Constructing typologies, a way to deal with farm diversity:

General guideline for Humidtropics

[Draft version for SRT1-workshop, 11-13 March 2014, Nairobi]

Stéphanie Alvarez, Wim Paas, Katrien Descheemaeker, Jeroen Groot

Plant Sciences Group, Wageningen University, The Netherlands
Summary
To capture variability of farming systems, typologies are an often used tool. Typologies in Humidtropics (HT) are used to arrive efficiently at best-fit farm adjustments and innovations in order to meet HT goals.

The objective of this document is to provide guidelines for the construction of typologies. The general typology approach proposed here is to combine expert knowledge obtained through participative approach with multivariate statistics.

After briefly sharing some background information on typology making, the different steps of the typology construction are described. The steps are:

1. precisely state the objective of the typology;
2. formulate a hypothesis on farm diversity;
3. select the variables;
4. design a sampling method for data collection;
5. use multivariate statistics;
6. compare the typology result with the hypothesis.

Contents
Summary .................................................................................................................................................. 2
Introduction .............................................................................................................................................. 3
Background on typology making ........................................................................................................... 4
    Purposes ............................................................................................................................................. 4
    Methods ............................................................................................................................................ 4
Typology guidelines ............................................................................................................................. 6
    1. Typology objectives ..................................................................................................................... 6
    2. Hypothesis on typology structure ............................................................................................... 6
    3. Variable selection ....................................................................................................................... 6
    4. Sampling ...................................................................................................................................... 9
    5. Multivariate statistics .................................................................................................................. 9
    6. Hypothesis verification ................................................................................................................ 10
Preliminary discussion ......................................................................................................................... 11
    General remarks ........................................................................................................................... 11
    Further development of this document .......................................................................................... 11
References .............................................................................................................................................. 12
Introduction
Farming systems in the humid tropics are found across a wide variety of cultures and landscapes. The biophysical, institutional, social and economic drivers change from place to place, resulting in different reactions of farmers and communities between and within areas. Over time, these different reactions lead to temporal and spatial variability between and within farming systems.

The existing variability that is encountered in farming systems is challenging to fully comprehend, leading to partial representation of reality. In developing countries, the possible explaining variables for variability in farming systems are diverse (e.g. Tittonell et al., 2010; Dossa et al., 2011). Tools and methods (e.g. wealth rankings, farm typologies, distributions) are developed to manage the farm diversity. When using these tools and methods, a trade-off is made between the quality of representing reality and the level of detail that can be managed. An often used methodology to deal with variability and diversity is typology making, i.e. the grouping of farmers/households.

In this document we provide guidelines for the construction of typologies (Figure 1). Our objective is to raise awareness of opportunities and pitfalls that arise during typology construction.
Background on typology making

Purposes

Typologies respond to research questions that require to take into account the agricultural heterogeneity within a region (e.g. Alary et al., 2002; Righi et al., 2011; cf. Figure 1). We consider three important aspects of typologies:

1. **Targeting**: the distinction between farming system types is aimed at identifying of best-fit options;
2. **Scaling-up**: typologies can enable the up-scaling of best-fit options;
3. **Selection**: typologies support the selection of representative farms for detailed analyses.

Methods

Farm typologies can be constructed using various methods:

- **Step by step comparison of farm functioning** (Capillon, 1993; Landais, 1998): for a delimited area, this classification method is based on extensive data about farm functioning (family, objectives, history, productions, management, techno-economic results, biophysical constraints, etc), which can be obtained from surveys of a stratified sample of farms. The grouping into types is made using a “step by step” comparison of neighbouring farms (for more details on the method, see Landais, 1998).

- **Expert knowledge**: the typology construction is based on aggregating farms around cluster poles defined by local experts, key informants, or farmers (Giller et al., 2011; Landais et al., 1998; Pacini et al., 2013). This approach leads to the establishment of a common reference base (Landais et al., 1998). Because the typology is built comparing the level of similarity in farm functioning between pre-defined clusters, the farm types do not change over time. However some farm type could be deleted, new types could be added, and, farms could move from one type to another (Alary et al., 2002). Generally, the typology approach based on expert knowledge requires little time and little costs (Landais, 1998).

- **Participatory rankings**: the ranking of households, mostly according to wealth (wealth ranking), by experts and/or farmers themselves in a participatory process. The observable aspect of assets is important when ranking is based on wealth (Kebede, 2007).

- **Multivariate analysis including ordination and clustering methods**: this method can be seen as the quantitative equivalent of the ‘expert knowledge approach’. The statistical methods (e.g. Principal components analysis, Multiple correspondence analysis, Multiple factorial analysis, Multidimensional scaling) are used to classify objects (here farms). In the ideal case no hierarchy or preconceptions are projected on the objects (Alary et al., 2002, Giller et al. 2011). This kind of methods are also called “dimension reduction” or “data-reduction” techniques (Pacini et al., 2013) because they have the advantage of capturing the complexity of farming systems through taking into account, at the same time, numerous farm dimensions and then highlighting some few dimensions that are more explanatory of global farm functioning (Alary et al., 2002).
In HT, we have to meet at least two important standards: (i) the standards of science in which accuracy, objectivity and reproducibility are important, and (ii) the standards of project outcomes, which are dependent on different needs, perceptions, interests, etc. of stakeholders.

Multivariate statics methods are often preferred to typology approaches based on expert knowledge due to reproducibility inherent to their statistical foundations” (Pacini et al., 2013). However, to enhance the success of HT the different typology methods could be used in a complementary way, i.e. using multivariate statics in addition to participatory approaches (Alary et al., 2002; Pacini et al., 2013; Righi et al., 2011).
Typology guidelines

1. Typology objectives

The farm typology is dependent on the research question (e.g. Köbrich et al., 2003). Typologies can be constructed for a precise research objective for a specific area (e.g. “to improve forage supply in the highlands of Madagascar”) or for a global objective for a broad zone (e.g. “to improve food security for humid tropics”). In both cases, keeping in mind the objective is important when a typology is constructed (Tittonell, personal communication). Moreover, because farms are moving targets (Giller et al., 2011), typologies give a snapshot of farm situations at a certain period of time (Kostrowicki, 1977).

Typologies could be regrouped into two main classes: (i) structural typologies based mainly on resources and asset levels, and, (ii) functional typologies based on livelihood strategies and household dynamics (Tittonell, 2014). In the end, the main purpose of the typology use should drive the typology development process, and hence the variables selection (Pacini et al., 2013).

2. Hypothesis on typology structure

As a starting point of the typology development, it is advised to establish a hypothesis on the farm diversity of the studied area (Tittonell, 2014). The hypothesis can be structured using expert knowledge, participatory methods, previous studies in the area or field observation. It would be good if the hypothesis is based on agricultural knowledge and theories (Whatmore et al., 1987).

We propose to use participatory approaches, in which local stakeholders (local research, actors, farmers) are included in order to formulate the hypothesis together, i.e. the statement on different farm types, and on the important drivers for variability. The effectiveness of the typology development could be improved by the participation of local stakeholders in this construction process (Righi et al., 2011). The hypothesis on the farm types should be expressed by the criteria that are selected by local stakeholders familiar with the local farming systems (Alary et al., 2002; Pacini et al., 2013). These selected criteria should be part of the questionnaire used for the farm surveys coming in a following step.

An externality of including local stakeholders is that communication and involvement can be increased.

3. Variable selection

The collection data from farms is an essential step in the typology construction by determining the data quality (amount, accuracy) required by the statistical methods (here multivariate analysis). Household survey data in development countries often is erroneous (Howe and McKay, 2007), undermining the reliability of the statistics. Therefore it is good to choose your indicators for certain variables wisely. It is advised to design the survey questionnaire to capture the whole farming system (Giller et al., 2011; Tittonell et al., 2010). As mentioned previously, the criteria selected by local stakeholders as defining the farm diversity should be included in the survey questionnaires.

In our view, the typology is based on the emerging properties of farms, i.e. the things we can observe. Besides that, there are external variables that can explain the diversity in those properties. Those external variables should not be used for the identification of the farm types.

1 Variable as a common attribute from farms in the farm population that we want to study
The variables used in the farms surveys could be grouped into categories that are more or less specific; for example:

- Variables of structural characteristics and variables of farm functioning (Sanago et al., 2010);
- Farm resources availability and Management (Pacini et al., 2013);
- Biophysical resources, Socio-economic aspects and Equipment (Righi et al., 2011);
- Cultivation, Chemical inputs, Harvest, density and fruit quality, Economic resources, Farming system nature, Physical context, Personal ambition, Social, Performance Agronomical, Performance Economic and Performance Environmental (Blazy et al., 2009).

Here, in order to ensure a systematic approach of farms, we advised to consider variables related to the main compartments of the farming system (i.e. household/family, cropping system, livestock system) and their interactions with the outside/environment (e.g. environmental context, economic context, socio-cultural context). As crop–livestock farming systems are the backbone of smallholder agriculture in developing countries (Thornton and Herrero, 2001), here we propose an example of a variable set commonly used in the context of crop-livestock farming systems (Table 1). Naturally, variables should be adapted to the area and the farming system context, and extra variables could be added (cf. example of used variables in Appendix); that would be done during the first step of exchanges with local stakeholders and hypothesis formulation.

Kostrowicki (1977) argued that “irrespective of the order and area concerned, to retain the comparability of the results, the identification of agricultural types should always be based on the same criteria, whether or not they differentiate a given territory”. This makes sense in theory, but this would be challenging to achieve in a complex program like HT. However, here we propose a set of common variables (last column of Table 1) in order to facilitate the comparison between typologies results.

It is advised to use “a small number of key variables” (Kostrowicki, 1977) and to make sure that the number of surveyed farms is larger than the number of variables used for the multivariate analysis. Hence the number of key variables required and selected for the multivariate statics could differ from variables asked during the survey. Table 1 provides some variable that might be useful to include in the multivariate analysis.

The number of key variables used in multivariate statics for a typology purpose could be highly variable; from 1 to 46 variables, with an average of about 15 variables2. Furthermore, other integrative variables (e.g. ratio Labour/Land, ratio LTU/Fodder area, ratio Food crops area/Cropped area) could be created from the variables asked during the survey (cf. example in Appendix).

Moreover, it should be noted that, as a result of the multivariate analysis, not all the key variables implemented in the multivariate analysis (input variables) would necessarily turn out as discriminating variables (output variable).

---

2 Based on 21 typologies studies
Table 1: Example of variables describing crop-livestock farming systems. The column ‘common set’ proposes a set of variables that should be included in typology making across sites in HT.

<table>
<thead>
<tr>
<th>Attr.</th>
<th>Category</th>
<th>Variable</th>
<th>Unit</th>
<th>Common set</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>Household</td>
<td>Family size</td>
<td>capita</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>Household</td>
<td>Household head age</td>
<td>year</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>Household</td>
<td>Family labour on farm activities</td>
<td>capita or man-day/year (^a)</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>Household</td>
<td>Labour hired</td>
<td>capita or man-day/year (^a)</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>Household</td>
<td>Months food self-sufficiency</td>
<td>months</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>Household</td>
<td>Total gross margin of the household</td>
<td>local currency</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>Household-Environment</td>
<td>Off-farm activities</td>
<td>classes (^b)</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>Household-Environment</td>
<td>Total gross margin of the household</td>
<td>local currency</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>Household-Environment</td>
<td>Off-farm income</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>Household-Environment</td>
<td>Food purchase</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>Cropping system</td>
<td>Area owned by the household</td>
<td>ha</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>Cropping system</td>
<td>Area farmed by the household</td>
<td>ha</td>
<td></td>
</tr>
<tr>
<td>R/O</td>
<td>Cropping system</td>
<td>Area with food crops</td>
<td>ha or % cropped area</td>
<td></td>
</tr>
<tr>
<td>R/O</td>
<td>Cropping system</td>
<td>Area with fodder crops</td>
<td>ha or % cropped area</td>
<td></td>
</tr>
<tr>
<td>R/O</td>
<td>Cropping system</td>
<td>Area with cash crops</td>
<td>ha or % cropped area</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>Cropping system-Environment</td>
<td>Crop production sales</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>Cropping system-Environment</td>
<td>Purchase of mineral or organic fertilizers</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>Cropping system-Environment</td>
<td>Purchase of pesticide</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>Livestock system</td>
<td>Total number of livestock</td>
<td>TLU</td>
<td></td>
</tr>
<tr>
<td>R/O</td>
<td>Livestock system</td>
<td>Number of local cattle</td>
<td>no.</td>
<td></td>
</tr>
<tr>
<td>R/O</td>
<td>Livestock system</td>
<td>Number of improved-bred cattle</td>
<td>no.</td>
<td></td>
</tr>
<tr>
<td>R/O</td>
<td>Livestock system</td>
<td>Number of small ruminants</td>
<td>no.</td>
<td></td>
</tr>
<tr>
<td>R/O</td>
<td>Livestock system</td>
<td>Number of small animal (pig and/or poultry)</td>
<td>no. or TLU</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>Livestock system</td>
<td>Milk production</td>
<td>L/year</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>Livestock system-Environment</td>
<td>Total animal products sales</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>Livestock system-Environment</td>
<td>Manure sales</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>Livestock system-Environment</td>
<td>Concentrate/Fodder purchases</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>Household-Environment</td>
<td>Production objective/strategies</td>
<td>classes (^d)</td>
<td></td>
</tr>
</tbody>
</table>

Attr.: Attribute; O: Orientations; R: Resources

\(^a\): man-day/year could allow to consider different kind of labour (e.g. full time person, children, woman)

\(^b\): classes to defined according the set of the survey results (e.g. Yes/No, Agricultural/Non-agricultural, Agricultural/Urban)

\(^c\): farmed area could include cropped, grazing or plantation areas

\(^d\): classes to defined according the set of the survey results (e.g. Increase/Maintain/Decrease production, Increase/Improve/Diversify/Change production)
4. Sampling

The farm sampling should cover the farm diversity of the studied area (Pacini et al., 2013). Therefore it should be done more or less randomly (e.g. Y-shaped method described by Tittonell et al., 2010), based on a stratification or on transects (e.g. “transect following an intensification gradient” used by Pacini et al., 2013). It should be noted that the selections based on space (e.g. Y-shaped method, transect line), might bias towards the selection of farms with larger land sizes, since they occupy more space (Tittonell, 2007). Besides, it is not recommended to ask to “all farmers” to come at a meeting place to make them fill the survey: the farm sample could be biased by the ability and/or motivation of farmers go or not to the meeting place.

Usually, for statistical reasons it advised to sample at least 30 farms. The sample of farms for typology studies could range from 18 farms to 2746 farms, with a median of 138 farm surveyed\(^3\).

5. Multivariate statistics

Multivariate and cluster analysis are used to identify explanatory variables (output variables) and to group farms in homogeneous types. Multivariate statistics allow reducing the number of variables and preserving the maximum of the total variability of the sample. According the nature of the selected key variables (quantitative and/or categorical) different multivariate statistics should be used:

- **Principal components analysis** (PCA) for quantitative (continuous or discrete) variables (e.g. Bidogoza et al., 2009; Sanogo et al., 2010; Tittonell et al., 2010);
- **Multiple correspondence analysis** (MCA) for categorical variables (e.g. Blazy et al., 2009);
- **Multiple Factorial Analysis** (MFA) for categorical variables organized in multi-table and multi-block data sets (Alary et al., 2002);
- **Multidimensional scaling** to build a classification configuration in a specific dimension (e.g. Pacini et al., 2013; Righi et al., 2011).

As it is mentioned previously, the first steps of the multivariate analysis concerns key variables selection. It could be necessary to check that a certain category of variables is not over-represented (i.e. the number of variables in this category is much larger than for the other categories); otherwise that could give more weight to this category of variables and so bias the analysis (Blazy et al., 2009; Kostrowicki, 1977).

Another precaution, for example for PCA, could be to standardize all the (quantitative) selected variables using percentages “to avoid the influence of different levels of variation due to the unit of measurement” (Pacini et al., 2013). The MCA and MFA methods are sensitive to low numbers of observations or unbalanced classes; hence it could be necessary to combine some classes. Furthermore, it would be required to test the independence of variables, with for instance Pearson’s Chi-squared Test. Moreover, MCA could be more difficult to interpret than the PCA or MFA analysis; hence, it would be useful to limit the number of selected variable (≤ 20 variables; Hervé, 2011).

‘Exceptional farms’ could be removed from the multivariate-analysis in order to reduce the noise/variability produced by these farms (Tittonell, personal communication).

---

\(^3\) Median observed on a sample of 22 typology studies
When it comes to choosing the number of axes (i.e. principal component or factors) for the PCA, MCA and MFA, it could be determined according to a criterion fixed before the analysis (e.g. the number of axes that explain a minimum of x % information or total inertia) or using the Kaiser’s criterion for the PCA (i.e. all axes with an eigenvalue higher than 1 are chosen; Hervé, 2011).

The Cluster Analysis (CA) aims to group farms into classes/types that are the most "homogeneous" possible. There are two main methods of CA commonly used:

- **Non-hierarchical clustering**, i.e. a separation of observations/farms space into disjoint groups/types, where the number of groups (k) is fixed;
- **Hierarchical clustering**, i.e. a separation of observations/farms space into disjoint groups/types at every moment (first each farm is a group all by itself, then at each step, the two "closer" groups are merged until only one group with all farms remains). The visual result of these steps (algorithm) is a dendrogram. Agglomerative Hierarchical Clustering algorithm is often used in the typology construction process (e.g. Alary et al., 2002; Blazy et al., 2009; Pacini et al., 2013; Sanogo et al., 2010). The height of the dendrogram branches represents the average distance between the groups. Different methods could be applied to compare the distance or inertia between the groups: “single linkage”, “complete linkage”, “average linkage” or “Ward method”.

![Dendrogram](image)

**Figure 2: Example of dendrogram from Agglomerative Hierarchical Clustering on 55 farms and showing on five farm types**

6. **Hypothesis verification**

The number of farm types could range from 3 to 7, with a median of 5 farm types. Farm groups should be selected on basis of their explanatory value, i.e. they have to be conceptually meaningful (Moreno-Pérez, 2011). The farm types resulting from the multivariate and cluster analysis should be compared with the initial hypothesis. It is necessary to discuss and to try to understand the observed differences between the hypothesis and the multivariate analysis. In case of unexpected results, the multivariate and cluster analysis may need to be repeated and/or to restart the discussion with local stakeholder with the typology feedback.

---

4 Data observed on a sample of 20 typology studies
Preliminary discussion

General remarks
To avoid a bias in the variable selection another option could be used: intermediate typologies per variable set can be constructed (e.g. Maton et al., 2005; Morena-Pérez et al., 2011). Especially for a global typology, that approach dissects the problem and will help to explain the relative importance of a typology component in the eventual typology. However, it implies that there is not a single hypothesis, but multiple hypotheses that eventually are combined in one.

Because farms are dynamic, typologies could quickly become obsolete and hence it is preferable to use updated typologies (Landais, 1998; Valbuena et al., 2014). Within HT we need to take that into account. Another point of attention in HT is that we might face data scarcity and narrow-timing. In that case a “simple classification” based mainly on resource endowment (structural typology), might after all be the best option (Giller et al., 2011).

For prioritizing within action sites and/or field sites and for comparisons between areas it is important to conduct an analysis to identify the household types that are below the poverty line. Some studies (Davis et al., 1997; Howe and McKay, 2007; Tittonell et al., 2010) estimate which household types are below the (local) poverty line. In addition, Howe and McKay (2007) make a distinction between types that are situated in a poverty trap, and those who are not. In regard to dynamics that gives at least some idea of (im)possibilities of households to escape poverty.

Several articles plead for spatially linked farm typologies (Landais, 1998; Carmona et al., 2010; Madry et al., 2013). However, doing spatial exercise, requires a data rich environment, which is often not present in developing countries (Carmona et al., 2010). However, land use planning in development countries would benefit from it greatly (Carmona et al., 2010).

Some scientists think that the further fine-tuning of either statistical or participatory methods will not result in better representation of reality. In other words they believe that the development of typology methodologies is a dead end, because the main objection of typologies is not counterfeited: the loss of information.

Further development of this document
This is a draft version for a typology guideline. Apart from improvements of the existing content, we would like to know what could be added.

- Should we dedicate some space for gender and/or nutrition specific typology making? Especially in regard to variable selection?
- Would you like to have some references of people in HT and/or CG-centres, to approach for typology making?
- Should we elaborate more on new developments in dealing with farm variability?
- Is it necessary to deal more with farm dynamics and the spatial link in this document?
References


Tittonell, 2014, Categorising diversity of smallholder farming systems Household typologies. FSE staff seminar, 30 Jan 2014, Wageningen.
