Stratification of sorghum hybrid testing sites in southern Africa based on grain yield

M.A. Mgonja a,*, S. Chandra b,2, A.B. Obilana a,1, E.S. Monyo c, S. Kudita d, M. Chisi e, H.M. Saadan f, E. Chinhema d

a International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Box 39063, Nairobi, Kenya
b International Crops Research Institute for Semi-Arid Tropics (ICRISAT), Patancheru, AP 502324, India
c International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Box 1096, Lilongwe, Malawi
d International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Box 776, Bulawayo, Zimbabwe
e Ministry of Agriculture & Cooperatives, C/o GART, P O Box 54, Fringila, Zambia
f Ministry of Agriculture and Food Security, P O Box 9071, Dar-es-Salaam, Tanzania

1. Introduction

Genotype-by-environment interactions (GEIs) complicate the development of improved genotypes for a crop in a targeted population of environments. Multi-environment trials (METs) provide an opportunity to assess the effects of GEI in order to discern the pattern of genotypic adaptation over sampled environments. GEI partly constitute heritable components of variation and can therefore be exploited through selection for broad and specific adaptation (Cooper and Fox, 1996). The strategy of dealing with GEI requires that, within the target population of environments, the environmental factors that discriminate among genotypes should have some degree of repeatability (Baker, 1988b; Cooper et al., 1993a). Characterization of sampled environments in the presence of GEI is therefore a necessary first step to evaluate specific adaptation. Cooper and Fox (1996) identified two broad approaches for environmental characterization: (a) direct characterization through measurement of environmental variables such as water availability, physical and nutritional status of soils and (b) indirect characterization accomplished in terms of the way the environments affect the relative performance of genotypes. It is the latter that plays a more important and direct role in breeding for improved genotypes.

Southern Africa Development Community (SADC) covers a large geographic area that spreads over 14 southern Africa countries. Characterization and stratification of testing sites in the SADC region could help in choosing appropriate future testing sites and in objective targeting of hybrids for maximizing production (Peterson, 1992). The availability of historical sorghum hybrid grain yield data from a large number of multi-location multi-year regional trials conducted in SADC region presents a unique opportunity to undertake this task.

The normal practice that over the years (Y), is that plant breeders change the number as well as the composition of both the genotypes (G) and the sites (L) in regional trials. This makes the combined analysis of these trials quite challenging due to the highly imbalanced structure of GLY data. For the purpose of
stratifying the testing sites according to their similarities in genotypic responses, statistical techniques developed over the last decade for this purpose attempt to account for this imbalance in GLY data through across-year averaging of suitably derived yearwise site proximity matrices. According to Peterson (1992), this approach is expected to minimize the effects of missing data and short-term weather events or rare disease epidemics on relationships among the testing sites. This approach has been used by Peterson and Pfeiffer (1989) and Peterson (1992) to stratify international winter wheat testing sites based on 17 years’ trial data. It has also been used by DeLacy et al. (1990) to classify Australian cotton testing sites based on 6 years’ data.

This paper reports on results of stratification of sorghum hybrid testing sites and their implications for hybrid evaluation in the SADC region using historical grain yield data.

2. Materials and methods

2.1. Structure and history of field trial

Sorghum growing areas in the SADC region fall under four broad agro-ecosystems: (1) short season and often drought stricken environments with less than 3 months growing season, (2) warm humid environments with less than 4-months growing season, (3) environments characterized by cool nights as a result of high altitude, and (4) environments characterized by a long season growing season (more than 4-months) with sporadic and unreliable rainfall regime. Each agro-ecosystem is characterized by a wide range of soil types, texture and water holding properties.

The Sorghum and Millets Improvement Program (SMIP) was initiated and implemented by ICRISAT from 1983 to 2003 and it was a major crop-breeding project for the SADC region. The program’s major objective was to develop improved sorghum and millet cultivars for the drought-prone SADC region, currently comprising of 14 southern Africa countries. In the process of identifying suitable cultivars for recommendation to farmers, elite lines were tested for 2–3 years across a range of locations in regional trials before the best lines were selected for further evaluation. However, the number and the composition of cultivars evaluated, as well as test-sites used, changed over successive years.

The major criteria for selection of the regional testing sites were their representativeness and proximity to major sorghum-growing areas. The number of testing sites within each country was mainly determined, by the importance of the crop, resource availability and the relative strength of the national research program. For example, nine sites were selected in Zimbabwe being the regional headquarters of SMIP; four sites in Tanzania; three sites in Namibia; two sites each in Botswana, Lesotho, Malawi and Zambia; and one site each in Mozambique and Swaziland. The major biophysical characteristics of the testing sites are presented in Table 1.

This study used grain yield data from 150 METs conducted in 28-sorghum hybrid testing sites over a period of 7 years (1988/1989–1999/2000). The 28 sites consisted of 26 normally planted and two late-planted locations denoted as Luc1 and Mat1. The late-planted sites were treated as different environments due to differences in the planting dates, cumulative rainfall during the cropping season as well as harvesting dates.

Genetic materials consisted of introduced parent materials as well as those developed to particularly suit the conditions in Southern Africa. These can be classified into two broad categories: hybrids that were mainly derived from parental line introductions as well as those developed to particularly suit the conditions in Southern Africa. These can be classified into two broad categories: hybrids that were mainly derived from parental line introductions during the early years of the program (1987/1988–1989/1990), and the hybrids developed in the region during the subsequent 5

Table 1

<table>
<thead>
<tr>
<th>Site (code)</th>
<th>Country</th>
<th>Soil typeb</th>
<th>SWHCc</th>
<th>pH</th>
<th>Drainagec</th>
<th>Longitude</th>
<th>Latitude</th>
<th>Altitude (m)</th>
<th>Annual rainfall (mm)</th>
<th>First monthb</th>
<th>Min temp. (°C)</th>
<th>Max temp. (°C)</th>
<th>LGPc</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sebele (seb)</td>
<td>Botswana</td>
<td>C</td>
<td>VL</td>
<td>6.4</td>
<td>ED</td>
<td>26.0</td>
<td>-24.6</td>
<td>976</td>
<td>495</td>
<td>11</td>
<td>12</td>
<td>28</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Pandamatenga (pand)</td>
<td>Botswana</td>
<td>M/F</td>
<td>H</td>
<td>6.4</td>
<td>MD</td>
<td>25.7</td>
<td>-18.3</td>
<td>1070</td>
<td>671</td>
<td>12</td>
<td>14</td>
<td>30</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Maseru (mas)</td>
<td>Mosaic (mas)</td>
<td>M/F</td>
<td>L</td>
<td>6.2</td>
<td>WD</td>
<td>27.6</td>
<td>-29.5</td>
<td>1501</td>
<td>669</td>
<td>12</td>
<td>8</td>
<td>23</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Leribe (ler)</td>
<td>Lesotho</td>
<td>M/F</td>
<td>L</td>
<td>6.2</td>
<td>WD</td>
<td>28.0</td>
<td>-28.9</td>
<td>1630</td>
<td>778</td>
<td>12</td>
<td>8</td>
<td>22</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Ngabu (nga)</td>
<td>Malawi</td>
<td>F</td>
<td>H</td>
<td>7.3</td>
<td>ID</td>
<td>34.9</td>
<td>-16.5</td>
<td>115</td>
<td>760</td>
<td>11</td>
<td>19</td>
<td>32</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Nampula (nampula)</td>
<td>Mozambique</td>
<td>M/F</td>
<td>VL</td>
<td>6.4</td>
<td>WD</td>
<td>39.3</td>
<td>-15.1</td>
<td>329</td>
<td>1045</td>
<td>10</td>
<td>19</td>
<td>31</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Mashare (mash)</td>
<td>Namibia</td>
<td>M</td>
<td>6.6</td>
<td>ID</td>
<td>20.2</td>
<td>-17.9</td>
<td>1061</td>
<td>568</td>
<td>14</td>
<td>11</td>
<td>31</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Mahanene (mah)</td>
<td>Namibia</td>
<td>M</td>
<td>6.4</td>
<td>ID</td>
<td>15.2</td>
<td>-17.5</td>
<td>1110</td>
<td>505</td>
<td>11</td>
<td>13</td>
<td>29</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Okashana (okashana)</td>
<td>Namibia</td>
<td>C</td>
<td>H</td>
<td>8.5</td>
<td>WD</td>
<td>16.5</td>
<td>-18.3</td>
<td>1097</td>
<td>446</td>
<td>1</td>
<td>15</td>
<td>31</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Malkorns (malkorns)</td>
<td>Swaziland</td>
<td>M/F</td>
<td>H</td>
<td>6.2</td>
<td>WD</td>
<td>31.2</td>
<td>-26.6</td>
<td>763</td>
<td>890</td>
<td>12</td>
<td>13</td>
<td>36</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Ilongo (ilo)</td>
<td>Tanzania</td>
<td>F</td>
<td>L</td>
<td>5.7</td>
<td>WD</td>
<td>37.0</td>
<td>-6.8</td>
<td>914</td>
<td>978</td>
<td>11</td>
<td>16</td>
<td>28</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Ukiriguru (ukiri)</td>
<td>Tanzania</td>
<td>C/M</td>
<td>M</td>
<td>5.4</td>
<td>ID</td>
<td>33.0</td>
<td>-2.7</td>
<td>1239</td>
<td>952</td>
<td>11</td>
<td>17</td>
<td>28</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Hombo (hom)</td>
<td>Tanzania</td>
<td>C/M</td>
<td>M</td>
<td>5.4</td>
<td>MD</td>
<td>35.9</td>
<td>-6.0</td>
<td>1019</td>
<td>562</td>
<td>12</td>
<td>16</td>
<td>30</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Nabendele (nabendele)</td>
<td>Tanzania</td>
<td>C/M</td>
<td>L</td>
<td>5.1</td>
<td>WD</td>
<td>38.8</td>
<td>-10.4</td>
<td>383</td>
<td>876</td>
<td>2</td>
<td>20</td>
<td>31</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Golden Valley (golden valley)</td>
<td>Zambia</td>
<td>M/F</td>
<td>M</td>
<td>6.4</td>
<td>MWD</td>
<td>28.1</td>
<td>-14.9</td>
<td>1189</td>
<td>909</td>
<td>11</td>
<td>14</td>
<td>27</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Lusitu (luscit)</td>
<td>Zambia</td>
<td>C</td>
<td>H</td>
<td>6.4</td>
<td>WD</td>
<td>28.8</td>
<td>-16.1</td>
<td>326</td>
<td>632</td>
<td>11</td>
<td>19</td>
<td>32</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Aisleby (aisleby)</td>
<td>Zimbabwe</td>
<td>F</td>
<td>H</td>
<td>5.8</td>
<td>MD</td>
<td>28.57</td>
<td>-20.04</td>
<td>1282</td>
<td>594</td>
<td>11</td>
<td>11</td>
<td>30</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Matopos (matopos)</td>
<td>Zimbabwe</td>
<td>M</td>
<td>H</td>
<td>6.4</td>
<td>MD</td>
<td>28.5</td>
<td>-20.4</td>
<td>1416</td>
<td>591</td>
<td>11</td>
<td>12</td>
<td>25</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Makholi (makholi)</td>
<td>Zimbabwe</td>
<td>M</td>
<td>M</td>
<td>6.4</td>
<td>MWD</td>
<td>30.8</td>
<td>-19.8</td>
<td>1111</td>
<td>628</td>
<td>12</td>
<td>13</td>
<td>26</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Kadoma (kadoma)</td>
<td>Zimbabwe</td>
<td>F</td>
<td>H</td>
<td>6.3</td>
<td>MWD</td>
<td>29.9</td>
<td>-18.3</td>
<td>1107</td>
<td>735</td>
<td>12</td>
<td>14</td>
<td>28</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Lucylade (lucylade)</td>
<td>Zimbabwe</td>
<td>M</td>
<td>6.4</td>
<td>MD</td>
<td>28.5</td>
<td>-20.4</td>
<td>1416</td>
<td>591</td>
<td>11</td>
<td>12</td>
<td>25</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Chiredzi (chiredzi)</td>
<td>Zimbabwe</td>
<td>C/M</td>
<td>M</td>
<td>6.4</td>
<td>WD</td>
<td>31.7</td>
<td>-21.1</td>
<td>388</td>
<td>544</td>
<td>11</td>
<td>15</td>
<td>30</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Muzarabani (b)</td>
<td>Zimbabwe</td>
<td>M</td>
<td>H</td>
<td>6.4</td>
<td>MWD</td>
<td>31.0</td>
<td>-16.4</td>
<td>427</td>
<td>665</td>
<td>12</td>
<td>17</td>
<td>32</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Gwere (gwe)</td>
<td>Zimbabwe</td>
<td>C</td>
<td>H</td>
<td>6.1</td>
<td>ID</td>
<td>30.8</td>
<td>-17.7</td>
<td>1418</td>
<td>831</td>
<td>11</td>
<td>16</td>
<td>26</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Panmure (pan)</td>
<td>Zimbabwe</td>
<td>M</td>
<td>M</td>
<td>6.6</td>
<td>MWD</td>
<td>31.6</td>
<td>-17.3</td>
<td>1037</td>
<td>817</td>
<td>11</td>
<td>13</td>
<td>27</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

a Southern Africa Development Community.

b M, medium; F, fine; C, coarse.

c Soil Water Holding Capacity; H, high; M, medium; L, low; VL, very low.

d WD, well drained; ID, imperfectly drained; MD, moderately drained; MWD, moderately well drained; PD, poorly drained; ED, excessively drained; SED, somewhat excessively drained.

e Officially declared month for beginning of planting.

f Length of growing period.
3. Results

Set 1 (1987/1988–1992/1993) contained 23 sites across 6 years (luc and luc1, and mat and mat1 treated as different sites due to late planting). The SeqRet analysis retained 10 sites, but 13 sites were eliminated due to lack of comparisons. A clustering of these 23 sites into six groups (Fig. 1), with the 13 eliminated sites assigned to site groups with the nearest centroid, delivered a model fit of $R^2 = 68\%$. The six groups were $\text{grp1} = \{\text{mas}, \text{ilo}, \text{gwe}, \text{kad}, \text{luc1}, \text{mat1}\}$, $\text{grp2} = \{\text{seb}, \text{pand}, \text{nga}, \text{pan}, \text{uki}\}$, $\text{grp3} = \{\text{kas}, \text{ais}, \text{hom}, \text{ler}, \text{mak}, \text{mal}\}$, $\text{grp4} = \{\text{gol}\}$, and $\text{grp5} = \{\text{lus}\}$, and $\text{grp6} = \{\text{muz}, \text{mat}, \text{chi}\}$.

Set 2 (1987/1988–1992/1993, 1999/2000) contained 28 sites across the 7 years, 23 sites being the same as in Set 1. The cumulative analysis retained 10 sites. The remaining 18 sites were eliminated due to lack of comparisons. A clustering of these 28 sites into six groups (Fig. 2), with the 18 eliminated sites assigned to site groups with nearest centroid, provided a model fit of $R^2 = 69\%$. The six groups were $\text{grp1} = \{\text{kas}, \text{ais}, \text{hom}, \text{kad}, \text{ler}, \text{mak}, \text{mal}, \text{mash}, \text{oka}\}$, $\text{grp2} = \{\text{gol}\}$, $\text{grp3} = \{\text{muz}, \text{mat}, \text{chi}\}$,
grp4 = {lus}, grp5 = {mas, ilo, gwe, luc1, mat1}, and grp6 = {seb, pand, nal, nam, nga, pan, uki}.

The site grouping from Set 1 (Fig. 1) was retained in Set 2 (Fig. 2). The group grp1 in Set 2 contains all the sites in-group grp3 in Set 1. Similar results were observed in other groups. Thus, despite a break of 7 years, the site grouping obtained in Set 1 remained stable after taking into account data from the 1999/2000 season.

The site proximity plots of 10 retained sites in Set 1 (Fig. 3) and Set 2 (Fig. 4) clearly indicate that when the 1999/2000 data was added to the analysis, the relative positions of the sites in the Euclidean space converged to almost fixed positions.

4. Discussion

The stratification analysis for 6 years data (1987/1988–1992/1993) partitioned the testing sites into six groups with $R^2 = 68\%$ (Set 1). Even with the addition of 1999/2000 data, the stratification remained essentially the same with $R^2 = 69\%$ (Set 2). Similar observations were made when the SADC regional sorghum testing sites were stratified based on grain yield of varieties (Mgonja et al., 2006).

The 6 years’ data stratified the testing sites with three clear groups of short duration environments (3–4-month LGP) namely group 2 (Sebele, Pandamatenga, Ngabu), group 3 (Kasinthula, Aisleby, Hombolo), and group 6 (Muzarabani, Matopos and Chiredzi). Exceptions to this stratification are five long duration sites (5–7-month LGP)—Ukiriguru and Panmure in group 2, Leribe, Makoholi and Malkerns in group 3. Panmure and Makoholi, both in Zimbabwe, are clearly border line sites between short to intermediate duration (just about 5-month LGP). Although Ukiriguru is characterized as a long duration site, sorghum is normally planted after everything else and is most often planted in February when actually the season begins in November. The sorghum crop in Ukiriguru therefore experiences a short season. The short duration sites in group 2 cut across two countries (Botswana and Malawi), those in group 3 fall in three countries (Malawi, Zimbabwe and Tanzania) and group 6 sites are all in Zimbabwe. Besides being short duration environments, the three groups also represent distinct environments. Groups 2 and 3 are characterized by hot, often drought stricken environments with severe midseason drought. The group 6 sites are characterized by deep and well-drained soils. Besides being short duration environments, the observation is that each of the three groups includes a site or sites in low altitude areas (Ngabu in group 2, Kasinthula in group 3 and Muzarabani and Chiredzi in group 6) all within 115–427 meters above sea level (masl). Whereas the low altitude sites would have been expected to be group together, the separate grouping indicates that the sites discriminate the hybrids differently.

The long season sites, with ≥5 months LGP, form separate groups by themselves; Golden Valley group 4 in Set 1 and group 2 in Set 2, with Illonga, Maseru, Gwebi and Kadoma falling under group 1 in Set 1, and group 5 in Set 2 without Kadoma. In most cases, the Lucydale environment was similar to the long duration sites.

The grouping together of Sebele and Pandamatenga in Botswana was unexpected given the known differences in soil texture in the two sites. In Sebele the soils are medium to fine and well drained while Pandamatenga has heavy deep soils that are somewhat poorly drained. A separate analysis of data from these two locations revealed significant GXL interactions within years, a phenomenon confirming their distinctiveness. The explanation for this observation might be that the stress caused by incidences of water logging during the season may be affecting the plants in a similar manner to the mid-season drought-induced stress experienced at Sebele. This explains why germplasm appropriately tested and selected at the Botswana’s main research center of Sebele normally performs quite well in Pandamatenga despite the known environmental differences. This implies that higher gains from selection can be obtained by concentrating testing and selection efforts in Sebele or Pandamatenga and another site and not on both of them simultaneously.
The low altitude environments (40–427 masl) with warm nights (17–22 °C) stratified together (Fig. 1), forming two groups; Muzarabani and Chiredzi (group 6) and Lusitu (group 5).

When the 1999/2000 data is considered, the site stratification remains essentially the same, with additional sites added to appropriate groups. For example, group 1 in Set 2 comprises all sites in group 3 of Set 1 plus Kadoma, Mahanene, Mashare and Okashana, all of which are short season environments which were not included in Set 1. Similarly, group 6 of Set 2 consists of all sites in group 2 of Set 1 plus Naliendele, and Nampula as additional sites which were excluded in Set 1. Just like Pandamatenga, Naliendele and Nampula are prone to water logging induced stress during the season. The 26 (excluding the second plantings in Lucydale and Matopos) SADC sorghum hybrid-testing sites have been stratified into six major groups. This was consistent with the stratification into six groups of 34 sorghum-testing sites based on grain yield of varieties. Likewise, the grouping together of Sebele, Pandamatenga and Ukiriguru on one hand and Mashare, Mahanene and Okashana on the other, based on stratification of sites using grain yield data from sorghum hybrids, was also similar to earlier stratification based on grain yields of sorghum varieties (Mgonja et al., 2006).

5. Conclusions

The stratification of testing sites is an approach that minimizes the within sites-group GXE interactions. A sampling of sites from each group would therefore represent the full spectrum of variation in the SADC sorghum testing environments and limit the number of testing sites that need to be used for regional cultivar evaluation without loss of scientific rigor. The application of SeqRet Pattern analysis stratified the 28 SADC regional sorghum hybrid sites representing 26 distinct locations (Table 1) into six major groups:

- **Group 1**: Kasinthula, Aislebyl, Hombolo, Kadoma, Leribe, Mahanene, Makoholi, Malkerns, Mashare, Okashana
- **Group 2**: Golden Valley
- **Group 3**: Muzarabani, Matopos, Chiredzi
- **Group 4**: Lusitu
- **Group 5**: Mashare, Ilonga, Lucydale, Gwebi, Lucydale1, Matopos1
- **Group 6**: Sebele, Pandamatenga, Naliendele, Nampula, Ngabu, Panmure, Ukiriguru

The results from indirect characterization in this study, in terms of the way environments affect the relative performance of genotypes, produced results that are consistent with those obtained earlier on using grain yield data of sorghum varieties. This indirect characterization is considered important, and has a direct role in breeding for improved genotypes. The stratification of sites will also contribute to increased efficiency and cost effectiveness in regional cultivar evaluation and registration that is currently being pursued by SADC member states to facilitate regional seed marketing.

Acknowledgements

The authors are grateful to the scientists of the National Agricultural Research System (NARS) in the SADC region for their collaboration in the implementation of the METs. We also appreciate the financial support provided to the scientists by the Sorghum and Millet Improvement Program (SMIP), a USAID supported sorghum program of the SADC member states. Appreciation is extended to ICRISAT staff who worked on the SMIP project over the years, my colleagues Dan Kiambi and also to the Field Crop Research reviewers for providing constructive criticisms during finalization of this manuscript.

References