The use of unit-level census data for research on poverty: a multiscale approach.

G. Leclerc, A. Nelson and E. B. Knapp.

International Center for Tropical Agriculture (CIAT), Colombia.

Abstract.

Unit-level data (individual, household or farm) from the 1988 population and 1993 agricultural censuses of Honduras have been integrated into a Geographic Information System (GIS). We showed how poverty indices can be computed for different scales of aggregation from village to country, and how they compare to other published figures. Indicators derived from the analysis of well-being ranking by local informants in 90 communities have been extrapolated to the entire country by means of proxy indicators computed from reasonably well correlated census data. We found that the choice of the indicator as well as the scale of analysis result different geographical representations of distributions of poverty, which may affect significantly the relevance and impact of poverty alleviation policies. We briefly introduce spatial statistical methods to process the data on a given scale, which allows analyses of correlation with other factors significant on the same scale. The same methods also help to detect errors in the data or to determine the optimum scale of a particular indicator.

1. Introduction

Since its inception, CIAT has focused on alleviating hunger primarily through increases in food production. Over the years, implicit assumptions linking progress in food production and the broader human conditions of "well-being" and "poverty" have been called into question[1]. The questions resulted in CIAT committing itself to a better understanding of the dynamics of well-being and poverty. Our specific needs include improved targeting of agricultural and NRM research, providing robust means to monitor project impact, and contributions to more informed decisions at all levels of agricultural land use planning and governance. This subject is so central to CIAT research that it is almost its raison d'être. In effect, CIAT mission statement reads: "To contribute to the alleviation of hunger and poverty in tropical developing countries by applying science to the generation of technology that will lead to lasting increases in agricultural output while preserving the natural resource base".

The focus of this paper is to present on-going work that examines methodological issues related to the measurement and geographical characteristics of poverty. Although a large body of literature can be found on the subject, traditional methods address facets of poverty not easily related to agriculture and NRM decision-making (Carvalho and White, 1997). In 1997, supported by a research grant from the International Development Bank, CIAT embarked on a research project that would define a unique approach to linking *ad hoc* measurements and geographical representations of poverty from community-level, locally constructed "well-being" rankings (Ravnborg *et al*, 1998) to standardized maps of national-level rankings. Contrary to proposing a single, unifying poverty index, we support the design of unique indexes targeting needs of specific decision-makers. However, a prerequisite for catalyzing collective action among all stakeholders is a shared vision, and shared visions cannot be created and communicated using unrelated component images.

This paper is organized in three parts. We begin by giving an example using household-level national census data supplied by our collaborating partner, the Honduran government agency DGEC. We show how the richness of this representative national census can be exploited to produce poverty indexes tailored to particular needs. We then introduce some results of an independent study that characterized 90 Honduran villages using locally identified indicators to derive locally relevant rankings of "well-being". We then "link" the two independent, *ad hoc* databases using the methodology of neural networks. The result is an example of a "common knowledge-base" that can bridge the communication gap from international and national perspectives to local community perspectives. Lastly, we demonstrate that different representations and interpretations of indexes can occur if consideration is not given to examining explicit relationships

---

[1] For reasons of simplicity, throughout this paper unless specifically noted, *poverty* and *well-being* will be used interchangeably to refer to a broadly defined but intuitively acknowledged human socio-economic condition. When "well-being" appears in parentheses, it will refer to the specific index proposed by Ravnborg (1999).

between census variables and possible scales of aggregation chosen to simplify analysis and presentation of results.

2. The Honduras Population, Housing, and Agriculture censuses.

The 1988 Honduras Population and Housing census is the most recent and complete data set about every single person and household in the country. It gives a panorama of the composition of the Honduran society and of the life conditions of its inhabitants in 1988. It contains answers that the 4,255,105 individuals gave to questions related to its education level, profession or vocation, family composition, age, mortality, migration, housing type and construction materials, ownership type, water supply, assets, etc… In total, 42 variables for 891,298 households, and 49 for each individual, in addition to 9 variables related to administrative localization of the household. The data collection phase of these censuses take only one day (it is done by an army of civilians, students, etc..), but then it takes over a year to prepare and another year before the results are published.

The 1993 Honduras Agricultural census is also the most recent data set to cover virtually every farm in Honduras (317,187 to be precise). In total, 161 variables covering land ownership, agricultural production, technology, and labor, as well as 6 variables about the farmer, and 8 variables related to administrative localization of the farm. The data collection, based on a statistical sampling, is done by government employees over a period of a few months. Many people state that agricultural censuses are error-prone, as farmers will avoid giving to government officials, detailed information that would give the government a chance to invade their privacy.

The census results are compiled at municipio level, in tables distributed within several thick books. This tradition is likely to change in the near future, as most developed Latin American countries can provide municipio-level census data on line or on CD. CIAT is currently implementing a project in 6 Central American countries, to help the governments to develop digital data products for public distribution.

In collaboration with de Estadística y Censos (DGEC) we have obtained access to the census at unit-level, and loaded the entire data set in an Oracle database. Confidentiality was ensured by omitting the names of the individuals.

3. Deriving indices from household census data.

1) Background.

The methodology we followed here draws from the traditional unsatisfied basic needs (UBN) approach, which has been the one followed for at least 11 countries in Latin America (UNDP, 1992; Boltvinik, 1996) because it incorporates important variables for the formulation of social policies. It involves the selection of a certain number of needs, the definition of a minimum criteria to satisfy for each need, and the combination into poverty indices. Therefore, according to this approach, poverty is linked to a state of necessity, a deficiency or deprivation of the goods and services necessary to sustain life to a minimum standard. In the Latin American practice, the UBNs are generally a set of poverty-related indicators: large number of people sharing a room; improvised or inadequate housing; inadequate water supply and inadequate sewer systems; low school attendance for children; and, household capacity to generate income. It is supposed that other factors such as lack of participation in collective decisions, social marginalization, discouragement, etc… are correlated to UBNs. We followed a scheme very similar to the one adopted in the elaboration of the "Mapa de Pobreza" (Republica de Bolivia, 1995), a multi-institutional effort that took advantage of unit-level census data to produce a very complete set of poverty data and maps for Bolivia. More details can be found in Oyana *et al*, 1998.

2) Methodology.

The UBNs are computed for each household, then aggregated at village, municipality or department by counting the fraction of the population in a particular UBN stratum. The variables considered to build the UBNs are labeled x_j, the subscript *j* representing the household, and *x* the variable. For certain variables, such as the education level of a household *j*, the value is computed for the household from the value for an each individual *i* forming the household.

First, we have to define x^*, the <u>acceptable value for variable *x*</u>. This is where the knowledge of the area and the local/national economy play a crucial role. It is also at this step where subjectivity (and gross errors) can occur and lead to conflicting conclusions. For the current example, the norm we used for a given variable was given by the average value of that variable for the country. I that sense, the poverty measure that we are developing here is more one of equity, which can help orient an internal social reform.

Second, we define an <u>indicator of success in obtaining,</u> for variable *x*, <u>the level defined for *x*</u>*. This indicator, lx_j, can be expressed as:

$$lx_j = \frac{x_j}{x^*} \qquad lx_j > 0$$

Third, an <u>index of failure in obtaining *x**</u> for household *j*, cx_j, is computed as follows:

$$cx_j = 1 - lx_j \qquad -1 < cx_j < 1 \text{ ideally}$$

The *cx* are normalized between –1 and +1 to allow comparison. To obtain this ideal range for the cx_j , each variable is normalized between its minimum and maximum value (for all households). If $cx_j < 0$, we divide cx_j by min(*cx*) and if $cx_j > 0$, cx_j is divided by max(*cx*). Put in other terms, one can interpret the cx_j as a distance between current conditions and the desired future condition defined by x^*.

3) Household-level indices derived from housing and population census.

The compound indices NBI_3 (a combination of 3 indices) and NBI_4 (a combination of 4 indices)[2] are obtained for each household by averaging several more specific indices, which themselves are the result of the combination of more basic ones (i.e. the *cx)*. This is detailed below (Figure 1).

For each household *j*, we define:

$$NBI\_3_j = (CV_j + CSIB_j + CIA_j)/3$$

$$NBI\_4_j = (CV_j + CSIB_j + CIA_j + RE_j)/4$$

where: CV_j = lack of housing size and quality
$CSIB_j$ = lack of basic services and energy
CIA_j = lack of non-land assets.
RE_j = lack of education

CV_j , the index of lack of housing size and quality, was derived from an index of the size of the house CEV_j,, and from an index of housing quality CMV_j:

$$CV_j = (CMV_j + CEV_j)/2$$

CMV_j is the average between lack of wall quality (cm_j), roof quality (ct_j), and floor quality (cp_j).

$CSIB_j$ the index of lack of basic services and energy source, is the average between the lack of basic services CSB_j, and of energy source CE_j :

$$CSIB_j = (CSB_j + CE_j)/2$$

CSB_j is computed as the average of water source quality cag_j, lack of water supply infrastructure ctu_j, and lack of latrines csa_j. CE_j is the average between the lack of lighting cal_j and of fuel cco_j.

CIA_j, the index of non-land assets, is derived from three indicators: the lack of househølod appliances (CBA_j), of telecommunication (CCA_j) and of means of transport (CTA_j). The first is the average of lack of sewing machine cm_coser_j, of refridgerator $crefrigerator_j$ and stove $cestufa_j$. The second is the average of the lack of radio $cradio_j$

[2] NBI is the Spanish acronym for UBN (Necesidades Basic as Insatifechas)

and television $ctelevisor_j$. The third is the average of lack of car $cautomovil_j$, of motorcycle/moped $cmotocicleta_j$ and of bicycle $cbicicleta$j. CIA_j is then computed as:

$$CIA_j = 0.25 \times CBA_j + 0.4 \times CTA_j + 0.35 \times CCA_j$$

The choice of weight in this equation is clearly a question of personal preferences or interests.

RE_j, the <u>index of lack of education</u> for each household, is computed from data from individuals i belonging to household j. The index of success of the individual within the household, $ane_{i,j}$, is computed as follows:

$$ane_{i,j} = (ap_{i,j} + as_{i,j}) \times al_{i,j} / (ap^* + as^*)$$

where:

$ap_{i,j}$ is the number of years of schooling
$as_{i,j}$ is the index of school attendance in function of age,
$al_{i,j}$ is the index of litteracy
ap^* is the norm for school attendance in function of age,
as^* is the norm for student status.

The index of education deficiency for each individual, $re_{i,j}$ is simply given by:

$$re_{i,j} = 1 - ane_{i,j}$$

Finally, RE_j is computed as the average of the $re_{i,j}$ for household j.

4) Aggregation of household-level indices.

Household indices georeferenced for each village can be aggregated at virtually any scale given predetermined boundaries: it can be village, watersheds, "eco-regions", municipalities, department, or country. We can produce mean or median values of the poverty index, or count proportions of the population considered as poor. For our example, we chose to define 2 indices, NBI_3, and NBI_4 as the proportion of household, for a given aggregation level, which poverty index is below 0.4. This was done only when more than 50% of the households poverty index could be characterized. In effect, there are cases where the data is not complete and do not allow to compute an index from all the variables. In section 6, we analyze the geographical distribution of missing data. The results of aggregation a village, municipio and department levels, is presented in Figure 2. One can immediately see that depending on the scale, or the poverty index, the map (and the message it conveys) changes drastically. We cannot emphasize enough how critical is our choice of poverty measure and the scale of analysis and action. There is no mechanism for government intervention at village level, but this doesn't mean that decision-maker shouldn't be aware of the implications of working with data aggregated to a scale imposed by administrative boundaries. What we have experienced, though, is that household and village-level databases are so extensive that they become very difficult to manage and interpret. In section 6 we introduce tools that can help extract important features from these data sets.

All the steps to process unit-level data into NBIs are realized through a series of Oracle procedures developed in PL/SQL language scripts that allow for full automation. There is very little to do to put the power of the raw census in the hands of any user through the Internet. A simple Java interface can easily provide to a remote user the capacity to produce a poverty index for a special-purpose thematic, through SQL queries with any variables of interest, any weights or ways of combining them, and the choice of any aggregation level, on a central computing facility (Openshaw, 1995).

But in addition, it is possible to address how certain variables such as the ones used for the definition of the NBIs are related. For example, we found that housing is an indicator that explain well other factors by analyzing the correlation (at municipio level ; n=291) between CV, CSIB, CIA, and CE. Table 1 shows the correlation coefficients obtained.

[TABLE 1]

We compared the NBIs to other published poverty measures, i.e. SECPLAN (1991) and FHIS (1992), department-level estimates (n=18). Although the NBIs figure are consistently higher, results are strongly correlated (Table 2). This is not entirely surprising since the same census data has provided part of the information used by SECPLAN and FHIS.

[TABLE 2]

4. The "well-being" index.

We briefly recall the work by Ravnborg (1999) on "well-being" ranking and on scaling-up "well-being" indicators. The author conducted a traditional participatory "well-being" ranking as a designed experiment, which allowed for extrapolation to areas different than the ones studied. This was a strategy to avoid what Rhoades (1999), in a discussion about participatory methodologies, describes as "the social under design of projects". Instead of seeking to identify "representativeness" i.e. find "standard" villages in which to conduct the study, the aim was to select a set of contrasting villages. This would maximize the chance of obtaining all possible indicators, but also would allow to conclude, if some indicators are found across all communities despite the dissimilarities, that these indicators could be valid for all communities from which the sample was taken.

First, assumptions were made with respects to factors that would influence well-being in Honduras, and a sampling was designed. Sites (villages) were selected so as to represent as many combinations of the 6 factors chosen: altitude, basic services (education and water), population density, ethnicity, gender composition, and accessibility to urban centers (>2000 inhabitants).

These factors were combined for every village, from census data and a GIS database, and a sample of 90 communities in 3 departments was obtained. In theory the indicators of "well-being" obtained are valid for all villages that have the same combination of factors as the ones used in the sample. In practice, since there was consistency I the indicators even for contrasting villages, it is likely that the extrapolation domain is much larger.

The well-being ranking, a technique for obtaining insights into local perceptions of "well-being" –and by inference, poverty- (Grandin 1988), was done in the 90 communities that formed the sample. For each community 3 to 5 informants, with different age, gender, occupation and ethnicity, are selected. This is to avoid the informant-related bias typical of this type of studies (Bergeron *et al*, 1998). They are asked to examine a set of cards, each of which representing a household, and group the cards into piles (maximum 3) according to their perception of the well-being or quality of life or the households. Generally, we end-up with one pile for the poor, one for the not-so-poor, and one for the non-poor, according to how the informant perceive poverty. These categories are of course only valid for the community, and not extrapolable to other ones. The informants are then asked to describe the content of each pile in terms of their differences with the other piles.

The descriptions are the base for the identifications of "well-being" indicators, which are reinterpreted and made quantifiable by means of a standard questionnaire. The authors obtained, from the 316 descriptions of well-being, almost 400 indicators, that were subsequently reinterpreted and reduced to 11, a priori valid at least within the set of communities from which the sample was drawn. These indicators were subsequently transformed into quantifiable ones, which are summarized in Annex 1. Once the score is given to each indicator for a household, the resulting "well-being" index is simply the average of the score of all indicators. The questionnaire is straightforward, and can be used to obtain quickly and inexpensively a poverty profile for a region of interest.

Ravnborg ends-up with an index which doesn't look that different than other published ones (such as NBIs), but which has a major advantage: it is entirely based on the perceptions that the poor have about poverty. In a way, it is a message from the poor about what really matters to them, that they are addressing to decision-makers. The extrapolation and mapping of the "well-being" index is no more than a translation of this message into a language more familiar to decision-makers.

2) A poverty profile for selected communities.

The questionnaire has been employed by Ravnborg and her team (Escolán Rodezno *et al*, 1998) to quantify the "well-being" of 768 Households, as part of a larger study to identify factors that lead to certain preferences related to agriculture and NRM (ref this conference). The households were selected at random, and belong to 12 communities, distributed among 3 hillsides watersheds located in distinct social and climatic environments. Important hypothesis can be drawn from detailed analysis of the distribution of indicators and well-being levels at various aggregation levels. We used *WBI*, the raw "well-being" index (before it is classified into 3 categories) to generate histograms, i.e. the number of households having *WBI* between certain ranges. The scale goes from 33 (higher well-being) to 100 (lower well-being). We found that the distributions were unimodal at all scale, have very similar ranges, and appeared skewed towards lower or higher well-being depending on the location and aggregation level. Therefore we do not observe, from this data set, curious poverty distribution that cannot be modeled simply (see next section).

On Figure 3, we can appreciate how the WBI compares to the NBI_3 and NBI_4. On the horizontal scale, we have the proportion of households which WBI is above 67 (i.e. the ones having the lowest "well-being" level) for the 12 communities. We do observe some correlation but no perfect match, which is expected as these indices measure different aspects of poverty. A word of caution: the problem of a low correlation between different poverty indicators has been observed in numerous cases (Henninger, 1998). This means that generalizations, which are very tempting because of the ready availability of nationwide data, may well be inconlusive.

5. Neural nets for extrapolation

In this section, we present a new methodology to extrapolate and map at country level indicators obtained at local level. To successfully apply the method to Ravnborg's WBI we have to consider 2 constraints.

First, the most detailed scale of nationwide georeferenced data is at the village level, which means that we cannot map census data at a finer scale.

Second, we cannot identify in the census the exact households that were surveyed by Ravnborg and her team, to respect confidentiality of the census. This means that we cannot calibrate our model at household level, which would have been straightforward. This is a similar situation as experienced by Bigman *et al* (1999) for poverty targeting in Burkina Faso. In this case, the well being was given from a Priority Survey (PS) of sample communities, and the only data available for extrapolation outside the PS sample were mean values of explanatory variables. In our case, we have all the data (survey as well as explanatory variables) at household level, so we can go a little deeper in our analysis by comparing the distributions of WB within a village to the distribution of explanatory variables.

1) Linking well-being to proxy variables

Taken together, population, housing and agricultural censuses should provide the equivalent of Ravnborg's 11 indicators. It is very unlikely that we find exactly the same indicators in the census, but we can find reasonable approximations. However, there is a rather strict definition of how these indicators are quantified into 2 or 3 categories and combined to give the WB index. In effect, the WB questionnaire has well defined questions that allow well defined calculations. Let us take the example of the indicator of Market Participation (PAGRICUL):

PAGRICUL	33	If the household grows coffee or cacao (café= 1 or cacao=1) or if the household does not buy basic grains and sells half or more of its production of basic grains (com_grab=2 and grab_uso>=3)
	67	If the household does not grow coffee or cacao (café=2 and cacao=2) but the household buys basic grains and at the same time sells at least part of their basic grain production (com_grab=1 and grab_uso>1) or if the household doesn't buy basic grains and less than less than half of its basic grain production is for sale (com_grab=2 and (grab_uso=1 or grab_uso=2)
	100	If the household does not grow either coffee nor cacao (café=2 and cacao=2) and the household buys basic grains at the same time as all what it produces is for home consumption (com_grab=1 and grab_uso=1)

Clearly, it is impossible to find this exact indicator in the censuses, neither is feasible to construct it by a linear combination of variables found in the censuses. This is the type of situation where artificial intelligence methods can be successfully applied. Economists usually account for non-linear relationships by using sub-models to fit the data, and then explain why the data is what it is by adjusting model parameters. An example of this approach, based on a consumption model, can be found in Hentschell *et al* (1998). In the case of poverty, we can imagine that such a model would be extremely complex, and we might not have enough data to calibrate it. By using artificial intelligence techniques to fit the data, we obtain an empirical model, which can be used to run simulations ("what-if" scenarios) with limited data availability.

We used an advanced neural net software package from Ward Systems (1999), which uses a strategy known as genetic algorithms (GA) to find optimum solutions. The advantage of Gas is their insensitivity to correlation between input variables, and their ability to find the "fittest" variables, i.e. the variables that resist the best to noisy data, or to the removal of one input. The neural net is constructed from a series of input variable, and one output variable (here a well being value or category), then is trained by being presented a series of cases, i.e. a combination of input variables and their associated output. Once the training is completed, new combinations of input variables are presented to the trained neural net, which turns this data into a predicted output.

Our hypothesis is that the proportion of poorer households in a village can be determined from the proportions, for every indicator, of households for which the value of the indicator correspond to the condition of the poorer. This is exactly equivalent to the so called Headcount index (Deaton, 1997). If we take the example of PGANADO (cattle ownership), the limit between the poorer and the richer correspond to 2 cows, so the proportion of poor farmers, according to this indicator only, is the number of farmers with less than 2 cows, divided by the total number of farmers. We repeat the procedure for all indicators, and end-up with a series of values for each village. The communities where the WB index has been obtained form our calibration set.

To start we decided to redefine each of the 11 indicators so that they will represent only two states: lower and higher "well-being". In other words, we can set a threshold for each indicator, that will result in a more sensitive model. We reinterpreted Ravnborg's homogeneity plots (Figure 4) to obtain and indication on what this threshold is for each indicator. We delimited the boundary between lower, middle and higher "well being" categories (thick solid line). As we can see, the distinction between the richer and middle is much better defined than between the middle and lower "well-being" categories. Prediction of middle "well-being" would certainly generate confusion with the lower "well-being". We can however draw a line between the two boundaries, that allow to determine what is the value of each indicator that separate the higher and lower "well-being" (dashed line). We found that most of the time, this division corresponds to the middle "well-being" category, but in some cases it is different.

The censuses were then screened to identify which variables would provide indicators (IC –"Indicator from Census") that most closely resembled the 11 of Table 1 (that we will denote as *I*), and ended with 9 summarized in table 2. For each indicator *IC*, we counted the proportion *P_IC* of people, households or farms, which correspond to the poorer condition (i.e. above the indicator threshold line), for each village. We calibrated the model for the 12 communities where the well-being index was computed. To do so, we computed P_WBI_j, the proportion of households in village *j* which *WBI* is above 67. To test the robustness of the method we also computed, from the 11 indicators *I* used to compute the *WBI*s, the proportions *P_I* of household for which the indicator *I* has a value above 67, as follows:

$$P\_I = [n(I{=}100) + n(I{=}67)/2] \,/\, N$$

Where *N* is the total number of households for which the *WBI* has been obtained, *n* is the number of household which satisfy the condition *I*=100 or *I*=67.

Now the big challenge is use the 11 indicators (the *P_I*) and subsequently the 9 indicators derived from census data (the *P_IC*) to predict the proportion of households which *WBI* is above 67 (the *P_WBI*). We expect these indicators to be correlated or redundant in certain situations, to have different weights depending on the social structure of the community (i.e. the combination is not unique), and that the relationship with respect to the *P_WBI* be non-linear. We also expect the data to be noisy, since neither the census nor the questionnaire is perfect.

We summarize in Table 3 the 9 proxy indicators found in the census. The two indicators that are missing from Ravnborg's 11, i.e. Food Security and Savings, are somehow embedded in the other 9 and should then indirectly contribute.

[TABLE 3]

2) Results.

We used the data set made of the 9 P_IC as inputs and of the P_WBI for the 12 communities to train the neural net. We obtained a model after 271 generations, which goodness of fit statistics are summarized in table 4.

[TABLE 4]

The scatter plot of actual P_WBI and predicted P_WBI is shown in Figure 5. The relative importance of the input indicators is given in table 5. We see that House, Health, and Income were the most important variables. The predictions were much worse, though, when only these three indicators were used. The importance of relative weights has to be interpreted with caution. A low weight may as well mean that the data is not reliable enough to have a good predictive power.

[TABLE 5]

Now that it is calibrated, we can apply the neural net to the entire set of P_IC obtained for all of Honduras where the 9 indicators could be computed (i.e. 3435 villages out of a total of 3730). From these 3435 villages, there are only 5 for which we couldn't predict the P_WBI because the data was too noisy. Note that a characteristic of this neural net /GA combination is that it cannot predict a value outside the range of output values used to train. In that sense we can say that it makes conservative predictions. For our case, the predicted values for *P_WBI* will always fall within the range 0.265- 0.714, which means that for villages with less than 27% or more than 71% poorer households, this proportion will be predicted as 27% and 71%, respectively. This is the case of at most 17% of Honduras villages, since there are only 465 villages which have a predicted P_WBI equal to 0.265 and 124 where this value is 0.714.

The village-level *P_WBI* were then aggregated to municipio and departamento level as follows:

$$P\_WBI = 1/N \, \Sigma \, n_j \, P\_WBI_j$$

Where n_j represents the number of households in village j, and N the total number of households in the aggregation unit.

Figure 6 presents the resulting maps of the P_WBI. The department and municipio maps are quite homogeneous, and highlights the marginalization of the Zona Norte. The village-level map presents visible clusters, which corresponds to micro-regional effects. There is little resemblance with the NBIs of Figure 2, but we have to keep in mind that these indices present different interpretations of what poverty is related to. Again this stresses the importance of the indicator used, and the necessity to enable the production of poverty measures adjusted to our specific needs, or to our capacity to induce change. In other words, poverty "measurement" and "policy" issues are inseparable.

6. Geographic analysis of poverty. First steps.

In this section, we briefly introduce two geographic analysis methods that we have tested on a range of problems, including poverty. As stressed above, the choice of aggregation scale matters as much as the choice of indicator. Aggregation at predefined scales such administrative boundaries impose a great deal of difficulty when cause-effects relationships have to be evidenced. Let us take a simple example to illustrate our point:

let suppose that we want to estimate the vulnerability of the poor to health risk, for example malaria in the coastal areas of Honduras. The distribution of mosquitoes doesn't follow administrative boundaries, so if poverty data is aggregated at department level, we expect a low correlation. With poverty data mapped at village level, we can examine only the villages that fall in malaria-affected areas, and compare with other similar areas where malaria is not present. Correlation, if any, will then appear clearly. This illustrates the concept of matching scales: data comparison should be done at a similar scale, e.g. global change may affect the entire country, infrastructure development may correspond to geomorphological units, etc…

Of course, if the central government transfers major decision power to the municipalities, this scale may be the appropriate one to analyze, say, policy reform.

In the next examples, we describe two approaches that help highlight the structure hidden in the data, as well as the matching of scales .

1) Poverty and Environmental risk

In this example, we start with a classic hypothesis: "poverty is related to environmental risk". We use the NBI_3 and NBI_4 (same as NBI_3 but includes also Education – see Figure 1) as poverty indicator, and Water Balance as an indicator of environmental risk. Our approach is inspired by Skidmore (1998) with additional considerations on random sampling. Essentially, we start with a map containing a certain number of categories (here areas corresponding to water balance ranges). We count the number of poor villages within each closed area, and compare with an estimate of what this number should be if there was no correlation (i.e. random distribution). If the number of villages exceeds what is expected from random sampling in this area, we can conclude that a village has a non-null probability of being poor if located in this area. Figure 7 shows the results obtained by averaging the probabilities computed from monthly water balance maps (Figure 7a). On the radar plots of Figure 7b, we find that probabilities, around 5% in the case of NBI_3, increases to around 10-15% when the poverty index chosen is NBI_4. This is surprising: education should not, a priori, correlate well with water balance! In addition we see that there are, more poor villages in areas and periods where the water risk is non-existent (the green lines); to the contrary: these are potentially very productive areas. This has a simple explanation: children form a good part of the working force in Honduras agricultural areas, and the number of dropouts is alarmingly low. In fact the World Bank financed a basic Education Project with the Honduras Instituto de Desarrollo Agrario in 1995, which strategy was: "…To reduce the high dropout rates among rural children, the project…..will adjust the school calendar to take into account the harvest period in agricultural regions. To serve indigenous children, it will offer bilingual programs". This spatial data analysis has permitted to highlight a much more complex phenomenon.

2) Spatial clustering: the Geographic Analysis Machine.

Our second example is the application to Honduras census data of the Geographic Analysis Machine (Openshaw, 1987a,b), a powerful public-domain tool that has been developed to identify significant spatial clustering from point data. It has been applied, for example, to the difficult problem of locating significant clusters of rare disease cases. It is essentially a multiple statistical testing on a population of points in space. GAM works by examining a large number of circles of varying sizes distributed on a regular grid covering the area of interest. The scale of analysis is therefore determined by the circles radius. For each circle, data is retrieved that represent a population "at risk", and a population of cases for which we want to determine clustering. In the case of poverty, we may use the number of households as population at risk (of being poor), and the number of poor households. Then a test of significance is applied to compare both distributions. If the number of poor, in our example, correspond to a sampling of the population of households, then we can say that there is no significant clustering. If there are significantly more poor than suggested by the sampled population, the degree of significance is assigned to the location corresponding to center of the circle. The procedure is repeated for all circles, and this generates a surface representing the degree of significance of clustering. Figure 8c,d show the results obtained for the village-level NBI_3 and NBI_4, for circles with up to 20 km radius. These maps highlight significant clusters localized towards the south-west, which may orient policy reform in theses areas. They can be used to study correlation of poverty with independent variables at a similar scale. On figure 8a,b we show the result of GAM applied to identify clusters of villages where data was not sufficient to compute the NBIs. In this case we see that the clusters correspond mostly to densely populated areas.

7. Conclusion.

Everyone agrees that poverty has many facets, is complex, and relative. Despite this apparent consensus, we are constantly torturing ourselves (and our databases) to obtain a number, one poverty measure, on which we can base our reforms, or target our investments. By default, we may agree on this number because it is too complicated to define a poverty measure tailored to our needs. We showed that it is possible to derive complex indices from unit-level census data. On the other hand poor farmers, which are living with another reality of poverty, may not see social investments (such as flushing toilets) as a way to end their poverty. The work that we have presented here, who links a measure of local indicators to nationwide databases, may contribute to bridge the knowledge gap between decision-makers and poor farmers. It is based on our belief in the use information with a clear purpose in mind, and in powerful methods and tools sufficiently flexible to permit linkages between and across scales.

8. Acknowledgements

We would like to thank the Dirección General de Estadística y Censo of Honduras for providing the unit-level census data used in this study; to Patrice Couillaud, Tonny Oyana, Hector Barreto and Pedro Jimenez who contributed to quality checking of the databases and related research; to SECPLAN who provided the georeferenced village database. We would like to acknowledge the contribution of Stan Openshaw who has generously put in the public-domain several spatial analysis tools that he developed.

9. References

Bergeron G., Sutkover Morris S., Medina Banegas J. –M. 1998. *“How reliable are group informant ratings? A test of food security rating in Honduras”*. Discussion Paper No 43, Food Consuption and Nutrition Division, IFPRI.

Bigman, D., Dercon, S., Guillaume, D., and Lambotte, M. (1999). *“Community targeting for poverty reduction in Burkina Faso”*. Manuscript (private communication).

Boltvinik, J. 1996. *“Poverty in latin America: a critical analysis of three studies”*. Inter. Soc. Sci. J. (GB) 148:245-260.

Carvalho, S., White, H. (1997) *“Combining the quantitative and qualitative approaches to poverty measurement and analysis. The practice and potential.”*. The World bank Technical Paper No. 366. The World bank, Washington. 30p.

Deaton, A (1997). *“The analysis of household surveys: a microeconomic approach to development policy”*, The John Hopkins University Press, Baltimore, 479p.

Escolán Rodezno, R., Méndez Castellanos, M. A., Mendoza, F, and Ravnborg, H.. 1998. *“Desarrollo de un perfil regional de pobreza para tres cuencas en Honduras: Rio Saco en Atlantida, tascalapa en Yoro, y Cuscateca en el Paraiso, y para el departamente de Yoro”*. Informe de Trabajo. CIAT. 81p.

FHIS (Fondo Hondureño de Inversión Social. 1992. *“Presupuesto de inversión 1992-1993”*. Presidencia de la Republica de Honduras. Tegucigalpa, onduras. 58p (annexes).

Grandin, B. 1988. *“Wealth ranking in smallholder communities: a field manual”*. London: Intermediate technology Publications.

Henninger, N. 1998. *“Mapping and Geographic Analysis of Human Welfare and Poverty. Review and assessment”*. World Resources Institute, Washington, D. C., USA. 97p.

Hentschel, J., Lanjouw, J., Lanjouw, P., and Poggi, J. 1998. *“Combining survey data with census data to construct spatially disaggregated poverty maps: a case study of Ecuador”*, mimeo, World bank.

Openshaw, S. (1995). *“The future of the census”*, in S. Openshaw (ed) The Census Users handbook, Longman, London, p.389-411.

Openshaw, S., 1987a, *“An automated geographical analysis system”*, Environment and Planning A 19, 431-436.

Openshaw, S., Charlton, M., Wymer, C. and Craft, A.W., 1987b, *“A mark I geographical analysis machine for the automated analysis of point data sets”*, Int. J. GIS, 1, 335-358

Oyana, T., Couillaud, P., Leclerc, G., Knapp, R., and Bell, W. 1998. *"Processing social indicators at individual and household level, and their aggregation at different scales"*. Technical document. CIAT. 66p.

Ravnborg HM. 1999 *(In Press)*. *"Assessing rural poverty: a practical method for identifying, extrapolating, and quantifying local perceptions of rural poverty"*. International center for Tropical Agriculture (CIAT), Cali, Colombia.

Republica de Bolivia. 1995. *"Mapa de Pobreza. Una guía para la acción social"*. Unidad de analisis de políticas sociales (UDAPSO), Instituto Nacional de Estadística (INE), Unidad de Políticas y Población (UPP), y Unidad de Análisis de Políticas Económicas (UDAP), Bolivia. 2nd Edn. LaPaz, Bolivia.

Rhoades, E. E. (1998). *"Participatory watershed research and management: where the shadow falls"*. Gatekeeper Series no. 81. Sustainable Agriculture and Rural Livelihoods Programme, IIED, London, 16p.

SECPLAN (Secretaría de Planificación, Coordinación y Presupuesto). 1991. *"Precenso y censo de población y vivienda 1988. Información básica de los municipios de Honduras"*. SECPLAN. Tegucigalpa, Honduras. 41p.

Skidmore, A. K. 1998. *"Non-parametric classifier for GIS data applied to kangaroo distribution mapping"*. Photogramm. Engineer. Remote Sens. 64:217-226.

UNDP (United Nations Environment Program). 1992. *"Proyecto regional para reducción de la pobreza en América Latina, 1990 y 1992. Desarrollo sin pobreza"*. 2nd Rev . Edn., Bogota, Colombia.

Table 1. Correlation coefficients between Population and Housing census indicators, computed from municipio level data.

	CV (housing)	CSIB (services)	CIA (non-land assets)	CE (education)
CV	1.0	0.79	0.76	0.7
CSIB	0.79	1.0	0.58	0.51
CIA	0.76	0.58	1.0	0.59
CE	0.7	0.51	0.59	1.0

Table 2. Correlation between Unsatisfied basic needs (NBIs –this work) and two other national poverty measures, computed from department level data.

	FHIS (1992)	SECPLAN (1991)
NBI_3 (1988)	0.89	0.94
NBI_4 (1988)	0.78	0.95

Table 3. Proxy to Ravnborg (1999) indicators, obtained from the 1988 Population and Housing census, and the 1993 Agriculture census of Honduras.

Indicator	Census variable	Census	Condition
Cattle ownership	Total cattle heads	Agriculture 93	≤ 1
Hire day labor	Total workers with pay	Agriculture 93	= 0
Land ownership	Size of exploitation	Agriculture 93	≤ 3mz
Health	Number of children dead/total number of children; Urban or rural area	Population 88	Continuous; Rural
Sell day labor	Relation to head of family; Activity; Class of activity; Urban or rural area; Total hours worked/number of people in household	Population 88	Categorical; Continuous; Rural
House	Ownership; Roof material, Walls material; Floor material; Urban or rural area	Housing 88	Categorical; Rural
Animal ownership	Total number of pigs, horses, oxen, mules, chicken, hens, sheeps, other poultry, rabbits	Agriculture 93	≤ 50 if chicken, ≤ 5 if sheep, rabbit or other poultry; ≤ 0 otherwise
Market participation	Production of permanent crops, other annual crops; Quantity of basic grains sold/Production of basic grains.	Agriculture 93	= 0 if permanent crops or other annual crops; ≤ 0.25 of basic grain production sold
Income	Occupation code; Urban or rural area.	Population 88	Categorical, each family member; Rural

Table 4. Goodness of fit statistics for the neural net model to predict P_WBI, the proportion of poorer households, for 12 communities.

R^2	0.78
Average Error	0.047
Correlation	0.89
Mean Square Error	0.0045
Root Mean Square Error	0.067

Table 5. Relative importance of proxy indicators selected for prediction of P_WBI, the proportion of poorer households, for 12 communities.

Indicator	**Weight**
House	0.344
Health	0.225
Income	0.210
Use of day laboring	0.089
Land ownership	0.055
Animal ownership	0.038
Cattle ownership	0.023
Day laboring	0.015
Market participation	0.002

Annex 1. Indicators of the Participatory Well-being Index (Ravnborg, 1999)

Variable	**Score**	**Condition**
Land Ownership	33	The household owns 4 manzanas or more, or has land in pasture or gives land in rent to other farmers
	67	Household owns land but fewer than 4 manzanas and doesn't have land in pasture nor land in rent to other farmers
	100	Household doesn't own land or only owns the house and land upon which it stands
Sell Day Labor	33	Nobody in the household works as a day laborer and the housewife does not do housework for other families nor prepare food to sell
	67	Someone in the household works as a day laborer but either for fewer than 9 months or for more than 9 months but fewer than 3 times a week
	100	Someone in the household works full-time for more than 9 months a year as a day laborer or if the housewife does house work for other families or sells prepared food
Income	33	Someone in the household is a professional, a businessman or a merchant or if children or other relatives send remittances
	67	Someone in the household is a skilled worker but no one in the household is a professional, businessman or merchant, and the household receives no remittances.
	100	No one in the household is a professional, businessman, merchant or skilled laborer, and the household receives no remittances.
Hire Day Labor	33	Household contracts day labor
	67	Household does not contract day labor
Cattle ownership	33	The household has cattle
	67	The household does not have cattle
Animals ownership	33	The household owns horses, pigs or oxen
	67	Household owns chickens but not horses, pigs nor oxen
	100	Household owns no animals
House	33	If the household owns its own house and the house is of good quality
	67	Household owns its own house but it is not of good quality
	100	Household owns its own house but it is of very poor quality or does not own its own house
Market participation	33	Household grows coffee or cacao or if household does not buy basic grains and sells half or more of its production of basic grains
	67	Household does not grow coffee but buys both buys and sells basic grains or if the household does not but basic grains and sells less than half of its production
	100	Household does not grow coffee or cacao and it buys basic grains in addition to using all of its production for home consumption
Money	33	Household has a savings account or makes loans to others
	67	Household does not save nor make loans
Health	67	No one in the house was sick or if someone were sick he/she paid for adequate health care either with own money or by selling assets
	100	Someone in the household has health problems and they were treated by asking relatives for money, borrowing money, or by going to the herbalist, or they were untreated for lack of money
Food Security	67	Household has not experienced a food shortage, or did for less than a week and solved it without having to ask others for food or money, to reduce number of meals, or to send the wife or children out to work
	100	Household experienced a food shortage for more than a week, or of less than a week but had to solve it by asking for food, by borrowing money or by sending wife and children out to work