Case Study No. 2

Good Practice in Survey Design and Analysis

Crop Protection Programme funded project titled *The development of a biocontrol strategy for the management of the alien perennial weed Mikania micrantha in tree crop based farming systems in India*, led by Dr. S.T. Murphy, CABI Bioscience.

1.   Background

The purpose of the project was to use biological methods by use of fungal pathogens, to control the perennial weed *Mikania micrantha* in three crops and agroforestry systems grown or used in tropical moist forest regions of southwest India..

The specific objectives were: (i) to determine the distribution, agricultural importance and socio-economic impact of the weed in selected areas of the Western Ghats; (ii) to identify the mycobiota of the weed and to assess the biocontrol potential of selected frugal pathogens collected in both its exotic range and native range; (iii) to develop a protocol for the import/release of selected coevolved biocontrol agent/s; and (iv) to train Indian scientists in weed management based on biocontrol.

This case study concentrates on the first objective above, for which the activities were an ecological and a socio-economic survey. It also considers general issues concerning the project as a whole (e.g. data management) to illustrate good biometric practices.

2.   Brief description of surveys conducted

2.1   Ecological Survey

A detailed survey was undertaken in the Western Ghat region of the states of Kerala, Karnataka and Goa to determine the distribution of the weed *Mikania micrantha*.

Mikania infestation was recorded only in Kerala. This state was selected for detailed studies since a preliminary survey indicated *Mikania* was most serious here.

In Kerala 90 forest ranges (the smallest administrative unit) were selected and 2-5 quadrats of 0.5 hectare were randomly chosen from each. These quadrats covered natural forests, forest plantations and agroforestry systems. In total 163 quadrats were sampled. A number of variables were measured on each of five 10m×10m quadrats, e.g. a 0-5 scale to indicate level of severity of *Mikania* (in Kerala only), canopy type (open or closed), altitude, and, in the case of natural forests, a measure of the degree of disturbance (on a 1-5 scale; 5 being no disturbance). Results were tabulated to show the severity of *Mikania* across the type of ecosystem, zones (North, Central and South), canopy type and altitude.

2.2   Socio-economic survey

Socio-economic aspects with respect to the effects of *Mikania* had been examined in Kerala, within each of the three systems, i.e. agroforestry, forest plantations and natural forests. The study data were collected through an interview schedule and observation as well as by the use of participatory rural appraisals.

The survey was conducted with one hundred households selected from each of seven districts. Much of the results were reported according to a subdivision of these households into 4 groups according to their holding size.

3.   Good biometric/computing practices

3.1   In the ecological survey concerning the distribution of *Mikania,* results were tabulated sensibly across other characteristics measured on the sampled quadrats. Chi-Square tests were conducted to investigate whether the severity of infestation by *Mikania* varied significantly across north, central and southern zones, across the canopy type and across altitude. This is a sensible approach, given that the objective was simply to acquire an understanding of the distribution of *Mikania* across the sampled areas, e.g. 68% of the sampled quadrats were found to have some *Mikania* infestation. In Section 4 below, we highlight ways in which the analysis could have been extended for a different but related objective.

Point No. 1: In a 2-way table where the tabled frequencies (numbers of quadrats here) represent sampled units classified by two categorical variables, chi-square tests may be applied, at an initial explanatory stage, to investigate whether the selected pair of classification variables are independent. Several classification variables can be explored simultaneously via log linear modelling techniques (see Section 4).

3.2   Procedures for the conduct of the socio-economic survey were well presented in a protocol prepared by a visiting scientist as a component of the visit report. The protocol clearly indicated that the primary responsibility for preparing the structured interview schedule and other survey implementation procedures lay with the local counterpart scientists. Among various suggestions made in the report were (a) the need to link results across the socio-economic and ecological surveys by use of a common identification code; (b) the need for information details to be collected at both household and farm levels; and (c) the need for relevant data to be recorded to enable a study of the impact of *Mikania* on crop production.

Selection of households for the socio-economic survey took place in two stages. Firstly, following a pilot survey of *Mikania* infestation, 7 districts (out of 14) where highest infestation occurred were selected. This is a good example of an instance where purposive sampling rather than simple random sampling is appropriate. To study the agronomic importance of *Mikania*, areas of high infestation **must** be selected so that the effects of *Mikania* on crop production can be investigated. This need was further emphasised by asking Range Officers in the Forest Department and Regional Agricultural Officers within the sampled districts for help in identifying areas of greatest *Mikania* infestation.

Farmers were selected at random from a list of farmers obtained from the respective village officers in the *Mikania* infested areas. This list also gave the farmers' holding size. A total of 100 households were selected. About 30 other households having the same cropping pattern as those selected, but without any *Mikania* infestation were also selected as "controls", so that their cost of cultivation and income from farming could be compared with farms having *Mikania* infestation.

Point No. 2: Prepare a protocol for each intended activity. This is invaluable in planning and implementation, as well as providing meta-data for the data archive.

Point No. 3: A pilot survey (as done in this project) is essential prior to actual data collection. Not only does it allow the practicalities and other aspects of the survey process to be clarified, it also provides relevant information to aid in any subsequent stages of sampling.

Point No. 4: Where more than one study is being conducted using overlapping sampling units, they must be linked via a unique identification code for each unit in the different studies.

Point No. 5: It is appropriate to use purposive sampling rather that simple random sampling in situations where the aim is to examine socio-economic aspects of the effects of a particular pest or disease. Areas of high incidence of the disease are needed to study such effects.

Point No. 6: It is appropriate to include additional sampling units as "controls", chosen so that the characteristics of the study units are similar to those of the control units, when comparisons between the study units and the "control" units are needed to fulfil research objectives.

4. Analysis extensions for alternative objectives

In section 3.1 we described the use of Chi-Square tests to study variation in the degree of infestation, with each of a number of factors such as altitude and canopy type, considered one at a time. If however, the objective was to investigate which *sub-set of factors* had an effect on the growth of *Mikania*, a combined analysis involving all the potential variates would be appropriate. This can be done by fitting **log-linear models** to the number of quadrats, including also the categorical variate identifying the main response, e.g. grades scored on a 0-5 scale for the degree of infestation. Log-linear models are an extension of the chi-square test that can be used on 2-way tables and is applicable to more complex tables of frequencies.

Another extension of the analysis arises if the main interest centres on estimating approximately the **percentage of areas** under natural forests, forest plantations and agroforestry systems **that are affected by *Mikania*.** This is different to determining the **percentage of quadrats** where *Mikania* occurs. The actual area covered by *Mikania* in each quadrat may be **approximately** calculated as follows by using information on the approximate number of stalks of *Mikania* found in each quadrat.

The information in Table 1 may be used.

Table 1. Number of quadrats showing different degrees of infestation by *Mikania*

Ecosystem	Degree of infestation							Totals	% quadrats infested
	none	1 to 10 stalks	<10 to 25 stalks	<25 to 50	<50 to 75	<75 to 100	>100		
Natural Forests	35	0	1	11	8	7	16	78	55.1%
Plantations	14	0	2	7	2	4	17	46	70.0%
Agroforestry systems	3	1	5	12	7	4	7	39	92.3%
Approx. area of a hectare covered	0%	1%	4%	10%	20%	25%	40%	100%	

Assume for simplicity that each quadrat covers an area of 1 ha. The grading on a 0-5 scale into different levels of infestation may be converted to areas infested by regarding the 1 ha area of the quadrat as the sum of 10000 one-square metre plots and arguing, for example, that the presence of 1-10 stalks approximates to 1% of the hectare, presence of <10 to 25 stalks is approximated to 4% of a hectare, etc. For example, we may use 10%, 20%, 25% and 40% as indicating the approximate percentage area of a hectare corresponding to the presence of <25 to 50 stalks, <50 to 75 stalks, <75 to 100 stalks and >100 stalks. These (guessed) figures are shown in the last row of Table 1. These percentages could be specified more accurately using the experiences of field staff who were responsible for conducting the survey.

Using the proposed percentages above, the area infested in Natural forests (say) will be

$$(35 \times 0) + (0 \times 0.01) + (1 \times 0.04) + (11 \times 0.01) + (8 \times 0.20) + (7 \times 0.25) + (16 \times 0.40)$$

$$= \quad 10.89 \text{ ha}$$

Similarly, area infested in Plantations $\quad = \quad$ 8.98 ha
and area infested within Agroforestry systems $\quad = \quad$ 6.61 ha

Hence percentage area covered in these three production systems are as follows:

$$\text{Natural Forests} \quad = \quad \frac{10.89}{78} \times 100 \quad = \quad 14.0\%$$

$$\text{Plantations} \quad = \quad \frac{8.98}{46} \times 100 \quad = \quad 19.5\%$$

$$\text{Agroforestry systems} \quad = \quad \frac{6.61}{39} \times 100 \quad = \quad 16.9\%$$

This compares with 55.1%, 70.0% and 92.3% corresponding to the percentage **of quadrats** showing some degree of infestation.

Both summaries are meaningful.

5. Concluding remarks

The project team are to be congratulated in their planning and implementation procedures undertaken with respect to the ecological and socio-economic surveys. The way the local collaborators played a major role in all research activities, including activities that are not reported here, was particularly impressive.

Discussions with the team leader indicated that the survey design was done in collaboration with the two statisticians from the Kerala forest Research Institute. We feel that the involvement of local biometric advice is excellent and is to be encouraged within all DFID-funded projects where such resource persons are locally available.

6. Acknowledgement