# Steps and information for realizing the variability study in East and South Africa by Mrs Everina LUKONGE

**Projet CFC/ICAC/33**

**Jean-Paul GOURLOT**

**Everina LUKONGE**

**Modeste ABOE**

# Steps and information for realizing the variability study in East and South Africa

## By Mrs Everina LUKONGE

Summary


1 - Expected output ... 3
2 - Preparation of situations list ... 3
3 - List of concerned countries ... 3
4 - Summary of the experiment to be made ... 4
4.1 - Hypothesis 1: The within bale variations are smooth enough to contemplate sampling by the lower (and upper?) sides only ... 4
4.2 - Hypothesis 2: the distribution remains unchanged (ideally Gaussian) during the whole ginning season and is not contaminated by abrupt changes ... 4
4.3 - Hypothesis 3: the between bales variations are smooth enough to contemplate a weighted moving average (the number of bales considered in this average can be increased up to the point that correlations decrease under a significative level) to improve the precision of the estimation of any given bale or to propose an estimate for the un-measured bales (risky) ... 4
4.4 - Hypothesis 4: one sample on the current bale + one sample on the next bale is a good proxy to 2 samples per bale (top and bottom) ... 4
5 - Practical organization ... 4
5.1 - General ... 4
5.1.1 - H1 ... 4
5.1.2 - H2 ... 5
5.1.3 - H3 ... 5
5.1.4 - H4 ... 5
5.2 - Definition of the situations ... 6
5.3 - Estimation of the total niumber of sample per situation ... 7
5.4 - Steps to perform when arriving in a situation? ... 9
6 - Selection of the situations ... 10
6.1 - Tanzania ... 10
6.2 - Uganda ... 11
6.3 - Zambia ... 13
6.4 - Zimbabwe ... 14
6.5 - Mozambique ... 15
6.6 - Sudan ... 15
7 - Travels of Everina ... 16
7.1 - Year 1 ... 16
7.2 - Year 2 ... 16
8 - Sample transfer form the gins to the RTC laboratory ... 16
8.1 - Tanzania ... 16
8.2 - Uganda ... 17
9 - Randomization of samples in the hypothesis ... 17
9.1 - For hypothesis 1 ... 17
9.2 - For hypothesis 2 ... 17

9.3 - For hypothesis 3 ... 18
9.4 - For hypothesis 4 ... 18
10 - Rules for the testing of the samples ... 19
10.1 - General ... 19
10.2 - Verification of measurement stability during the testing part of this experiment ... 19
10.2.1 - Calibration of HVTE ... 19
10.2.2 - Other checks of collected data ... 20
10.2.3 - Number of consecutive measurements per sample ... 20
10.2.4 - Randomization of the testing of samples ... 20
10.2.5 - What is a result? ... 20
11 - Analysis of the results obtained on batches of 100 consecutive bales in various ginning factories in Tanzania (preliminary data set) ... 20
11.1 - 1 Upper Half Mean Length (UHML) ... 20
11.1.1 - 1.1 UHML OLAM B 6782 B.pdf ... 20

## 1 - **Expected output**

Expected output: confidence in testing results from any laboratory in Africa; this requires the setting of operatings methods including sampling and testing procedures.

- a typical protocole to be applied by countries where we made the experiments
- a typical protocole to be applied by countries we did not consider at first for making the variability study.
- The typical protocole will be applied next year to
    - Simplify the tested hypothesis
    - Check the feasibility of its use on several crop years in various conditions
    - To include the test of other hypothesis (more situations in year 2? …)

## 2 - **Preparation of situations list**

We decided at first to prepare the list of situations according to the records of the following informations:

- RefSituation Code
- Country
- Variety
- Agro conditions
- Harvesting
- Gin name
- Gin type
- Gin alimentation
- Gin lint cleaning
- Gin equipment
- Gin remarks
- A1= Several varieties in border regions
- A2= One variety in region cores
- B1=Heterogeneous growing conditions in small farmers production
- B2= Homogeneous growing conditions within a bale in big farmers production

## 3 - **List of concerned countries**

Mozambique, Tanzania, Uganda, Zambia, Zimbabwe and Sudan (Upland)

## 4 - Summary of the experiment to be made

All the following hypothesis have to be proved in order to fill the objective of measuring the within-bale variability of cotton fibre characteristics and of designing operating methods in sampling and measuring those fibres:

### 4.1 - Hypothesis 1: The within bale variations are smooth enough to contemplate sampling by the lower (and upper?) sides only

➔ 10 bales taken at random * 8 layers per bale = 80 samples
➔ for every situation

### 4.2 - Hypothesis 2: the distribution remains unchanged (ideally Gaussian) during the whole ginning season and is not contaminated by abrupt changes

➔ 75 couples of two successives bales * 1 sample per bale every 2 days during the whole season = 150 samples
➔ for every situation

### 4.3 - Hypothesis 3: the between bales variations are smooth enough to contemplate a weighted moving average (the number of bales considered in this average can be increased up to the point that correlations decrease under a significative level) to improve the precision of the estimation of any given bale or to propose an estimate for the un-measured bales (risky)

➔ 200 consecutive bales * 1 sample per bale
➔ for every situation

➔ 430 samples per situation in total in every situation

### 4.4 - Hypothesis 4: one sample on the current bale + one sample on the next bale is a good proxy to 2 samples per bale (top and bottom)

➔ 75 extra-samples on the sample for hypothesis 2 at the same time as hypothesis 2 or during an extra experiment in the most difficult situation (small scale, variable situation), meaning to be investigated only on the cotton produced by small holders = worst situation.
➔ some situations only

## 5 - Practical organization

### 5.1 - General

#### 5.1.1 - H1

For testing H1, ideally the 10 bales should be taken at random in the production from the entire ginning season. This is not possible for 2008-2009, the ginning season comes to its end in West Africa, but we rely on the fact that the production becomes more and more heterogeneous as the season progresses. So if the sample is biased by taking all the samples at the end of the season, the variability estimate will be biased upwards and we will end up with a conservative confidence interval and an oversized sample recommendation. Anyway it is safer than the other way round, and the results will be adjusted according to the next year results
In order the sampling to span as diverse situations as possible, it would be sensible to train the local technicians on the 8 samples/bale operation. However, it is best for Everina LUKONGE and Modeste ABOE (E&M) to be there for this particular sampling operation. So we rely on the assumption that cotton from different trucks come from different villages, and should show different variabilities, if the variability is different from one village to another. As a

truck usually carries a module of about 10 tons of seed cotton per trip, with a 40% ginning yield we expect 4 tons of lint from one truck, which is a bit less than 18 bales of 225 kg. So by picking 1 bale every 20 bales we should expect each bale in our sample to come from a different village.

In most ginning factories the production if about 500 bales in 23 hours, that is 21 bales an hour. To sample 10 bales from different villages one would then have to stay in the factory for 10 hours

5.1.2 - H2

Every 2 days in a given factory, we need 1 (one) ordinary sample per bale on 2 consecutive bales, regardless of whether the bales come from the same truck or not. Whether the 2 consecutive bales come from the same truck or not should be recorded. This means that it would be necessary to record, together with the samples, the gin name, the bale numbers, the origin ( village name) and the variety.
As the bales are numbered in sequence in every ginning factory, the samples necessary for testing the H2 hypothesis can be collected from the classing room. Because most of the countries have only one classing room, collecting the necessary samples can be made with one single travel to the classing room.
If this collecting method is not feasible, M&E will have to ask in advance to each factory manager to provide them with the necessary samples.

5.1.3 - H3

For each of the 200 bales, the gin name, the bale numbers, the variety and the origin ( village) should be recorded. Changing variety within one 200 bales sample should be avoided, if possible. At most one variety change within a 200 bales sample is possible.

5.1.4 - H4

We only test one situation per country for this hypothesis.
Each situation combines variety homogeneity (A1 vs A2) growing homogeneity (B1 vs B2), and ginning technique (roller vs saw). Other factors such as factory feeding device (module feeder or telescope) will be considered next year. There is then 8 situations overall for a given year. Not every situation is present in every country. For each of the situations the set of countries where it is encountered is listed; then a set of countries is selected at random, by sorting the country names according to a pseudo-random generated with Excel, and selecting the first n countries in each set. If the 8 different sets do have common countries, each country is retained only in the set where it appears first. In case of ex-aequo the set where the country is retained is selected at random. In the sets from which the country is removed, replacement is obtained by taking the first non-selected country in the random sorted list. The common countries elimination is then continued until the different sets do not intersect.

For the second year, in each country when there is several situations, the one that was selected the first year will not be proposed the second year, in order this situation not to be sampled twice in the same country.

Factory feeding device (module feeder or telescope) will be considered the second year . In each situation, if there is only one sort of feeding device the factory will be taken at random with the same method used for the first year.
The ideal solution would be to list all the possible existing stratas to define all situations in every country. However, as we do not have access to all information from the countries, we

will rely on the following definitions of situations while recording information to more closely represent the actual situation during Year 2 experiments:

### 5.2 - Definition of the situations

The strata or situations have been defined according to the factors that can influence variability and that can be measured or known for any given bale. That is : size of farm and possibility of a mix between 2 varieties. In addition, type of gin (roller or saw) is a further important factor. The factory feeding device (telescope or module feeder) may also influence the variability and should be include as a stratum defining factor. If in a given year not all the strata can be investigated, the telescope and module feeder equipped factories can be studied on different years.

> Note 1: Lint cleaning : only short range smoothing : not studied.
>
> Note 2: Even though we categorized the gins in Tanzania according to 4 types (old roller, old saw, new roller and new saw), we will only consider roller vs saw ginning during the first crop year for the moment.

When only one type of ginning is existing in one country, we will use the following simplified table:

| Country name | A1= several varieties in border regions | A2 = one variety in region cores |
|---|---|---|
| B1 - Heterogeneous growing conditions in small farmers production | | |
| B2 - homogeneous growing conditions within a bale in big farmers production | | |

But when necessary, we will keep three factors crossed together to built the statas:

- A1 : several varieties vs A2 : one variety
- B1 - Heterogeneous growing conditions in small farmers production vs B2 - homogeneous growing conditions within a bale in big farmes production
- C1: Roller ginning vs C2: Saw ginning

| Country name | A1= several varieties in border regions | | A2 = one variety in region cores | |
|---|---|---|---|---|
| | C1 Roller | C2 Saw | C1 Roller | C2 Saw |
| B1 - Heterogeneous growing conditions in small farmers production | | | | |
| B2 - homogeneous growing conditions within a bale in big farmers production | | | | |

In every country and for every existing situation, we then need to record:

- Gin name
- Gin type (roller or saw)
- Gin feeding (telescope vs module feeder)
- Variety
- Where goes the extra seed cotton on the gin (before gin 1 or after)
- Humidification of Seed cotton at the gin?
- Humidification of fibre at the gin

- lint cleaner type (airjet or saw or both)

### 5.3 - Estimation of the total niumber of sample per situation

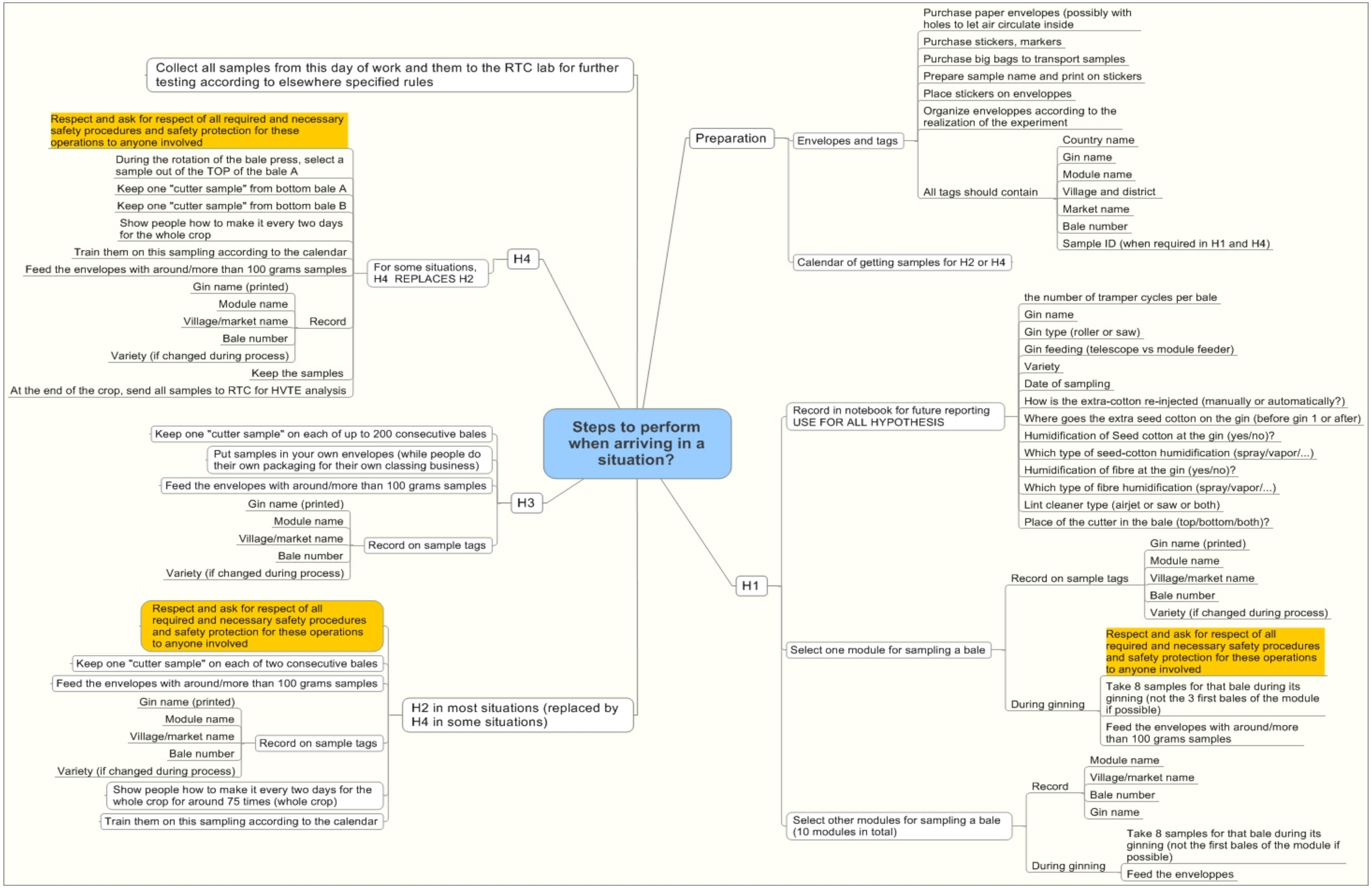
| Hypothe sis | Sampling | Situation | Sample collected from | Time required | Specific countries | Requirement |
|---|---|---|---|---|---|---|
| H1 | 8 samples/ bale; at least 10 bales / situation | ALL | Everina on site | 10 hours | ending crop for Sudan and Uganda => Start with these 2 countries | ON SITE |
| H2 | 1 sample/bale, 2 consecutive bales, every 2 days for 75 times | ALL | (Everina on sites + gins alone) OR from classing rooms | 1 hour for demonstr ation | ending crop fir Sudan and Uganda => part of the crop only (or postpone Y2) | |
| H3 | 1 sample/bale on 200 (if possible) consecutive bales | ALL | (Everina on sites) OR Classing rooms | (6 hours) | Start at earliest Sudan and Uganda | ON SITE |
| H4 | 1 sample/bale, 2 consecutive bales (A and B), + 1 sample on top of bale A, every 2 days for 75 times | some | Everina on site + gins alone | 1 hour demonstr ation | ending crop for Sudan and Uganda: => part of the crop only (or postpone Y2) | ON SITE |

In 2009, Uganda and Sudan cannot be tested for H1 and H4, because these demand specific samples to be collected and the season is about to be finished at the time this document is written. *However 0.25 of samples can be collected for the time remained before season closed.*

| Normal years | | Most situations | Specific situations |
|---|---|---|---|
| | Number of samples | Deliver | Number of samples |
| H1 | 80 | During visit | 80 |
| H2 | 150 (40 in year 1) | End of crop | |
| H3 | 200 | During visit | 200 |
| H4 | | End of crop | 225 (57 in Year 1) |
| | | | |
| Total / situation | 430 | | 505 |

| RTC | Countries | Hypothesis | Total situation per country | Nbre situations | Nbre of samples | Remain in crop | H1 | H2 | H3 | H4 | All Hyp without |
|---|---|---|---|---|---|---|---|---|---|---|---|
| East | Sudan (end of crop) | H1 | 3 | 3 | 80 | 1 | 240 | | | | 240 |
| East | Sudan (end of crop) | H2 | 3 | 2 | 150 | 0.25 | | 75 | | | 75 |
| East | Sudan (end of crop) | H3 | 3 | 3 | 200 | 1 | | | 600 | | 600 |
| East | Sudan (end of crop) | H4 | 3 | 1 | 225 | 0.25 | | | | 57 | 57 |
| East | Uganda (end of crop) | H1 | 3 | 3 | 80 | 1 | 240 | | | | 240 |
| East | Uganda (end of crop) | H2 | 3 | 2 | 150 | 0.25 | | 75 | | | 75 |
| East | Uganda (end of crop) | H3 | 3 | 3 | 200 | 1 | | | 600 | | 600 |
| East | Uganda (end of crop) | H4 | 3 | 1 | 225 | 0.25 | | | | 57 | 57 |
| East | Mozambique (normal crop) | H1 | 3 | 3 | 80 | 1 | 240 | | | | 240 |
| East | Mozambique (normal crop) | H2 | 3 | 2 | 150 | 1 | | 300 | | | 300 |
| East | Mozambique (normal crop) | H3 | 3 | 3 | 200 | 1 | | | 600 | | 600 |
| East | Mozambique (normal crop) | H4 | 3 | 1 | 225 | 1 | | | | 226 | 226 |
| East | Zambia (normal crop) | H1 | 3 | 3 | 80 | 1 | 240 | | | | 240 |
| East | Zambia (normal crop) | H2 | 3 | 2 | 150 | 1 | | 300 | | | 300 |
| East | Zambia (normal crop) | H3 | 3 | 3 | 200 | 1 | | | 600 | | 600 |
| East | Zambia (normal crop) | H4 | 3 | 1 | 225 | 1 | | | | 226 | 226 |
| East | Zimbabwe (normal crop) | H1 | 3 | 3 | 80 | 1 | 240 | | | | 240 |
| East | Zimbabwe (normal crop) | H2 | 3 | 2 | 150 | 1 | | 300 | | | 300 |
| East | Zimbabwe (normal crop) | H3 | 3 | 3 | 200 | 1 | | | 600 | | 600 |
| East | Zimbabwe (normal crop) | H4 | 3 | 1 | 225 | 1 | | | | 226 | 226 |
| West | Mali (end of crop) | H1 | 3 | 3 | 80 | 1 | 240 | | | | 240 |
| West | Mali (end of crop) | H2 | 3 | 2 | 150 | 0.25 | | 75 | | | 75 |
| West | Mali (end of crop) | H3 | 18 | 18 | 200 | 1 | | | 3600 | | 3600 |
| West | Mali (end of crop) | H4 | 3 | 1 | 225 | 0.25 | | | | 57 | 57 |
| West | Burkina Faso (end of crop) | H1 | 3 | 3 | 80 | 1 | 240 | | | | 240 |
| West | Burkina Faso (end of crop) | H2 | 3 | 2 | 150 | 0.25 | | 75 | | | 75 |
| West | Burkina Faso (end of crop) | H3 | 17 | 17 | 200 | 1 | | | 3400 | | 3400 |
| West | Burkina Faso (end of crop) | H4 | 3 | 1 | 225 | 0.25 | | | | 57 | 57 |
| West | Côte d'Ivoire (end of crop) | H1 | 3 | 3 | 80 | 1 | 240 | | | | 240 |
| West | Côte d'Ivoire (end of crop) | H2 | 3 | 2 | 150 | 0.25 | | 75 | | | 75 |
| West | Côte d'Ivoire (end of crop) | H3 | 13 | 13 | 200 | 1 | | | 2600 | | 2600 |
| West | Côte d'Ivoire (end of crop) | H4 | 3 | 1 | 225 | 0.25 | | | | 57 | 57 |
| West | Sénégal (enf of crop) | H1 | 3 | 3 | 80 | 1 | 240 | | | | 240 |
| West | Sénégal (enf of crop) | H2 | 3 | 2 | 150 | 0.25 | | 75 | | | 75 |
| West | Sénégal (enf of crop) | H3 | 6 | 6 | 200 | 1 | | | 1200 | | 1200 |
| West | Sénégal (enf of crop) | H4 | 3 | 1 | 225 | 0.25 | | | | 57 | 57 |
| West | Togo (enf of crop) | H1 | 3 | 3 | 80 | 1 | 240 | | | | 240 |
| West | Togo (enf of crop) | H2 | 3 | 2 | 150 | 0.25 | | 75 | | | 75 |
| West | Togo (enf of crop) | H3 | 3 | 3 | 200 | 1 | | | 600 | | 600 |
| West | Togo (enf of crop) | H4 | 3 | 1 | 225 | 0.25 | | | | 57 | 57 |
| West | Bénin (end of crop) | H1 | 3 | 3 | 80 | 1 | 240 | | | | 240 |
| West | Bénin (end of crop) | H2 | 3 | 2 | 150 | 0.25 | | 75 | | | 75 |
| West | Bénin (end of crop) | H3 | 18 | 18 | 200 | 1 | | | 3600 | | 3600 |
| West | Bénin (end of crop) | H4 | 3 | 1 | 225 | 0.25 | | | | 57 | 57 |
| West | Cameroun (end of crop) | H1 | 3 | 3 | 80 | 1 | 240 | | | | 240 |
| West | Cameroun (end of crop) | H2 | 3 | 2 | 150 | 0.25 | | 75 | | | 75 |
| West | Cameroun (end of crop) | H3 | 8 | 8 | 200 | 1 | | | 1600 | | 1600 |
| West | Cameroun (end of crop) | H4 | 3 | 1 | 225 | 0.25 | | | | 57 | 57 |
| West | Tchad (end of crop) | H1 | 3 | 3 | 80 | 1 | 240 | | | | 240 |
| West | Tchad (end of crop) | H2 | 3 | 2 | 150 | 0.25 | | 75 | | | 75 |
| West | Tchad (end of crop) | H3 | 8 | 8 | 200 | 1 | | | 1600 | | 1600 |
| West | Tchad (end of crop) | H4 | 3 | 1 | 225 | 0.25 | | | | 57 | 57 |

## 5.4 - Steps to perform when arriving in a situation?



## 6 - Selection of the situations

At a first step, we list all possible situations. In each cell of the table, situations are then sorted in a random order of priority to appear in the experiments for any reason of budget limitations. Are only retained situations listed first as a function of the total number of retained situations.

### 6.1 - Tanzania

Among all possible situations, all are relative on one single variety, all rainfed, all picked by hand, all small farmers.
However, we have four types of equipments in ginneries (design 1 to 4).
By random sorting per design (using the function rand () under Excel), we would retain the next gins situations by decreasing order of final appearance in the study.
Chosen on 13/03/2009

| Country name | A1= several varieties in border regions | | A2 = one variety in region cores | |
|---|---|---|---|---|
| Tanzania | C1 Roller | C2 Saw | C1 Roller | C2 Saw |
| B1 - Heterogeneous growing conditions in small farmers production | None | None | SHIRECU-Sola (H1,2,3), MSK (H1,2,3), S&C Bunda (H1,2,3), Afrisian Shinyanga (H1,2,3), Alliance Kasoli (H2) | Jambo Shinyanga (H1,3,4), Cargill Lalago (H1,2,3), Alliance Kasoli (H1,2,3), Afrisian Sangu (H1,2,3) |
| B2 - homogeneous growing conditions within a bale in big farmers production | None | None | Info: source is from ginners at the specific area there is no Seed multiplication farms | Manawa Ginnery (?) |

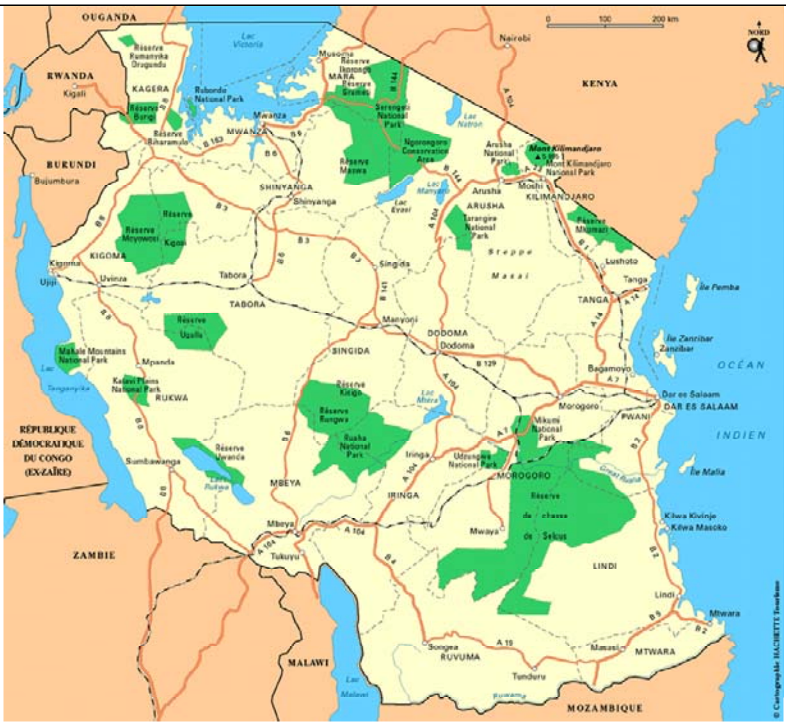



**Figure 1: Map of Tanzania**

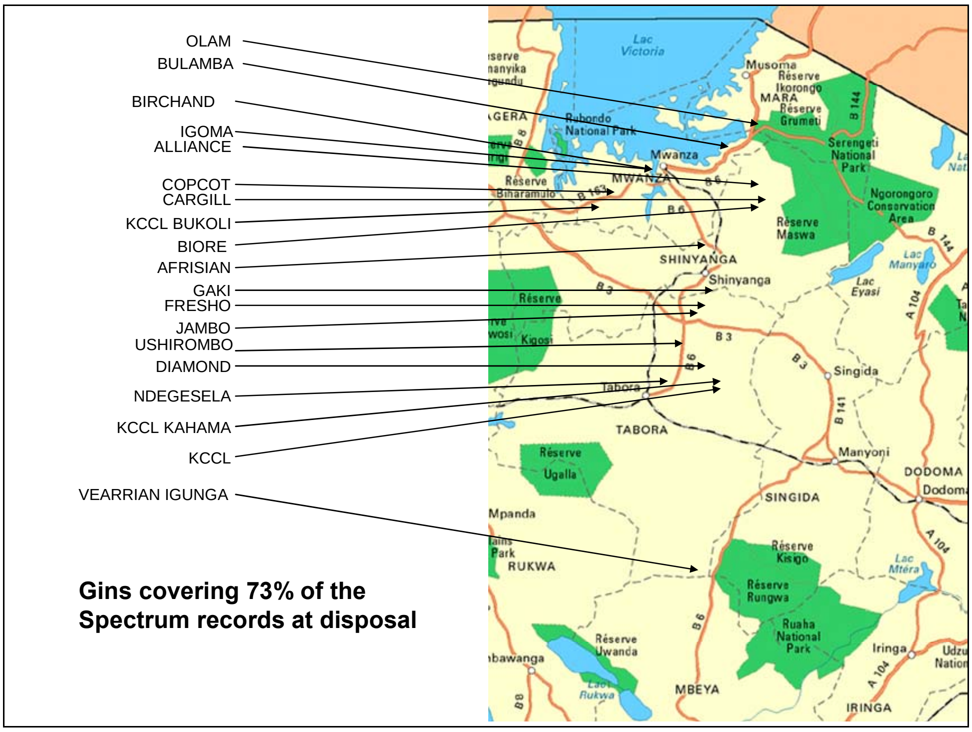



**Figure 2: Gin locations in North-West Tanzania.**

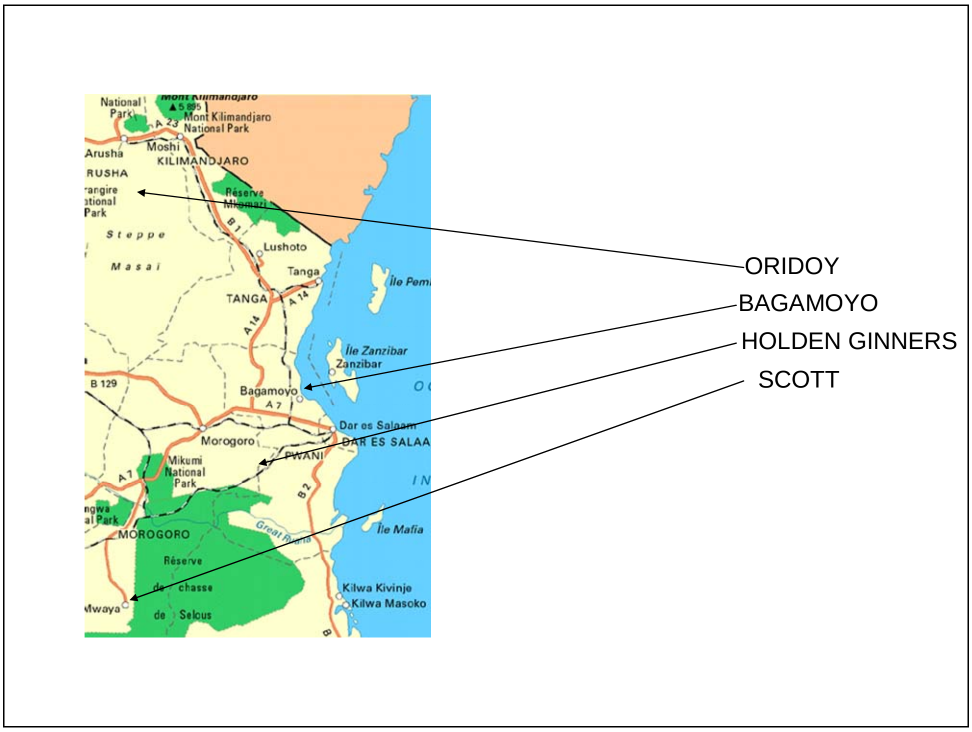



**Figure 3: Gin location in East Tanzania**

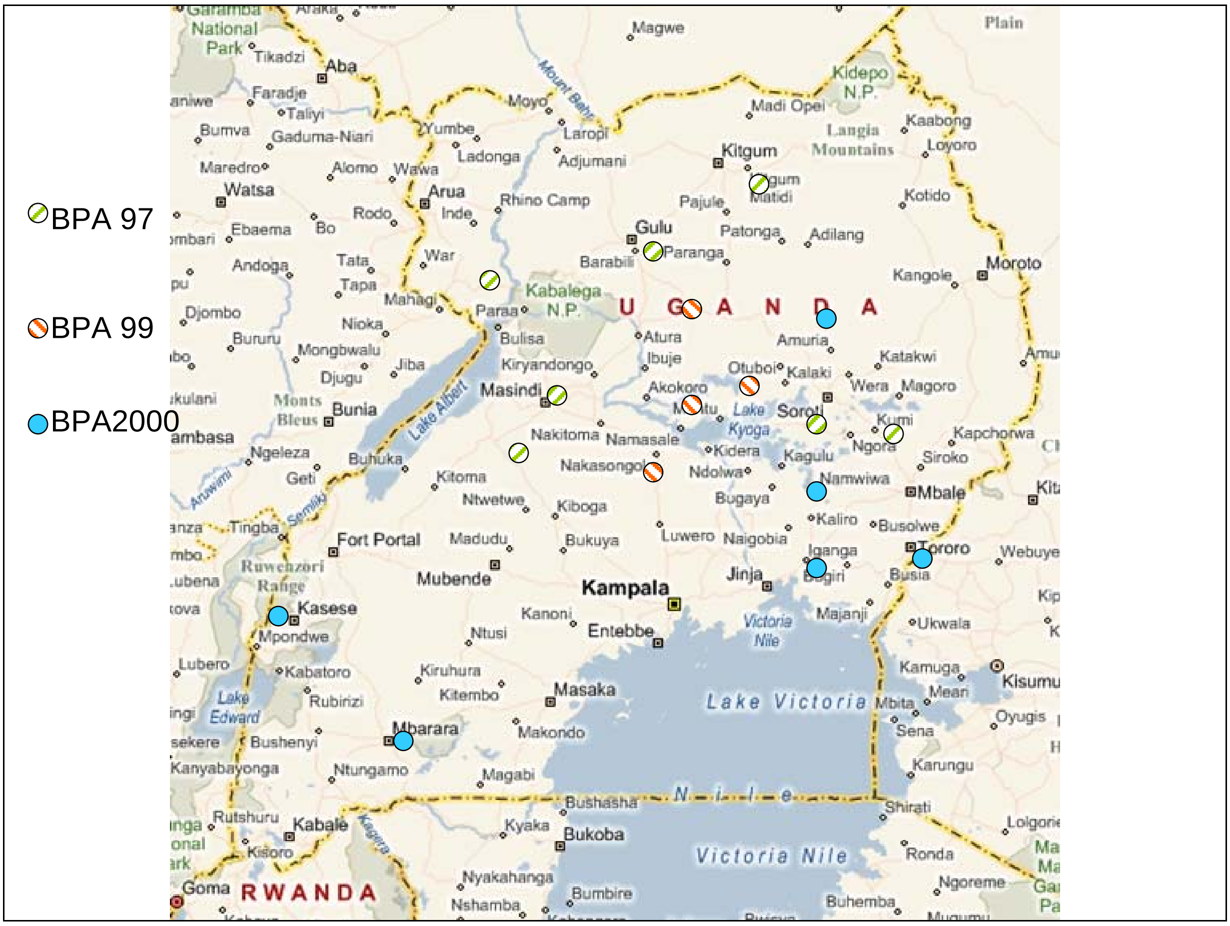
## 6.2 - Uganda

Among all possible situations, all are picked by hand, all small farmers. However, most are rainfed, but it exist some irrigated fields, with 3 varieties.

However, we have one type of equipments in ginneries (Double Roller gin)
By random sorting per design (using the function rand() under Excel), we would retain the next gins situations by decreasing order of final appearance in the study.

Chosen on 13/03/2009

| Uganda | A1= several varieties in border regions | A2 = one variety in region cores |
|---|---|---|
| B1 - Heterogeneous growing conditions in small farmers production | Rwenzori Kasese WESTERN Irrigated, Busolwe Tororo SEAST (H1,3,4), Kasese Nyakatonzi (WEST) Irrigated<br>Nakasongola (MW&CE)(H1,2,3),<br>Pramukh Iganga industries (SEAST)(H1,2,3),<br>Aduku APAC (North), Balawory Cotton ginnery Ltd Kamuli (SEAST), Aboke Rafiki cotton Industry Apac (NORTH), Bamunanika ginnery Ltd Ruweero (NEASTERN), Kachumbala N.C. cotton enterprises Kumi (NEAST), Clark cotton Uganda ltd Soroti (NEAST) (organic | None |
| B2 - homogeneous growing conditions within a bale in big farmers production | None<br>Info : Seed source is from ginners at the specific area there is no Seed multiplication farms | None |



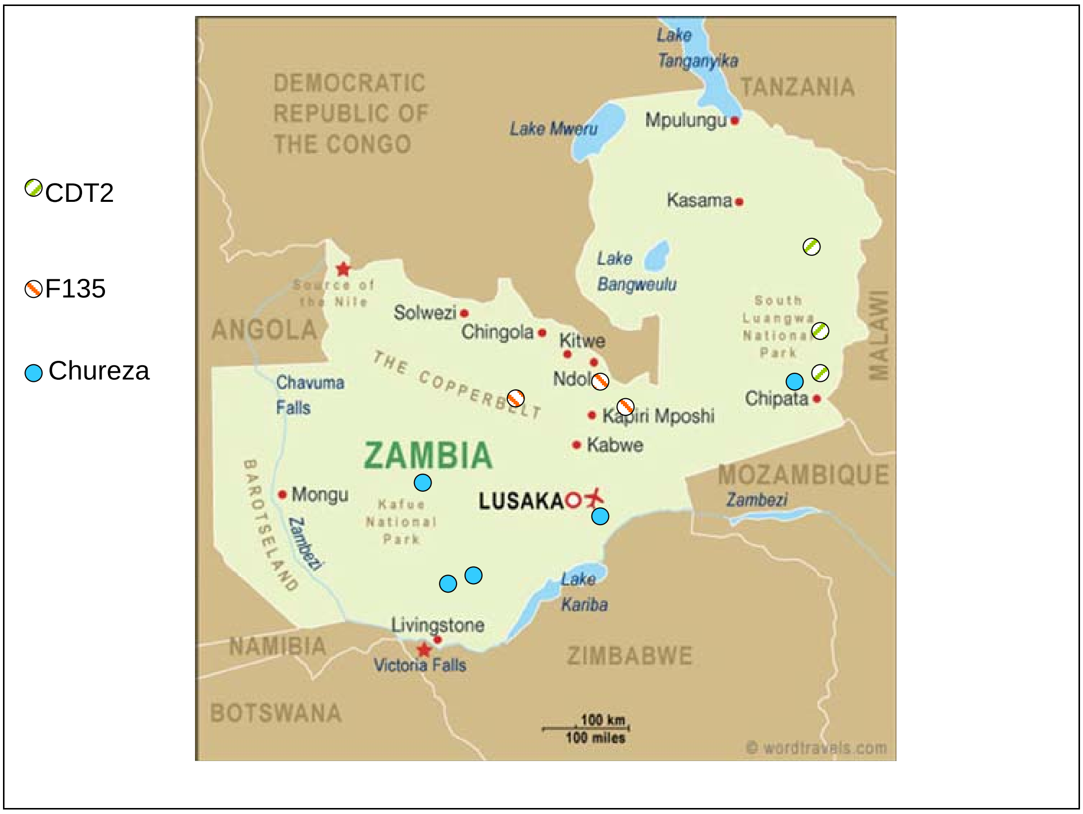
### 6.3 - Zambia

Among all possible situations, there are three varieties, all rainfed, all picked by hand, all small farmers, and one type of ginnery (Roller gin?).

By random sorting per design (using the function rand() under Excel), we would retain the next gins situations by decreasing order of final appearance in the study.

Chosen on 13/03/2009

| Zambia | A1= several varieties in border regions | A2 = one variety in region cores |
|---|---|---|
| B1 - Heterogeneous growing conditions in small farmers production | **Eastern:** Dunavant Katete (H1.2.3), Continental Petauke (Sinda), Dunavant Lundazi, Chipati petauke, Chipata Chinese Ginnery, Chipata Cargil Chipati, | **Central:** Dunavant Kabwe, Mukuba Ndolla Copperbelt Province (H1,2,3), Mulungushi Chinese Cotton Co. Ginnery Kabwe (ZCMT),<br><br>**Lusaka:** Alliance Cotton Kafue Lusaka (H1,2,3), Dunavant Lusaka Lusaka<br><br>**Southern:** Dunavant Gwembe Southern |
| B2 - homogeneous growing conditions within a bale in big farmers production | Info: Seed source is from ginners at the specific area there is no Seed multiplication farms | None |



### 6.4 - Zimbabwe

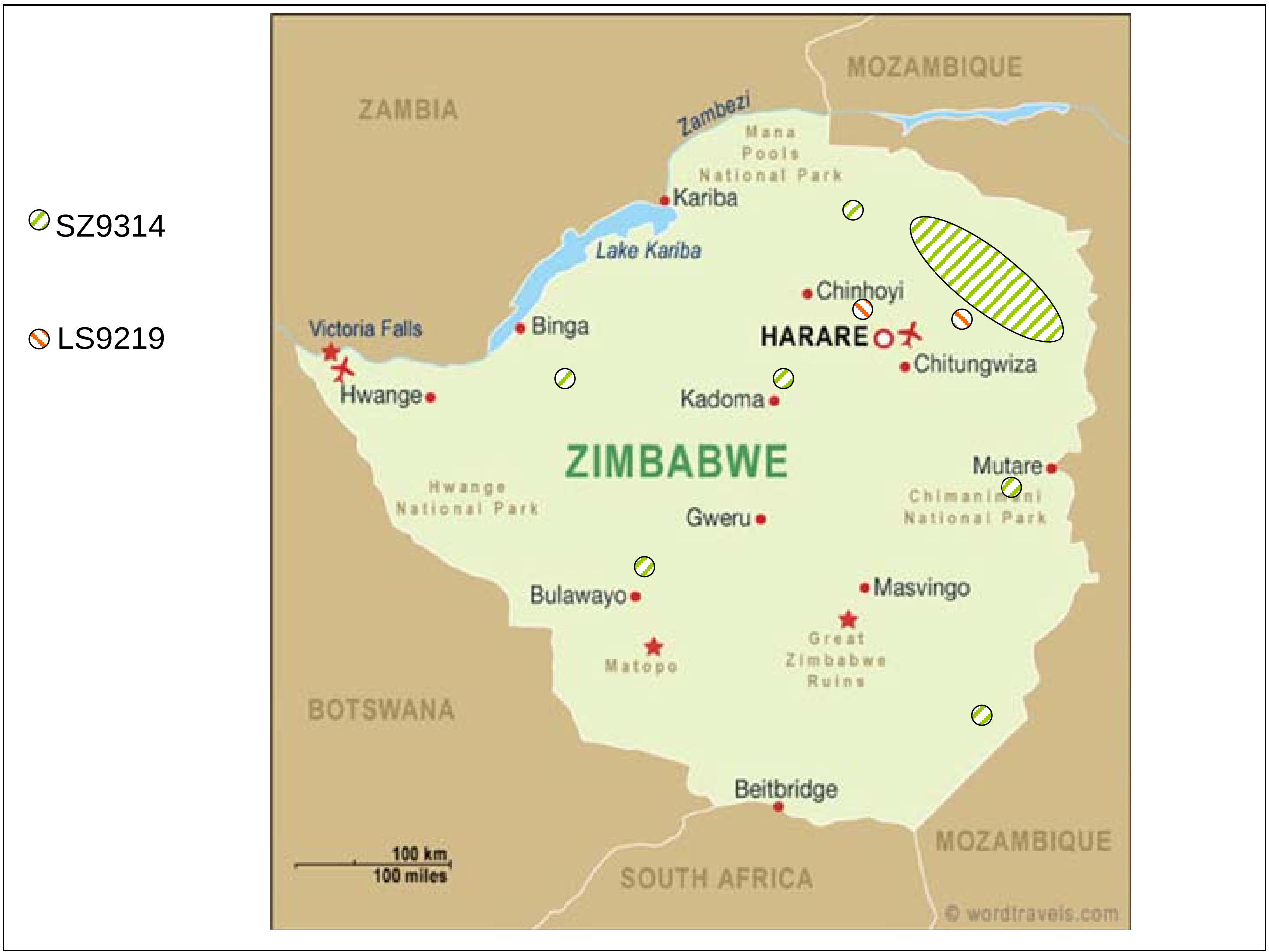
Among all possible situations, all rainfed, all picked by hand, grown by small farmers and few large farms.

However, there are two varieties and 9 saw gins and one Roller gin.

By random sorting per design (using the function rand() under Excel), we would retain the next gins situations by decreasing order of final appearance in the study.

To be concluded because of missing information

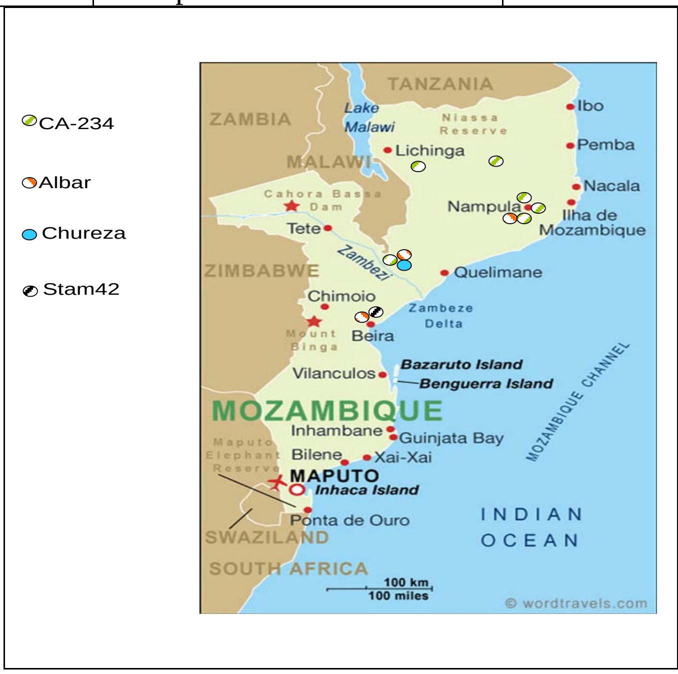
| Country name | A1= several varieties in border regions | | A2 = one variety in region cores | |
|---|---|---|---|---|
| Zimbabwe | C1 Roller | C2 Saw | C1 Roller | C2 Saw |
| B1 - Heterogeneous growing conditions in small farmers production | Cottco Bindura | Cottco Sanyati, Cottco Gokwe, Cottco Kadoma, Cottco Muzarabani, Cottco Glendale, Chiredz | | |
| B2 - homogeneous growing conditions within a bale in big farmers production | None | | | Cottco (H1,3) |



### 6.5 - Mozambique

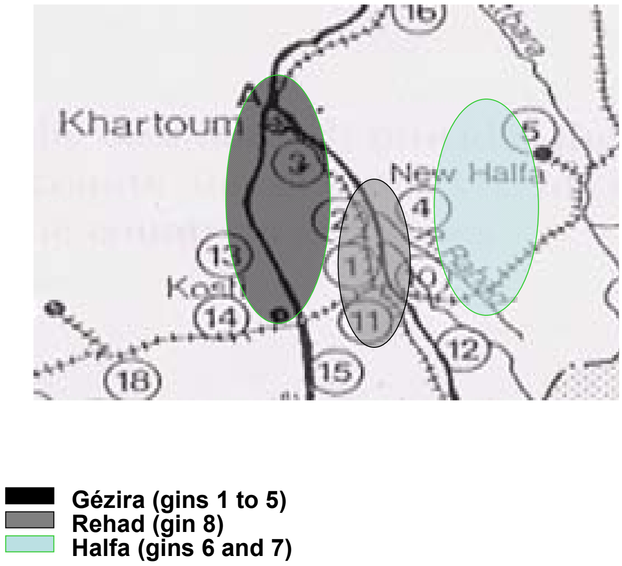
Among all possible situations, all rainfed, all picked by hand, all small farmers, one type of ginnery (Saw??).
However, there are four varieties
By random sorting per design (using the function rand() under Excel), we would retain the next gins situations by decreasing order of final appearance in the study.
Chosen on 13/03/2009

| Mozambique | A1= several varieties in border regions | A2 = one variety in region cores |
|---|---|---|
| B1 - Heterogeneous growing conditions in small farmers production | Nametil Nampula Mogovolas (H1,3,4), Zambezia – Morrumbala (H1,2,3), Manga - Sofala – Beira, Namapa Nampula – Erati | Montepuez Cabo Delgado (H1,2,3), Niassa Cuamba, |
| B2 - homogeneous growing conditions within a bale in big farmers production | None<br>Info: Seed source is from ginners at the specific area there is no Seed multiplication farms | None |




### 6.6 - Sudan

To be concluded because of missing information


# 7 - Travels of Everina

## 7.1 - Year 1

## 7.2 - Year 2

# 8 - Sample transfer form the gins to the RTC laboratory

## 8.1 - Tanzania

### 8.2 - Uganda

## 9 - <u>Randomization of samples in the hypothesis</u>

### 9.1 - For hypothesis 1

The first objective is to measure the within-bale variability, and maybe next, the differences in mean values between the bales. The randomization for this hypotheses should take care of a possible drift along the time in the results. So, for every single bale studied, we will randomize the sample measurement by random. In addition, in order to evaluate the measurement error, we will make two blocks of measurements nested into the bale factor with a different randomization in each block. As an example, we performed a randomisation and obtained the following table:

| Order | Bale | Sample | Replicate of measurement | Reference material |
|---|---|---|---|---|
| | | | | X |
| 1 | 6 | 8 | 1 | |
| 2 | 6 | 4 | 1 | |
| 3 | 6 | 3 | 1 | |
| 4 | 6 | 2 | 1 | |
| 5 | 6 | 7 | 1 | |
| 6 | 6 | 6 | 1 | |
| 7 | 6 | 5 | 1 | |
| 8 | 6 | 1 | 1 | |
| 9 | 6 | 1 | 2 | |
| 10 | 6 | 6 | 2 | |
| 11 | 6 | 3 | 2 | |
| 12 | 6 | 4 | 2 | |
| 13 | 6 | 5 | 2 | |
| 14 | 6 | 2 | 2 | |
| 15 | 6 | 8 | 2 | |
| 16 | 6 | 7 | 2 | X |
| 17 | 2 | 2 | 1 | |
| 18 | 2 | 5 | 1 | |
| 19 | 2 | 6 | 1 | |
| 20 | 2 | 4 | 1 | |
| 21 | 2 | 8 | 1 | |
| 22 | 2 | 3 | 1 | |
| 23 | 2 | 7 | 1 | |
| 24 | 2 | 1 | 1 | |
| 25 | 2 | 6 | 2 | |
| 26 | 2 | 7 | 2 | |
| 27 | 2 | 3 | 2 | |
| 28 | 2 | 2 | 2 | |
| 29 | 2 | 8 | 2 | |
| 30 | 2 | 4 | 2 | |
| 31 | 2 | 5 | 2 | |
| 32 | 2 | 1 | 2 | X |
| etc. | | | | |

### 9.2 - For hypothesis 2

All samples will be analyzed together at the end of the crop within a short time. To avoid any drift effect on the data, the randomization will nest a block of measurement effect into a "day of collection of the couple of samples" factor. In addition we will look at the reference material results to check if any existing drift is limited enough for assuring that all results from any given day is comparable as given in the following example:

Example

| Order | Day | Sample bale | Replicate of measurement | Reference material |
|---|---|---|---|---|
| | | | | every 20 samples |
| 1 | 68 | 70 | 1 | |
| 2 | 68 | 71 | 1 | |
| 3 | 68 | 71 | 2 | |
| 4 | 68 | 70 | 2 | |
| 5 | 20 | 50 | 1 | |
| 6 | 20 | 51 | 1 | |
| 7 | 20 | 51 | 2 | |
| 8 | 20 | 50 | 2 | |
| 9 | 75 | 86 | 1 | |
| 10 | 75 | 85 | 1 | |
| 11 | 75 | 86 | 2 | |
| 12 | 75 | 85 | 2 | |
| 13 | 3 | 1 | 1 | |
| 14 | 3 | 1 | 1 | |
| 15 | 3 | 2 | 2 | |
| 16 | 3 | 2 | 2 | |
| etc | | | | |

## 9.3 - For hypothesis 3

All samples will be analyzed together at any moment when all samples are available and over a short testing period. To avoid any drift effect on the data, the randomization will nest a block of measurement effect into a gining factor. All samples from a situation will be tested in a random order in first block, and in a different order in block 2 so that an estimation of factory the variance of the error of measurement can be made. In addition we will look at the reference material results to check if any existing drift is limited enough for assuring that all results from any given day are comparable (see following example).

| Order | Balle | Replicate of measurement | Reference material |
|---|---|---|---|
| | | | every 20 samp |
| 1 | 7 | 1 | |
| 2 | 6 | 1 | |
| 3 | 2 | 1 | |
| 4 | 5 | 1 | |
| 5 | 4 | 1 | |
| 6 | 10 | 1 | |
| 7 | 11 | 1 | |
| 8 | 1 | 1 | |
| 9 | 3 | 1 | |
| 10 | 9 | 1 | |
| 11 | … 100 | 1 | |
| … | … | … | |
| 100 | | 1 | |
| 101 | 8 | 2 | |
| 101 | 7 | 2 | |
| 101 | 6 | 2 | |
| 101 | 2 | 2 | |
| 101 | 5 | 2 | |
| 101 | 4 | 2 | |
| 101 | 10 | 2 | |
| 101 | 11 | 2 | |
| 101 | 1 | 2 | |
| 101 | 3 | 2 | |
| 101 | 9 | 2 | |
| 101 | 8 | 2 | |
| … | … | … | |
| 200 | … 100 | 2 | |

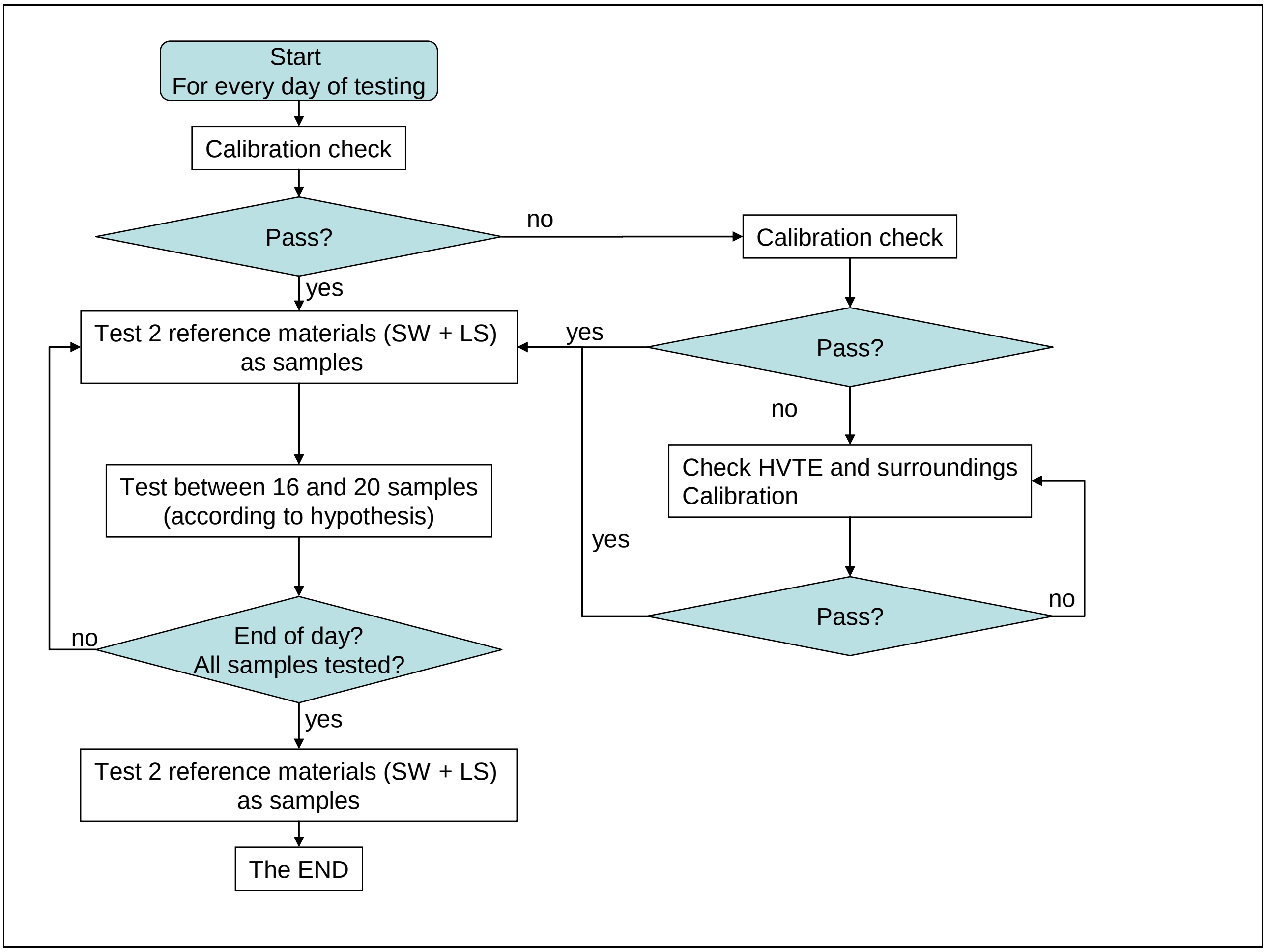
## 9.4 - For hypothesis 4

All samples will be analyzed together at the end of the crop within a short time. To avoid any drift effect on the data, the randomization will nest a block of measurement effect into a "day of collection of the couple of samples" factor. In addition we will look at the reference material results to check if any existing drift is limited enough for assuring that all results from any given day are comparable (see following example).

| Order | Day | Sample | Replicate of measurement | Reference material |
|---|---|---|---|---|
| | | | | every 20 samples |
| 1 | 56 | 86bottom | 1 | |
| 2 | 56 | 85top | 1 | |
| 3 | 56 | 85bottom | 1 | |
| 4 | 56 | 85bottom | 2 | |
| 5 | 56 | 86bottom | 2 | |
| 6 | 56 | 85top | 2 | |
| 7 | 34 | 85bottom | 1 | |
| 8 | 34 | 86bottom | 1 | |
| 9 | 34 | 85top | 1 | |
| 10 | 34 | 86bottom | 2 | |
| 11 | 34 | 85top | 2 | |
| 12 | 34 | 85bottom | 2 | |
| 13 | | | | |
| 14 | | | | |

## 10 - Rules for the testing of the samples

### 10.1 - General




### 10.2 - Verification of measurement stability during the testing part of this experiment

10.2.1 - Calibration of HVTE

Calibration first day, and then calibration checks; if pass=> no problem, if fail => check everything and make a new calibation check if Pass ok unless make a calibration + records + include reference materials periodically

10.2.2 - Other checks of collected data

- Include reference material tested as samples on a periodical basis ➔ protection against unstable within-day conditions
- Include another reference material (which one?) in addition to Short/Weak and Long/Strong Upland HVICC tested as samples on a periodical basis ➔ Protect against off-limits conditions (day-to-day offset change compared to expectations)
- Register conditions of testing

10.2.3 - Number of consecutive measurements per sample

The idea is tio statrt with the USDA protocol:

- one IM,
- two measurements of Length/ uniformity index,
- two measurements of strength,
- two measurements of Color Rd and yellowness (possibly Trash measurement in addition),
- Insure that all data information for every measurement are collected in the database for statistical analysis.

10.2.4 - Randomization of the testing of samples

For every situation, the order of analyzing samples should be randomized ➔ protection against drifts of results along time (because of testing or lab conditions or human fatigue …)

10.2.5 - What is a result?

Result = true value + lab bias + sampling error + measurement error (supposed independent from sampling error) + random error;
➔ the final operating method will have to balance between sampling and measurements errors to help in defining the number of samples to be taken out of bales and the number of measurements to be made per sample on any cotton crop in Africa (extrapolation from the studied situations).

## 11 - Analysis of the results obtained on batches of 100 consecutive bales in various ginning factories in Tanzania (preliminary data set)

The results have been checked for normality and plotted against time; a variogram has been computed to detect any correlation between neighbouring bales that could be exploited for module averaging.
Data are analysed and stored in a pdf file separately for each variable and each ginning factory. Data analysis results are commented in the present document.

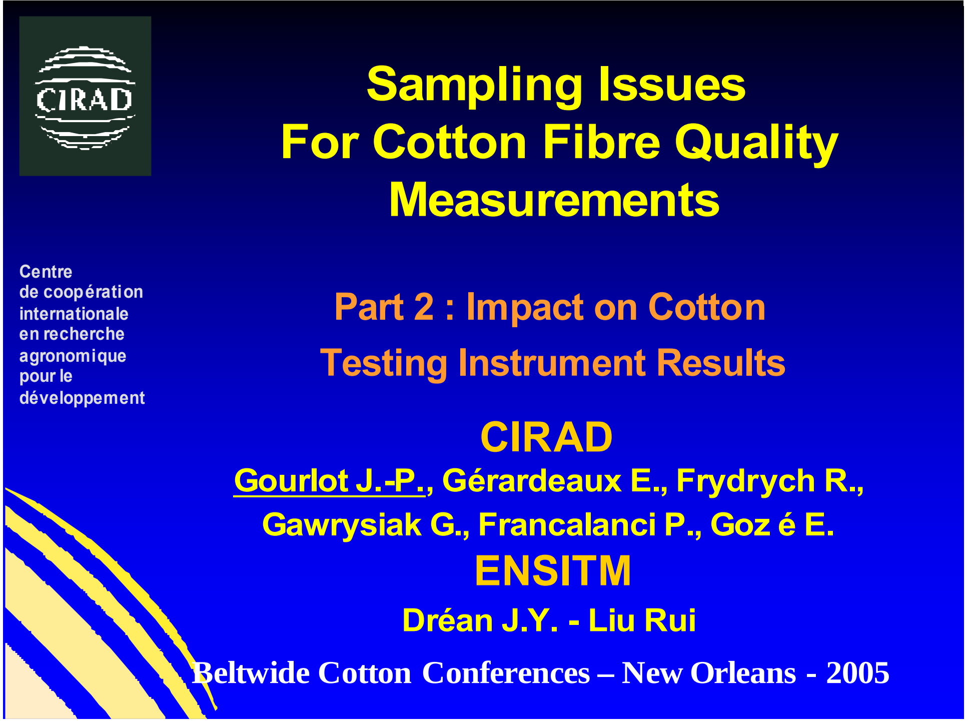
### 11.1 - 1 Upper Half Mean Length (UHML)

11.1.1 - 1.1 UHML OLAM B 6782 B.pdf

Fact: The data does not follow a Gaussian distribution.
Interpretation: in the case there was no difference between bales, we would expect the observations to be Gaussian, as it would result solely from the sampling and measurement errors that are usually Gaussian. If we suppose so, that we can infer that this 100 bales sample comes from a heterogeneous population of bales (e.g. coming from several modules).

Fact: The results plotted against bale ID do not show any trend. Nor does regression of the observations against bale id, either with a linear regression or smoothing spline.
The variogram does not show any clear trend, so there is no visible correlation between neighbouring bales;
Interpretation: either this shows erratic quality variations from one bale to another; or no real quality variations but only measurement errors that are essentially erratic, not correlated.

Fact: the observed standard deviation is 0.0231 inches, which is 0.59 mm. This is very similar to the within bale standard deviation of UHML measured on 8 layers within bales from different origins (Gourlot *et al*, 2005).
Interpretation: so provisionally we can suppose that these erratic variations result from measurement and sampling errors, and that all the bales of this particular batch had the same quality. However, this is somewhat contradictory with the non-Gaussian distribution.



## Statistical results

**Main result 2 : The variance components depend on the origin of the coton, and the error component is dominant**

| UHML | $\sigma_{Layer}$ | $\sigma_{Column}$ | $\sigma_{Error}$ |
|---|---|---|---|
| Origin 1 | 0.11 | 0.03 | 0.59 |
| Origin 2 | 0.00 | 0.18 | 0.55 |
| Origin 3 | 0.08 | 0.11 | 0.51 |
| Origin 4 | 0.21 | 0.16 | 0.68 |