Characterizing vertical forest stands structure using data mining methods

Marko Debeljak[1] and Jana Babič[2]

[1] Department of Knowledge Technologies, "Jožef Stefan" Institute, Ljubljana, Slovenia,

[2]Nova Gorica Polytechnic, School of Environmental Sciences, Nova Gorica, Slovenia

Vertical forest stands structure

DEFINITON:
"… the bottom to top configuration of above ground vegetation within a forest stand"

"distribution of tree heights within a forest stand"

Changes during natural development of stand:
- the number, height and biomass of trees,
- arrangement of branches
- leaf area on trees,
-understory composition and structure, etc.

Vertical forest stands structure

Influences working on stand:

- **natural forest development processes** (cyclical stand development, successions)
- **silviculture treatments** (part of forest managemt).

Factors that **affect** the development of vertical structure:

- physiological and morphological properties of individual trees,
- size of the trees
- spatial position of their neighbors,
- stand density,
- disturbance history,
- site conditions etc.

Vertical forest stands structure

Changes in vertical forest structure affect:

- both **microclimatic** factors and **processes** in the system,
- tree **growth**,
- **understory** community structure,
- suitability of the stand for **wildlife**,
- **hydrologic** response,
- fire **hazard**,
- susceptibility to **pest and disease**,
- **aesthetic** value etc.

Vertical forest stands structure

Stand structure is the OUTER reflection of the processes.

The study of vertical structure may make a significant contribution to the knowledge about growth and developmental processes of forest ecosystems.

Methods for vertical structures quantification

- **arbitrarily** defined and do not represent natural stratification patterns of forests
- are **too time consuming** for landscape analyses
- can be **too expensive**
- **inappropriate** at dense stands

To improve the methodological shortages we conducted a very detailed study of the vertical stand structure.

Goals

To identify properties of the vertical structure in both managed and virgin forest.

Case study

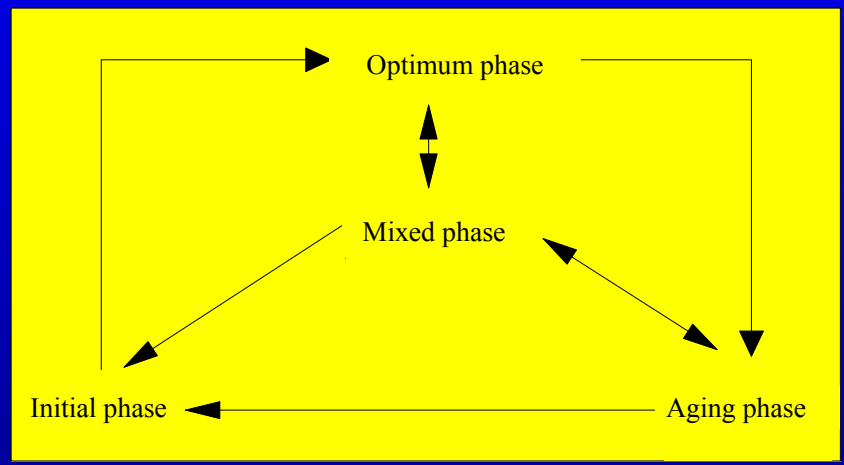
- study plots : - the **virgin forest remnant Rajhenavski Rog**,
- lightly **managed forest in its vicinity**.

- the selection of **study plots** was restricted to the most dominant forest plant community ***Omphalodo-Fagetum omphalodetosum*** => high karst **Dinaric forests with silver fir** (*Abies alba* Mill.) and **beech** (*Fagus sylvatica* L.) as the most dominant tree species.

Case study

Stand dynamics was **described** by **four** indicative forest **cyclical developmental** phases, which have distinctive vertical and horizontal stand structures:

- a **juvenile** phase,
- an **optimal** phase,
- a **mixed** phase,
- a **regeneration** phase.



Sampling and data

Four research plots (35m by 35m) were **randomly** selected within optimal, mixed and regeneration developmental phase in **both** managed and virgin forest.

Stand structure was described by the following **attributes**:

- **tree species,**
- **dimager at breast high (DBH),**
- **tree height,**
- **layer,**
- **depth of the crown,**
- **width of the crown,**
- **social position,**
- **vitality.**

Data analysis

Structural patterns of vertical stand structure were **studied by automated data analysis** using **machine learning** techniques:

- **classification trees**

- **regression trees**

Classification trees

They predict the value of a **discrete dependent variable** with a **finite** set of values (called **class**)
from
the values of a set of **independent variables** (called **attributes**), which may be either **continuous or discrete**.

Data mining analysis was performed by the **Weka** machine learning package.

We used **J4.8 algorithm**, which is Weka's implementation of **C4.5** decision tree algorithm - one of the most widely used decision tree system.

Regression trees

They predict the value of a **continuous dependent variable** (called **class**) or **linear function of some attributes** from the values of a set of **independent variables** (called **attributes**), which may be either **continuous or discrete**.

Data mining analysis was performed by the **Weka** machine learning package. We used **M5 algorithm** as one of the most widely used regression tree system.

Results – Classification trees

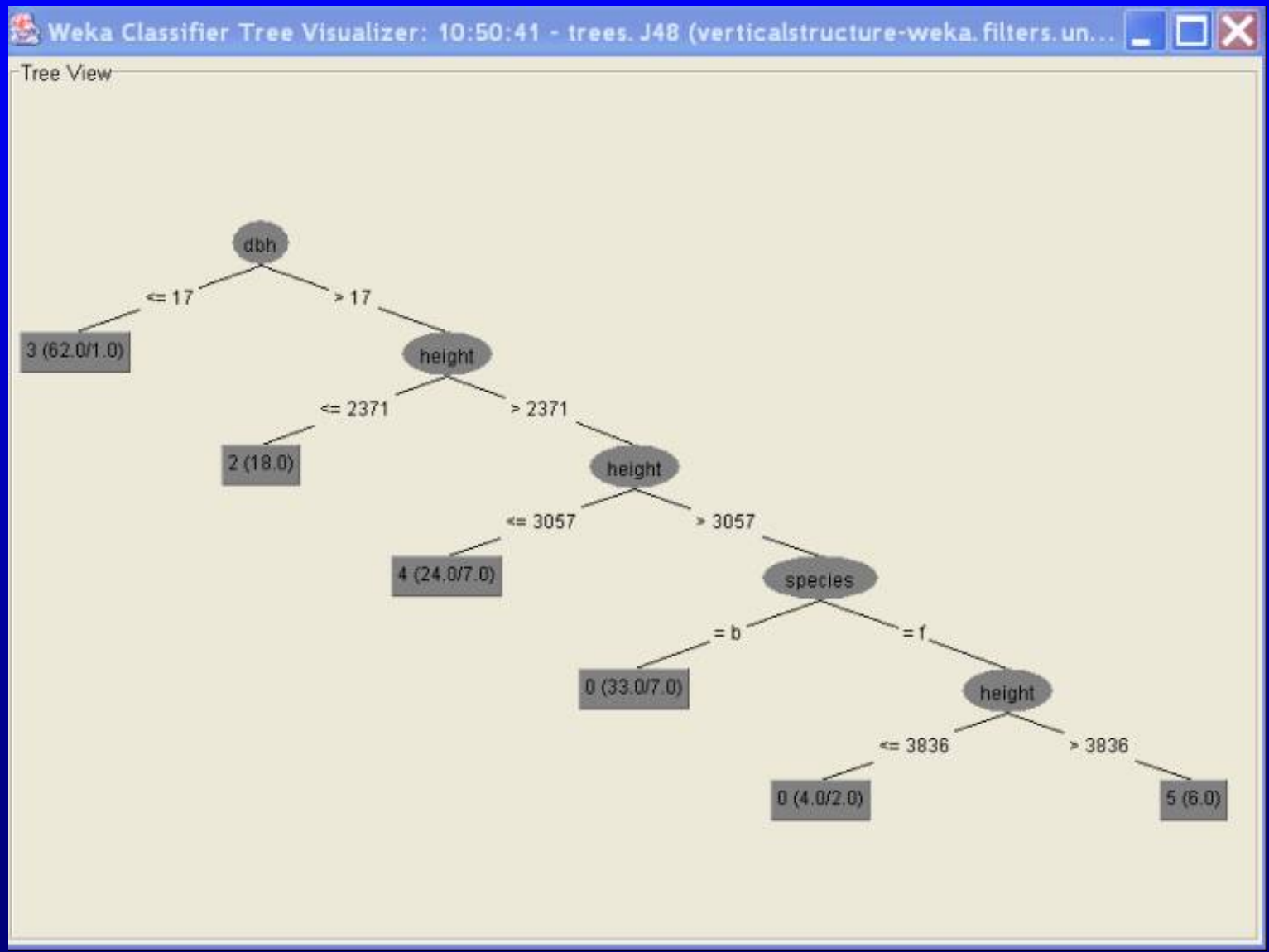
Induced models 24 different models:

	Optimal phase	Mixed phase	Regeneration phase
VIRGIN forest	layer crown depth crown width 195 instances	layer crown depth crown width 303 instances	layer crown depth crown width 147 instances
MANGED forest	layer crown depth crown width 336 instances	layer crown depth crown width 325 instances	layer crown depth crown width 218 instances

VIRGIN forest	layer crown depth crown width 645 instancs
MANGED forest	layer crown depth crown width 879 instances

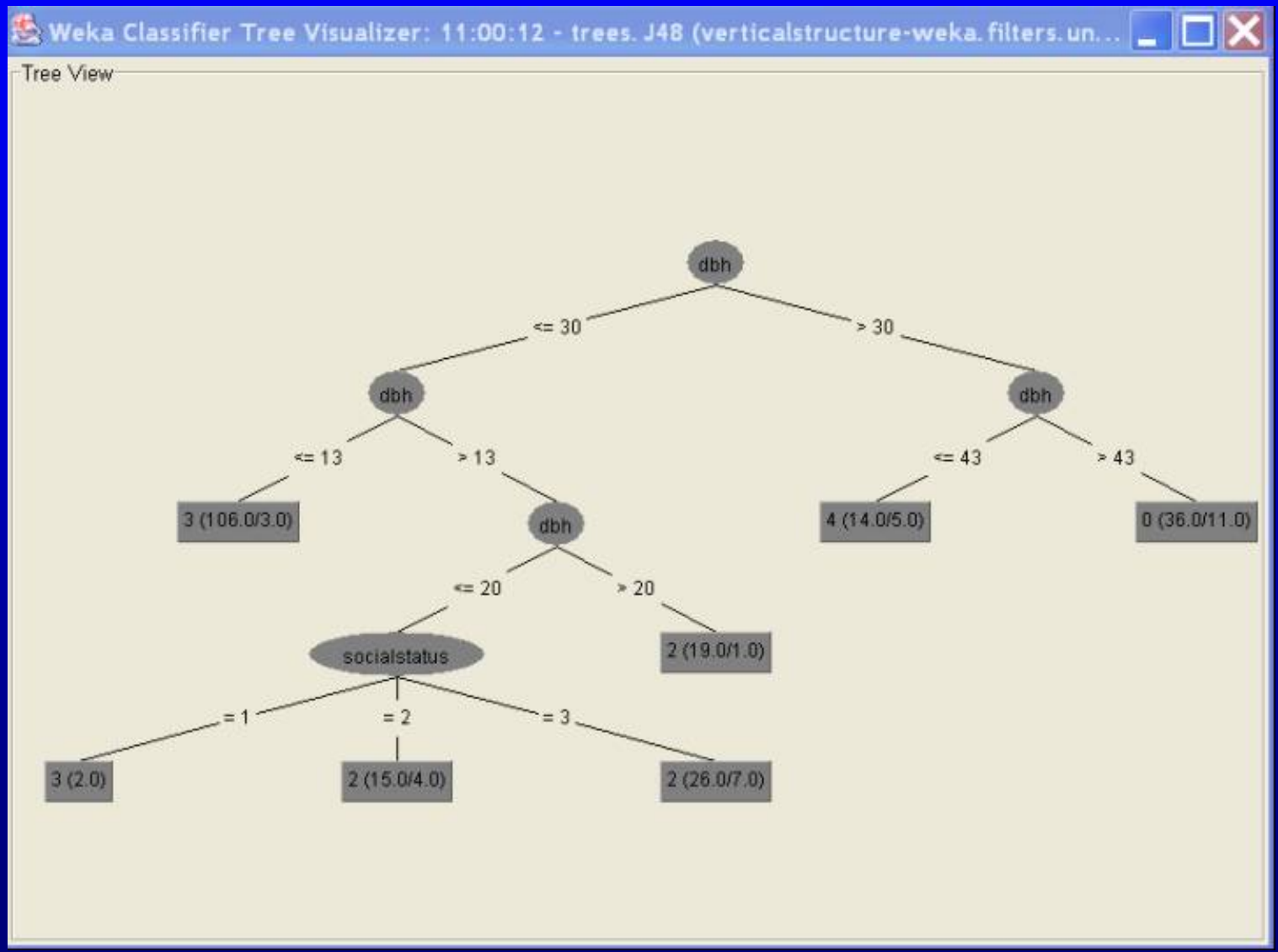
Results – Classification trees

Layer: virgin f.- regeneration ph.



Results – Classification trees

Layer: managed f.- regeneration ph.



Results – Classification trees

Layer m.: virgin f.: regeneration

```
dbh <= 17: 3 (62.0/1.0)
dbh > 17
|   height <= 2371: 2 (18.0)
|   height > 2371
|   |   height <= 3057: 4 (24.0/7.0)
|   |   height > 3057
|   |   |   species = b: 0 (33.0/7.0)
|   |   |   species = f
|   |   |   |   height <= 3836: 0 (4.0/2.0)
|   |   |   |   height > 3836: 5 (6.0)
```

Correctly Classified Instances	124	84.3537 %
Incorrectly Classified Instances	23	15.6463 %
Kappa statistic	0.7853	
Mean absolute error	0.0866	
Root mean squared error	0.2299	
Relative absolute error	29.4685 %	
Root relative squared error	60.065 %	
Total Number of Instances	147	

Layer m.: managed f. regeneration

```
dbh <= 30
|   dbh <= 13: 3 (106.0/3.0)
|   dbh > 13
|   |   dbh <= 20
|   |   |   socialstatus = 1: 3 (2.0)
|   |   |   socialstatus = 2: 2 (15.0/4.0)
|   |   |   socialstatus = 3: 2 (26.0/7.0)
|   |   dbh > 20: 2 (19.0/1.0)
dbh > 30
|   dbh <= 43: 4 (14.0/5.0)
|   dbh > 43: 0 (36.0/11.0)
```

Correctly Classified Instances	177	81.1927 %
Incorrectly Classified Instances	41	18.8073 %
Kappa statistic	0.7105	
Mean absolute error	0.0951	
Root mean squared error	0.233	
Relative absolute error	37.1597 %	
Root relative squared error	65.3369 %	
Total Number of Instances	218	

Results – Classification trees - conclusions

24 models : the **independent attributes** that **correlates** the most with the **selected (dependent) class**-attribute

Layer models: DBH, crown depth, social status

Crown depths models: social status, layer, tree height

Crown width models: vitality, crown depth, social status

Results – Classification trees - conclusions

Patterns of relationships among attributes:

lower DBH => lower layer => the thresholds between layers are lower in managed then in virgin forest

narrow crowns => lower vitality or/and trees are groups

low crown depths => trees are in groups

Results – Classification trees - conclusions

VIRGINE forest: high diversity in vertical structures between development phases

MANAGED forest: homogeneous vertical structure => no differences between development phases

Results - regression trees

Virgin forest height model:
Instances: 645

Managed forest height model:
Instances: 879

Results - regression trees
Virgin forest: model of tree's height

```
dbh <= 24.5 :
|   dbh <= 7.5 : LM1 (209/6.253%)
|   dbh >  7.5 : LM2 (198/3.491%)
dbh >  24.5 :
|   dbh <= 44.5 : LM3 (88/6.504%)
|   dbh >  44.5 :
|   |   dbh <= 61.5 : LM4 (68/4.559%)
|   |   dbh >  61.5 : LM5 (82/6.126%)
```

Correlation coefficient	0.9962
Mean absolute error	0.9469
Root mean squared error	2.1788
Relative absolute error	4.5358 %
Root relative squared error	8.7022 %
Total Number of Instances	645

LM num: 1
height = 4.9316 * phase=3,2 - 53.6045 * species=f + 118.5593 * dbh - 1.555 * crowndepth=2 - 5.7753 * socialstatus=1,2 + 6.7971 * socialstatus=2 + 4.3573 * vitality=1,2 + 52.7343

LM num: 2
height = 4.9316 * phase=3,2 - 723.1937 * species=f + 91.5362 * dbh - 1.555 * crowndepth=2 - 5.7753 * socialstatus=1,2 + 6.7971 * socialstatus=2 + 4.4243 * vitality=1,2 + 338.7362

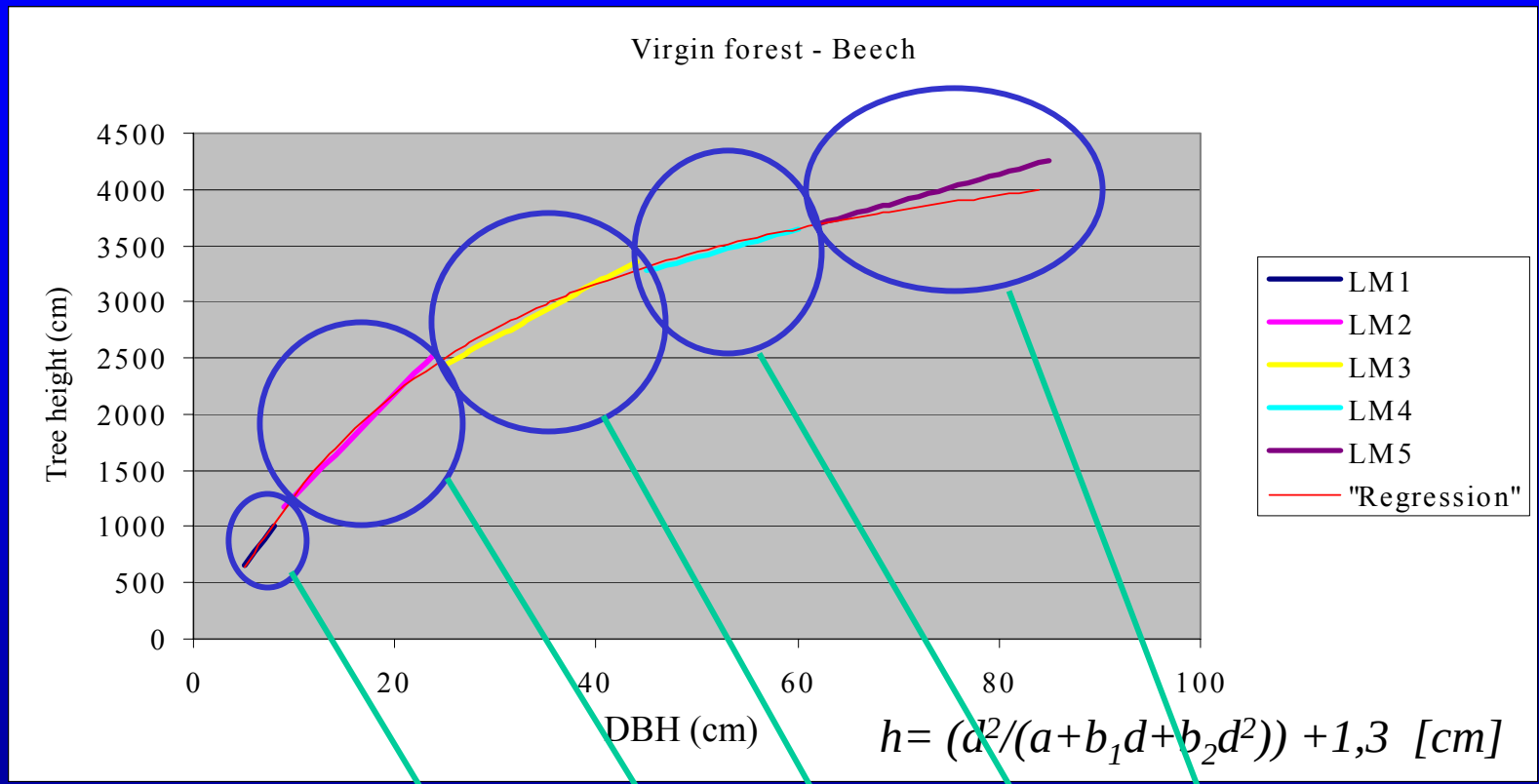
LM num: 3
height = 8.2259 * phase=3,2 - 391.9149 * species=f + 48.393 * dbh - 2.5937 * crowndepth=2 - 9.6331 * socialstatus=1,2 + 11.3375 * socialstatus=2 + 5.1068 * vitality=1,2 + 1220.2757

LM num: 4
height = 8.2259 * phase=3,2 - 28.015 * species=f + 24.6605 * dbh - 2.5937 * crowndepth=2 - 9.6331 * socialstatus=1,2 + 11.3375 * socialstatus=2 + 5.1068 * vitality=1,2 + 2150.9432

LM num: 5
height = 8.2259 * phase=3,2 - 28.015 * species=f + 16.9728 * dbh - 2.5937 * crowndepth=2 - 9.6331 * socialstatus=1,2 + 11.3375 * socialstatus=2 + 5.1068 * vitality=1,2 + 2665.7544

Results - regression trees

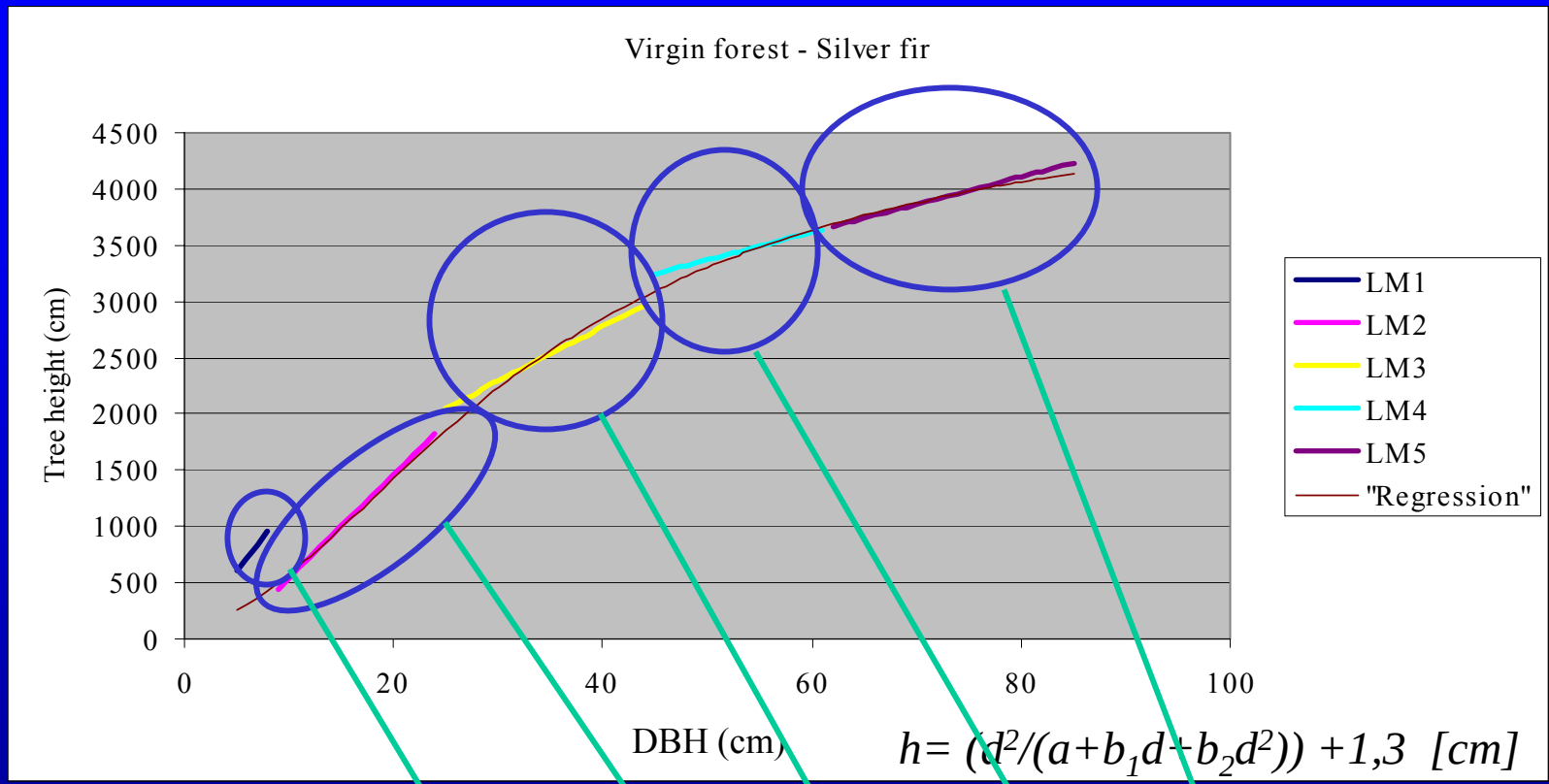
Virgin forest: model of tree's height



Virgin f.	*LM1*	*LM2*	*LM3*	*LM4*	*LM5*
phase=3,2	4.932	4.932	8.226	8.226	8.226
species=f	-53.605	-723.194	-3[illegible]1.915	-28[illegible]015	-2[illegible].015
dbh	118.559	91.536	48.393	24.661	16.973
crowndepth=2	-1.555	-1.555	-2.594	-2.594	-2.594
socialstatus=1,2	-5.775	-5.775	-9.633	-9.633	-9.633
socialstatus=2	6.797	6.797	11.338	11.338	11.338
vitality=1,2	4.357	4.424	5.107	5.107	5.107
n	52.734	338.736	1220.276	2150.943	2665.754

Results - regression trees

Virgin forest height model



$h = (d^2/(a+b_1 d+b_2 d^2)) + 1{,}3$ [cm]

Virgin f.	*LM1*	*LM2*	*LM3*	*LM4*	*LM5*
phase=3,2	4.932	4.932	8.226	8.226	8.226
species=f	-53.605	-723.194	-3[illegible]1.915	-28[illegible]015	-2[illegible].015
dbh	118.559	91.536	48.393	24.661	16.973
crowndepth=2	-1.555	-1.555	-2.594	-2.594	-2.594
socialstatus=1,2	-5.775	-5.775	-9.633	-9.633	-9.633
socialstatus=2	6.797	6.797	11.338	11.338	11.338
vitality=1,2	4.357	4.424	5.107	5.107	5.107
n	52.734	338.736	1220.276	2150.943	2665.754

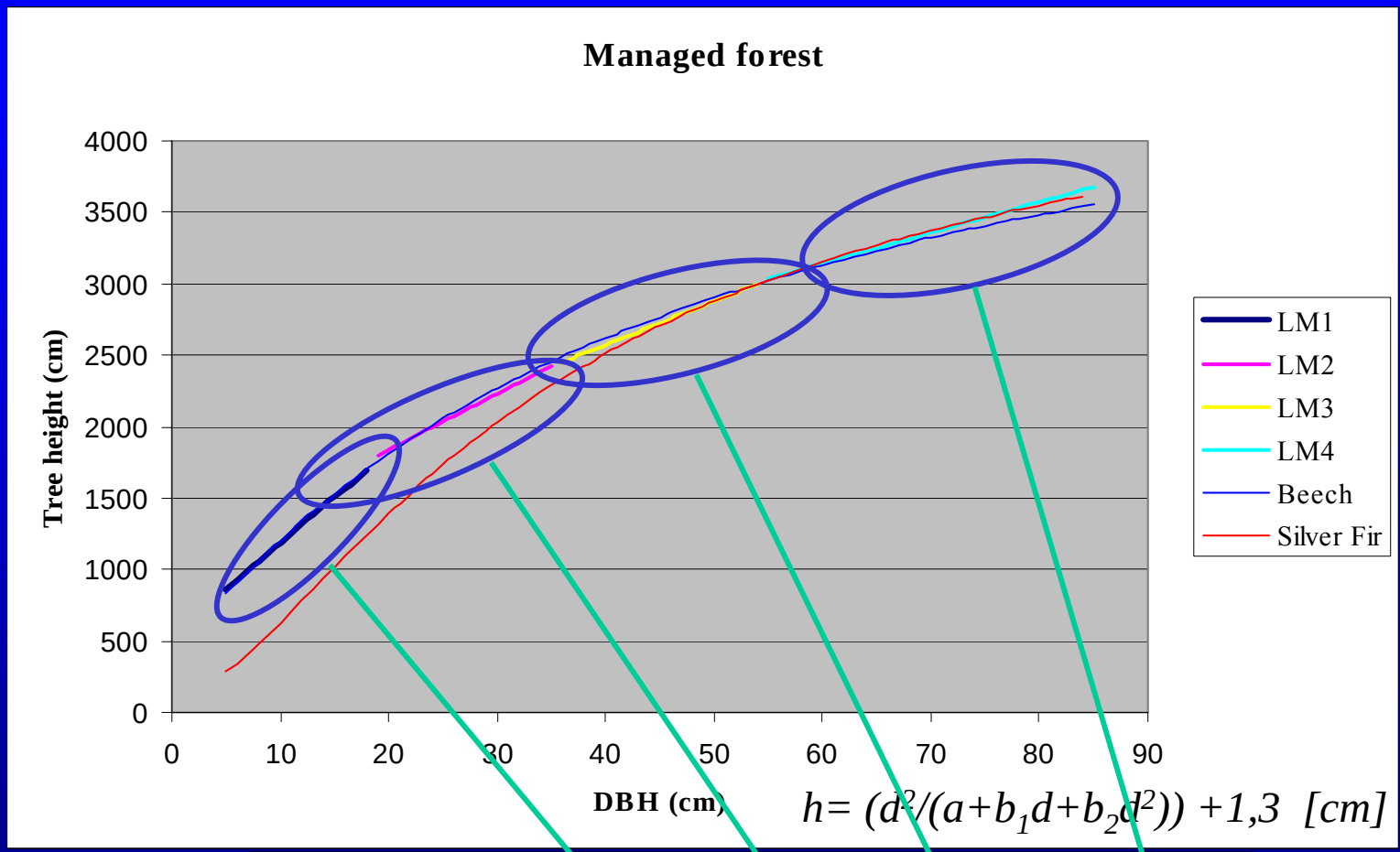
Results - regression trees
Managed forest height model

dbh <= 17.5 : LM1 (559/15.305%)
dbh > 17.5 :
| dbh <= 35.5 : LM2 (194/18.573%)
| dbh > 35.5 :
| | dbh <= 54.5 : LM3 (69/7.654%)
| | dbh > 54.5 : LM4 (57/4.948%)

Correlation coefficient	0.9923
Mean absolute error	41.7166
Root mean squared error	87.8429
Relative absolute error	7.3238 %
Root relative squared error	12.3599 %
Total Number of Instances	879

LM num: 1
height = 64.7499 * dbh + 4.9658 * layer=4,0,5 + 0.7611 * crowndepth=2 + 2.6777 * vitality=2,1 + 525.534
LM num: 2
height = 39.7414 * dbh + 77.4803 * layer=4,0,5 + 1.304 * crowndepth=2 + 4.588 * vitality=2,1 + 954.3475
LM num: 3
height = 30.2793 * dbh + 24.2647 * layer=4,0,5 + 1.304 * crowndepth=2 + 4.588 * vitality=2,1 + 1329.8797
LM num: 4
height = 21.4912 * dbh + 24.2647 * layer=4,0,5 + 1.304 * crowndepth=2 + 4.588 * vitality=2,1 + 1823.0698

Results - regression trees
Managed forest height model



$h = (d^2/(a+b_1 d+b_2 d^2)) + 1{,}3 \ [cm]$

Managed f.	*LM1*	*LM2*	*LM3*	*LM4*
dbh	64.7499	39.7414	30.2793	21.4912
layer=4,0,5	4.9658	77.4803	24.2647	24.2647
crowndepth=2	0.7611	1.304	1.304	1.304
vitality=2,1	2.6777	4.588	4.588	4.588
n	525.534	954.3475	1329.88	1823.07

Results – Regression trees - Conclusions

Explanations of the shape of the Prodan's tree height curve.

M5 didn't distinguish between beech and silver fir tree height curves.

General conclusions

Classification trees: identification of the most descriptive attributes of vertical structure and their hierarchic relationship.

Regression trees: identification of the attributes which have the strongest effect on the Prodan's tree-height regression curve.

Managed forest has unified vertical structure while virgin forest show distinguished differences among development phases.

Thank you!