# Transect sampling strategies for semi-detailed characterization of inland valley systems.

P.N. WINDMEIJER<sup>1,\*</sup>, T.J. STOMPH<sup>2</sup>, A. ADAM<sup>3</sup>, R. COPPUS<sup>4</sup>, N. DE RIDDER<sup>2</sup>, M. KANDEH<sup>5</sup>, M. MAHAMAN<sup>3</sup> AND M. VAN LOON<sup>2</sup>

- DLO-Winand Staring Centre, Inland Valley Consortium, c/o WARDA, 01 BP 2551, Bouaké 01, Côte d'Ivoire.
- 2. Wageningen Agricultural University, Department of Agronomy, P.O. Box 341, 6700 AH Wageningen, The Netherlands.
- 3. West Africa Rice Development Association, 01 BP 2551, Bouaké 01, Côte d'Ivoire.
- 4. University of Amsterdam, Laboratory of Physical geography and Soil Science, Nieuwe Prinsengracht 130, 1018 VZ Amsterdam, The Netherlands.
- 5. Land and Water Development Division, PMB 187, Freetown, Sierra Leone.

\* Corresponding author (fax:+ 225 634714; e-mail: p.windmeijer@cgnet.com) Received 26 September 1997; accepted 3 April 1998

## Abstract

For the semi-detailed characterization of inland valleys systems, the Inland Valley Consortium (IVC) uses transect sampling to describe valley morphology, soils, and land use characteristics. Initially, for each valley one transect with a width of 200 to 400 m was proposed. More recently, use of a larger number of transects has been suggested. However, there is no statistical basis for either the number of transects to be described or their optimal width.

Two sets of aerial photographs (1979, scale 1:50,000, and 1996, scale 1:20,000) of the Boundiali key area were processed and the results digitized. Using GIS, sets of transects with different width were made, and used for the determination of the sample size for estimating the land use ratio and the valley bottom ratio at valley system level with given confidence intervals. These results were validated with data from the field and with sets of pre-selected transects.

The calculated sample size is based on simple random sampling and depends on both transect width and variability in the study area. For the characterization of inland valleys with reasonable confidence intervals, the sample size for random sampling is too large for the IVC semi-detailed characterization method.

Smaller samples of transects selected on the basis of sub-recent aerial photographs gave comparable results for land cover characteristics estimation as the larger samples for random sampling. It is, therefore, concluded that there is scope for reducing sample size by formalizing criteria for transect placement on the basis of sub-recent aerial photographs.

## Résumé

Dans le cadre de la caractérisation semi-détaillée des systèmes de bas-fonds, le Consortium Bas-Fonds (CBF) a recours à une approche par transects pour décrire la morphologie de la vallée, les sols et les caractéristiques de l'utilisation des terres. Initialement, il a été proposé de faire un transect d'une largeur de 200 à 400 m dans chaque vallée. Plus récemment, il a été suggéré d'augmenter le nombre de transects. Cependant, on manque de bases statistiques pour déterminer le nombre de transects à décrire et leur largeur optimale.

Deux jeux de photographies aériennes (1979, échelle 1:50,000, et 1996, échelle 1:20,000) du site-clé de Boundiali ont été analysés et les résultats numérisés. En utilisant un SIG, des jeux de transects de largeurs différentes sont élaborés, puis utilisés pour déterminer la taille de l'échantillon nécessaire, pour un intervalle de confiance donné, pour l'estimation du ratio d'utilisation de la terre et le ratio bas-fond/bassin versant. Ces résultats sont calibrés avec des données collectées sur le terrain, et avec des jeux de transects pré-sélectionnés.

Le calcul de la taille de l'échantillon est basés sur un échantillonnage aléatoire simple et dépend à la fois de la largeur du transect et de la variabilité de la zone étudiée. Pour une estimation correcte des caractéristiques du basfond, avec un intervalle de confiance raisonnable, la taille de l'échantillon nécessaire dans le cas un tirage aléatoire est incompatible avec la méthode de caractérisation semi-détaillée du CBF.

Des échantillons plus petits de transects, sélectionnés sur la base de photographies aériennes pas trop anciennes,

ont donné des résultats équivalents en matière de caractéristiques de l'utilisation des terres, comparé avec les échantillions plus large nécessaire dans le cas un tirage aléatoire simple. En conséquence, on peut conclure qu'il est possible de réduire les dimensions de l'échantillon, en ayant des critères rigoureux de positionnement du transect, grâce à l'utilisation de photographies aériennes pas trop anciennes.

#### Keywords:

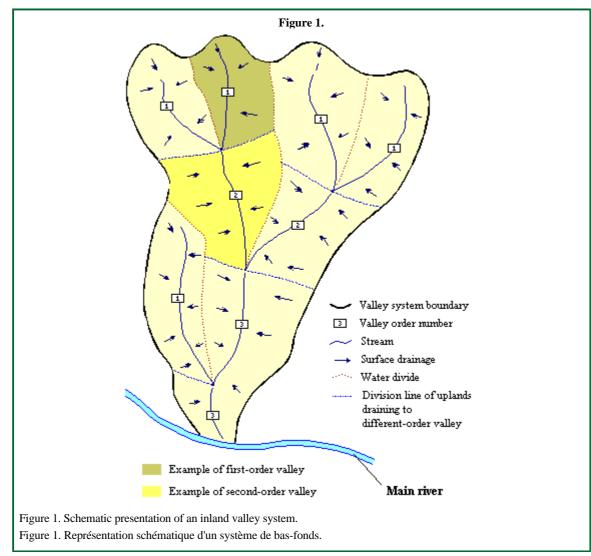
inland valley, characterization, transect sampling, GIS, acceptable error, minimum sample, Côte d'Ivoire *Mots-clés*:

system de bas-fonds, caractérisation, échantillons de transects, SIG, limit d'erreur, échantillion minimal, Côte d'Ivoire

## Introduction

The multi-scale agro-ecological characterization of inland valleys (Andriesse *et al.*, <u>1994</u>) is used by the Inland Valley Consortium (IVC) to quantify major constraints to agricultural production in inland valleys, and to target research and technology generation to overcome these constraints. In this way, IVC strives to attain its major objective: the intensification and diversification of sustainable agricultural production in inland valley lowlands. Data produced by the characterization will be used to develop decision support systems for technology selection and transfer of technologies.

Inland valley systems (inland with respect to main rivers and tributaries) can be defined as the upper reaches of river systems where river alluvial processes are of only minor importance. An inland valley system has only a minor floodplain and a levee system. An inland valley system consists of individual inland valleys of different order (Figure 1; Windmeijer & Andriesse, 1993).



The multi-scale characterization of inland valley agro-ecosystems as described by Andriesse *et al.* (1994) has four levels. For each level the scale, the unit of analysis, and a minimum data set of bio-physical and socio-economical characteristics to be described are defined. The unit of analysis decreases in size with each step, while the level of detail in the characteristics increases.

The semi-detailed characterization (scale 1:50,000) is the third level of characterization. The objective of this level of characterization is to describe relatively fast the global characteristics of the inland valley systems and their spatial variability. For this purpose a key area ( $50 \times 50$  km) is chosen in an agro-ecological unit as described at reconnaissance level. Using aerial photographs, four valley systems are selected. To describe the inland valleys systems, it has been proposed to describe one transect across a first, second, third, etc. order valley for land cover, morphology, and soils. The width of the transect for land cover characterization is between 200 and 400 m. The assumption behind this approach is that we can assess, within acceptable margins, the land cover and morphological characteristics at valley system level, using a limited number of transects.

The proposed number of the transects to be described and their width where based on expert knowledge, but no statistics were applied so far to determine the minimum sample size and transect width for a proper inland valley characterization. In this article we present both a method to determine minimal sample sizes with the use of aerial photographs and GIS, and the optimum in number and width of transects to be sampled to distinguish inland valley systems on the basis of land cover and morphological characteristics.

## **Materials and methods**

#### Site

One of IVC's and WARDA's (West Africa Rice Development Association) key areas in Côte d'Ivoire is found near Boundiali (6° 20' WL, 9° 25' NL). It is situated in the Guinea Savanna zone and covers an area of about 625 km<sup>2</sup>. Here, the annual precipitation is about 1200 mm and the lithology is formed by schists. The Boundiali key area covers parts of 22 valley systems, of which 15 are situated entirely within the area. These 15 valley systems were used for the statistical analysis.

In 1993, IVC conducted the semi-detailed characterization of this key area. Based on this survey, the Poundjou valley system, with a total surface of 4500 ha, was selected as a key site for IVC's and WARDA's research on technology generation and testing.

#### Data collected and parameters quantified

During the semi-detailed characterization a number of land cover, soil, and morphological characteristics of the valley systems are described. To facilitate the comparison between valley systems, a number of land cover and morphological characteristics are quantified. Only these quantified characteristics allow calculations of minimum sample sizes.

At semi-detailed level, land cover observations result in land cover types like land under annual crops, perennial crops, young fallow, old fallow and natural vegetation. The differences in land cover between valley systems are analyzed on the basis of the Land Use Ratio (LUR). The LUR is defined as the area actually used for agricultural production and under young fallow, expressed as a percentage of the total area of the valley system. A high LUR value indicates a high land use pressure. Other land cover indicators have been suggested (Van Duivenbooden & Windmeijer, <u>1995</u>) but it is often more difficult to calculate these accurately from aerial photographs.

Observations on valley morphology result in data on the form, length, and width of the different land subelements like valley bottom, slopes, and crests. Morphological differences between valley systems are analyzed on the basis of the Valley Bottom Ratio (VBR). The VBR is defined as the area occupied by crests, slopes, and fringes divided by the surface of the valley bottom. A high VBR indicates a relatively small valley bottom.

#### Approach

The Boundiali key area is covered by aerial photographs, scale 1:50,000, made in 1979. Aerial photographs of the Poundjou valley system, scale 1:20,000, were made in 1996. Both sets of photos were interpreted for land cover (cropped land, natural vegetation, etc.) and morphology (valley bottom, slopes, etc.).

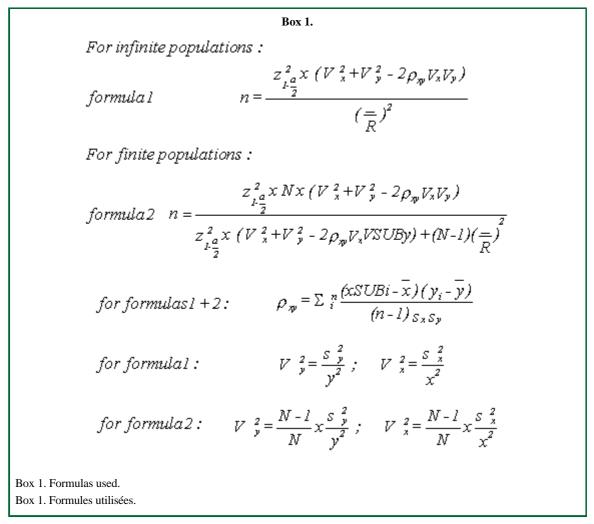
In the actual study, the digitized interpretation of the 1:50,000 aerial photographs was used for several purposes. Firstly, LUR and VBR were calculated for the key area and for all valley systems, i.e. the values we have to estimate using a limited number of transects. The variables defining LUR and VBR were used to calculate their variances

using the formula for the variance of a ratio estimate. From these, the sample size (n), or number of valley systems to be sampled for estimation of the LUR and VBR at key area level is calculated. Secondly, using GIS, the maximum number of 400 m wide transects (N) are placed in the same four valley systems as described in the field during the 1993 campaign. Using the same statistical approach, the sample size (n) or number of transects needed for estimates of the LUR and VBR at the valley system level, to be within specified limits of the true value with given acceptable errors, was calculated.

The same steps were repeated for the 1:20,000 aerial photographs made in 1996. The LUR and VBR at valley system level were derived from an interpretation of the complete valley system. In this case the sample sizes were calculated using transects of different width, (i.e. line transects and band-transects of 100, 200, and 400 m wide).

#### Statistical analysis

Both the LUR and VBR are ratios of which during transect sampling the numerator as well as the denominator are subject to sampling variation. Moreover, the value of the numerator can be supposed to be partly dependant on the value of the denominator. These two considerations point towards the use of the statistical techniques relevant for ratio estimation (Levy & Lemeshow, <u>1991</u>). In Box <u>1</u>, the formulas are given which are used to calculate the minimum sample size.



When line transects are used to calculate the LUR and VBR the maximum number of transects, the population, that can be taken from a unit larger than a valley is infinite. When band transects are used, the population, determined by the length of the valley and the width of the transects, is finite. Therefore, minimum sample sizes for line and band transects are calculated using Formula 1 and Formula 2, respectively.

The original formulas from Levy & Lemeshow (1991) have been modified slightly. We replaced the maximum relative difference allowed between the estimated ratio and the true ratio (\_) by the maximum absolute difference \_/R (in which R is the estimated ratio). This was done because using \_ results in relatively large minimum samples in cases where values of LUR and VBR are low, and relatively small samples in case where LUR and VBR are high. By using \_/R, the minimum sample sizes calculated are independent of the value of the ratio.

The sample sizes were calculated for a 95% confidence level, and for estimates to be within 20, 15, 10, 5 and 1% of the true value (acceptable errors). This means that the sample sizes are determined so that, for instance, if the estimated value of LUR is 30%, then there is a 95% chance for the true value to be between 20 and 40% using the calculated sample size, in case an acceptable error of 10% was chosen.

Sample sizes were calculated at two different scales. The first scale was the sample size taken for estimation of LUR and VBR at valley system level. However, as indicated in Figure 1, valley systems are complex watersheds, composed of different valleys. The second way of calculating the sample size was taking the sum of transects required to estimate the LUR and VBR for each valley within the watershed with a given acceptable error. In the following sections, the second way of calculating the sample is referred to as "valley level".

#### Comparison with field data

The results of the GIS exercise were compared with available field data. For the 1993 semi-detailed characterization of the Boundiali key area, 3 to 6 transects were selected in four valley systems using the 1:50,000 scale aerial photographs (Windmeijer *et al.*, <u>1994</u>).

Secondly, a limited number of line transects were selected on the basis of the 1979 aerial photographs. These were projected on the 1:20,000 aerial photographs and were used, with GIS, to estimate the LUR and VBR of the Poundjou valley system. These line transects were described in the field for validation of the photo interpretation and to estimate time requirements for describing line transects as compared to band transects of 400 m wide (as was done during the 1993 campaign).

### Results

Using GIS, we found that the VBR had a lower variability than the LUR. This indicates that the LUR determines the sample size, assuming that we have one sample for both characteristics. Therefore, we only present the minimal sample sizes for the estimation of the LUR.

#### Sample sizes of valley systems and 400 m wide transects

Using the sub-recent 1:50,000 aerial photographs, the LUR of the key area was 28%, with a range from 13 to 45% for individual valley systems. To estimate the LUR of the key area with an acceptable error of 5%, we calculated that 4 valley systems should be sampled.

Table <u>1</u> shows the minimum number of 400 m wide transects to be sampled for an estimation of the LUR with different acceptable errors at the valley system level for the four selected valley systems (6, 8, 10, and 19). Between 50 and 85 % of the population N had to be sampled to be 95% confident that the estimated value was within 0.05 of the true. To be 95% confident that the estimated value was within 0.2 of the true value, between 7 and 25% of the maximum number of transects (N) has to be sampled.

**Table 1**. The number of 400 m wide transects (n) required to estimate LUR with different acceptable errors (a.e. in %) at valley system level.

**Tableau 1**. Nombre de transects d'une largeur de 400 m requis (n) pour estimer le ratio d'utilisation des terres (LUR), à différents limits d'erreur (a.e. in %), au niveau des systèmes de bas-fonds.

Valley system	No. of transects	LUR	n				
			a.e. 20%	a.e 15%	a.e. 10%	a.e. 5%	a.e. 1%
6	17	32	4	6	10	14	17
8	88	27	6	9	18	45	85
10	77	38	4	8	14	36	74
19	68	23	5	8	15	36	66

In case we calculate the number of transects at valley level, a much higher percentage of the transect population needs to be sampled to estimate the LUR assuming the same acceptable error. In all cases, we observed that the sum of the minimum samples calculated for individual valleys within valley systems is more than twice the minimum sample as calculated for that valley system.

#### Minimum sample size for transects of different width

For the Poundjou valley system (which is the same as valley system 10 in Table 1), using the recent 1:20,000 aerial photographs, the LUR was calculated to be 39,5%. Table 2 shows a decreasing sample size required with increasing width of the transects to estimate the LUR and VBR at valley system level. This increase is most accentuated as the acceptable errors are small and the transects are narrow. In case an acceptable error of 10% is chosen, 14 band transects of 400 m wide have to be described, or 19 line transects. With an acceptable error of 5 % these values are 36 and 75, respectively.

**Table 2**. The effect of the type of transect on the required sample size (n) to estimate LUR with different acceptable errors (a.e. in %) at valley system level of Poundjou inland valley.

**Tableau 2**. Impact du type de transect sur la taille de l'échantillon requis (n) pour estimer le ratio d'utilisation des terres à différents limits d'erreur (a.e. in %), au niveau du système de bas-fonds de Poundjou.

Type of transect		No. of transects	n					
			a.e. 20%	a.e. 15%	a.e. 10%	a.e. 5%	a.e. 1%	
Line transect	0 m wide	infinite	5	10	19	75	1876	
Band transect	100 m wide	320	4	9	17	59	271	
Band transect	200 m wide	158	4	8	15	46	144	
Band transect	400 m wide	77	4	8	14	36	74	

The aerial photographs of scale 1:20,000 are also used to calculate the sample size for line transects at the valley level. To estimate the LUR and VBR at the valley level with an acceptable error of 5%, it is necessary to describe 436 line transects within the Poundjou valley system. To estimate the LUR and VBR at the valley system level, the sample size was 75 line transects, i.e. six times less than at the valley level.

#### Comparison with field data and selective sampling

For the key area, the LUR<sub>FIELD</sub> (established on the basis of the four valley systems described during the 1993 field survey) and LUR<sub>GIS</sub> (calculated over the total surface of the key area using the 1:50,000 aerial photographs) were both 28%. At the level of the valley systems, the absolute differences between the 'LUR<sub>FIELD</sub>' and 'LUR<sub>GIS</sub>' only varied between 4 to 8%. This means that using a lower number of 400 m wide band transects than required (3 to 6 per valley system), we were able to estimate the LUR within 10% of the true value.

**Table 3**. Comparison of LUR values estimated at valley system level by using digitized aerial photographs ( $LUR_{GIS}$ ) or transect sampling ( $LUR_{FIELD}$ ) using expert knowledge to select location of the transects.

**Tableau 3**. Comparaison des valeurs estimées du Ratio d'utilisation des terres au niveau des systèmes de bas-fonds en utilisant des photographies aériennes numérisées (LUR<sub>SIG</sub>) ou un échantillon de transects (LUR<sub>TERRAIN</sub>) basé sur un positionnement des transects à dire d'expert.

		LUR	GIS	LUR <sub>FIELD</sub>		
Valley system		scale 1:50,000	scale 1:20,000	400 m transects	line transects	
		Values per v	alley system			
6	(Poundjou valley system)	32.0	-	24.0	-	
8 (Doundiou v		29.0	-	35.0	-	
10 (Poundjou V		37.0	39.5	31.0	37.0	
19		25.0	-	21.0	-	
		Values for	the key site	·		
		28.0 <sup>1</sup>	-	28.0 <sup>2</sup>	-	

The same results are found for the 17 selected representative line transects placed in the Poundjou valley system. The sample sizes calculated for this system were 19 and 75 line transects with acceptable error of 10 and 5%, respectively, and a LUR of 39,5% was determined. Using 17 pre-selected line transects only, we estimated the LUR at 37%, well within 5% of the true value.

## **Discussion and conclusion**

The required minimum sample sizes calculated in this paper are based on simple random sampling of the unit of analysis. The number of valleys systems (4) needed to estimate the LUR and VBR of the Boundiali key area, with an acceptable error of 5%, corresponds with the currently used protocol (Van Duivenbooden & Windmeijer, <u>1995</u>).

With the occurring spatial variability of LUR and VBR in the key area, a large number of transects (36 band transects of 400 m wide or 75 line transects) has to be sampled to estimate the LUR and VBR at valley system level with the same acceptable error of 5%. Accepting an error of 10%, only 14 band transects of 400 m have to be described, or 19 line transects. With an acceptable error of 5%, twice as many line transects must be described in comparison to 400 m wide band transects, while only 1,5 times as many line transects have to be described if we accept an error of 10%.

The number of transects to be sampled depends on the accuracy with which we like to estimate the different parameters. Accepting errors of 20 and 15%, the number of transects to be sampled is low, but we consider this level of accuracy too low for the semi-detailed characterization. The numbers of transects to be described, accepting errors of only 5 and 1 %, are too high for IVC's semi-detailed characterization. We conclude that with the methodology proposed for the semi-detailed characterization, we may aim at estimating the values of LUR and VBR within 10% of the true values.

The calculated number of transects needed for simple random sampling to estimate of the LUR and VBR within 10, 5, and 1% of the true values is rather high for the semi-detailed characterization. As explained in the Section statistical analysis, we replaced the maximum relative difference allowed between the estimated ratio and the true ratio by the maximum absolute difference. The result of this is an underestimation of the number of transects forming the minimal simple random sample.

However, simple random sampling is not often practiced in soil, landscape, and land cover mapping at scale 1:50,000 or smaller (Brus, <u>1995</u>). Existing information like topographic maps, geological maps, and especially remote sensing tools like satellite images and aerial photographs are used to develop a sampling strategy aiming at the highest level of accuracy required with the lowest number of observations.

The results of our study prove that a selective sampling strategy, accepting a given error, reduces the sample size compared to random sampling calculated. Or, in other words, the number of transects calculated for simple random sampling with a chosen acceptable error, but placed in a selective way, results in a more accurate estimation.

During the semi-detailed characterization conducted in Boundiali, we have described up to 6 band transects of 400 m wide, carefully selected, using aerial photographs. According to our statistical analysis for simple random sampling, this sample size allows only an estimation of LUR and VBR with an error of 20%. But with our selective sampling strategy, we were able to estimate the LUR and VBR within 10% of the true value. The same conclusion can be drawn for the 17 selectively placed line transects. This number is required for a simple random sampling accepting an error of 10%, but we were able to estimate LUR and VBR well within 5% of the true value.

We conclude that to estimate the LUR and VBR of the Poundjou valley system within 10% of the true value, the minimum sample size is about 6 band transects of 400 m wide. Because 1,5 times the number of line transects has to be described, about 9 line transects have to be sampled.

Based on our experience in the field, two to three line transects can be described in the same time needed for one 400 m wide band transect. This means that for estimates within 10% of the true value, it is faster to describe the higher number of line transects. Another advantage is that there are less data to be processed, because the total area of the valley system covered is smaller using line transects instead of band transects.

It is interesting to know, though, that the land use ratio in the Boundiali key area does not seem to have changed between 1979 and 1996 (37% and 39,5%, respectively). This is probably due to the low population density in the area, less than 16 persons/km<sup>2</sup> (Becker & Diallo, <u>1992</u>).

Our results confirm that at semi-detailed level, the unit of analysis is the valley system. The number of transects to be sampled for a sound estimate of LUR and VBR at the valley level is much larger than the number required for a good estimation at the valley system level, and too large for the semi-detailed characterization approach used by IVC. Transect data can be seen as indicators of LUR and VBR values of individual valleys, but no estimates at the individual valley level can be made from semi-detailed level data.

However, the data and results from the semi-detailed characterization can be successfully used for extrapolation from the valley system level to the level of the key area. Using the LUR figures from the four valley systems described, the LUR of the key area was estimated well within 5% of the true value.

For areas which differ from the Boundiali key area in land cover distribution, overall LUR, land use variability, morphological variability, and the size of the valley systems, we consider the presented minimum sample sizes as poor indicators only. For such areas we advocate adopting the same methodology we used to quickly establish the minimum sample size from sub-recent aerial photographs.

In a further study, in an effort to formalize expert knowledge, we hope to determine rules for the selection of transects using older aerial photographs to enable the reduction of the minimum random sample size.

## References

Multi-scale Characterization of Inland Valley Agro-ecosystems in West Africa. <i>Netherlands Journal of Agricultural Science</i> 42: 159-179. Becker, L., & R. Diallo, 1992.
Rocker J & D Diallo 1007
Deckel, L., & K. Diano, 1772.
Characterization and Classification of Rice Agro-ecosystems in Côte d'Ivoire.
West Africa Rice Development Association, Bouaké, 135 pp., 33 maps, 49 tables.
Brus, D.J., 1994.
New quantitative inventory methods (in Dutch).
In: P. Buurman & J. Sevink (Eds.), Van Bodemkaart tot Informatie systeem. Wageningen Pers,
Wageningen, pp. 85-110.
Jamin, J.Y., & P.N. Windmeijer (Eds.), 1998.
Characterization of inland valley agro-ecosystems: a tool for their sustainable use. Proceedings of the first
IVC scientific workshop, November 1995.
Inland Valley Consortium, Bouaké, 325 pp.
Levy, P.S., & S. Lemeshow, 1991.
Sampling of populations, methods and applications.
John Wiley and Sons, New York, 420 pp.
Van Duivenbooden, N., & P.N. Windmeijer, 1995.
Manual for Semi-detailed Characterization of Inland Valley Agro-ecosystems. Characterization of Rice-
growing Environments, Technical Report 4.
DLO Winand Staring Center, Wageningen, 85 pp.
Van Duivenbooden, N., P.N. Windmeijer, L.O. Fresco, & W. Andriesse, 1996.
The Integrated Transect Method as a tool for land use characterization, with special reference to inland
valley agro-ecosystems in West Africa.
Landscape and Urban Planning 34: 143-160.
Windmeijer, P.N., & W. Andriesse (Eds.), 1993.
Inland valleys in West Africa: an agro-ecological characterization of rice-growing environments.
ILRI publication 52, International Institute for Land Reclamation and Improvement, Wageningen, 160 pp.
Windmeijer, P.N., N. Van Duivenbooden, & W. Andriesse, 1994.
Semi-detailed characterization of inland valleys in Côte d'Ivoire. Characterization of rice-growing
environments, Technical Report 3, vol 1 and 2.
DLO Winand Staring Center, Wageningen, 129 pp.
© NJAS 611, 1998 - Comments to: J.J.Neetson@AB.DLO.NL