

A Note on the Normalized Definition of Shannon's Diversity Index in Landscape Pattern Analysis

Habib Ramezani ¹

¹ Department of Forest Resource Management, Swedish University of Agriculture Science, Sweden

Correspondence: Habib Ramezani, Department of Forest Resource Management, Swedish University of Agriculture Science, Sweden. Tel: 46-907-868-151. E-mail: Habib.Ramezani@slu.se

Received: August 7, 2012 Accepted: September 25, 2012 Online Published: October 20, 2012

doi:10.5539/enrr.v2n4p54

URL: <http://dx.doi.org/10.5539/enrr.v2n4p54>

Abstract

A common approach for quantifying landscape pattern through landscape metrics is to use categorical maps of entire landscape. However, a new interest is to use sampling data where the data are collected for only a small fraction of the entire landscape. In sample based approaches some currently used landscape metrics may not be estimated since these metrics are defined based on mapped data. Shannon's diversity index is a frequently used metric in landscape pattern analysis. In this study, the performance of the normalized Shannon's diversity index is demonstrated when using sampled full-coverage maps and then point sampling on the maps. Artificial and real landscapes have been employed for this purpose. The results showed that calculation of the normalized Shannon's diversity index based on the number of land cover types in the entire classification system is more appropriate than based on the number of land cover types present within landscape. There was a strong and positive correlation between reference and estimated values but the estimator of Shannon's diversity index was slightly and negatively biased. In conclusion, it is needed to slightly redefine some currently used landscape metrics to accommodate sampling data.

Keywords: landscape pattern quantification, sampling techniques, landscape pattern metrics

1. Introduction

Landscape pattern has received much attention by landscape ecologists over the past decades. A fundamental assumption is that many ecological phenomena such as population dynamics and biodiversity can be affected by landscape pattern (e.g., van Dorp & Opdam 1987; Turner, 2005; Hernandez-Stefanoni & Dupuy, 2008). To understand pattern-process relationships, landscape pattern needs to be quantified first. For this purpose, a set of spatial indices so-called landscape metrics, have been developed (McGarigal et al. 2012) since direct measurement of landscape pattern is difficult (Traub & Kleinn, 1999). Definition of these metrics is based on measurable patch attributes such as the number, size and perimeter length of patches; a patch is defined as a relatively homogenous area which is different from its surrounding (Forman, 1995). Landscape metrics are classified into two general categories (i.e., composition and configuration) which can capture different aspects of landscape pattern (Gustafson, 1998). Configuration refers to the geographical distribution of patches whereas composition refers to the variety and abundance of different land cover types within the landscape. Shannon's diversity index is a typical example of the category of composition. This index is frequently used in landscape ecological studies to describe landscape diversity (e.g., Turner et al., 2001; Ramezani et al., 2010; Corona et al., 2011).

A common approach for quantifying landscape pattern through landscape metrics is to use categorical maps, for instance, land cover/land use maps, of an entire landscape (O'Neill et al., 1988; Wu et al., 2002; Li et al., 2005) and software FRAGSTATS (McGarigal et al. 2012) is usually used for this purpose. However, a competitive alternative is to use sampling data for the estimation of the metrics (Hunsaker et al., 1994; Kleinn, 2000a; Kleinn, 2000b, Loveland et al. 2002, Griffith et al. 2003a, Ramezani et al. 2010, Hassett et al. 2011, Ramezani&Holm 2011, 2012). Sample based approaches takes less time (low cost) and it might be possible to achieve more accurate result when using a well – designed and – executed sample survey. For instance, Corona et al. (2004) demonstrated that in the estimation of the total length of linear features within a landscape line intersects sampling method (LIS) appears to be more reliable than the polygon delineation approach (complete assessment). It was argued that in the LIS procedure, the assessment is carefully conducted on along sampling units (line

transects) only. Thus the possibilities of polygon delineation errors can considerably be reduced (Carfagna & Gallego, 1999). In sample based approaches, the data are collected for only a small fraction of the entire landscape, for instance, on a sample of aerial photos (Dramstad et al., 2002; Ståhl et al., 2011), or satellite images (e.g., Hunsaker et al., 1994; Griffith et al., 2003b; Stehman et al., 2003). Kleinn (2000a) demonstrated that some landscape metrics can be derived from field-based forest inventories. Landscape metrics currently in use were originally defined based on full-coverage map data; hence, it is possible that it is not applicable to calculate some metrics using sample based data. Furthermore, some of the metrics may produce similar numerical values for different landscape patterns (Tischendorf, 2001). Thus, development of new metrics is an urgent demand (Kleinn, 2000a) or to slightly redefine currently used landscape metrics for use with sample based data. Recently, a such modification has been made for contagion metric (Ramezani & Holm, 2012) which was adapted for use with point sampling data, as it was originally defined based on the full-coverage maps (O'Neill et al., 1988; Li & Reynolds, 1993).

In this study, the performance of the estimator of normalized Shannon's diversity index is demonstrated when using sampled full-coverage maps and point sampling method on the sampled full-coverage maps. The performance of the index is demonstrated for twenty 1 km² real landscapes (with seven land cover types) and two artificial landscapes.

2. Shannon's Diversity Index (H)

This index refers to the variety and abundance of different land cover types within a landscape (Shannon, 1948; Turner et al., 2001; McGarigal et al., 2012). This index is a general index of diversity and its value, if normalized, ranges from 0 to 1. The value of the index tends to 1 when the land cover types present have roughly equal proportion or a high number of cover/use types actually being present. A low value means that the landscape is dominated by one land cover type. In other words, the landscape has less diversity. The possibility to compare landscapes with a different number of land cover types is an advantage of the normalized Shannon's diversity over the usual definition (i.e., without normalized factor $\ln(s)$). The estimator of H is defined as

$$\hat{H} = -\frac{\sum_{j=1}^s \hat{p}_j \cdot \ln(\hat{p}_j)}{\ln(s)} \quad (1)$$

where s is the number of land cover types and p_j is the area proportion of the j th land cover type. In Eq. 1 s can either be treated as the number of land cover types considered in a given classification system (assumed fixed) or the number of land cover types present within the landscape (as a random variable). The p_j will be 0 for those land cover types that are not present with $p_j \cdot \ln(p_j)$ simply set to 0.

3. Examples

In large spatial scale survey, a combination of field-based inventory methods and remote sensing data are frequently used, for instance, two-stage sampling design (Thompson, 2002; Gregoire & Valentine, 2008). In such design, remote sensing data (e.g., aerial photos) are sampled in the first stage and then in the second stage one of basic sampling methods (e.g., point sampling) is used within each first stage sampling units. Note that sampled remote sensing data at the first stage can be treated as a full-coverage data in the second stage. In this study, 1 km² real landscapes was served as a full-coverage data where point sampling method was conducted over them in the second stage.

3.1 Artificial Landscape

With this example, I attempt to illustrate how different definitions of the normalized factor $\ln(s)$ influence the value of the normalized Shannon's diversity index in full-coverage mapping and sample based approaches. Figure 1 shows two artificial landscapes with the same extent but having a different number of land cover types present within landscape. The normalized Shannon's diversity index of these landscapes was calculated based on both land cover types actually being present and on the number of land cover types within the classification system, which is a fixed number (in this example, seven land cover types).

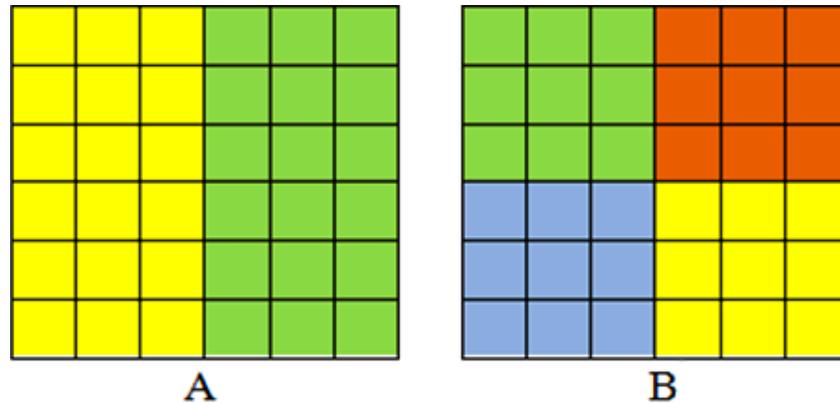


Figure 1. Two landscapes with the same extent but a different number of land cover types; landscape *A* with two land cover types and landscape *B* with four land cover types

3.2 Real Landscapes with Point Sampling Method in Second Stage

The analysis was conducted on twenty 1 km² real landscape (sampling units) from the National Inventory of Landscape in Sweden (NILS, Ståhl et al., 2011). These sampling units were randomly selected and distributed across Sweden. The already delineated land cover maps (in GIS environment) were used for the calculation area proportion of different land cover types and thus the reference value of the index. In order to estimate the index value in second stage a systematic point sampling design was applied on each already delineated land cover maps (see Figure 2). The bias and variance of Shannon's diversity estimator was derived through a large number of independently replicated samples (1000 times) for each sample size (i.e., 49, 100, 225, and 400 points) and each real landscape. There was convergence in the mean of the estimates since samples were independently selected. Note that each systematic sampling was conducted with a random starting point.

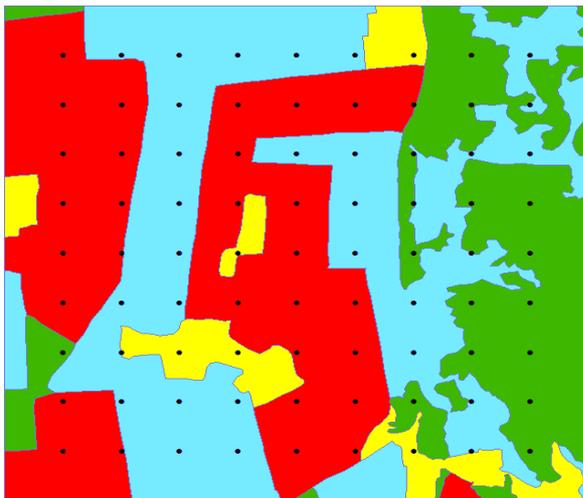


Figure 2. An example of 1 km² real landscape with a systematic point sampling design and a random starting point

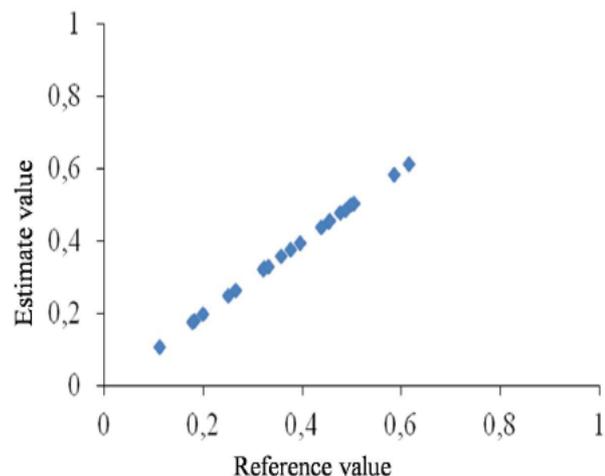


Figure 3. Relationship between reference and estimated values of Shannon's diversity index for twenty real landscapes with sample size 225

4. Results

The value of the normalized Shannon's diversity index was calculated for the two artificial landscapes. The value of the index was the same (i.e., 1) for both landscapes when the number of land cover types present within landscape was used. But the index value was 0.356 (for landscape *A*) and 0.712 (for landscape *B*) when the number of landscape cover types in classification system (here seven land cover types) was applied. Relationship

between reference and estimated values of the index for twenty 1 km² real landscapes with sample size 225 (points in point grid) is shown in Figure 3. There was a high and positive correlation between the reference and estimated values of the index and this was true for all four different sample sizes.

Detailed information on reference values (true values), bias and variance of the Shannon's diversity estimator for twenty real landscapes are provided in Table 1. As expected, the estimated variance of Shannon's diversity estimator tended to decrease with increasing sample size. In some cases, however, variances were not exactly inversely proportional to sample sizes. This can be explained by the landscape pattern and systematic sampling design applied. There was also a very small and negative bias for the estimator. Ramezani et al. (2010) has already explained that this bias is due to non-linear definition of Shannon's diversity index. Similar to variance, bias of the estimator tended to decrease when sample size increased.

Table 1. The reference value of the normalized Shannon's diversity index, estimated bias and variance for the twenty 1 km² real landscapes based on the fixed number of land cover types in classification systems (i.e., 7), four different sample sizes (49, 100, 225, 400) and systematic sampling design

ID of 1 km ² map	Reference Value	Sample size							
		Bias ^a				Variance			
	49	100	225	400	49	100	225	400	
1	0.35	-0.013	-0.007	-0.002	-0.001	0.047	0.036	0.02	0.013
2	0.322	-0.009	-0.006	-0.002	-0.002	0.033	0.025	0.012	0.01
3	0.672	-0.018	-0.006	-0.001	0	0.056	0.02	0.013	0.011
4	0.447	-0.016	-0.008	-0.004	-0.002	0.047	0.044	0.024	0.017
5	0.025	-0.007	-0.005	-0.002	-0.001	0.033	0.021	0.015	0.009
6	0.432	-0.018	-0.008	-0.001	-0.002	0.041	0.033	0.02	0.014
7	0.572	-0.011	-0.007	-0.002	-0.001	0.039	0.03	0.019	0.012
8	0.427	-0.011	-0.006	-0.002	-0.002	0.052	0.034	0.019	0.013
9	0.557	-0.02	-0.007	-0.001	-0.001	0.05	0.028	0.021	0.009
10	0.428	-0.012	-0.004	-0.003	-0.001	0.048	0.026	0.016	0.013
11	0.336	-0.016	-0.004	-0.002	-0.001	0.069	0.036	0.025	0.011
12	0.734	-0.022	-0.006	-0.004	0	0.04	0.027	0.02	0.013
13	0.508	-0.021	-0.014	-0.003	-0.001	0.056	0.046	0.022	0.015
14	0.398	-0.014	-0.007	-0.001	-0.001	0.033	0.022	0.015	0.008
15	0.645	-0.008	-0.001	-0.001	-0.001	0.022	0.014	0.01	0.007
16	0.788	-0.015	-0.007	-0.004	0	0.027	0.019	0.014	0.01
17	0.647	-0.009	-0.003	-0.003	-0.001	0.027	0.018	0.013	0.008
18	0.522	-0.012	-0.004	-0.001	-0.001	0.039	0.022	0.015	0.009
19	0.629	-0.015	-0.006	-0.004	-0.001	0.047	0.029	0.015	0.012
20	0.37	-0.012	-0.006	-0.002	0	0.058	0.032	0.02	0.012

^a Bias is deviation between estimator and the true parameter.

4. Discussion

This paper demonstrates the need for caution when calculating currently used landscape metrics in sample based approaches. A given landscape metric will be proper when differentiating landscapes with different patterns. Depending on whether landscape pattern analysis is conducted based on a full-coverage mapping approach or sample based approaches, an appropriate scaling factor ($\ln(s)$) should be chosen.

In any stage of multi-stage sampling design some land cover types may be missed, as demonstrated by Hunsaker et al. (1994), particularly when using a detailed classification system (where a general land cover type may be

divided into several more detailed sub-types). The results show that some metrics need slight modification to accommodate sampling data. Such modifications can lead to straightforward interpretation of metrics and also enable the differentiation of two different landscapes with different patterns. The results also show that with a moderate sample size a reasonable precision can be achieved. In a high fragmented landscape, where a large and continuous land cover type is broken into many small and isolated patches, a larger sample size is needed in order to achieve an acceptable precision.

In this study, whereas in the second stage point sampling was conducted on 1 km² already delineated land cover map, in practice such a sampling survey can be performed on non-delineated maps (raw remotely sensed data). In such situation in addition to saving time of survey (cost), the associated polygon delineation errors can be eliminated. The potential of sampling survey in eliminating the errors has been demonstrated by Corona et al. (2004) and Ramezani and Holm (2011) where line intersect sampling (LIS) method has been applied for estimating Shannon's diversity and forest edge length on non-delineated aerial photo. See Figure 4 a systematic point sampling design with a random starting point over a raw aerial photo where polygons are not delineated.

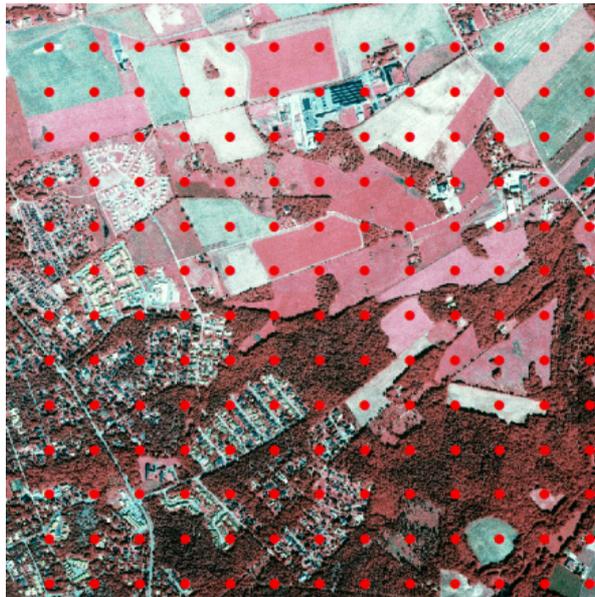


Figure 4. An example of systematic distribution of point grid (point sampling) on raw aerial photos (non-delineated)

Acknowledgments

The author thanks Dr. Heather Reese for English correction. I am also grateful for the constructive comments by two anonymous reviewers of the paper.

References

- Carfagna, E., & Gallego, F. (1999). *Thematic maps and statistics*. In Land cover and land use information systems for European Union policy needs, 219-228. Office for Official publications of the European Communities, Luxembourg.
- Corona, P., Chirici, G., McRoberts, R. E., Winter, S., & Barbati, A. (2011). Contribution of large-scale forest inventories to biodiversity assessment and monitoring. *Forest Ecology and Management*, 262, 2061-2069. <http://dx.doi.org/10.1016/j.foreco.2011.08.044>
- Corona, P., Chirici, G. & Travaglini, D. (2004). Forest ecotone survey by line intersect sampling. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, 34, 1776-1783. <http://dx.doi.org/10.1139/x04-045>
- Dramstad, E., Fjellstad, W. J., Strand, G.-H., Mathiesen, H. F., Engan, G., & Stokland, J. N. (2002). Development and implementation of the Norwegian monitoring programme for agricultural landscapes. *Journal of Environmental Management*, 64, 49-63. <http://dx.doi.org/10.1006/jema.2001.0503>
- Forman, R. T. T. (1995). *Land mosaics: the ecology of landscapes and regions*. Cambridge: Cambridge

University Press.

- Gregoire, T. G., & Valentine, H. T. (2008). *Sampling Strategies for Natural Resources and the Environment* Boca Raton, Fla. London: Chapman & Hall/CRC.
- Griffith, J. A., Stehman, S. V., & Loveland, T. R. (2003a). Landscape trends in Mid-Atlantic and southeastern United States ecoregions. *Environmental Management*, 32, 572-588. <http://dx.doi.org/10.1007/s00267-003-0078-2>
- Griffith, J. A., Stehman, S. V., Sohl, T. L., & Loveland, T. R. (2003b). Detecting trends in landscape pattern metrics over a 20-year period using a sampling-based monitoring programme. *International Journal Of Remote Sensing*, 24, 175-181. <http://dx.doi.org/10.1080/01431160305009>
- Gustafson, J. E. (1998). Quantifying landscape spatial pattern: What is the state of the art? *Ecosystems*, 1, 143-156. <http://dx.doi.org/10.1007/s100219900011>
- Hassett, E. M., Stehman, S. V., & Wickham, J. D. (2011). Estimating landscape pattern metrics from a sample of land cover. *Landscape Ecology*, 1-17.
- Hernandez-Stefanoni, J. L., & Dupuy, J. M. (2008). Effects of landscape patterns on species density and abundance of trees in a tropical subdeciduous forest of the Yucatan Peninsula. *Forest Ecology and Management*, 255, 3797-3805. <http://dx.doi.org/10.1016/j.foreco.2008.03.019>
- Hunsaker, C. T., O'Neill, R. V., Jackson, B. L., Timmins, S. P., Levine, D. A., & Norton, D. J. (1994). Sampling to characterize landscape pattern. *Landscape Ecology*, 9, 207-226. <http://dx.doi.org/10.1007/BF00134748>
- Kleinn, C. (2000a). Estimating metrics of forest spatial pattern from large area forest inventory cluster samples. *Forest Science*, 46, 548-557.
- Kleinn, C. (2000b). Sample based assessment landscape metrics.
- Li, H. & Reynolds, J. (1993). A new contagion index to quantify spatial patterns of landscapes. *Landscape Ecology*, 8, 155-162. <http://dx.doi.org/10.1007/BF00125347>
- Li, X. Z., He, H. S., Bu, R. C., Wen, Q. C., Chang, Y., Hu, Y. M., & Li, Y. H. (2005). The adequacy of different landscape metrics for various landscape patterns. *Pattern Recognition*, 38, 2626-2638. <http://dx.doi.org/10.1016/j.patcog.2005.05.009>
- Loveland, T. R., Sohl, T. L., Stehman, S. V., Gallant, A. L., Sayler, K. L., & Napton, D. E. (2002). A strategy for estimating the rates of recent United States land-cover changes. *Photogrammetric Engineering and Remote Sensing*, 68, 1091-1099.
- McGarigal, K., Cushman, S. A., Neel, M. C., & Ene, E. (2012). FRAGSTATS: spatial pattern analysis program for categorical maps, version 4.0.
- O'Neill, R. V., Krumme, J. R., Gardner, H. R., Sugihara, G., Jackson, B., DeAngelis, D. L., ... Graham, L. R. (1988). Indices of landscape pattern. *Landscape Ecology*, 1, 153-162. <http://dx.doi.org/10.1007/BF00162741>
- Ramezani, H. & Holm, S. (2011). Sample based estimation of landscape metrics: accuracy of line intersect sampling for estimating edge density and Shannon's diversity. *Environmental and Ecological Statistics*, 18, 109-130. <http://dx.doi.org/10.1007/s10651-009-0123-2>
- Ramezani, H., & Holm, S. (2012). A distance dependent contagion functions for vector-based data. *Environmental and Ecological Statistics*, 19, 161-181.
- Ramezani, H., Holm, S., Allard, A., & Ståhl, G. (2010). Monitoring landscape metrics by point sampling: accuracy in estimating Shannon's diversity and edge density. *Environmental Monitoring and Assessment*, 164, 403-421. <http://dx.doi.org/10.1007/s10661-009-0902-0>
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379-423.
- Stehman, S. V., Sohl, T. L., & Loveland, T. R. (2003). Statistical sampling to characterize recent United States land-cover change. *Remote Sensing of Environment*, 86, 517-529. [http://dx.doi.org/10.1016/S0034-4257\(03\)00129-9](http://dx.doi.org/10.1016/S0034-4257(03)00129-9)
- Ståhl, G., Allard, A., Esseen, P. A., Glimskär, A., Ringvall, A., Svensson, J., ... Inghe, O. (2011). National Inventory of Landscapes in Sweden (NILS) - Scope, design, and experiences from establishing a multi-scale biodiversity monitoring system. *Environmental Monitoring and Assessment*, 173, 579-595.

- <http://dx.doi.org/10.1007/s10661-010-1406-7>
- Thompson, S. K. (2002). *Sampling*. New York: Wiley.
- Tischendorf, L. (2001). Can landscape indices predict ecological processes consistently? *Landscape Ecology*, *16*, 235-254. <http://dx.doi.org/10.1023/A:1011112719782>
- Traub, B., & Kleinn, C. (1999). Measuring fragmentation and structural diversity. *Forstwissenschaftliches Centralblatt*, *118*, 39-50. <http://dx.doi.org/10.1007/BF02768972>
- Turner, M. G. (2005). Landscape ecology: What is the state of the science? *Annual Review of Ecology Evolution and Systematics*, *36*, 319-344. <http://dx.doi.org/10.1146/annurev.ecolsys.36.102003.152614>
- Turner, M. G., Gardner, R. H., & O'Neill, R. V. (2001). *Landscape ecology in theory and practice : pattern and process*. New York: Springer.
- van Dorp, D., & Opdam, P. E. (1987). Effects of patch size, isolation and regional abundance on forest bird communities. *landscape Ecology*, *1*, 59-73. <http://dx.doi.org/10.1007/BF02275266>
- Wu, J., Shen, W. J., Sun, W. Z., & Tueller, P. T. (2002). Empirical patterns of the effects of changing scale on landscape metrics. *Landscape Ecology*, *17*, 761-782. <http://dx.doi.org/10.1023/A:1022995922992>