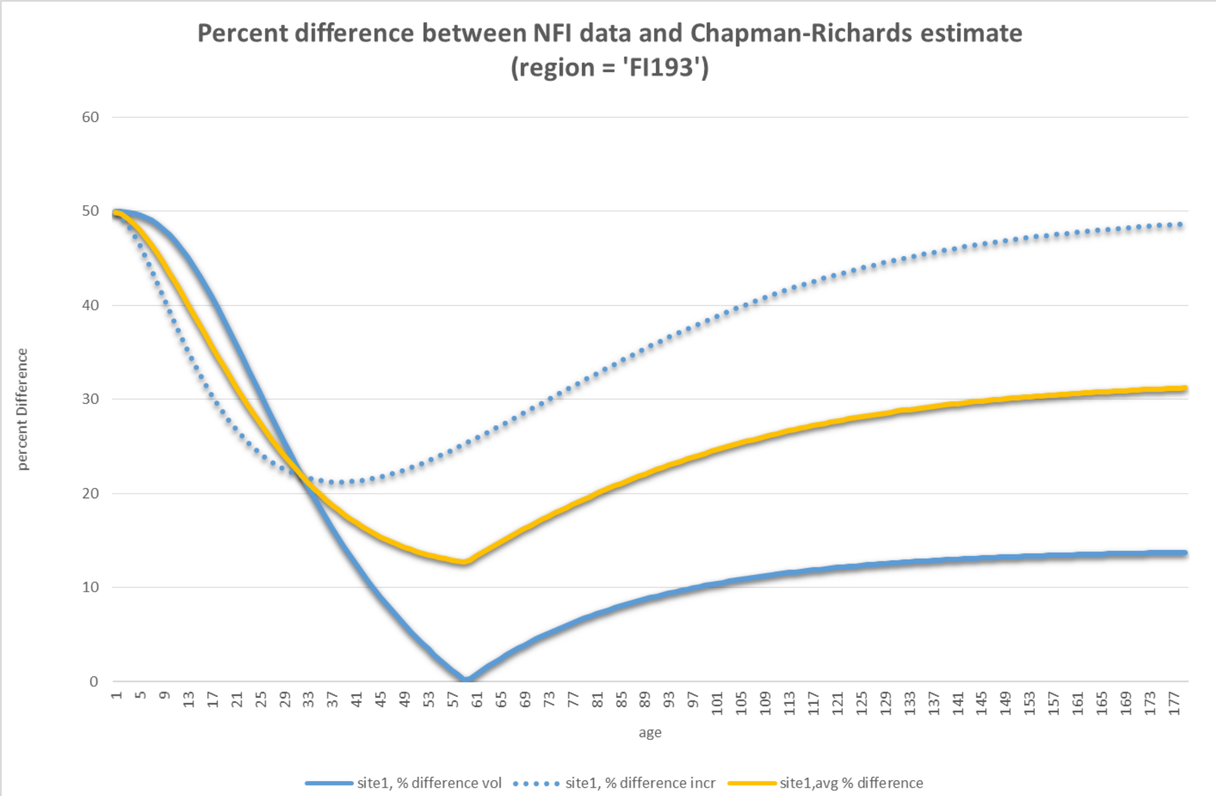


Figure 10. Percent difference between NFI measurement and Chapman-Richards curve over a 180-year span, showing that the average percent difference is not sufficient to use as a proxy to estimate age of species (example in NUTS3 region of Finland).

Figure 11 shows that, for an example of a single species group in a single region, there are points of convergence between percent differences in reported volume and increment and calculated volume and increment. This minimum distance occurs at different ages depending on the site used to estimate the GF equation. For site #1, for example, this convergence occurs at the age of 40; for site #4, this convergence occurs at the age of 25 years (x-axis, Figure 11).

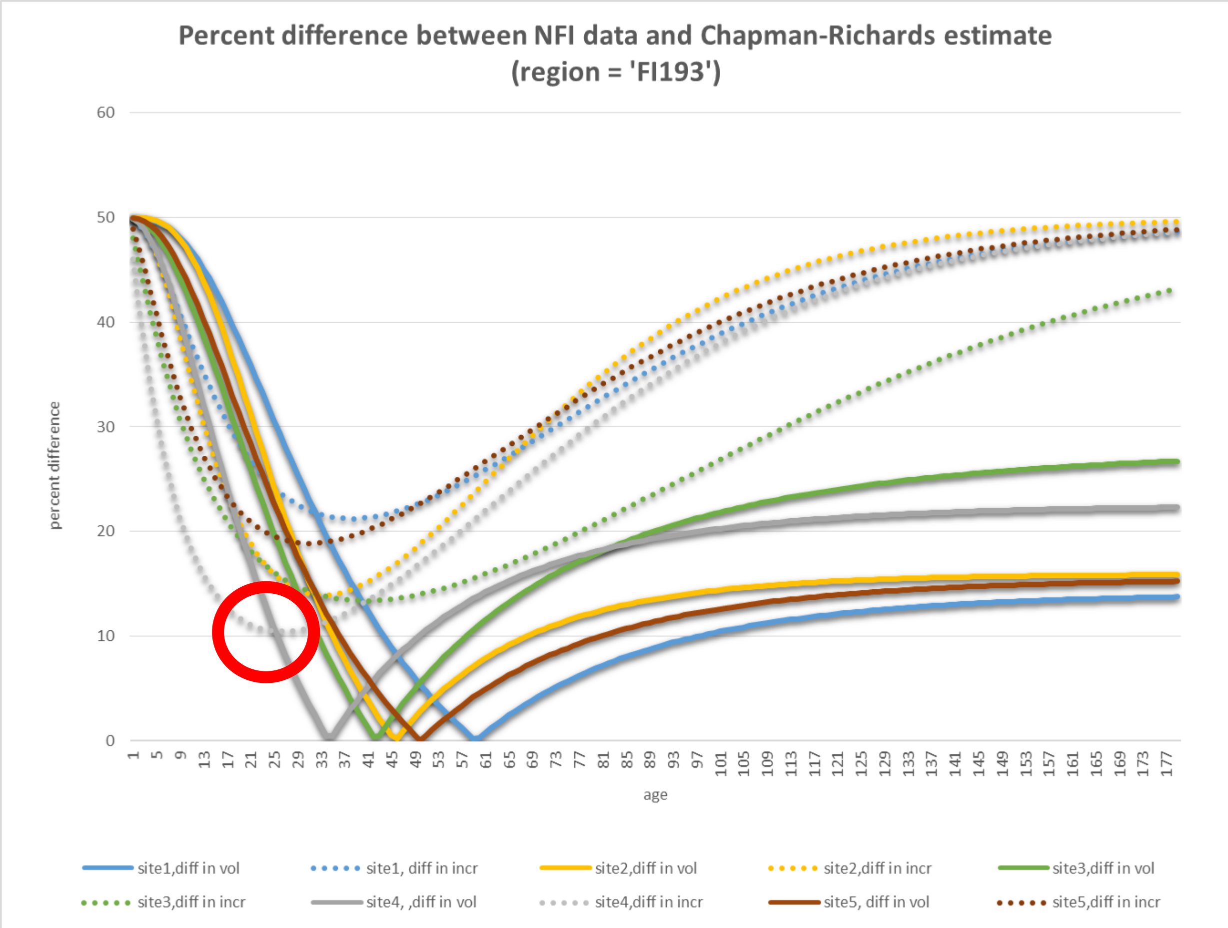


Figure 11. Differences between calculated increment and calculated volume, and NFI reported volume and increment for one region.

There are therefore as many locations on the x-axis where there is a "best fit" in terms of volume and increment, as there are sites from which the growth function coefficients were derived. A third test is therefore one of assessing the best site to choose to use as a proxy for estimating age. The best site is considered the one with the lowest overall difference between NFI and textbook volume and increment. As shown in Figure 12, this test must be applied to a restricted number of years because the overall average difference for any given site varies according to age.

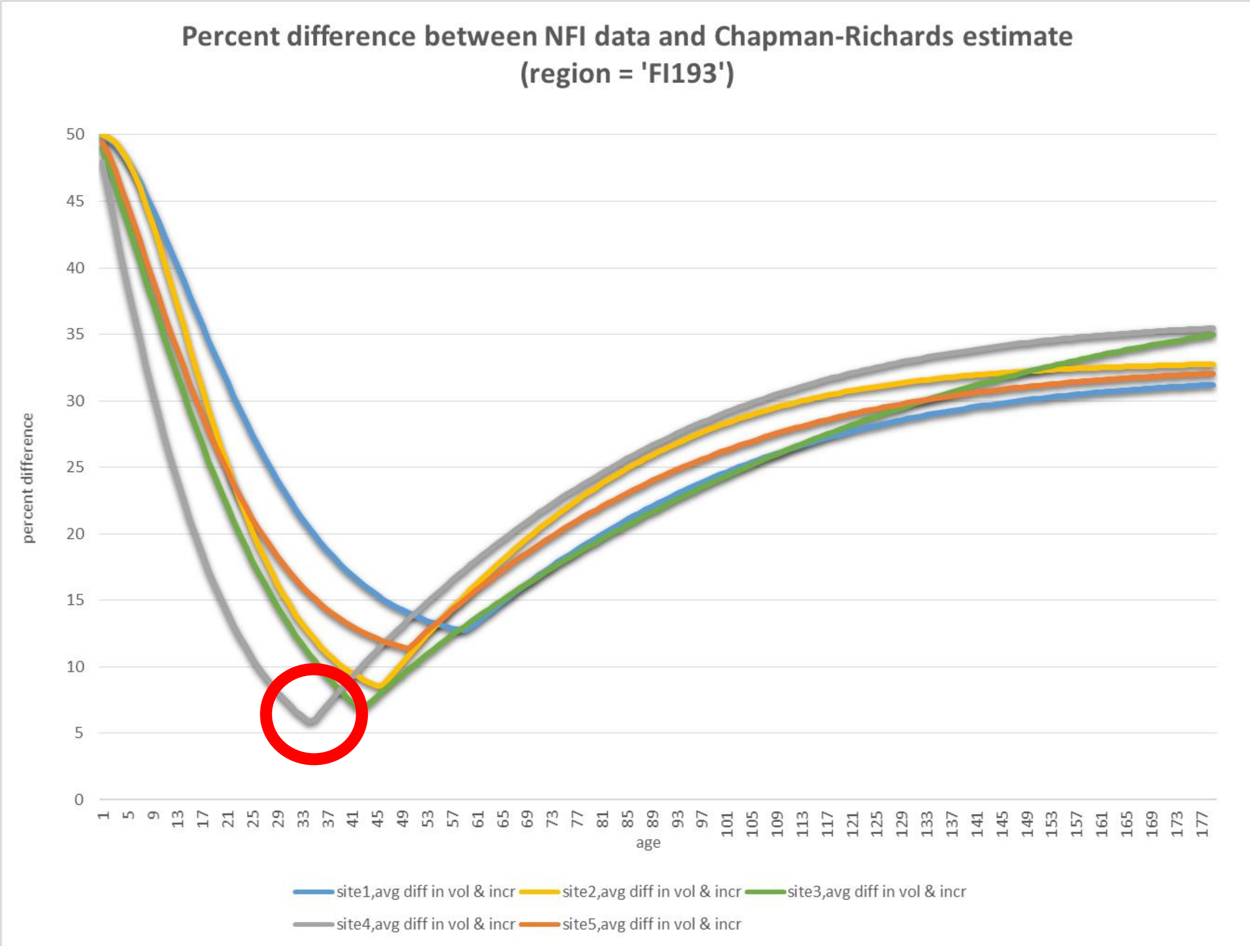


Figure 12. Average percent difference between NFI volume and increment measures and Chapman-Richards measures for one NUTS 3 region in Finland (FI193).

The process of narrowing down the best estimate for age based on species, volume and increment (in the absence of any knowledge of site quality), is therefore stepwise. In summary, the data used to estimate age per species group includes:

1. ForeStats NFI data on volume per species (${}^{r,s}vNFI$)
2. ForeStats NFI data on increment per species (${}^{r,s}iNFI$)
3. UBald growth functions, estimated from yield tables for as many sites as are available (e.g. equations 10 & 11), yielding arrays for the age range of interest ($\{{}^{r,s}vGF\}^{0-180}$, $\{{}^{r,s}incrGF\}^{0-180}$)

If the input is disaggregated to grid level, we may process each grid cell (i) instead of processing each region (r). For simplicity's sake, we assume calculations at grid level in denoting equations 12-18. The data array for each species group and cell is processed as an array ${}^{s}data_i$ in which columns contain variables derived from the growth functions; and differences between the measured NFI volume and increment for that cell or region and those growth functions outputs estimated for the 0-180 year time span (Table 5).

Table 5. Structure of input table needed for equations 12-18.

Age *0-180 years*	${}^{r,s}\,vGF$	${}^{r,s}\,incrGF$	${}^{v,s}\,\varepsilon_i$	${}^{incr,s}\,\varepsilon_i$	*Site ID*
0					
..					
180					

Ranges of elements of ${}^{s}\,data_i$, corresponding either to columns or the result of calculations of the data in columns, are represented with curly brackets. The percent difference between the NFI volume and volumes estimated using the growth functions (εv) are computed using equation 12. The same is done for increment ($\varepsilon incr$, eq. 13). The results will also be an array of data, over a span of 180 years, for volume and for increment. The smallest overall difference between εv and εi is calculated as their average ($avg\varepsilon$, eq. 14). To visualize these large amounts of data (181 years multiplied by number of sites) more easily, they are sorted in order of smallest to largest percent differences between both volume and increment (eq. 15), resulting in a sorted dataset per region and species group: ${}^{s}\,data_i$, whereby it is possible to index the site number and age because the rows (shown in Table 5) are shifted as a unit.

$$\{\varepsilon v\} = \left|\frac{\{\mathrm{vGF}\} - {}^{r,s}\,vNFI}{(\{vGF\} + {}^{r,s}\,vNFI\ )/2}\right| \qquad \text{eq. 12}$$

$$\{\varepsilon i\} = \left|\frac{\{incrGF\} - {}^{r,s}\,incrNFI}{(\{incrGF\} + {}^{r,s}\,incrNFI\ )/2}\right| \qquad \text{eq. 13}$$

$$\{avg\varepsilon\} = (\{\varepsilon incr\} + \{\varepsilon v\})\ /2 \qquad \text{eq. 14}$$

$${}^{s}\,data'_i\ = sortAscend\ (\{avg\varepsilon\}\ ) \qquad \text{eq. 15}$$

A minimum threshold for acceptable average differences ($avg\varepsilon$), is set to 25% to reduce the dataset (eq.16). This threshold serves two purposes: First, it removes unreasonably distant data points; second, it removes artefacts that are visible in the very young ages.

$$\{\ {}^{s}\,data'_i\ \} = \ {}^{s}\,data'_i\ \ : \{\mathrm{avg\varepsilon}\}\ > 0.25 \qquad \text{eq. 16}$$

The best age estimate for each region/ cell and species group (${}^{s}\,G_i$) is now considered to be the point in the array ${}^{s}\,data'_i$, which corresponds to the compounded minimum value of $\mathrm{avg\varepsilon}$, and the minimum difference between εv and εi (εvi, eq. 17; eq. 18) :

$$\varepsilon vincr = \varepsilon v + \varepsilon incr \qquad \text{eq. 17}$$

$${}^{s}\,G_i\ = index[\ \min(\mathrm{avg\varepsilon}^{\{\ {}^{s}\,data'_i\ \}}, \varepsilon vincr^{\{\ {}^{s}\,data'_i\ \}})] \qquad \text{eq. 18}$$

We order twice: first on $avg\varepsilon$, then on $\varepsilon vincr$. ${}^{s}G_i$ is an estimated average age per cell per species. This is a single value per species group per 1km x 1km cell.

We have so far generated a grid of the average age per square kilometer. We have not yet considered however, the proportion of land to be allocated to each age group, given by the NFIs (${}^{r}gNFI$), an example is shown in Figure 13. ${}^{r}gNFI$ is a range of forest area for a series of ranges of age and is not species-specific, nor is it spatially-explicit. The proportion of forest area for each age class can vary a great deal from region to region. In the example figure, the most area is attributed to the age groups 21-40 years and 41-60 years. There is also a significant amount of young forests.

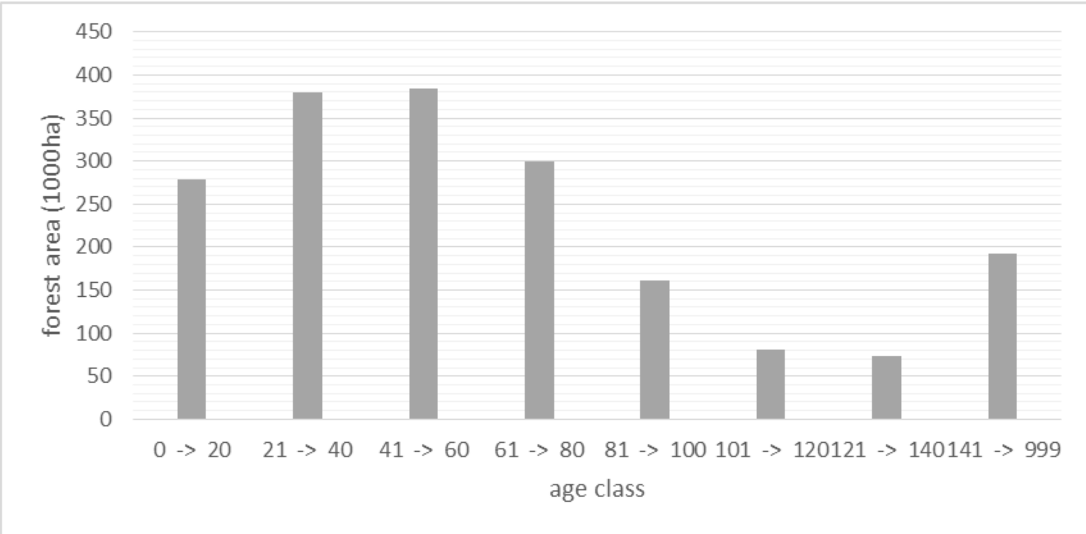


Figure 13. Example of forest area distribution per age class for a region in Finland

To allocate the area of forestland related to each age class, this range of ages per cell is translated to a probability set ${}^{s}pG_i$. To do this, we have to use the estimated age as a proxy for probability of being assigned into a particular age group, but not as an absolute value. Otherwise we risk not allocating the NFI data properly. Furthermore, we must consider only that fraction of the cell that is effectively covered in forest. We thus have a double constraint. The probability of an age being assigned from a set of ages can be assigned as shown in Figure 14.

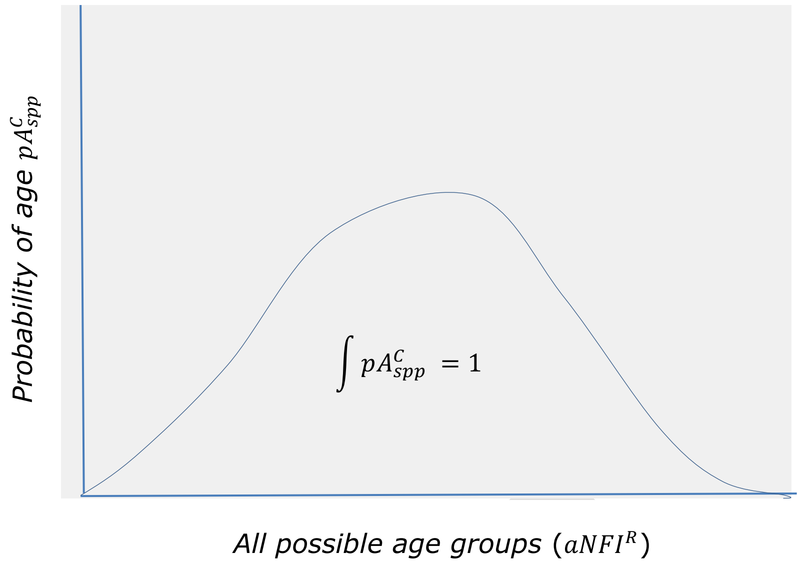


Figure 14. Range of probabilities of age being assigned to a cell

The peak in the Gaussian curve is $^{s}G_i$. The standard deviation is assigned to be one age class, which is variable from country to country in this application. Values gradually decrease around the peak, with gradually decreasing probabilities that the cell will be assigned this age. To obtain the probabilities under the curve, each possible age within the range is divided by the sum of the area under the curve. In this way, we have two advantages: First, we "soften" the otherwise rigid allocation of an age, allowing for a margin of distance from the mean to be allocated. Second, we can translate this transformation into a probability that can be inserted into a doubly constrained model. Thus, the total area of forest per age is spread out among the cells and species groups through an allocation procedure that distributes age based on $^{s}pG_i$, while respecting two constraints: The total amount of forested area per cell f_i , and the total area to allocate per age class per region ($^{r}aNFI$). In the first step of this allocation process, the area available per cell to “receive” a likely age range is equal to the area covered by forest for each species group ($^{s}f_i$). This area per cell is given for $^{s}X_i$ and the sum of all the area covered by all species group equals f^{C}:

$$^{s}f_i = \sum {}^{s}X_i \qquad \text{eq. 19}$$

When f^{C}_{spp} is multiplied by the probability it will receive an age class (pA^{C}_{spp}), we obtain the surface area for the age class per species group:

$$^{s,g}X_i = {}^{s}pG_i * {}^{s}f_i \qquad \text{eq. 20}$$

Thus, the sum of the surface area per age per species group in a cell is contained within the sum of all forested area for the cell: $^{s}X_i \in f_i$. This value, however, may not be in agreement with the total area per age class as given by ForeStats $^{r}gNFI$.. This means that although $^{s,g}X_i$ is the "best" answer in terms of age per cell, it may not sum to the correct proportions of the age groups. We must therefore aim for a value whose sum is close to ${}^{r}gNFI$. We denote this new variable as ${}^{g}M_i$ (eq. 21):

$$ {}^{g}M_i = {}^{g}\alpha * \beta_i * {}^{s}pA_i \qquad \text{eq.21} $$

In equation 21, we have introduced two coefficients, α and β. These correspond to factors that are iteratively adjusted until the constraints are satisfied. ${}^{g}\alpha$ is initially assigned a value of 1.

$$ \beta_i = \frac{f_i}{\sum_{\mathrm{A}} {}^{g}\alpha * {}^{s}pG_i} \qquad \text{eq.22} $$

$$ {}^{g}\alpha = \frac{{}^{r}aNFI}{\sum_{\mathrm{i}} \beta_i * {}^{s}pA_i} \qquad \text{eq.23} $$

We can establish an acceptable threshold for convergence between ${}^{g}M_i$ and ${}^{r}aNFI$. Once this accepted threshold is met, we can solve equation 21 using the best values for ${}^{g}\alpha$ and β_i. At the time of writing, this allocation procedure is done using GeoDMS outside of EFDM-geo, and it is foreseen to eventually be fully embedded.

Figure 11 shows that several site types are yield different growth curves. Since the site number (analogous to the curve identifier) is recorded in Table 5, the same equation used to estimate age at the point in the dataset that most closely corresponds to the NFI data values for volume and increment, can be used to identify the best site number (eq. 24):

$$ {}^{r,s}site = index[\min(\mathrm{avg}\varepsilon^{\{{}^{r,s}data'\}}, \varepsilon vi^{\{{}^{r,s}data'\}})] \qquad \text{eq. 24} $$

3.1.4 Resulting initial state

The resulting spatially-distributed species, age and volume maps are used to generate the initial state matrix. This can be visualized graphically as shown in Figure 15.

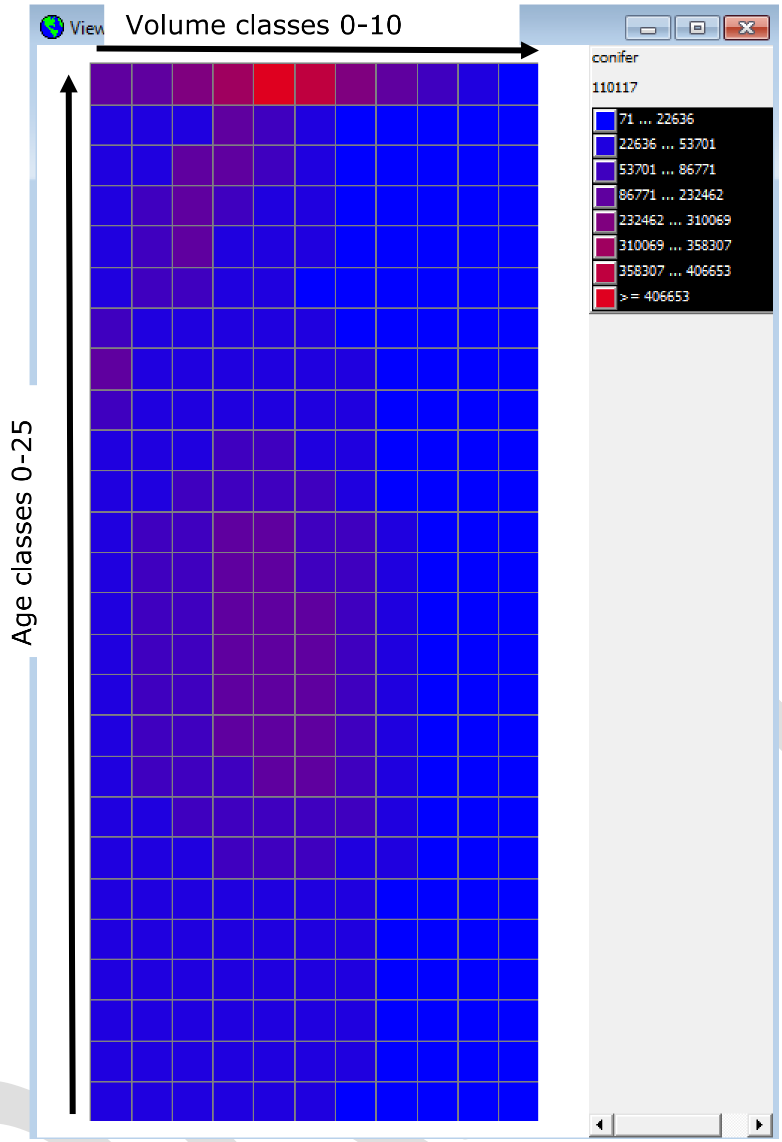


Figure 15. Graphical view of area distribution over age and volume classes (blue -> red = lowest -> highest areas of forest in the vol/age class combination).

In this example for a given region and for conifers only, we have 11 volume classes along the x-axis, and 26 age classes along the y-axis. Different regions of Europe will have different age and volume classes. From this graph, we can see immediately that there is a large amount of forest area in middle-range volume and age classes and in the highest age and mid-volume classes for conifers.

3.2 Transition Matrices

The transitions that are modelled with EFDM-geo are those between age-volume classes: as the forest ages, it gains volume. The rate at which the volume increments according to age differs between the factors such as species and site qualities. Often this will be non-linear. Volume increments more quickly at young ages. Figure 16 illustrates this concept if one matrix exists for one species group, as is the case in the EFDM-geo configuration.

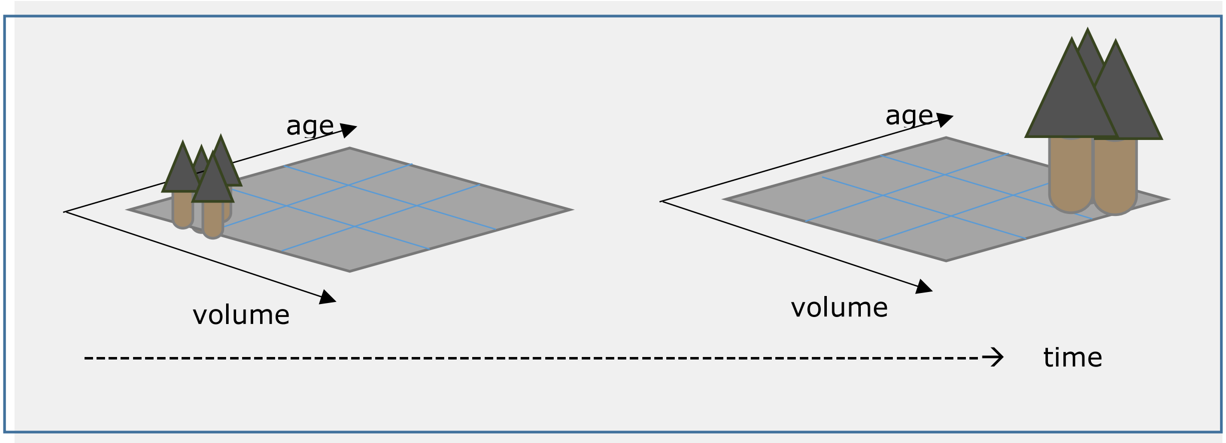


Figure 16. The progress of a group of trees in a "no management" scenario whereby volume class may increase with age class.

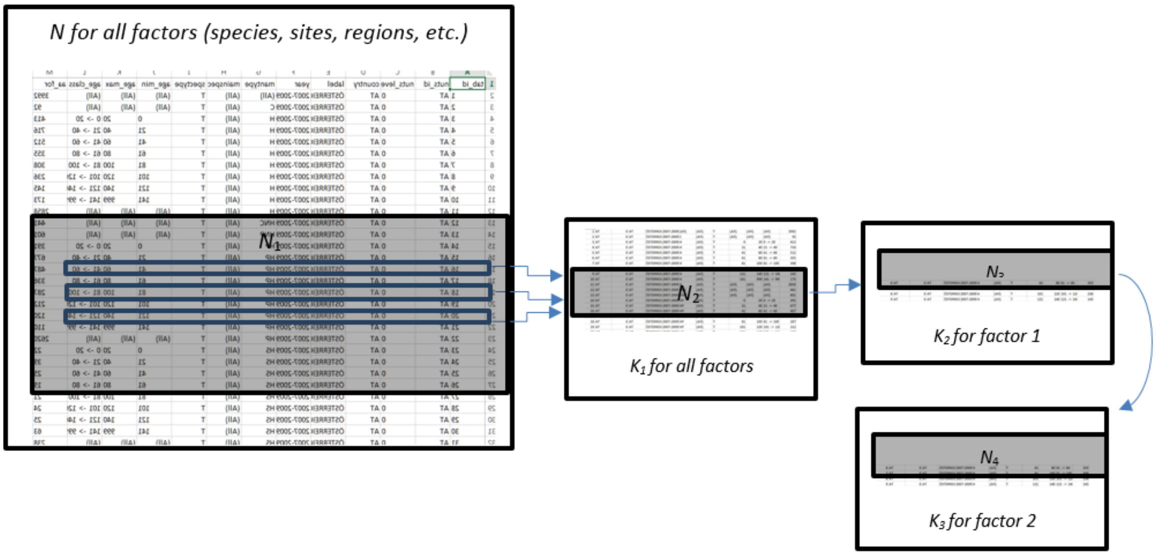
The above schema is an example of natural succession in the absence of management activities. In EFDM-geo, the user determines the number of transition matrices, as well as how they are generated, as the result of how many dimensions should be modelled.

The transition matrices are generated using the R-code of EFDM, as shown in Figure 4. They are generated using data from two consecutive NFI campaigns. In an ideal situation, sufficient plot data are available to estimate the transition probabilities based on the proportion of times a transition occurs out of the total number of times it is tried and observed. The best predictor that an event will happen is the proportion of times that it happens out of a total number of trials. This boils down to how many successes are witnessed vs. how many failures; or the fact that the outcome is either true or false. However in this situation, it is almost impossible to have a sufficient number of observations to rely on this approach. The next section describes a work-around for these situations.

3.2.1 Iterative Bayesian Estimation

In the document written by S. Sirkia ("Some mathematics behind EFDM", October 2012[11]), she describes that in the absence of sufficient plot data, a recursive Bayesian filter can be applied. This approach involves using aggregated data to sketch a rough prior, then iteratively narrowing down the hypothesis among all possible states to the most likely outcome state given the more refined data available. In practical terms, EFDM will count the frequency of transitions between the volume class from a point in time in the past, and the volume class for a later point in time from the field data. The count will correspond to the fraction k1 observations out of a total of N1 observations whose start step was volume class j for all pooled data (Figure 17).

11 https://webgate.ec.europa.eu/CITnet/stash/projects/FISE/repos/efdm/browse/documents/EFDMinstructions/Seija_Mathematics_behind_EFDM.pdf



$$p(H|O) = \frac{\alpha+t}{\beta+f+\alpha+t}$$

$$p(H|O) = \frac{\alpha+k1}{\beta+(N1-k1)+\alpha+k1}$$

$$p(H|O) = \frac{\alpha+k1+k2}{\beta+(N1-k1)+(N2-k2)+\alpha+k1+k2}$$

$$p(H|O) = \frac{\alpha+k1+k2+k3}{\beta+(N1-k1)+(N2-k2)+(N3-k3)+\alpha+k1+k2+k3}$$

Figure 17. Illustration of how NFI data is used to describe transition probabilities.

The subscripts indicate the iteration level. In the first order estimate, k1 corresponds to the number of cases where, for all factors pooled together, the individuals moved up to volume class v+1. Within this data pool, there are different factors to contend with: The species, site or region classes may have an influence on the increment. EFDM then splits the k1 data into factors. In the illustration above, a subset of k1 is extracted based on the factor level, corresponding to N2 samples. The number of samples that correspond to the size class pair increment under examination (from v to v+1) with the additional criteria of a specific factor level, corresponds to k3. Since N3 ∊ N2 ∊ N1 and k3 ∊ k2 ∊ k1, the frequencies of "truths" are actually repeating; they are summed. Imagine a tree of species *s* and region *r* that increments from v to v+1. The number of truths are summed 4 times, one for each time it was included in the data pool N. If there is an unlikely situation within the state space, such as an increment from class 1 to class 4 directly, N1 will probably equal to 0. The very low prior, in this case, will remain intact. EFDM will therefore be more likely to shift size class 1 areas to another more probable class, such as class 2 (Figure 18).

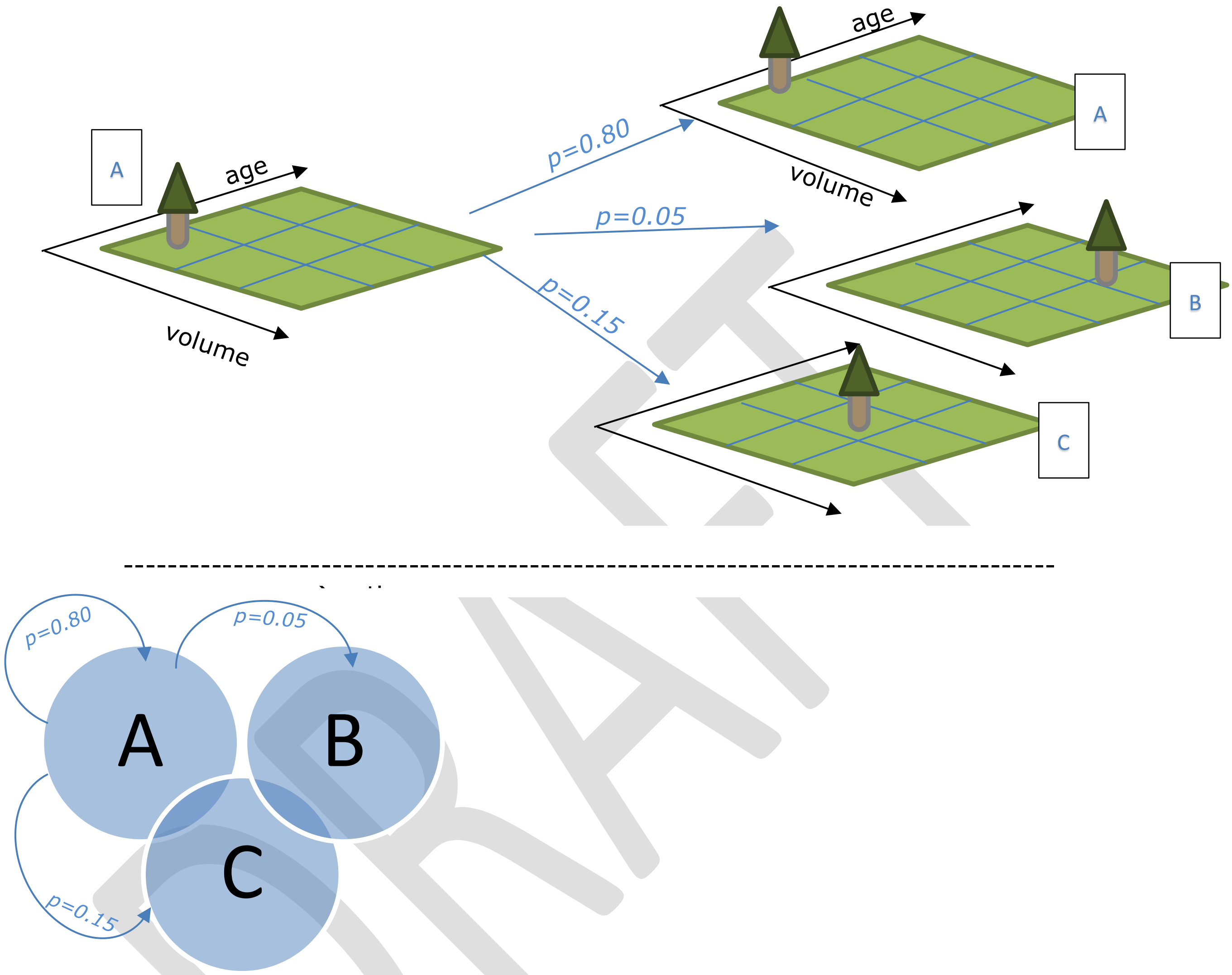


Figure 18. Illustration of effect of transition probabilities.

For the purposes of developing the EFDM-geo prototype, we only use one additional point of information to refine the prior: the species. For real modelling exercises, it would be possible to add other factors, such as site quality, ownership or topography. These are intended to be region-specific, meaning that different factors can be attributed to different regions within a same country.

3.2.2 Application of transition matrices

The transitions correspond to a second-order Markov assumption whereby current transition rates are estimated based on the state of an individual and on its previous transition[12]. The transitions are related to the probability of remaining in same volume

[12] This differs from most matrix models in which the first-order Markov assumption is accepted. The first order assumption relies only on the state of an individual at time *t*, and

class within the age class; the probability of increase by one volume class within the same age class; the probability of increase by more than one volume class within the same age class. The forest cells are, in a sense, maintaining a "memory" of their previous state because of the state they find themselves at present. The probability of transiting into another state in the age-volume dimension is based on their actual state, which in turn, was due to their preceding state. So, contrary to a first-order Markov assumption whereby the probabilities are applied to the state space irrespective of what happened in previous time steps, the probabilities will be applied to a specific state which has succumbed to an activity in the previous state.

EFDM follows the assumption of stationarity. With such an assumption, transition matrices are invariable over time, which means the probabilities neither increase nor decrease as a function of time.

The transition probabilities must be included in EFDM-geo ready-made, after having generated them using the "hackfunctions[13]" in EFDM. This same set of functions is also used to generate the rough prior. Instructions on usage published in the EFDM repository[14].

3.3 Activity matrices

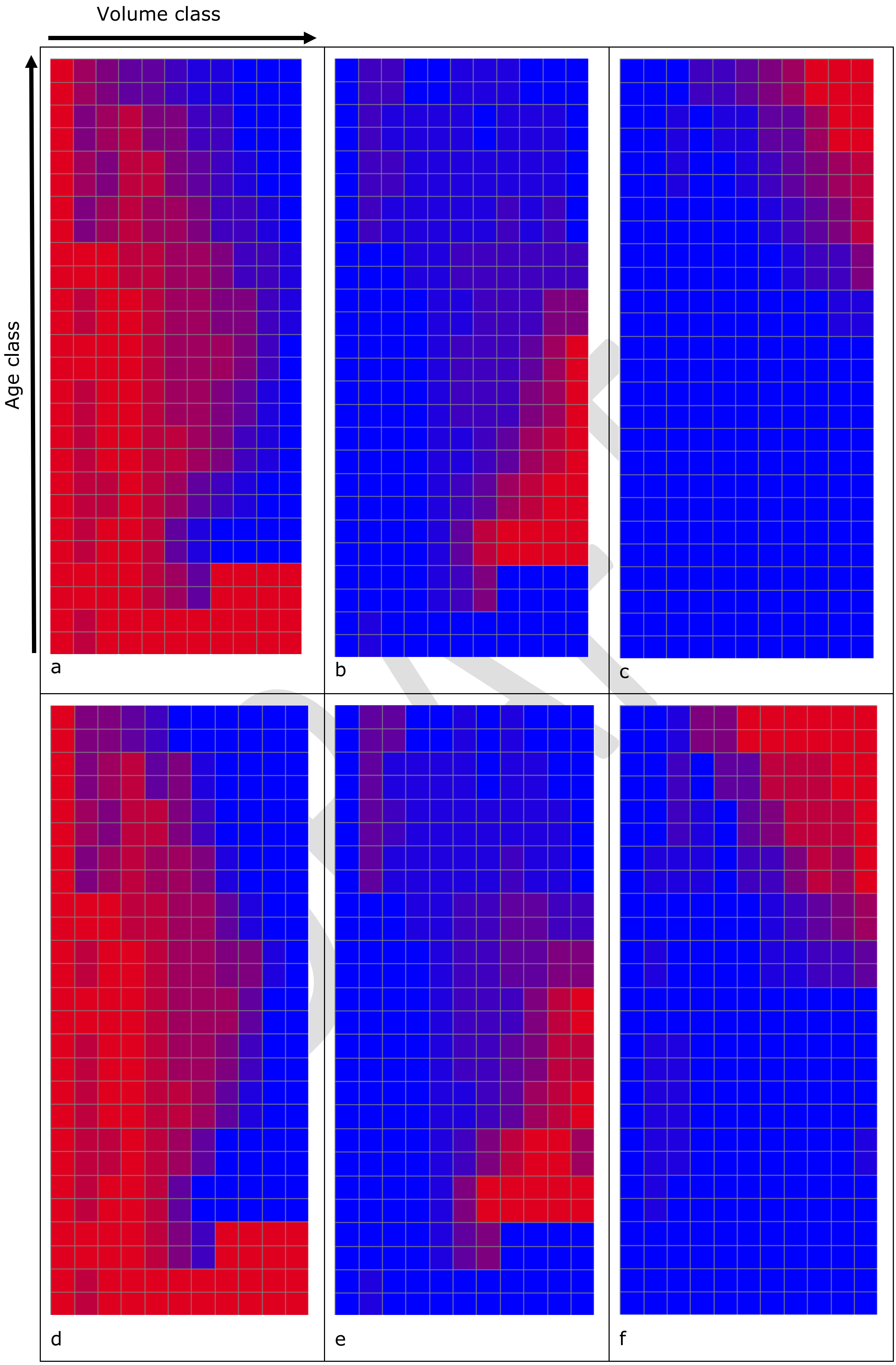
In EFDM-geo, the following activities are modelled: thinning, final felling and no management. These are based on an associated probability of the activity occurring should be quantified. For example, based on statistics, the user should estimate the likelihood that any class of forest be managed in any particular way. The age, volume and species type will play a role in the likelihood of thinning or harvesting. Generally, no management prevails in very young forests, thinning prevails in medium-aged forests and final felling prevails in older forests (Figure 19). Ingrowth, is also, in a sense, included as a modelled activity but can only occur in volume and age class 0. Depending on the transition matrix, the model could assume that only a part of this area in bare land is actually recruited and moves up one volume class in the next time step; or it may be so well managed that it even moves up two volume classes in one age step. The fraction of the area that is not afforested, remains in the same volume state '0', although it ages.

---

is independent of the previous states – in this way it has no "memory" of previous treatments.

13 https://webgate.ec.europa.eu/CITnet/stash/projects/FISE/repos/efdm/browse/EFDMcode/v2.0/hackfunctions.r

14 https://webgate.ec.europa.eu/CITnet/stash/projects/FISE/repos/efdm/browse/documents/EFDMinstructions/instHACK.pdf


Volume class
Age class
a
b
c
d
e
f

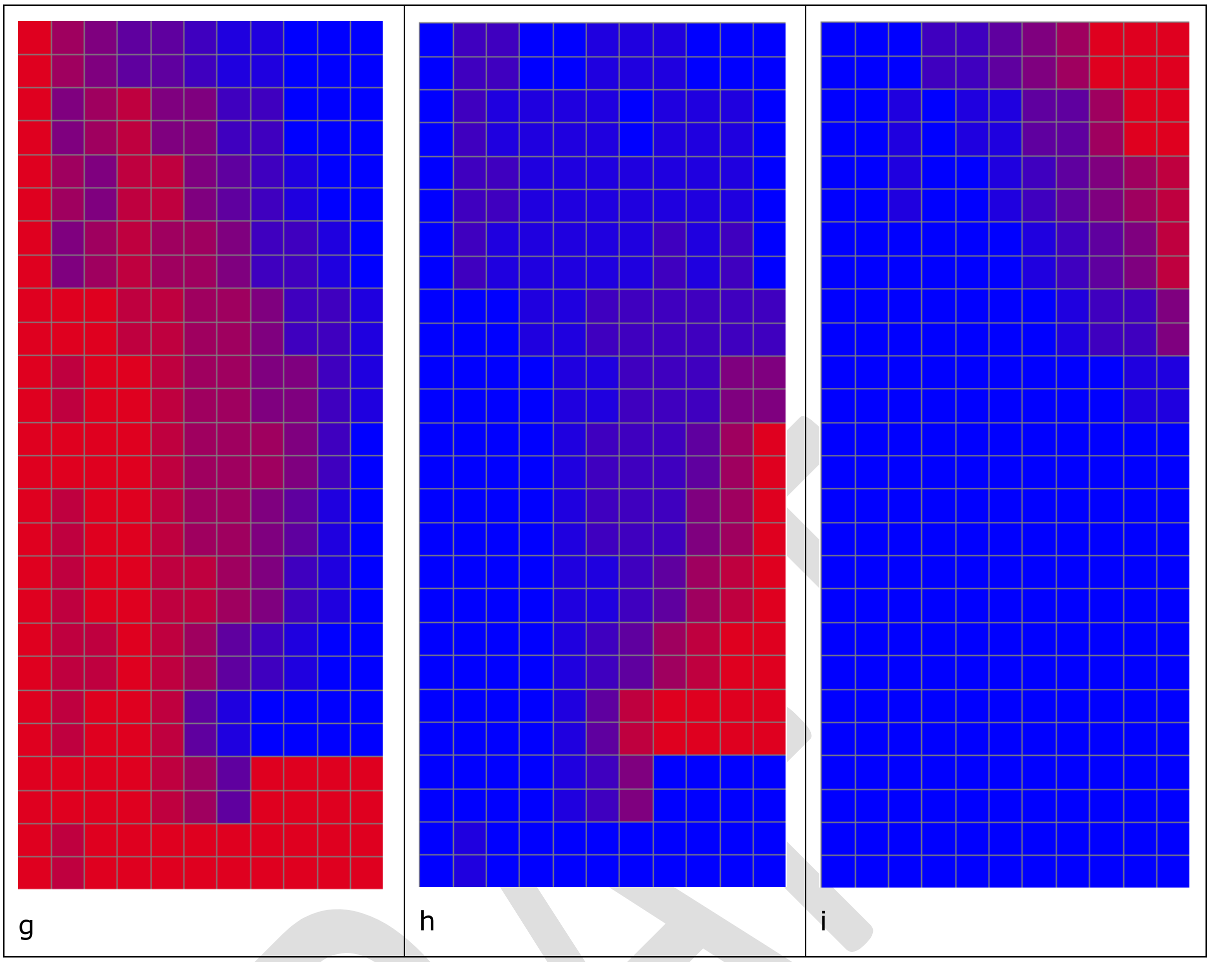


Figure 19. Management activities associated to age and volume classes

a. No management, 2015 (probability)
b. Thinning, 2015 (probability)
c. Final fell, 2015 (probability)
d. No management, 2030 (allocated)
e. Thinning, 2030 (allocated)
f. Final fell, 2030 (allocated)
g. No management, 2100 (allocated)
h. Thinning, 2100 (allocated)
i. Final fell, 2100 (allocated)

The activity probability matrices (a-c) are ideally provided by national experts. In the absence of expert opinion, the activities matrices are generated automatically in an R code following a textbook solution, an example is shown in Figure 19, a-c. The proportion of each activity is taken from a distribution over the forest types. The activities are then rescaled according to demand, as described in detail in section 2 and shown in Figure 19, d-i. Activities probabilities are used in the model in conjunction with the cost to perform the activity, described in section 2.2.2.

4 Output

The model output is numerous, generous and flexible. Processing time depends on the size of the country. The model must read in all the rasters for the country, as well as the associated datasets. For a large country such as Finland, the model will spend approximately 3 minutes to process 85 years (2015-2100) on a Windows machine with 2 Intel Xeon 2.40 GHz processers (only one is used in a run for one country), 32 Gb RAM on a 64-bit Windows 7 operating system. The fixed input maps are on a solid-state drive, which accelerates the runs. Many maps can be generated to estimate the spatial allocation of different attributes of forests. These maps, overlain with other thematic information such as bio-geo-climate variables, can result in interesting indicators. Beyond the spatially-explicit output is the fundamental information on the volume of wood removed from forests over the time span of the simulation, and the volume of wood remaining in the forests after meeting demand. These are discussed in the following sections.

4.1 State of the forest

We have seen that as the simulation runs in EFDM-geo, forest area moves through the volume/age matrix. This is visible as a spatially explicit output, whereby each 1-ha cell carries a forest type, and it is also possible to visualize the distribution of forest area within the matrix. The next series of figures shows an example of output for the ITF NUTS 1 region of Italy (Figure 20).

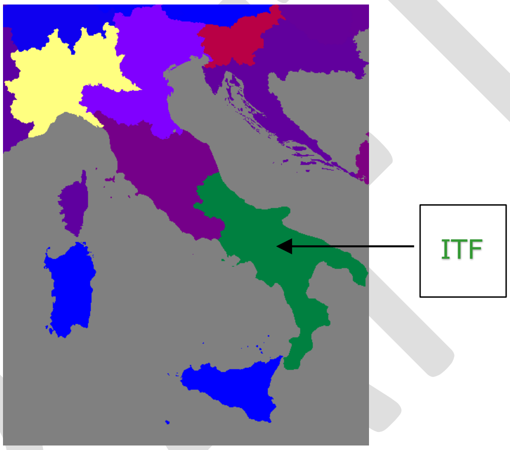


Figure 20. Location of ITF in Italy.

The resulting maps show the spatial distribution of the forests in different age/volume classes for broadleaves and conifers; and the resulting matrices show the sum of the forest area in each volume/age class (Figure 21 & Figure 22). These are meant as examples to show the format of the output; the data are not real.

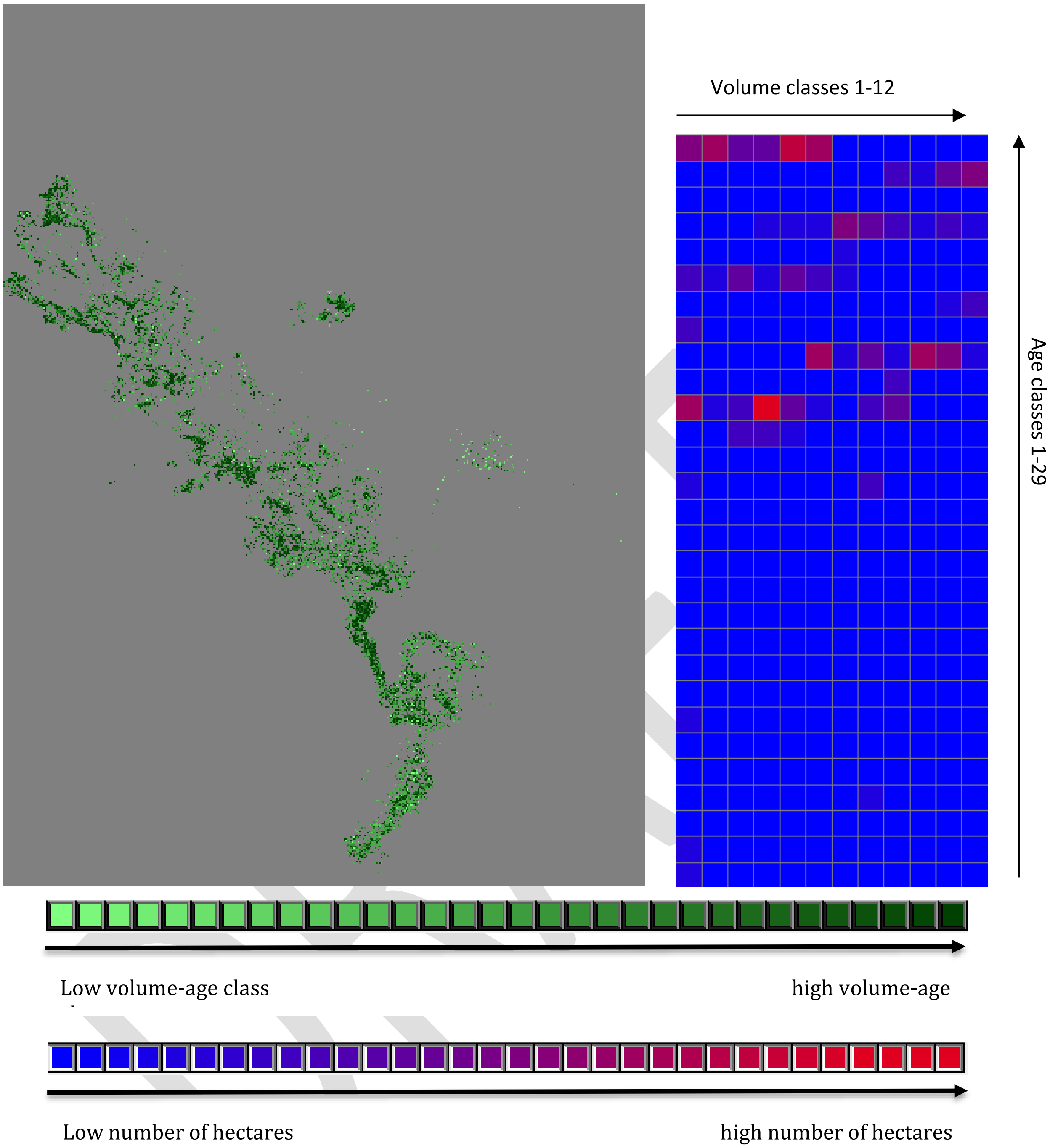


Figure 21. Example of resulting map and corresponding forest area distribution on the volume/age matrix for broadleaves in ITF in 2045.

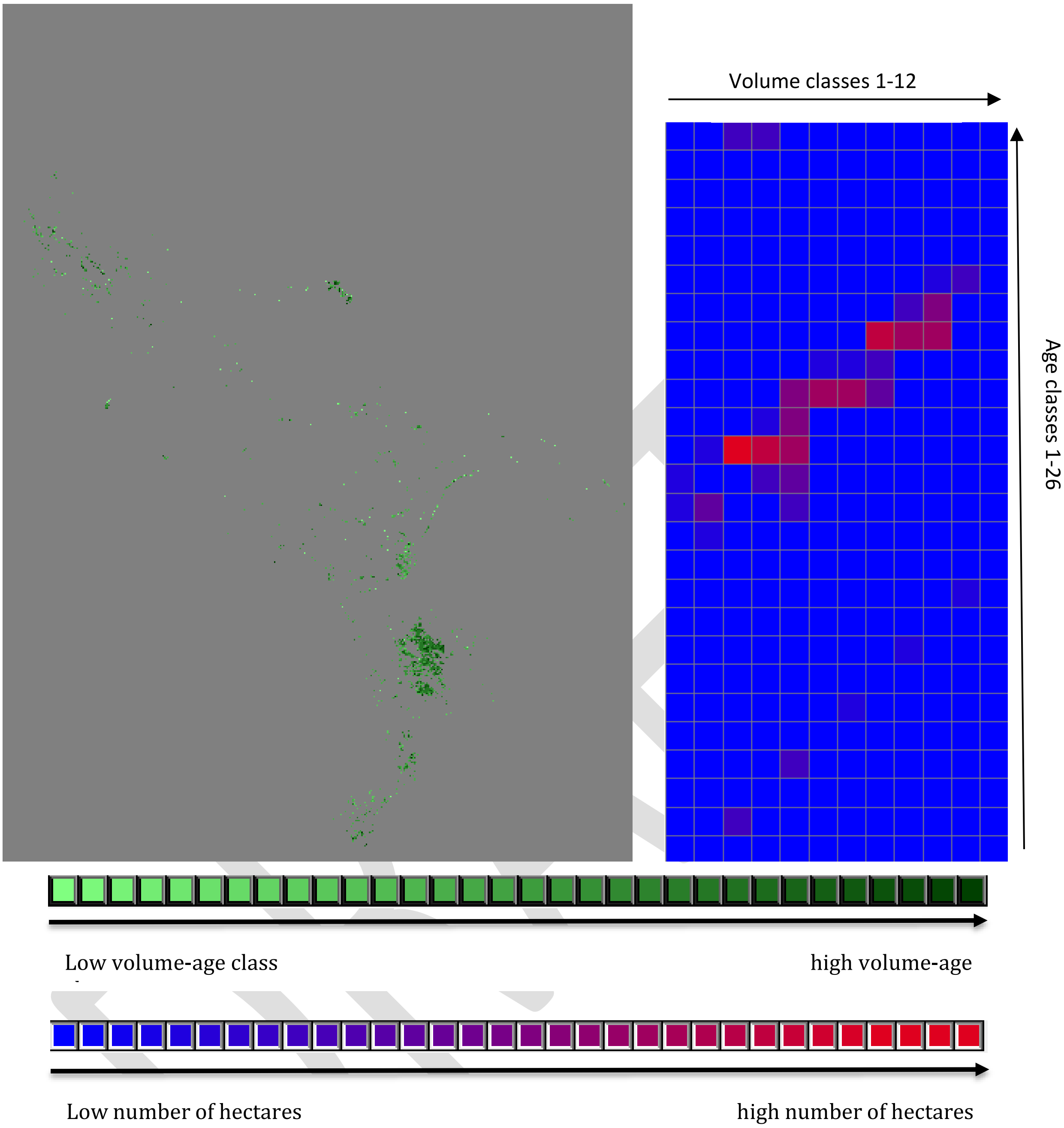


Figure 22. Example of resulting map and corresponding forest area distribution on the volume/age matrix for conifers in ITF in 2045.

4.2 Activities & Drain

The activities probabilities, shown in Figure 23 describes the likelihood of an activity occurring at time=0, so before any rescaling to meet demand occurs. If we look at the first year of the simulation, a rescale factor of roughly 0.26 is given as the outcome of the model Figure 25, with an outcome shown in Figure 24. The sum of the columns in Figure 24 is used to compute the total production for each age/volume class. We can compute the total production of broadleaves and conifers by multiplying the harvest from each activity per age/volume class (e.g. Figure 26) by the activity that the model expects will occur in each class. This figure will always sum to the demand (see section 2.1.1). The volume extracted from thinning, for example, is calculated based on the mean value of the volume class from which the volume is taken, minus the mean value of the volume class to which the forest area falls after thinning. We assume a drop of one volume class in all regions and species groups after thinning in this prototype. Figure 26 shows a sample of this output, whereby the result of the subtraction is calculated in the final column "yield" (referring to "harvest"). Similarly, the volume extracted from final felling of conifers is calculated based on the mean value of the volume class from which the volume is taken. We assume a drop to the first volume class and first age class after final felling in all regions and species groups in this prototype. This is analogous to a clear cut of the interested area.

id	regions	spcs	vol	age	finalfell	thin	nomgmt
0 (FI1.conifer	FI1	conifers	1	1	0	0	1
1 (FI1.conifer	FI1	conifers	2	1	0	0.111790079	0.888209921
2 (FI1.conifer	FI1	conifers	3	1	0	0	1
3 (FI1.conifer	FI1	conifers	4	1	0	0	1
4 (FI1.conifer	FI1	conifers	5	1	0	0	1
5 (FI1.conifer	FI1	conifers	6	1	0	0	1
6 (FI1.conifer	FI1	conifers	7	1	0	0	1
7 (FI1.conifer	FI1	conifers	8	1	0	0	1
8 (FI1.conifer	FI1	conifers	9	1	0	0	1
9 (FI1.conifer	FI1	conifers	10	1	0	0	1
10 (FI1.conife	FI1	conifers	11	1	0	0	1
11 (FI1.conife	FI1	conifers	1	2	0	0	1
12 (FI1.conife	FI1	conifers	2	2	0	0.111790079	0.888209921
13 (FI1.conife	FI1	conifers	3	2	0	0	1
14 (FI1.conife	FI1	conifers	4	2	0	0	1
15 (FI1.conife	FI1	conifers	5	2	0	0	1
16 (FI1.conife	FI1	conifers	6	2	0	0	1
17 (FI1.conife	FI1	conifers	7	2	0	0	1
18 (FI1.conife	FI1	conifers	8	2	0	0	1
19 (FI1.conife	FI1	conifers	9	2	0	0	1
20 (FI1.conife	FI1	conifers	10	2	0	0	1
21 (FI1.conife	FI1	conifers	11	2	0	0	1
22 (FI1.conife	FI1	conifers	1	3	0	0	1
23 (FI1.conife	FI1	conifers	2	3	0.003914976	0.029390527	0.966694496
24 (FI1.conife	FI1	conifers	3	3	0	0.005849295	0.994150705
25 (FI1.conife	FI1	conifers	4	3	0	0.02144147	0.97855853
26 (FI1.conife	FI1	conifers	5	3	0	0.06330743	0.93669257
27 (FI1.conife	FI1	conifers	6	3	0	0.225043954	0.774956046
28 (FI1.conife	FI1	conifers	7	3	0	0.439946019	0.560053981
29 (FI1.conife	FI1	conifers	8	3	0	0	1
30 (FI1.conife	FI1	conifers	9	3	0	0	1
31 (FI1.conife	FI1	conifers	10	3	0	0	1
32 (FI1.conife	FI1	conifers	11	3	0	0	1
33 (FI1.conife	FI1	conifers	1	4	0	0	1
34 (FI1.conife	FI1	conifers	2	4	0.003914976	0.029390527	0.966694496

Figure 23. Activities probabilities table, t.

id	finalfell	thin
0 (a0_v0: 0..5	0	0
1 (a0_v1: 0..5	0	0.0291591
2 (a0_v2: 0..5	0	0
3 (a0_v3: 0..5	0	0
4 (a0_v4: 0..5	0	0
5 (a0_v5: 0..5	0	0
6 (a0_v6: 0..5	0	0
7 (a0_v7: 0..5	0	0
8 (a0_v8: 0..5	0	0
9 (a0_v9: 0..5	0	0
10 (a0_v10: 0	0	0
11 (a1_v0: 5..	0	0
12 (a1_v1: 5..	0	0.0291591
13 (a1_v2: 5..	0	0
14 (a1_v3: 5..	0	0
15 (a1_v4: 5..	0	0
16 (a1_v5: 5..	0	0
17 (a1_v6: 5..	0	0
18 (a1_v7: 5..	0	0
19 (a1_v8: 5..	0	0
20 (a1_v9: 5..	0	0
21 (a1_v10: 5	0	0
22 (a2_v0: 10	0	0
23 (a2_v1: 10	0.00102118	0.00766618
24 (a2_v2: 10	0	0.00152572
25 (a2_v3: 10	0	0.00559276
26 (a2_v4: 10	0	0.016513
27 (a2_v5: 10	0	0.0587001
28 (a2_v6: 10	0	0.114755
29 (a2_v7: 10	0	0
30 (a2_v8: 10	0	0
31 (a2_v9: 10	0	0
32 (a2_v10: 1	0	0
33 (a3_v0: 15	0	0
34 (a3_v1: 15	0.00102118	0.00766618

Figure 24. Updated activities, t+1

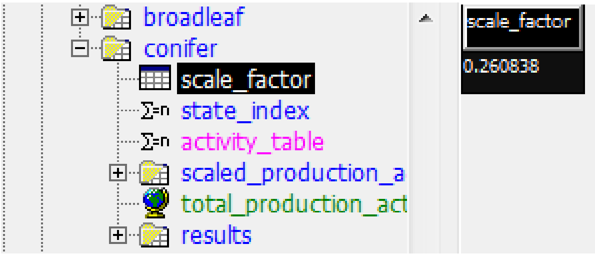


Figure 25. Factor to rescale activities to meet commodities demands for coniferous wood for t+1.

id	from	prev_volume	to	Probability	old_volume	new_volume	yield [100 mi
239 (a21_v8:	239 (a21_v8:	7	249 (a22_v7:	1	238.189	192.405	45.7848
240 (a21_v9:	240 (a21_v9:	8	250 (a22_v8:	1	291.3	238.189	53.1103
241 (a21_v10	241 (a21_v10	9	251 (a22_v9:	1	359.8	291.3	68.4999
242 (a22_v0:	242 (a22_v0:	0	253 (a23_v0:	1	7.5	7.5	0
243 (a22_v1:	243 (a22_v1:	0	253 (a23_v0:	1	23.7	7.5	16.2
244 (a22_v2:	244 (a22_v2:	1	254 (a23_v1:	1	42.492	23.7	18.792
245 (a22_v3:	245 (a22_v3:	2	255 (a23_v2:	1	64.2907	42.492	21.7987
246 (a22_v4:	246 (a22_v4:	3	256 (a23_v3:	1	89.5772	64.2907	25.2865
247 (a22_v5:	247 (a22_v5:	4	257 (a23_v4:	1	118.91	89.5772	29.3324
248 (a22_v6:	248 (a22_v6:	5	258 (a23_v5:	1	152.935	118.91	34.0255
249 (a22_v7:	249 (a22_v7:	6	259 (a23_v6:	1	192.405	152.935	39.4696
250 (a22_v8:	250 (a22_v8:	7	260 (a23_v7:	1	238.189	192.405	45.7848

Figure 26. Example of data showing yield from thinning per age/volume class in conifers in FI1.

The location in which the activity took place throughout the simulation can also be verified, as shown in Figure 27.

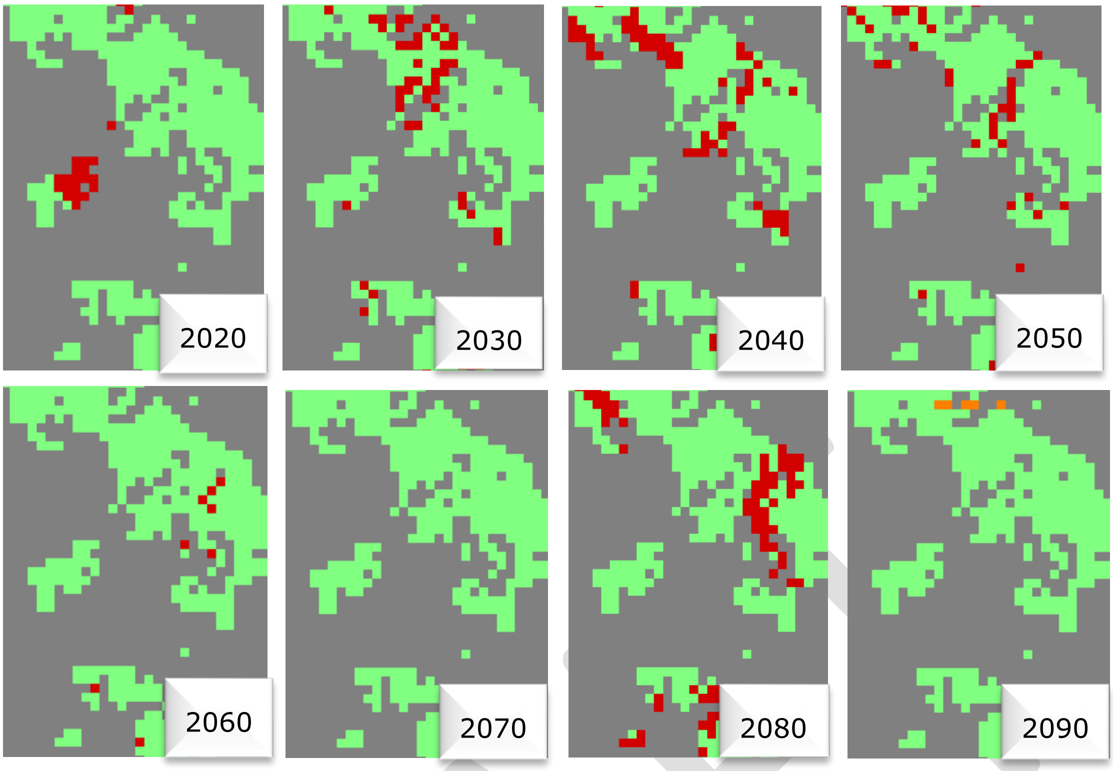


Figure 27. Example of a time series of management treatments.

	No management
	Thin
	Harvest (final fell)

5 Overview of Graphical User Interface

The Graphical user interface (Figure 28), is divided into four main panels. In the first, the tree structure allows the user to visualize the individual containers of the model. All aspects of the model, from details on the configuration to the containers that are used to execute the run can be visualized in this panel. Each container is described in further sections. The second panel is the data panel. When an element in the tree in panel 1 is clicked, this second panel will display either the map or table that is being requested. Similarly, it may also display calculation rules if requested through the third panel. The third panel has several tabs. Each tab displays slightly different information about the tree item that is highlighted in the first panel. The main tab is labelled "General", and allows the user to read important details about the tree item at a glance. Other tabs include "Explore", which shows the relative paths of the data in the tree item; "Properties", which defines the item according to its properties, as intended for the GeoDMS language; Metadata, which can be a pdf uploaded by the user to give further background information on the tree item, such as its history; "Statistics" summarise the raster data by showing basic statistics such as minimum and maximum values of the raster cells for the entire raster, and calculated values such as standard deviation, variance etc; "Configuration" summarises the expression used to obtain the tree item; "XML" summarises the tree item in XML format; and "Source Descr" shows paths of the datasets used to compute the tree item. The fourth panel is analogous to a descriptive progress bar when a tree item that requires computation is selected. Depending on the complexity of the calculation, the tree item may not appear immediately in the second panel (viewer). The user can follow the progress of the calculation through the messages give in this fourth panel.

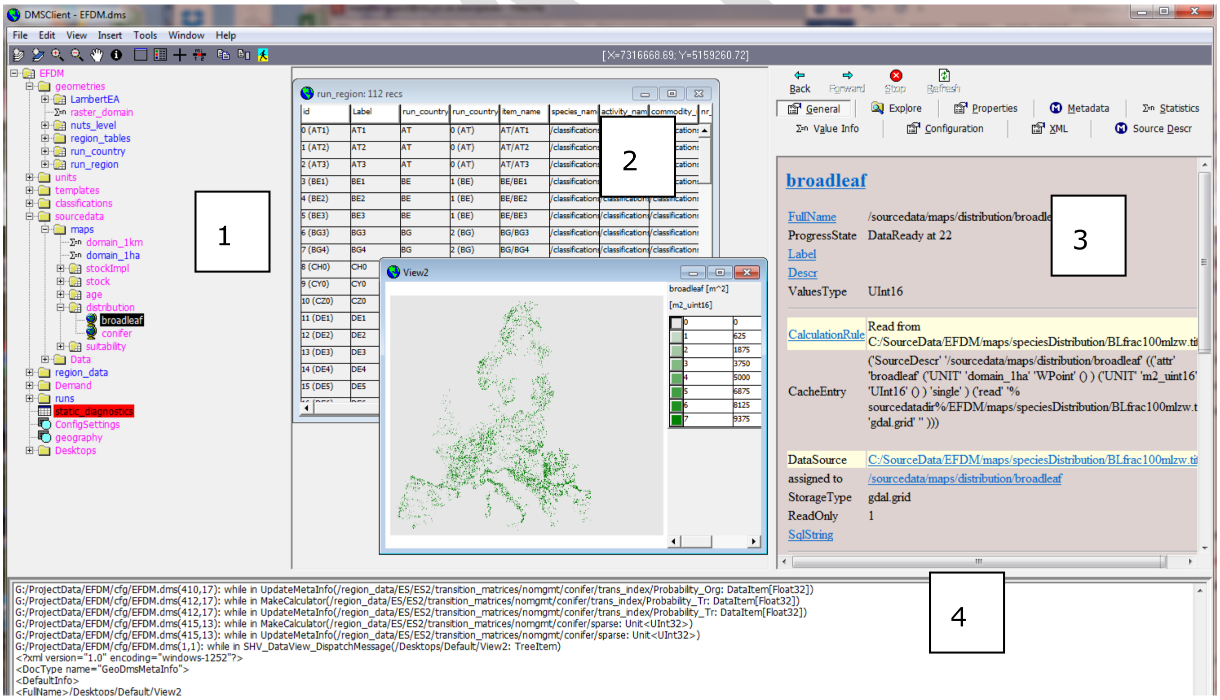


Figure 28. Graphical User Interface of EFDM-geo includes a tree structure (1), data panel (2), details pages (3), progress messages (4).

5.1 Detailed contents

The detailed contents of the model are visible to the user in the first panel called "tree items". All items below the main container consist of the configuration of the implementation of the EFDM model in GeoDMS.

Geometries. The first item under the main container is the "geometries" container. In this container, the map projection is defined. As an INSPIRE-conform tool, all input datasets are in Lambert-LAEA89 projection. This projection is defined within EFDM-geo and all rasters or vectors is assigned this projection. Furthermore, the maps of the regional divisions are set up in this container. Up to the time of writing, the NUTS regions NUTS0, NUTS1, NUTS2, NUTS3) are defined in this container. In the future, national-level regional divisions based on forest inventories could be added. This is relevant because the model will process data within the administrative boundaries defined by the maps in this container. Thus if forest inventory data corresponds to a particular administrative region, only the area within the boundary of this region will be processed with that data. Furthermore, each region is part of a larger region. When highlighting the tree item for run regions, a table appears showing this dependency (Figure 29).

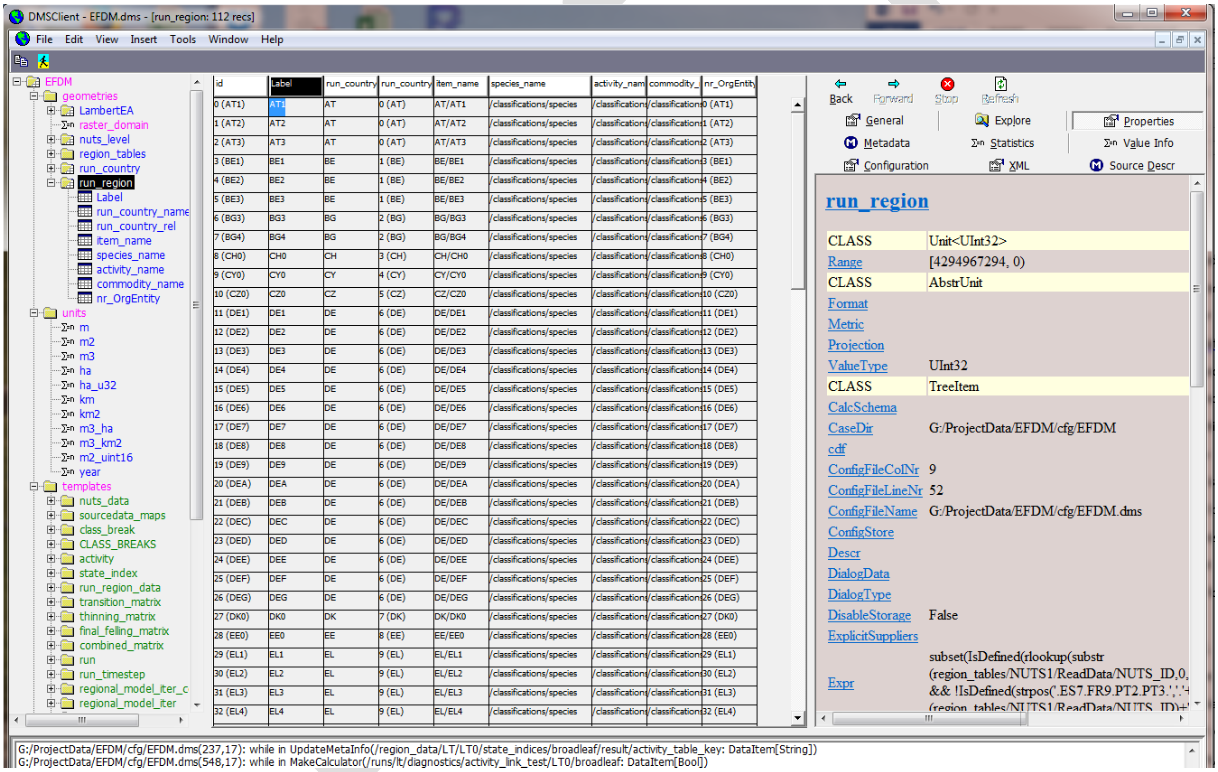


Figure 29. Regions modelled in EFDM-geo.

Units. The Units container defines basic units used in the model. This is necessary because each item can have an associated unit. This facilitates integrity checks, whereby the model will only compute a result of the units are compatible. The container contains very basic units, such as meters squared and hectares, and other units may be added, which is especially important for complex calculations that may be required, for example, to compute indicators.

Templates. Templates are a fundamental concept in GeoDMS in that they allow for the definition of basic data and calculations that will be called through other tree items. These are contained in one place.

Classifications. Classifications split continuous data into categories. This container describes how the raw data is categorised and the colour ramps associated to the classes.

In EFDM-geo, this is fundamental because the model works with age and volume classes.
Source Data. The source data container hosts the raw datasets used in the calculations of the model. It refers to templates container for the source and geographical projection of the base raster maps. Both maps and a tabular view of the data at regional level are given in this container.
Region Data. All data related to each region in Europe is contained in this mother container. Sub-containers are divided by country and each country container is further divided by region. Maps and data for the regions are visible at this finest geographical level. The main sub containers of direct interest to the simulations are described below.

- ➔ *CLASS_BREAKS*. The regional division consents a definition of class breaks on a regional level. This is important since regions in large countries in particular, will have different definitions of forest type. Age classes may be different from country to country, but will not vary from region to region as they are based on a common NFI strategy, which is set at national level (Figure 30). The class breaks for volume and age are read from the templates container, which in turn read from the source data, which is a cvs file, whose source is defined in the code. The source data can be changed by the user to change the class breaks.

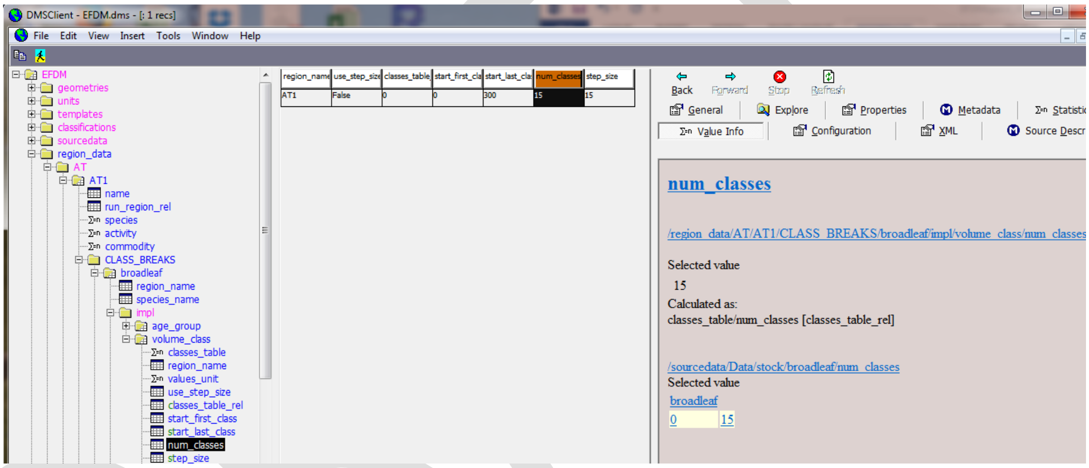


Figure 30. Volume class breaks can be defined per NUTS region in EFDM-geo

- ➔ *raster_impl.* This sub-container includes the components that are directly involved in the computation of the raster model. This is what can be considered as the core of the model computation because it is where the allocation of activities occurs. This allocation is dependent on the density of forests in each 1ha cell.
- ➔ The "balanced_alloc" sub-container is the first container we have come to where allocation takes place. In this step, the discrete allocation is used to assign a preferential species to a cell. The preference is given to the species with the highest coverage within the cell. The code is the following:

```
container balanced_alloc :=
discrete_alloc(extended_species/label, domain_100m_sel,
suitabilities, ID(extended_species), extended_species/label,
atomic_regions, atomic_regions/UnionData,min_claims,
max_claims, 0i, raster_impl);
```

The command "discrete_alloc" is a powerful one. This example is a very simple application. It simply assigns the species with the highest value, or highest presence, in any given cell. As we can read from the code above, it is reading the "suitability" for each species. This corresponds to what we saw in Figure 4.

➔ *initial_state*. The initial state matrix can be visualized as a vector within these regional containers (Figure 31).

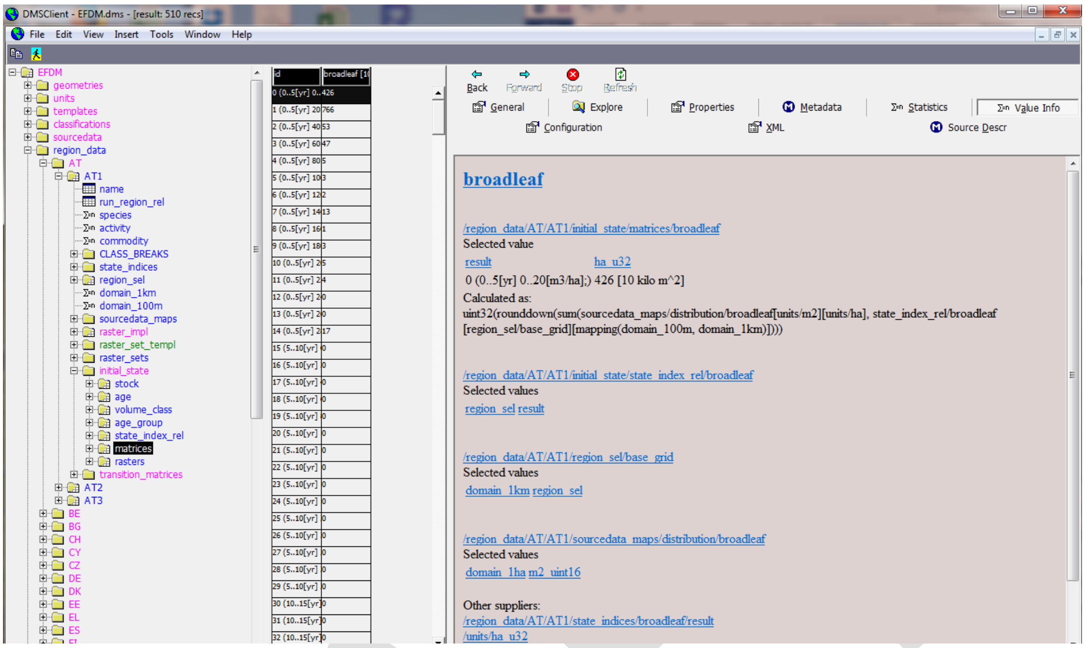


Figure 31. Initial state matrix per NUTS 1 region and per species group.

Transition matrices for non management situations, that is, for normal growth without immediate management, can be viewed in the regional containers as a table, ordered from-> to (Figure 32).

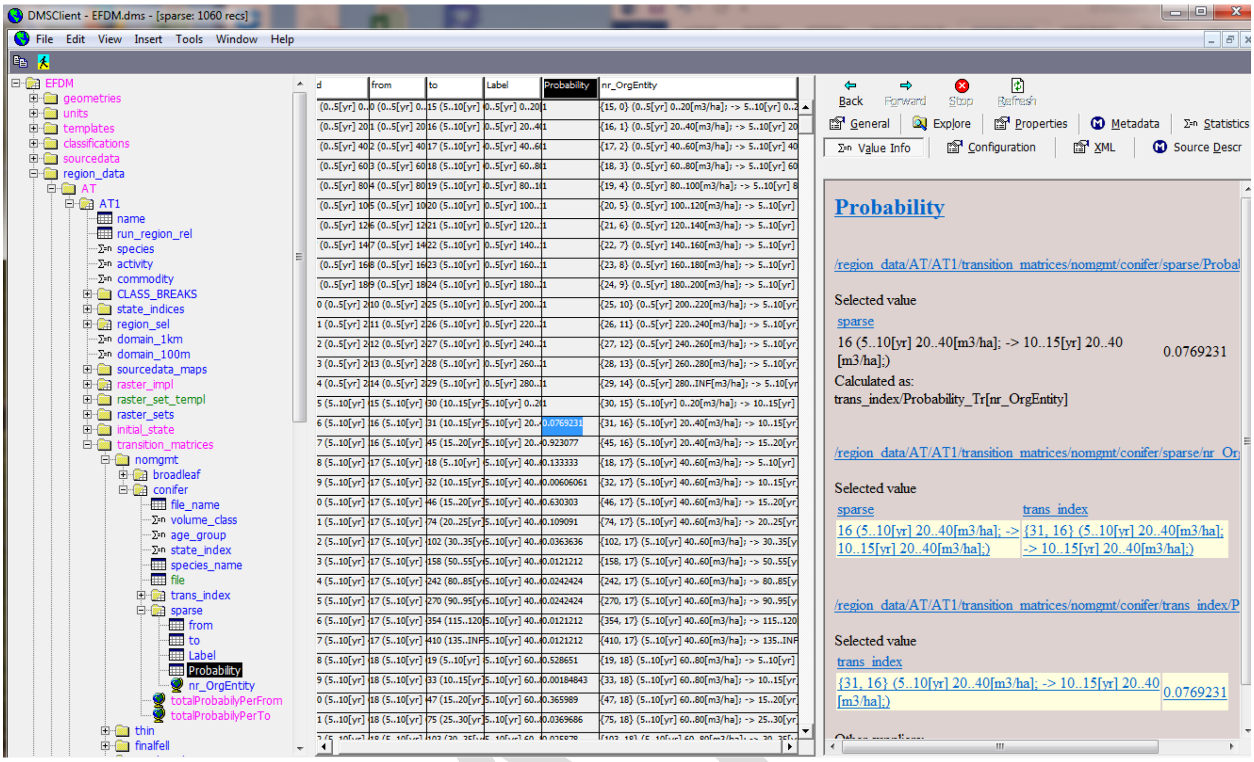


Figure 32. Transition matrix at NUTS 1 level per species group.

The no-management transition probabilities can also be visualized as a graph. The graphical representation allows the user to see if the transitions are realistic. If a linear diagonal pattern is seen, the transitions follow an expected pattern, that is: the ages systematically increase by one class and volume classes usually increase by one class, may remain the same or may increase by two classes. An example of a good matrix is shown in Figure 33.

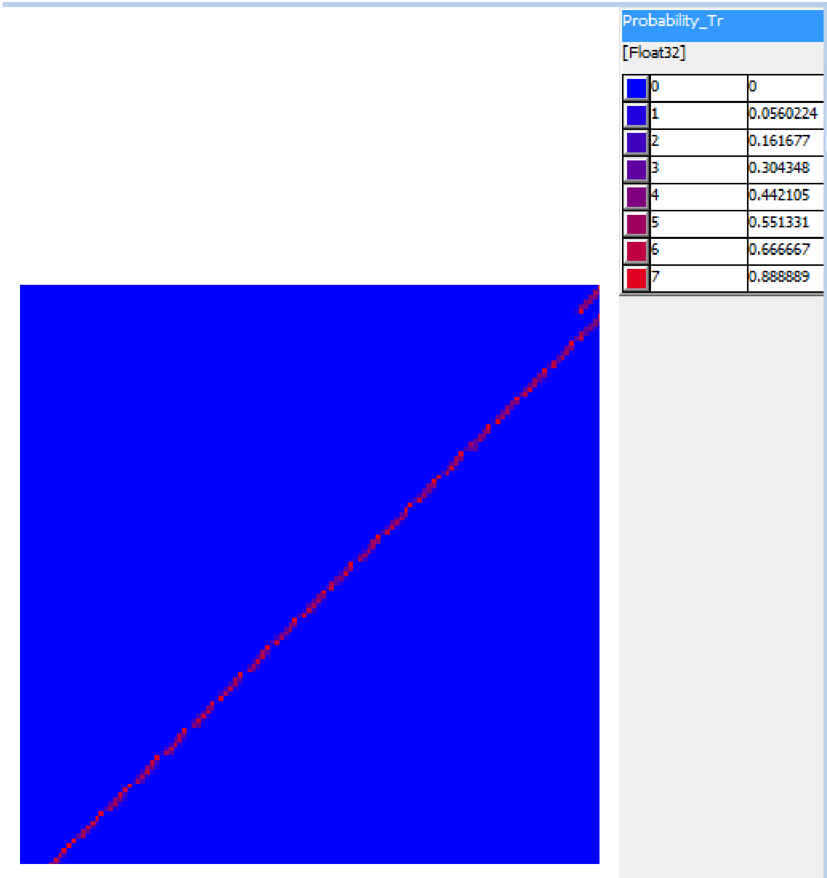


Figure 33. Example of transition probabilities matrices for FI1.

Figure 34 shows a zoom of Figure 33. The central square, highlighted by a shaded box, lies on the diagonal, so up one age step and up one volume step. This box usually has the highest probability of occurring. The boxes directly above and below this central box represent exceptions that could occur. The age group will always increase as time passes, but volume may remain the same (box below the diagonal); or increase by two volume classes (box above the diagonal.

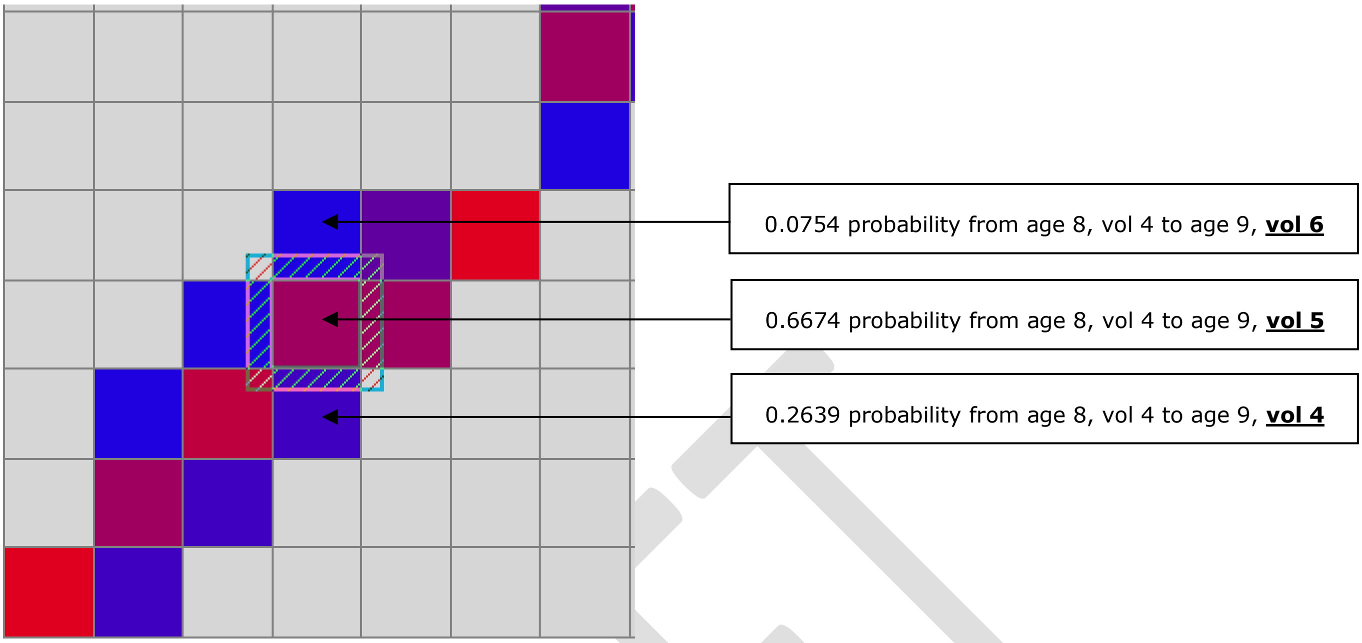


Figure 34. Zoom in to explain the transition matrix graphic.

Demand. The demand container houses the figures given by the Global Forest Sector Model (GFTM). This model output is read directly into EFDM-geo and mapped (Figure 35).

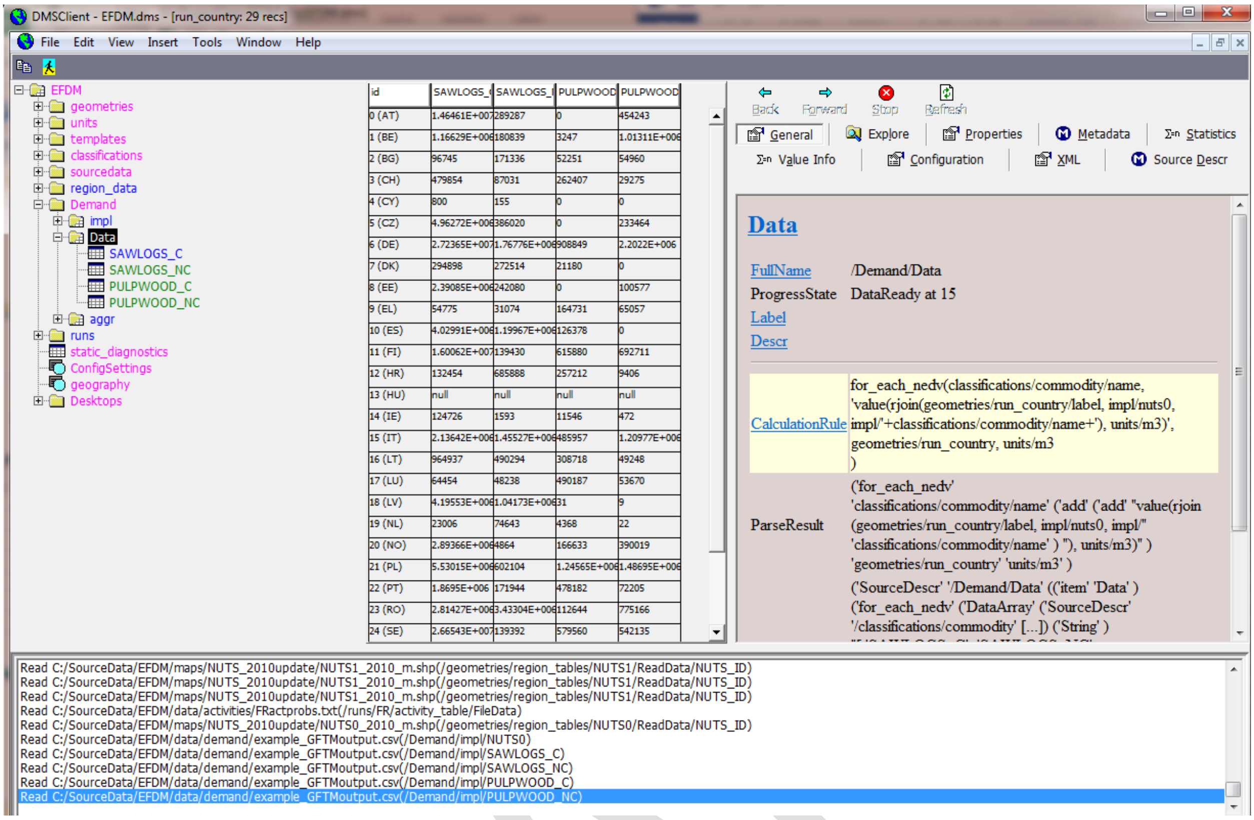


Figure 35. Demand for material use of wood from the GFTM model.

Runs. This is the container in which the actual simulations happen. In this container, the regional and raster levels come together and the operations described throughout this document, are computed.

Here the activities tables are used as a probability that final felling, thinning or no management will take place for any given forest type (Figure 36).

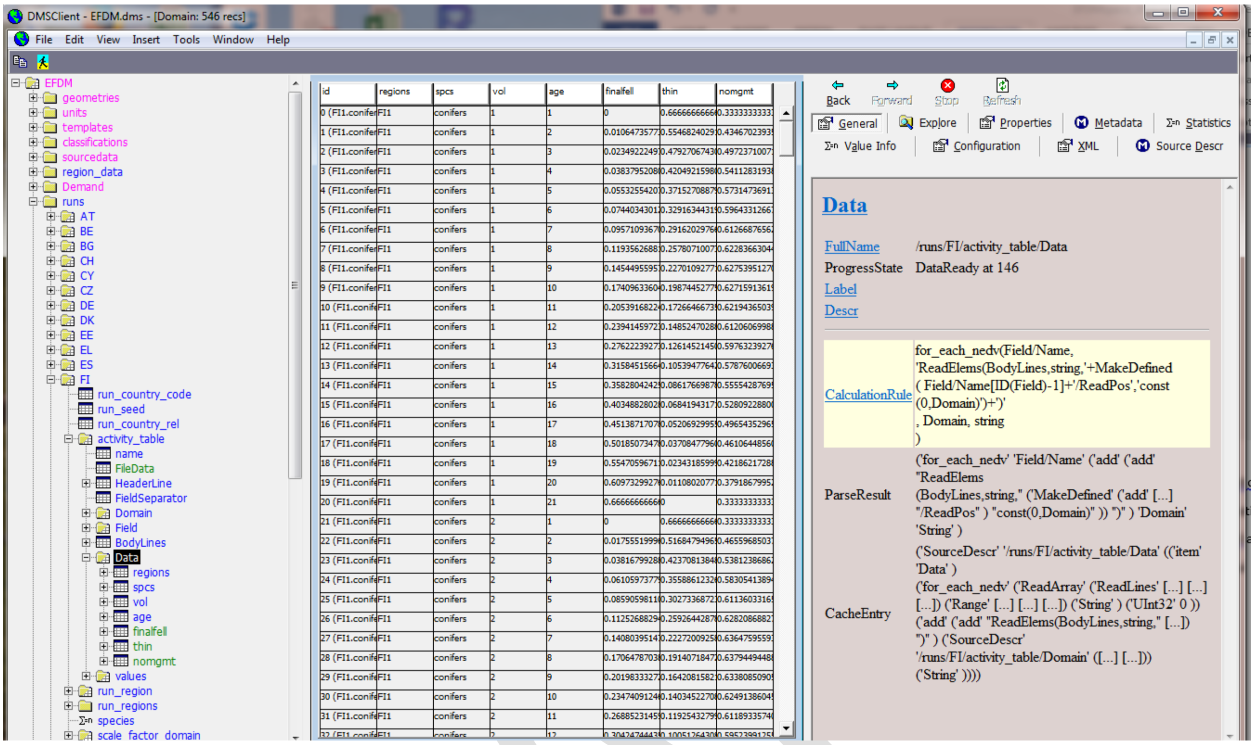


Figure 36. Activities probabilities matrix per forest type at regional level.

6 Conclusions and way forward

The approach described in this document has resulted in a very powerful and fast modelling environment that is ready and able to integrate further information and data in an organised and homogeneous way for forward-looking scenario assessments in the forest-based sector. The spatially-explicit component of this model allows for the computation of environmental indicators. When overlain with data such as waterways, soil type and areas prone to natural disaster, the impact of forest growth and activities in the forest can be studied. Furthermore, the impacts of spatially-explicit policy such as zoning and protection of certain areas; or the simulation of new biomass processing plants can be assessed through scenarios. Plans for the modelling system fall into three broad categories: sensitivity analysis, improvement and expansion.

Quantifying the uncertainty, especially that associated with the spatial disaggregation, is a priority. Among improvements are the refinement of the input layers used in the model; the integration of a capacity to shift species; split commodities in a more accurate manner; model non-even-aged forests; and develop the feedback mechanism between the forest trade model and EFDM-geo. Among plans to expand the model are the inclusion of short-rotation forestry, especially coppice for fuelwood; and a simple land use model to allow for changing forest area and expanding the model to span a larger area, beyond Europe. Furthermore, as data and information become available on forest ownership or forest-owner reactions to market triggers in terms of wood mobilisation from their woodland, these aspects should also be integrated. At this stage the shortage of data and information are restricting the modelling capabilities in EFDM-geo with respect to ownership.

6.1 Refinement & validation of the input layers

We have seen that the initial state is the basis of all the modelling that takes place in EFDM-geo. An accurate as possible estimation of the initial state is therefore clearly crucial in order to avoid the propagation and amplification of errors throughout the modelling process. The volume and species are taken directly from satellite-based products, and the estimation of the age is a derivative of the volume and increment estimates. This implies a margin of error is inevitable, but to reduce this margin, a lengthy validation procedure should take place. This process is preferably undertaken with the help of local expertise.

6.2 Species shifts and death

As described under section 3.3 on Activities Matrices, ingrowth only occurs in volume and age class 0. The model assumes that the same species will return, however it would technically be possible to recruit bare land with a different species in the model. Furthermore, there is no net loss or gain of forested area, which is also something that could be changed, as discussed further in sections 6.5 & 6.6.

Schematically we can represent the different path options for artificial recruitment as such (Figure 37):

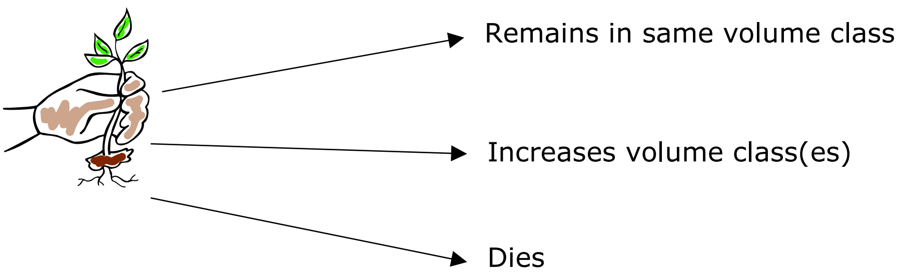


Figure 37. Possible paths for newly and artificially recruited seedlings for each size class

This model configuration does not consider the probability of tree mortality. Natural mortality should also be taken into account. For example, Liang (2010) considers a presumed mortality of 20%.

6.3 Timber assortments

In order to better estimate the supply side of the forest-based sector, some work needs to be done on the split of harvested volumes of wood to the different assortments. In EFDM-geo, the split is made between coniferous and non-coniferous commodities, but there is no split between pulpwood and sawlogs, which means that all volume-age classes from which volume is removed, whether that be by thinning or by final felling, is equally capable of supplying the demand for both. As described earlier however, the model is equipped to handle a more sophisticated and complex rule-set. This is not currently feasible because of lack of research in this field. Furthermore, it would be interesting to integrate the common situation whereby some forests labelled as coniferous are contributing to hardwood commodities to a small extent, and vice versa. The forests are usually mixed to a slight degree, and harvesting will pick up both species at times.

6.4 Non-even-aged forests

EFDM is available in a version that was conceived to deal with non-even-aged forests, using volume and stem numbers (instead of age) as dynamic variables when describing forest type (Sallnäs et al, 2015). The model was applied by national experts in the following countries: Austria, France, the Netherlands, Portugal, Slovenia, Spain, Switzerland. The model performed well in some situations, but thre are were nevertheless points to consider for further development. For instance, dependence on plot data for estimation of transition matrices. Sometimes low numbers of sample plots for these types of forests, particularly unmanaged plots in smaller countries, poses a problem. Another issue to contend with is a situation as was described in Switzerland, whereby the memory of management of the plot is important because these plots tend to have a much higher growth rate. This would imply generating a specific transition probability matrix for these areas and applying them once to the areas that have been managed. Another concern to address is the age window. It is currently equivalent to the time step width, which is too large to capture management practices on coppices as was the case in Spain. These issues will probably have to be addressed on a country-by-country basis, however in general terms, the fine granularity of the raster in this system, and the possibility to move from discrete allocation to probabilistic allocation[15] for forest types would allow for modelling non-even-aged forests. We should also consider running conifers and broadleaves models as described throughout this document separately, allowing each species group to occupy the same cell with their own independent age class, and summing the results at the end of the simulation.

6.5 Short-rotation forests

Short rotation forests (SRF) has always been a bit tricky for both the agriculture and forestry sectors to handle because they could be considered as one or the other. On the one hand, they supply the forestry sector but on the other, they usually occupy agricultural land. Given the mandate of EFDM-geo to assess biomass availability given demand for wood products at a pan-European scale, it makes sense to integrate SRF into the model. The challenge is in the volatile land take of SRC in very short time steps. This can be

[15] http://objectvision.nl/geodms/operators-a-functions/allocation

resolved by integrating a simple land use model in which parcels of agriculture land may be made available for SRF, which in turn, would make the products from SRC available to contribute to the overall supply of wood to specific markets.

Further to SRC, we may introduce more forests (afforestation) or fewer forests (permanent deforestation) to this concept, which would then lead to the creation of a simplistic land use model for biomass estimation from both forestry and agriculture (see section 6.6).

6.6 Land use modelling for biomass assessment

This spatially-explicit environment is conducive to the net land exchange between the agricultural and forestry sectors. Since the initial state is generated "on the fly", meaning the underlying maps contributing to the initial state could theoretically be modified at every time step. Thus the forest species, age and volume maps used to generate the initial state vector could be modified based on exogenous factors governing net areas of certain species. We can therefore imagine an introduction of a new "species class" on top of broadleaves and conifers, which would be short-rotation broadleaves. This new class would have its own characteristics in terms of growth rate, age and volume classes (just as conifers and broadleaves do). From a theoretical point of view, the area-based matrix set-up in a volume/age dimension, based on the Markov principles governing the shift of forest areas within the matrix, is analogous to probabilistic cellular automata within the spatial dimension. In addition to hectares of forest moving through the volume/age dimensions, we could have hectares of SRC moving through the x-y dimensions. The main conceptual difference is that we would have to add the ability of a cell to maintain the memory of the previous states in order to allow the model to make good decisions about the land use conversions; and instead of transition probabilities, we would depend on conversion probabilities of land use.

Given the magnitude of importance associated to biomass as the core of the bio-based economy and the subsequent demand for bioenergy from both within and outside of Europe, it may also be appropriate to seek Global-modelling capabilities for biomass availability. This would probably be possible with a degraded spatial resolution for the "rest of the world" and simplistic rules regarding growth and activities.

References

1. Mubareka, S., Jonsson, R., Rinaldi, F., Fiorese, G., San-Miguel-Ayanz, J., Sallnäs, O., Baranzelli, C., Pilli, R., Lavalle, C., Kitous, A. 2015. An Integrated Modelling Framework for the Forest-based Bioeconomy. IEEE Earthzine doi: http://dx.doi.org/10.1101/011932
2. Packalen, T., Sallnäs, O., Sirkiä, S., Korhonen, K., Salminen, O., Vidal, C., Robert, N., Colin, A., Belouard, T., Schadauer, K., Berger, A., Rego, F., Louro, G., Camia, A., Räty, M., San-Miguel-Ayanz, J. 2014. The European Forestry Dynamics Model. EUR 27004. Publications Office of the European Union, doi 10.2788/153990.
3. Liang, J., Picard, N. 2013. Matrix Model of Forest Dynamics: An Overview and Outlook. Forest Science 59(3): 359-378.
4. Sallnäs, 0.(1990). A matrix growth model of the Swedish forest. Studia Forestalia Suecica 183. 23 p.
5. Jonsson, R., Rinaldi, F., San-Miguel-Ayanz, J. 2015. The Global Forest Trade model (GFTM). EUR 27360. Publications Office of the European Union, doi 10.2788/666206.
6. Verkerk, P.J., Zanchi, G., Lindner, M. 2008. Impacts of Biological and Landscape Diversity Protection on the Wood Supply in Europe. EFI Technical Report 27. European Forest Institute, Joensuu, Finland.
7. de Rigo, D., Caudullo, G., Houston Durrant, T., San-Miguel-Ayanz, J. 2016. The European Atlas of Forest Tree Species: modelling, data and information on forest tree species. In: San-Miguel-Ayanz, J., de Rigo, D., Caudullo, G., Houston Durrant, T., Mauri, A. (Eds.), European Atlas of Forest Tree Species. Publication Office of the European Union, doi 10.2788/038466 (online); 10.2788/4251 (print).
8. 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 11 (3), 61pp. doi 10.2903/sp.efsa.2014.EN-434.
9. Gallaun, H., Zanchi, G., Nabuurs, G., Hengeveld, G., Schardt, M., Verkerk, P. 2010. EU-wide maps of growing stock and above-ground biomass in forests based on remote sensing and field measurements. Forest Ecology and Management. 260:252-261.
10. Santoro, M., Beer, C., Cartus, O., Schmullius, C., Shvidenko, A., McCallum, I., Wegmüller, U., Wiesmann, A. 2011. Retrieval of growing stock volume in boreal forest using hyper-temporal series of Envisat ASAR ScanSAR backscatter measurements. Remote Sensing of Environment. 115:490-507.
11. Kempeneers, P., McInerney, D., Sedano, F., Gallego, J., Strobl, P., Kay, S., Korhonen, K., San-Miguel-Ayanz, J. 2013. Accuracy assessment of a remote sensing-based pan-European forest cover map using multi-country national forest inventory data. IEEE Journal of selected topics in applied Earth observation and remote sensing, 6(1):54-65.
12. Pekkarinen, A, Reithmaier, L., Strobl, P. 2009. Pan-European forest/non-forest mapping with Landsat ETM+ and CORINE Land Cover 2000 data. ISPRS Journal of Photogrammetry and Remote Sensing 64:171-183.
13. Pilli, R., Grassi, G., Kurz, W., Smyth, C., Blujdea, V. 2013. Application of the CBM-CFS3 model to estimate Italy's forest carbon budget, 1995–2020. Ecological Modelling 266:144-171.
14. State of Europe's Forests 2015. Ministerial Conference on the Protection of Forests in Europe. FOREST EUROPE Liaison Unit Madrid. www.foresteurope.org
15. Sirkia, S. "Some mathematics behind EFDM", October 2012. Available at https://webgate.ec.europa.eu/CITnet/stash/projects/FISE/repos/efdm/browse/documents/EFDMinstructions/Seija_Mathematics_behind_EFDM.pdf.

16. Liang, J. 2010. Dynamics and Management of Alaska Boreal Forest: An All-aged Multi-species Matrix Stand Growth Model. Forest Ecology and Management 260: 491-501.
17. Sallnäs, O., Berger, A., Räty, M., Trubins, R. 2015. An Area-Based Matrix Model for Uneven-Aged Forests. Forests 6, 1500-1515; doi 10.3390/f6051500.

List of figures

List of tables

Europe Direct is a service to help you find answers to your questions about the European Union
Free phone number (*): 00 800 6 7 8 9 10 11
(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server http://europa.eu

How to obtain EU publications

Our publications are available from EU Bookshop (http://bookshop.europa.eu),
where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents.
You can obtain their contact details by sending a fax to (352) 29 29-42758.

KJ-NA-xxxxx-EN-C

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub

Publications Office

doi:xx.xxxx/xxxx

ISBN xxx-xx-xx-xxxxx-x