

Randomized branch sampling to estimate fruit production in Pecan trees cv. 'Barton'

Filemom Manoel Mokochinski^{1*} Luciano Farinha Watzlawick¹
Renato Vasconcelos Botelho¹ Vanessa Moreira¹

¹Departamento de Agronomia, Universidade Estadual do Centro-Oeste (UNICENTRO), 85040-080, Guarapuava, PR, Brasil. E-mail: filemom_mom@hotmail.com. *Corresponding author.

ABSTRACT: Sampling techniques to quantify the production of fruits are still very scarce and create a gap in crop development research. This study was conducted in a rural property in the county of Cachoeira do Sul - RS to estimate the efficiency of randomized branch sampling (RBS) in quantifying the production of pecan fruit at three different ages (5, 7 and 10 years). Two selection techniques were tested: the probability proportional to the diameter (PPD) and the uniform probability (UP) techniques, which were performed on nine trees, three from each age and randomly chosen. The RBS underestimated fruit production for all ages, and its main drawback was the high sampling error (125.17% - PPD and 111.04% - UP). The UP was regarded as more efficient than the PPD, though both techniques estimated similar production and similar experimental errors. In conclusion, we reported that branch sampling was inaccurate for this case study, requiring new studies to produce estimates with smaller sampling error.

Key words: *Carya illinoensis*, mathematical model, crop production, post-harvest.

Amostragem aleatória de ramos para estimativa de produção de frutos de Nogueira-pecã (cv. 'Barton')

RESUMO: Técnicas de amostragem para quantificar a produção de frutos ainda são muito raras e geram uma lacuna no desenvolvimento das culturas. O presente estudo foi realizado em uma propriedade rural no município de Cachoeira do Sul - RS, com o objetivo de verificar a eficiência da amostragem aleatória de ramos (AAR) para estimar a produção de frutos de nogueira-pecã com três diferentes idades (5, 7 e 10 anos). Foram testadas duas técnicas de seleção: probabilidade proporcional ao diâmetro (PPD) e probabilidade uniforme (PU), aplicadas em nove árvores, sendo três de cada idade escolhidas através de sorteio. A AAR subestimou a produção de frutos nas idades avaliadas, sendo sua principal desvantagem o alto erro amostral (125,17% - PPD e 111,04% - PU). A PU foi considerada mais eficiente que a PPD nas árvores amostradas, mas com estimativas de produção e erros experimentais semelhantes. Conclui-se que a amostragem em ramos foi imprecisa para o caso estudado, sendo necessária a realização de novos estudos para geração de estimativas com menor erro de amostragem.

Palavras-chave: *Carya illinoensis*, modelo matemático, produção vegetal, pós-colheita.

INTRODUCTION

In recent decades, there has been an increased demand for healthier foods that have the nutritional elements recommended for health maintenance, such as nuts. The pecan tree (*Carya illinoensis* (Wangenh.) K. Koch) provides fruits that are considered an important raw material for agro-industry. Its almonds can be used in the food industry and for the production of oil. Shells can be used to obtain tea and fertilizers, and its wood can be used for furniture. The species also enables a consortium with other cultures and the integration of livestock.

Despite these benefits, the pecan crop has not been encouraged, since little research has been conducted

on its implementation and management (DUARTE & ORTIZ, 2001). Studies that seek to assist producers in pecan orchard planning should be prioritized.

Methods that quantify the productive potential of orchards, based on reliable sampling techniques, are practically nonexistent. Sampling methods may provide an alternative to the census by estimating the attributes of interest in individual trees. This may prove to be an extremely useful tool, allowing producers to be informed on whether the chosen species will have good productive output.

JESSEN (1955) developed the technique of randomized branch sampling (RBS) to estimate the attributes of interest in individual trees. The technique is based on the selection of branches in

the tree and randomly assigning probabilities to that selection of branches. In the uniform probability technique (UP), all sampled branches have the same probability, independent of its diameter, whereas in the probability proportional to the diameter technique (PPD), the branch diameter is taken into account in the selection of the final branch.

In the RBS technique, a single tree is regarded as the population, and the target being measured is the total amount of the quantified attributes in the sampled branches. Branch sampling has been used to assess the amount of fruits, seeds, insect attack and for the quantification of volume, biomass and carbon stocks (BORGES, 2009; RYALL et al., 2011; BARBEIRO, 2012; LÔBO, 2012).

So, the goal of this study was to determine the efficiency of probability sampling of branches to estimate pecan production at three different ages by comparing techniques of uniform probability and probability proportional to the diameter, with the overall aim to support farmers in the assessments of their orchards.

MATERIALS AND METHODS

The study area is located in the county of Cachoeira do Sul - RS, Jacuí Valley region, with the geographical coordinates of 30°20'16" and 30°20'25" South and 52°00'23" and 52°03'27" West, at an average altitude of 60 meters. According to the climatic classification of KÖPPEN (1948), the region has a Subtropical Climate (CFA), whose main characteristics include winters with the occurrence of severe and frequent frosts, an average annual temperature of 19°C and an annual rainfall of approximately 1500mm year with no dry season.

Pecan plantations were of three different ages (5,7 and 10 years) and had been cultivated under the same management practices since their implementation, with a spacing of 7x7 meters. To select the trees, three individuals from each age were randomly selected. Each sampled tree was regarded as a population, which was evaluated using the techniques uniform probability (UP) and probability proportional to the diameter (PPD). Fruits from the trees were the estimated attributes.

Branch sampling technique uses the natural offshoot from the inside of the crown to compose its sample, requiring prior establishment of branches, nodes, segments of branches and paths. Branches form an entire system that has its source in a single bud, lateral or terminal, and comprises a subpopulation within the tree. The node is the point

on the tree where a branch splits into two or more (bifurcation). Segments are the part of a branch between two consecutive nodes, and the path is a series of segments between the basal segment and the terminal bud. Number of possible paths in a tree is equal to the number of terminal buds.

The first segment of the path extends from the base of the stem, defined as the first node, to the second node. Most of the time, the first node of the path leads to one branch (in this case, the stem), where the probability of selection will be equal to one ($q_1 = 1$). In the second node, the probability of selection will be assigned to each branch, randomly choosing one. The choice for the second branch, with the probability of selection q_2 , defines the second segment of the path. This procedure is repeated until a lower branch or an end segment is selected at the final node, with probability q_{ri} .

To assign probabilities proportional to the diameter for each branch, the following formula was used, adapted from GREGOIRE & VALENTINE (2007). The diameter of the branches was taken near the node and branches were measured with a digital caliper:

$$q_{ri} = \frac{d^2_{ri}}{\sum_{i=1}^n d^2_{ri}}, i = 1, 2, \dots, n$$

where:

q_{ri} = probability of selection assigned to the i -th branch from the r -th node;

d^2_{ri} = squared diameter (d^2) of the i -th branch from the r -th node;

$\sum_{i=1}^n d^2_{ri}$ = sum of the squared diameters of the i -th branch from the r -th node.

Thus, the branch is selected if $u \leq q_{ri}$, where u is a random number between 0 and 1 that will determine which branches at a particular node will be chosen. For the selection of branches, u values with three decimal places were determined, and the branches were randomly assigned a number between 0 and 1 (e.g., $u = 0.123$). The draw of the three decimal places was done with 10 cubes, numbered from 0 to 9. When the raffled u value was less than q_{ri} , the branch was chosen.

To assign a uniform probability (UP) for selecting branches at each node of the path, the following formula was used:

$$q_{ri} = \frac{1}{n}$$

where:

n = number of branches from the r -th node.

To estimate the number of fruits on each tree, the number of fruits were counted in the segments

of five paths on the selected terminal branch, and the calculation of unconditional probability of selection (Q_{mi}) was performed. These calculations allowed the estimation of the total amount of fruit on each tree through the sum of all sampled segments using the following formula:

$$\bar{y}_{Qm} = \frac{Y_{mi}}{Q_{mi}}$$

where:

\bar{y}_{Qm} = estimate of the number of fruits of the tree given by the path m ;

Y_{mi} = number of fruits from the branch i in the path m ;

Q_{mi} = unconditional probability of the selection of the branch i in the path m .

After obtaining an estimate of the number of fruits of m paths ($m \geq 2$), these estimates were combined to estimate an average number of fruits per tree (\bar{y}_Q), which is given by:

$$\bar{y}_Q = \frac{1}{m} \sum_{j=1}^m \bar{y}_{Qm}$$

The joint probability of the selection of all segments S forming the m -th path is represented by Q_{mi} . GREGOIRE & VALENTINE (2007) stated that, though the number of segments (S) can vary between paths, if there are M possible paths with different terminal segments, the sum of the unconditional selection probabilities of all possible branches in a tree will equal one. Thus, we assumed that Q_{mi} is the probability of obtaining the estimate (\bar{y}_{Qm}), which represents one of M possible estimates of fruits for the sampled tree.

For all sampled trees, the coefficient of variation (cv_{y1} - UP and cv_{y2} - PPD) and the relative sampling error ($Er1$ - UP and $Er2$ - PPD) were calculated:

$$cv_y = \frac{s_y}{\bar{y}_Q} \times 100$$

where:

cv_y = estimate of the coefficient of variation;

\bar{y}_Q = average number of fruits;

s_y = estimate of the standard deviation.

$$r(\%) = \left(\frac{s_{\bar{y}_Q}}{\bar{y}_Q} \right) \times 100$$

where:

$Er(\%)$ = relative sampling error.

RBS efficiency was assessed by comparing the methods of assigning probabilities (PPD and UP) in different sampling intensities, taking into account the time spent by path and the coefficients of variation (cv). To determine the

time spent on measurements for each path, a timer was used, starting at the moment the first node was measured to the quantification of the fruits in the terminal branch.

$$ER(\%) = \left(\frac{cv_{y1}}{cv_{y2}} \times \frac{t_1}{t_2} \right) \times 100$$

where:

$ER(\%)$ = relative efficiency

RESULTS AND DISCUSSION

Estimates of pecan fruit production using the PPD and UP techniques regarded the tree as a population of possible paths, which resulted in five production estimates for each technique. In all sampled paths, both techniques underestimated fruit production averages for trees (Table 1).

The number of fruits reported in the sampled terminal branches averaged 3.5 fruits per tree at 5 years, 18.6 fruits at 7 years and 19.2 fruits at 10 years. There was a 81.2% increase in the number of fruits as the age increased from 5 to 7 years and a 3.12% increase from 7 to 10 years. The small difference in the average of fruits on the sampled branches at 7 and 10 years demonstrates the stabilization trend in the number of fruits in the terminal branches as trees age.

The time required to perform the PPD was, on average, five-fold higher than the UP. The time difference for carrying out the PPD is related to the difficulty in measuring the diameters. According to WOLF (2009), the specific characteristics of growth and the arrangement of branches and crown affect the sampling time due to trees irregularity.

In some paths, few fruits were reported in the terminal branch, resulting in an underestimation of the real fruit production. Terminal branches without any fruits were also selected, which contributed to the overall underestimation. A similar situation was reported by BORGES (2009), who quantified fruit production of pequi and reported that branches with no fruit on fruiting trees resulted in the underestimation of mean production.

Trees at 10 years had the highest production, at approximately 1,559.30 units. Special attention was given to some of the terminal branches that presented the same number of fruits, but had different estimates because of the probability of selection (Q_{mi}). Coefficients of variation (cv) for both sampling techniques showed oscillation in different paths. For trees at 7 years using the UP, and at 10 years using the PPD, the coefficient of variation decreased by increasing the sample size.

Table 1 - Mean variables of random sampling of branches performed on pecan trees.

	M	Fruits/ branch	Q_{mi}	$\bar{y}Q_m$	$\bar{y}Q$	Time (s)	cv_y (%)	Er (%)	Real Mean
5 years	PPD	1	0.0263	38	~	378.9	~	~	96.5 fruits
		2	0.0211	94.5	66.2	294.9	61.02	548.10	
		3	0.0813	86	72.8	195.6	22.54	57.11	
		4	0.0707	56.5	68.7	247.8	15.88	25.24	
		5	0.0645	31	61.2	318.3	33.40	41.43	
	Mean	3.2	61.2	287.1	33.21	167.97			
	UP	1	0.0509	78.5	~	55.5	~	~	
		2	0.0740	27	52.7	49.5	40.11	360.24	
		3	0.0344	87	64.1	53	27.93	70.98	
		4	0.0457	87.5	70	112	28.50	45.29	
5		0.0550	109	77.8	70.5	15.20	18.99		
Mean	3,8	77.8	68.1	27.93	123.87				
7 years	PPD	1	0.0209	1048.33	~	499.6	~	~	1118.5 fruits
		2	0.0379	342.67	695.5	432.6	18.82	169.12	
		3	0.0195	715.33	702.11	551	21.39	54.47	
		4	0.0311	706.67	703.25	492.8	16.79	26.58	
		5	0.0657	289	620.4	400.6	24.90	30.91	
	Means	18	620.4	475.3	20.48	70.28			
	UP	1	0.0359	666.67	~	91.67	~	~	
		2	0.0639	406.67	536.67	58.33	46.45	417.27	
		3	0.0312	447.33	506.89	76	36.78	65.16	
		4	0.0355	506.33	506.75	62.67	32.07	35	
5		0.0445	314	468.2	69.67	8.54	9.59		
Means	19.2	468,2	71.67	30.96	131.75				
10 years	PPD	1	0.0172	868.67	~	285.6	~	~	1559.3 fruits
		2	0.0185	1237.67	1053.17	318	45.03	404.57	
		3	0.0294	679.33	928.55	427	37.70	95.81	
		4	0.0380	605	847.66	235	30.45	24.06	
		5	0.0241	745.67	827.26	361.2	26.93	33.43	
	Means	19.8	827.27	325.36	35.03	139.47			
	UP	1	0.0322	589	~	67	~	~	
		2	0.0233	941	765	69	18.35	164.84	
		3	0.0353	396.33	642.1	63	21.30	58.61	
		4	0.0341	526.33	613.15	69,3	53.39	80.60	
5		0.0308	648.33	620.18	56,3	4.89	6.07		
Means	18.6	620.20	64,93	24.48	77.53				

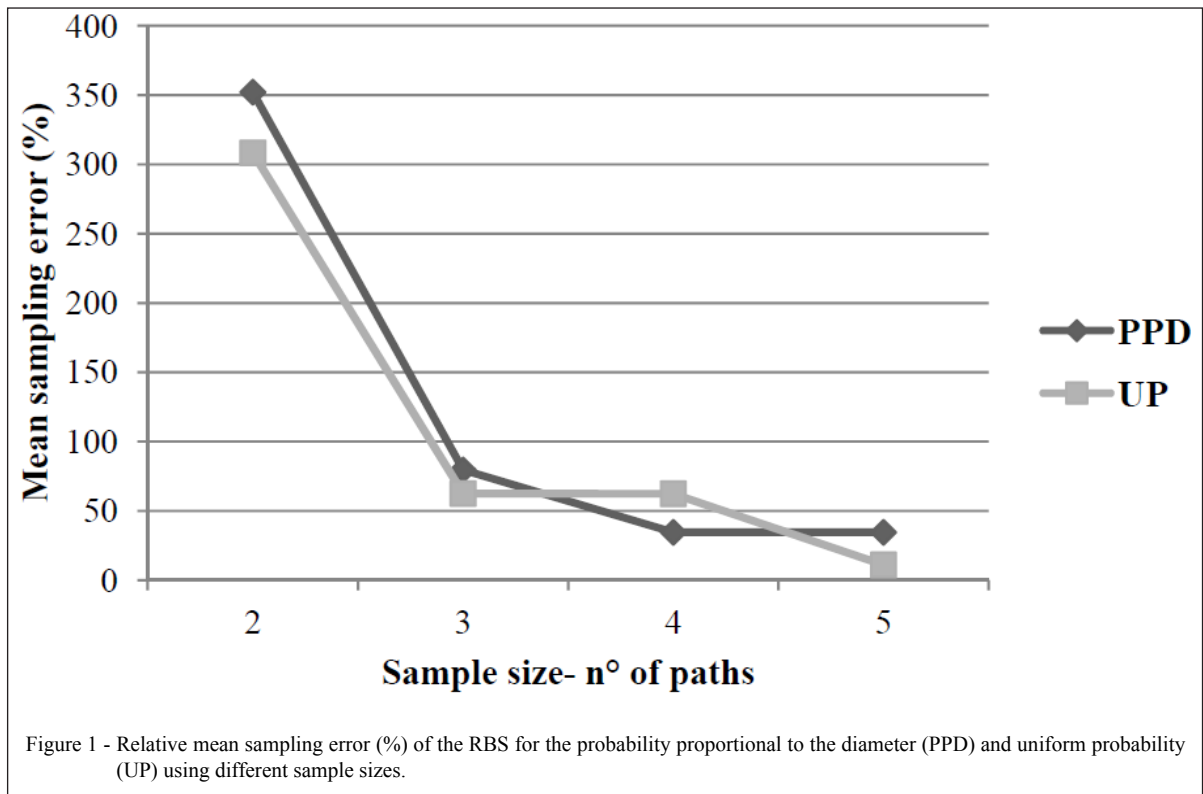
UP = uniform probability; PPD = probability proportional to the diameter; m = number of paths taken or sampling intensity; Fruits/branch = Number of fruits in the terminal branch; cv_y = estimate of the coefficient of variation; $\bar{y}Q$ = average number of fruits; Er = Relative sample error; $\bar{y}Q_m$ = estimate of the number of fruits of the tree given by the path m ; Q_{mi} = unconditional probability of the selection of the branch i in the path m .

Analyzing real fruit production using the census, trees at 7 years produced approximately twelve-fold more fruit than trees at 5 years, indicating that the beginning of pecan production takes 4 or 5 years, and at 7 years, production reaches commercial levels.

By analyzing only the sampling errors produced by the estimates from RBS, we can observe that the values were high. We noticed that the sampling error (%) decreased with an increase in the sample size. Regardless of the technique used (PPD or UP), the more paths sampled, the smaller the errors produced by the RBS (Figure 1). Similar results were reported by other authors, who suggested increasing the sample size to reduce the variance and errors (GOOD et al., 2001; CANCINO, 2005; BORGES, 2009).

The mean sampling error was 125.17% for the PPD and 111.04% for the UP. However, when the UP was used on trees at 7 and 10 years, assuming a sample size of 5 paths, the sampling errors were less than 10%. That is, in larger trees taking five paths, the sampling errors are smaller.

The largest errors were reported in a sample size of two paths (352.15% and 308.35% for the PPD and UP, respectively) and minor errors were reported with five paths (34.48% and 10.99% for the PPD and PU, respectively). The largest error reduction was gained by increasing sample size from two to three paths. BORGES (2009) reported errors ranging from 44.65% to 1,270.60% using a sample size of 10 and does not recommend using the techniques to quantify pecan fruits, due to high error values.



CANCINO & SABOROWSKI (2007) stated that the RBS provided estimates with minor errors when used in trees with regular crowns, assuming that the possible paths are equal in size. According to these authors, trees with irregular crowns and different sizes may present larger errors as a result of the probability of selecting branches between the paths.

LAMIEN et al. (2007), analyzing the production of Shea (*Vitellaria paradoxa*) in Africa, reported that fruit production occurs unevenly on trees with cross-pollination and that such differences were the result of the wind. The pecan tree also showed differences in the number of fruits in the terminal branches, contributing to the increase of errors in sampling.

WILLIAM (1989), estimating biomass of *Pinus* sp. with the branch sampling technique, described that the error tends to decrease with increasing sample size. BORGES (2009), estimating the production of pequi fruit with RBS using 10 paths, reported that the two sampling techniques (PPD and UP) underestimated the real fruit production, which occurred due to the great variability of fruit distribution inside the crown.

The relative efficiency resulted in mean values above 100% (588.5% - 5 years; 641.7% - 7 years;

1,120.7% - 10 years), indicating that the UP was more efficient than the PPD. Range inefficiencies of variation (cv) and the time differences became key factors in analyzing the efficiency of the techniques. For the PPD, the difficulty of measuring the diameter near the node, makes the technique less efficient.

The decision to determine which sampling technique is more efficient for quantifying an attribute of interest must take into account the time spent on sampling and the coefficients of variation. The relative efficiency is not necessarily an indication that the estimate is close to the real value obtained from the census. The value of the relative efficiency is the basis for choosing which technique is more efficient.

For trees at 5 and 7 years, the relative efficiency values were similar. The highest value was reported for trees at 10 years. Due to the increased number of nodes in larger trees, there is an increase in the sampling time, resulting in an increase in the relative efficiency.

Since time is a factor that affects the relative efficiency, when repeated paths are sampled again, the time is reduced. The sampling time for pecan trees was important in the efficiency of the UP technique because the measurement of diameters

was not required, making this technique faster and affecting its relative efficiency.

CONCLUSION

Random sampling of branches did not prove efficient enough to estimate the production of pecan fruit in the evaluated ages (5, 7 and 10 years). The two evaluated techniques (uniform probability and probability proportional to the diameter) showed similar estimates of production and similar experimental errors.

ACKNOWLEDGEMENTS

The authors thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for financial support.

REFERENCES

- BARBEIRO, L. S. **Sistema de amostragem para quantificar a produção de sementes de *Bertholletia excelsa* h. b. k (Castanha do Brasil) na região de Oriximiná - PA.** 2012. 130f. Dissertação (Mestrado em Ciências Florestais) - Universidade Federal do Paraná, Curitiba, PR.
- BORGES, L. M. **Amostragem aleatória de ramos como técnica para quantificar a produção de frutos de *Caryocar brasiliense* camb. (*caryocaraceae*).** 2009. 168f. Dissertação (Mestrado em Ciências Florestais) - Universidade de Brasília, DF.
- CANCINO, J. Determinación del tamaño de la muestra em el muestreo RBS com selección sin reposición em la primera etapa. **Bosque**, v.26, n.1, p.65-75, 2005. Available from: <http://www.scielo.cl/scielo.php?pid=S0717-92002005000100007&script=sci_arttext>. Accessed: Apr. 14, 2015. doi: 10.4067/S0717-92002005000100007.
- CANCINO, J.; SABOROWSKI, J. Comparison of randomized branch sampling with and without replacement at the first stage. **Silva Fennica**, v.39, n.2, p.320-333, 2007. Available from: <<http://www.silvafennica.fi/article/384/keyword/randomized+branch+Sampling>>. Accessed: Apr. 16, 2015. doi:10.14214/sf.384.
- DUARTE, V.; ORTIZ, E. R. N. **Podridão de *Phytophthora da amêndoa e casca da noqueira-pecã no Brasil*.** Campinas: Livraria Rural, 2001. p.493-508.
- GREGOIRE, T. G.; VALENTINE, H. T. **Sampling strategies for natural resources and the environment.** New York: Chapman Hall, 2007. p.474.
- GOOD, M. et al. Estimating tree component biomass using variable probability sampling methods. **Journal of Agricultural**, v.6, p.258-267, 2001. Available from: <<http://link.springer.com/article/10.1198/108571101750524599#page-1>>. Accessed: Apr. 13, 2015. doi: 10.1198/108571101750524599.
- JESSEN, R. J. Determining the fruit count on a tree by randomized branch sampling. **Biometrics**, v.11, n.1, p.99-109, 1955. Available from: <http://www.jstor.org/stable/3001484?seq=1#page_scan_tab_contents>. Accessed: Apr. 14, 2015. doi: 10.2307/3001484.
- KÖPPEN, W. **Climatologia: com um estudio de los climas de La tierra.** México: Fondo de Cultura Económica, 1948. 478p.
- LAMIEN, N. et al. **Variations in dendrometric and fruiting characters of *Vitellaria paradoxa* populations and multivariate models for estimation of fruit yield.** Agroforestry System, v.69, p.1-11, 2007. Available from: <<http://link.springer.com/article/10.1007/s10457-006-9013-x>>. Accessed: Apr. 15, 2015. doi: 10.1007/s10457-006-9013-x.
- LÔBO, R. N. **Método de Amostragem Aleatória de Ramos (AAR) como técnica para estimar volume, biomassa e estoque em cerrado sensu-stricto.** 2009. 80f. Monografia (Graduação em Engenharia Florestal) - Universidade de Brasília, Brasília, DF.
- LÔBO, R. N. **Amostragem aleatória de ramos como método para estimar os estoques em volume, biomassa e carbono em Floresta Estacional no estado de Goiás.** 2012. 127f. Dissertação (Mestrado em Ciências Florestais) - Universidade de Brasília, Brasília, DF.
- RYALL, K. L. et al. **Detection of emerald ash borer in urban environments using branch sampling.** Toronto: Forestry Research Applications, 2011. 118p. (Nota técnica 111).
- WILLIAM, R.A. Use of randomized branch and importance sampling to estimate loblolly pine biomass. **Southern Journal of Applied Forestry**, v.13, p.181-184, 1989. Available from: <<http://www.ingentaconnect.com/content/saf/sjaf/1989/00000013/00000004/art00007>>. Accessed: Apr. 15, 2015.