Discussiones Mathematicae
Probability and Statistics 29 (2009) 155–168

A DOMINANT HEIGHT GROWTH MODEL FOR EUCALYPTUS PLANTATIONS IN PORTUGAL

Ayana Mateus

CMA, Departamento de Matemática,
Faculdade de Ciências e Tecnologia FCT
Universidade Nova de Lisboa, UNL
Quinta da Torre, 2829–516 Caparica, Portugal

e-mail: amf@fct.unl.pt ; +351 21 2948388

and

Margarida Tomé

CEF-Centro de Estudos florestais
ISA- Instituto Superior de Agronomia
UTLisboa-Universidade Técnica de Lisboa
Tapada da Ajuda, 1349–017 Lisboa, Portugal

e-mail: magatome@isa.utl.pt ; +351 21 3653100

Abstract

Eucalyptus globulus Labill is one of the most important economic forest species in Portugal, occupying an area of $875.10^3 ha$ in a total forest area of $3346.10^3 ha$ (Tomé *et al.,* 2007). The main goal of this study is to develop a dominant height growth model for Eucalyptus, applicable throughout the country, representing an improve of the curves that are part of the whole stand model existing in Portugal, the GLOBULUS model (Tomé *et al.,* 2001). The dominant height growth model will be built on a biological function formulated as a difference equation (Algebraic Difference Approach) to an all possible growth intervals data structure. The dynamic model proposed, obtains non biased and efficient estimates for the parameters. Comparing to the already available model, the last has the advantage of expressing the asymptote in weather variables functions, meaning that it is possible to reproduce the eucalyptus growth according to weather changes.

Keywords: algebraic difference equation, dynamic model, dominant height, forest planning, non linear model.

2000 Mathematics Subject Classification: Primary 62J02; Secondary 62P12.

1. Introduction

Eucalyptus globulus Labill is one of the most important economic forest species in Portugal, occupying an area of 875.10^3ha in a total forest area of 3346.10^3ha (Tomé *et al.,* 2007). It is a fast growing specie mainly used by the pulp industry; the trees are planted at final density-thinning and pruning practices are unusual at first rotation stands. The stands are intensively managed in a short rotation coppice system in which the first cycle of plantes seedlings (single stem) is followed by two or three coppiced stands, with an average cutting cycle of 10-12 years. In order to contribute to a balanced and resourceful management of Eucalyptus stand in Portugal, it is necessary to acquire models that simulate their growth under different environment conditions and treatments.

Estimating forest productivity is both necessary for effective forest management and useful for evaluating basic site conditions for ecological field studies. Site quality is therefore influenced by factors such as available light, heat, moisture and nutrients along with other soil characteristics such as soil depth and aeration (Wang and Klinka, 1996). Although it would be best to directly measure and predict these factors, some of them fluctuate widely over the course of a day, month or year, whereas others require precise measurements that may be difficult extrapolate across scales. Therefore, indirect methods for evaluating site quality are more frequently used in forest management (Monserud, 1984).

Site index, defined as dominant height at some fixed base age, is one of the most commonly used indicators of site productivity because there exists a close correlation between volume and dominant height growth, and it is generally accepted that height of dominant trees is only slightly affected by competition (Clutter *et al.,* 1983).

The main goal of this study is to develop a dominant height growth model for Eucalyptus, applicable throughout the country, representing an improve of the curves that are part of the whole stand model existing in Portugal, the GLOBULUS model (Tomé *et al.,* 2001). This model based on environmental factors would be seen as a boom for a site quality one. In particular, these curves should be totally independent on any global climatic classification and parameterized directly from local climate variables. A global climatic classification has the disadvantage of not considering each plot unique characteristics, which may lead to wrong conclusions. Nowadays, the development of geographical information systems facilities the access to local climatic variables used to parameterize the curves.

The dominant height variable is defined as the average height of the thickest trunk trees in the plot (keep in mind that the dominant height quantifies/qualifies the productivity of the field).

Many mathematical functions are available to model dominant height growth. These functions should satisfy some conditions. Based in biologically behavior, meaning have a universal application. In fact, these functions are obtained in the differential form, establishing an hypothesis over absolute or relative growth rates, obtaining an integration expression for production. In this way one can assign absolute parameter meanings for this functions. Consistent behavior with the biological growth, i.e. allowing null height at the beginning of production and maximum finite height at an advance stage (existence of a horizontal asymptote). Allow a sigmoidal growth (S curve): curve slope increases with increasing production at an early stage and should decrease in the final stage (existence of a inflection point). These requirements are achieved depending on both the construction method and the mathematical function used. Among the three general methods for site index curve construction, the algebraic difference approach (ADA), has the following advantages: short observations periods can be effectively used and the structure of equations is base-age invariant (Clutter *et al.,* 1983).

The algebraic difference approach formulates a function as a difference equation in order to express the dominant height at age $t_i(hdom_i)$ according to dominant height at age $t_{i-1}(hdom_{i-1})$ and measurement ages t_i and t_{i-1}, i.e.

$$hdom_i = f(hdom_{i-1}, t_{i-1}, t_i, \beta) + e_i$$

where, β is the parameters vector to estimate and e_i represents the stochastic part of model.

In this study dominant height growth model will be built on a biological function formulated as a difference equation (Algebraic Difference Approach). For the adjustment all observed pairs emerged from the available measurements will be used. The dynamic model proposed, obtains non biased and efficient estimates for the parameters. Comparing to the already available model, the last has the advantage of expressing the asymptote in weather variables functions, meaning that it is possible to reproduce the eucalyptus growth according to weather changes.

1. Data

The data used to model dominant height growth of eucalyptus plantations in Portugal is slightly different from those used in the existing model, GLOBULUS (Tomé *et al.,* 2001). In fact, several validations were performed in order to eliminate or correct data unreliable and using new observations available from recent experiments.

In a previous work, Portugal was divided in eight climatic regions concerning the productivity of eucalyptus (Ribeiro and Tomé, 2000). The data used has cover this regions, although have a weak representation about the relationship between site index in successive rotations (coppiced stands). Rotations represents the number of cuts that a plot has been subject. Also, measurements for ages above 10 years are little representative in the data. This problem is more pronounced in coppiced stands. This fact may lead to low precision of the asymptotes estimates, and consequently an underestimation of growth. We must take this in account in the fase of adjustment. Table 1 and 2 shows some characteristics of data used in first rotation and coppiced stands.

Table 1. Characteristics of plots used at first rotation.

	Region							
	1	2	3	4	5	6	7	8
N° of measurements	98	1538	8194	1111	11364	836	3600	628
Age (years)								
Minimum	1,0	0,6	0,8	2,8	1,1	2,6	2,8	3,0
Mean	5,4	5,6	7,8	6,5	6,7	7,2	6,5	5,4
Maximum	12,1	19,4	34,3	14,2	20,8	11,9	13,5	8,0
hdom (m)								
Minimum	2,6	2,8	1,5	4,8	2,1	4,0	3,5	3,0
Mean	16,8	13,7	16,6	14,4	12,3	12,0	11,7	8,2
Maximum	29,2	28,1	36,6	27,9	34,6	25,4	27,4	15,0
Density (ha)								
Minimum	901	350	200	421	200	250	200	200
Mean	2094	1297	1075	997	848	949	765	583
Maximum	4745	5000	3750	1850	2600	1875	1829	1575
Site Index								
Minimum	16,0	5,2	4,2	8,3	7,0	5,7	9,5	7,8
Mean	24,5	21,5	20,1	19,2	17,0	16,0	15,9	13,6
Maximum	29,6	28,5	35,9	28,1	32,0	25,9	27,3	21,2

Table 2. Characteristics of plots used at coppiced stands.

	Region							
	1	2	3	4	5	6	7	8
Nº of measurements	0	242	1029	131	2477	38	130	0
Age (years)								
Minimum		1,7	1,1	4,4	1,9	6,1	2,3	
Mean		8,0	7,5	7,8	6,8	7,8	5,8	
Maximum		14,9	19,1	11,7	17,1	10,4	10,4	
hdom (m)								
Minimum		4,8	4,2	9,0	5,1	5,5	5,8	
Mean		16,9	16,3	17,1	14,4	13,0	11,1	
Maximum		28,2	30,8	25,4	29,5	19,9	21,0	
Density (ha)								
Minimum		825	328	800	221	550	275	
Mean		2048	1594	1320	1374	1405	755	
Maximum		3563	4852	2225	4179	2550	1650	
Site Index								
Minimum		12,2	8,5	11,2	8,1	8,0	10,2	
Mean		19,8	19,6	19,0	17,9	15,2	14,5	
Maximum		29,3	29,1	25,1	29,0	20,5	22,5	

1.2. Selecting a growth function

In Table 3 we present the biological functions, formulated as an algebraic difference equation, that we will be used in our work. These functions were selected based on their accurate achievements in previous studies (Tomé *et al.,* 2001).

Table 3. Candidate functions formulated as an algebraic difference equation (y_i represents dominant height at age t_i).

Growth function	Analytical expression	
Lundqvist-Korf	$y_i = A\,(y_j/A)^{(t_j/t_i)^n}$	$A > 0,\ n > 0$
Richards-m	$y_i = A^{1-\ln\left(1-e^{-kt_i}\right)/\ln\left(1-e^{-kt_j}\right)}\, y_j^{\ln\left(1-e^{-kt_i}\right)/\ln\left(1-e^{-kt_j}\right)}$	$A > 0,\ k > 0$
McDill-Amateis	$y_i = A\left(1-\left(1-\frac{A}{y_j}\right)\left(\frac{t_j}{t_i}\right)^b\right)^{-1}$	$A > 0,\ b > 0$

1.3. Data Structure and model fitting

Modeling dominant height growth involves two processes:

- estimating height at base age ($hdom_p$), given height at some other age. The main objective is obtaining a site index prediction equation:

$$hdom_p = f(t_p, t_i, hdom_i)$$

- estimating height at some desired age ($hdom_i$) given height at base age. The main objective is obtaining is obtaining a height prediction equation, having a prior knowledge about site index

$$hdom_i = f(t_i, t_p, hdom_p).$$

These two processes may be modeled by individual functions for each process, or by one equation that predicts height at any desired age, given height at any other known age. When using individual functions, height is assumed to be measured without error when on the right-hand of equation but with error used on the left- hand side. This assumption causes a bias in the parameters of the curves; neither the height-prediction equation nor the site index prediction equation will have a shape that represents the true functional relationship between height and age across levels of site index.

When a unique function is used, the purpose is to simultaneously optimize the regression of Y on X and X on Y and avoid parameter bias. In this study, the model to dominant height growth will be built using a unique function, in the form of an algebraic difference equation (ADE) to an all possible growth intervals data structure. For example if plot height is measured at ages t_1, t_2 and t_3, we would have the following set of observations for the adjustment ($hdom_1$, t_1, $hdom_2$, t_2), ($hdom_1$, t_1, $hdom_3$, t_3), ($hdom_2$, t_2, $hdom_3$, t_3), ($hdom_2$, t_2, $hdom_1$, t_1), ($hdom_3$, t_3, $hdom_1$, t_1), ($hdom_3$, t_3, $hdom_2$, t_2). With the all possible growth intervals, there is no independence between observations. In order to avoid this problem we used an auto-regressive structure to errors, such as:

$$hdom_{i,j} = f(hdom_j, t_i, t_j, \beta) + e_{i,j} \;\; com$$

$$e_{i,j} = \rho e_{i-1,j} + \gamma e_{i,j-1} + \epsilon_{i,j} \;\; |\rho| < 1, \; |\gamma| < 1$$

where $hdom_{i,j}$ represents the obtained value for dominant height at age t_i using as explanatory variables dominant height at age t_j and also measurements age t_i and t_j. β is the parameter vector to be estimate and $e_{i,j}$ is the correspondent error term. The parameter ρ takes into account the correlation between the current residual and the residual associated to estimating dominant height at age t_{i-1} using dominant height at age t_j as a predictor variable. Finally, the parameter γ refers to the autocorrelation between the current residual and what you get from estimating dominant height at age t_i using the dominant height at age t_{j-1} as predictor variable. $\epsilon_{i,j}$ are independent and identically distributed variables.

Models like this are usually referred as dynamic models and its main objective is getting unbiased and efficient estimates to the parameters model (Gallant, 1987) (p. 405–426).

By using all possible differences, the number of observations is artificially inflated, although no additional information is obtained. As a consequence a short standard deviation is obtained and needs to be corrected. This can be achieved multiplying the obtained deviation by the factor $\sqrt{N(apd/N(fd))}$, where $N(apd)$ is the number of observations using all possible differences and $N(fd)$ is the number of observations using only the first differences. If k represents the number of trees and n_i represents the number of measurements on tree i, then

$$N(apd) = \sum_{i=1}^{k}(n_i(n_i-1)) \text{ and } N(fd) = \sum_{i=1}^{k} n_i.$$

We deal in this way because if a data set is replicated, for instance each observation reappears four times, the mean standard deviation will have a factor decrease of $\sqrt{1N/4N}$, where N is the original number of observations.

The fitting were carried using the generalized least square method in non linear equations (Generalised Least Squares, GLS) (Seber and Wild, 1989). The PROC MODEL procedure on the SAS/STAT software (SAS, 2005) was used.

1.4. Model selection criteria

A three-step procedure is used to evaluate and select the most appropriate model, which include qualitative and quantitative examinations.

The first step evaluates the model fitting statistics based on mean sum square residual and coefficient of determination, represented respectively by:

$$MSSR = \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{n} \quad \text{and} \quad R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \overline{Y})^2}$$

where $\widehat{y_i}$ represents the predict value and $\overline{Y}$ mean of observations.

The model validation proceeds in the second step. An independent data set testing was not available, so the the validation criteria were: in plots with measurements above six, the second and penultimate were removed for validation. The plots that met this criterion were in number of 178. From the second observation the penultimate observation was predicted, and both mean predicted residual (r) and mean absolute predicted residual ($|r|$) were computed. The same procedure was used to predict the second observation based on the penultimate. To evaluate the dispersion of residuals percentiles 5% and 95% were calculated.

The selection was also based on the analysis of parameters obtained with the expected biological behavior, including: Signs and values of the coefficients in the model components, especially the asymptotes and height curve development.

2. Results

An iterative process always requires initial values for the parameters in order to avoid local solutions and to quickly achieve convergence.

The initial parameters related with growth used in the three selected functions, were those obtained in the GLOBULUS model (Tomé *et al.,* 2001). A regression like,

$$e_{i,j} = \rho e_{i-1,j} + \gamma e_{i,j-1} \quad |\rho| < 1,\ |\gamma| < 1$$

was used in order to obtain initial information about the parameters, ρ and γ, related with the stochastic part of the model.

The results achieved by each of the stared functions were the following ones:

(1) Function Lundqvist-Korf MSSR= 1.369, R^2= 0.9606;

(2) Function McDill-Amateis MSSR= 1.412, R^2= 0.9595;

(3) Function Richards-m MSSR= 1.497, R^2= 0.9569.

Table 4 presents some validation statistics on selected models, namely characterization of the error (bias and precision). By the same table, the use of Richards-m leads to a less precise model. Taking in to account this together with the fact that this function provides a not suitable curve development, we decided to not consider this function.

Table 4. Evaluation of the predictive ability of models.

Growth function	Projection order	Mean r	Abs. mean $\|r\|$	Percentile 5%	Percentile 95%
Lundqvist-Korf	$t_i < t_j$	0,1005	0,3590	-0,5725	1,0428
	$t_j < t_i$	-0,0767	0,2933	-0,8524	0,4926
McDill-Amateis	$t_i < t_j$	0,0909	0,3520	-0,6121	0,9861
	$t_j < t_i$	-0,0741	0,3005	-0,8208	0,5069
Richards-m	$t_i < t_j$	0,0982	0,5053	-0,7801	1,3354
	$t_j < t_i$	-0,0824	0,4791	-1,2347	0,7278

As dominant growth of a certain specie is related with it's location, we tried to relate the function parameters with the specifies of the region, in order to obtain a model reflecting how geography influences growth.

There is a biological relation between temperature (TM) and vegetation growth though it is not linear as there is an exact temperature associated to an optimal growth, and above this level the growth decreases substantially. Below we present the mathematical expression utilized: $a + bTM + cTM^2$.

In order to get a proper biological adjustment, coefficient b has to be positive and c negative, with the maximum curve slope around the maximum temperature supported by the tree. The same criteria was applied to the insolation (INS) and radiation (RAD).

The remain variables, such as, frost, evapotranspiration, precipitation and humidity were incorporated linearly. All the coefficients should be positive regarding evapotranspiration, precipitation and humidity variables. Regarding frost it is expected a negative coefficient.

In order to avoid over-parametrization problems, we did not introduce both radiation and insolation as well as temperature and humidity, because this variables have high correlation. So the variables were picked up according to their sign compatibility with the biological growth and the best adjustment.

Once we have data from stands in first rotation and coppice, we decided to study the influence of rotation in the dominant height growth, expressing the rotation as a dummy variable (1 if coppice).

Basically parameter A, related with asymptote value, was estimated in function of climatic variables. The effect of rotation (Rot) has only been tested, in a linear way, at growth rate parameter (n, b and k) because the asymptote expressing the potential of a species is the same in successive rotations. Table 5 and 6 shows the main results obtained.

In Portugal the conditions that most affect Eucalyptus plantations productivity are related to water and frost. Conclusions taken after intense analysis to the country climatic conditions state there's a negative correlation between precipitation, temperature and radiation. So, the effects of precipitation are quantified indirectly in the estimates of the coefficients $a2$ and $a4$ used in the models (see Table 5). Inferior asymptote values will represent low levels showers associated to hight temperatures and radiations.

Growth rate parameter, estimated in function of rotation, always came significant with a negative signal compatible with the biological growth in coppiced stands. That is, in coppice stands the shape of growth curve is different from stands in first rotation. Initially, there is a faster growth, consequence of the root system already be installed, becoming slower then.

Table 5. Results from estimating parameters in function of climatic variables.

Growth function	Parameters estimates	Standard error (correct)	Statistics t *
Lundqvist-Korf			
	$n1$=-0,081	0,003	-27,99
$A = a0 + a1\,TM + a2\,TM^2$	$a0$=-1978,11	492,721	-4,01
$+a3\,RAD + a4\,RAD^2$	$a1$=32,389	7,231	4,48
	$a2$=-1,134	0,247	-4,59
	$a3$=24,067	6,603	3,65
	$a4$=-0,0796	0,022	-3,56
	ρ=0,287	0,007	44,23
	γ=0,621	0,007	88,79
MSE=1,3301			
McDill-Amateis			
	$b1$=-0,233	0,008	-30,16
$A = a0 + a1\,TM + a2\,TM^2$	$a0$= -971,874	212,173	-4,58
$+a3\,RAD + a4\,RAD^2$	$a1$=17,859	3,020	5,91
	$a2$=-0,604	0,104	-5,82
	$a3$=11,809	2,856	4,13
	$a4$=-0,039	0,010	-4,09
	ρ=0,325	0,006	51,13
	γ=0,606	0,007	87,90
MSE=1,3696			

* p-value $< 0,0001$

Table 6. Evaluation of the predictive ability of models expressed in function of climatic variables.

Growth function	Projection order	Mean r	Mean $\|r\|$	Percentile 5%	Percentile 95%
Lundqvist-Korf	$t_i < t_j$	0,0997	0,3530	-0,5426	0,9888
	$t_j < t_i$	-0,0765	0,2941	-0,8006	0,4997
McDill-Amateis	$t_i < t_j$	0,0849	0,3437	-0,5829	0,9183
	$t_j < t_i$	-0,0676	0,2943	-0,7281	0,4961

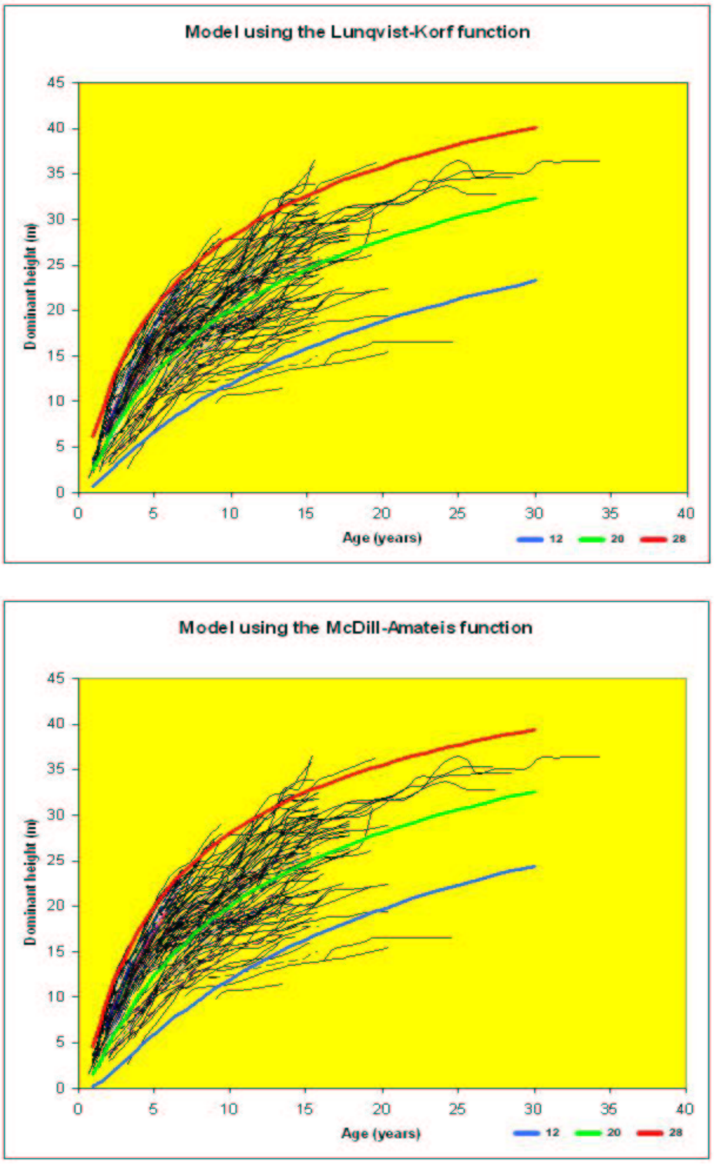

Figure 1. Graphs obtained with different models parameterize in function of climatic variables.

Figure 1 above represent the growth obtained with different models parameterize in function of climatic variables. The three main bold curves represent the growth quality according to the season. Observing the results above we immediately notice the Lundqvist-Korf function has a far better asymptotic behavior regarding old ages plots than the McDill-Amateis one. We can also notice the tendency of each curve to converge at latter ages regardless the climatic conditions (McDill-Amateis function). To finalize we choose the Lundqvist-Korf function based of the facts stated above plus it's adjustment and predicting capacities.

Acknowledgements

This work was supported by the FP6 EFORWOOD IP project (contract 518128) and the FCT project CarbWoodCork (POCI/AGR/57279/2004). The authors also thanks Celbiana Silvicaima pulp company that provide the data used in this study.

References

[1] M. Tomé, S. Barreiro, A. Cortiçada, A. Meyer, T. Ramos and P. Malico, Inventário florestal 2005-2006. Áreas, volumes e biomassas dos povoamentos florestais. Resultados Nacionais e por Nut´s II e III. Publicações GIMREF RT 5/2007 Universidade Técnica de Lisboa. Instituto Superior de Agronomia, Centro de Estudos Florestais, Lisboa, Portugal 2007.

[2] M. Tomé, F. Ribeiro and P. Soares, O modelo GLOBULUS 2.1-Relatórios técnico científicos do GIMREF n° 1/2001. Universidade Técnica de Lisboa, Instituto Superior de Agronomia, Centro de Estudos Florestais, Lisboa, Portugal 2001.

[3] G.G. Wang, K. Klinka and W. John, *Use of synoptic variables in predicting white spruce site index*, For. Ecol. Manage **80** (1996), 95–105.

[4] R.A. Monserud, *Height growth and site index curves for inland Douglas-firbased on stem analysis data and forest habitat type*, For. Sci. **30** (1984), 943–965.

[5] L. Clutter, J. Fortson, L. Pienaar, G. Brister and L. Bailey, *Timber Management: A Quantitative approach*, John Wiley and Sons 1983.

[6] F. Ribeiro and M. Tomé, *Classificação climática de Portugal continental, baseada em informação do atlas do ambiente*, Revista de Ciẹncias Agrárias 2000.

[7] A.R. Gallant, *Nonlinear Statistical Models*, John Wiley & Sons (1987), 624p.

[8] The SAS System for windows, 9.1 edition.

[9] G.A.F. Seber and C.G. Wild, *Nonlinear Regression*, New York, John Wiley 1989.

Received 17 September 2009