

Use of Lorenz curves and Gini coefficients to assess yield inequality within paddocks

Victor Sadras^{a, *} and Rodolfo Bongiovanni^b

^a*CSIRO Land and Water & APSRU, PMB2 Glen Osmond 5064, Australia*

^b*National Institute of Agricultural Technology (INTA), 5988 Manfredi, Argentina*

Abstract

The opportunity for site-specific management of crops depends on both the magnitude and spatial structure of yield variation. This study explored the applicability of Lorenz curves and Gini coefficients (G) to characterize the magnitude of the variation in grain yield. Maize crops were grown in farmers fields in Córdoba, a semi-arid region of central Argentina; major sources of yield variation between and within paddocks included season, soil type and topography, rate of nitrogen (N) fertiliser (nil to 132 kg N ha⁻¹), and the interactions among these factors. Grain yield was measured in 9.8 x 9.8 m field sections using an AgLeader yield monitor. Average yield in 0.7-2.8 ha field sections ranged from 1.6 to 7.0 t ha⁻¹. Gini coefficients ranged from 0.027 to 0.191 whereas its theoretical limits are 0 for a perfectly uniform population, and 1 for a theoretical population of infinite size, where all units but one yield 0. Conditions conducive to high yield, e.g. adequate availability of nitrogen and water, reduced crop yield inequality, were quantified with G. The agronomic relevance of G was summarised in an inverse relationship with yield. Lorenz curves and Gini coefficients provide a potentially useful extension tool, a complement to yield maps and other statistical indices (which ones?) of yield variation, and further contact points between site-specific management, economics and ecology.

Key words: Bimodality, Hierarchy, Inequality, Maize, Nitrogen, Population Ecology

1. Introduction

Increasing interest in site-specific management is largely driven by satellite aided geo-positioning and information technology. In the definition of Lowenberg-DeBoer and Swinton (1997), site-specific management is electronic monitoring and control applied to data collection, information processing and decision support for the temporal and spatial allocation of inputs for crop production. The variogram, as used in geo-statistics, is the central concept in characterising the scale and pattern of spatial variation in precision agriculture (Oliver, 1999). The variogram measures the spatial correlation or dependence between sampling points; it is based on the notion that properties of the soil or the yield of crops are more likely to be similar at nearby locations than at distant ones. On the other hand, spatial econometrics –created by economists to model spatial dependence in social sciences, has recently been adapted by Bongiovanni (2002) to study the economics of site-specific management in yield crops. Spatial econometrics assumes that spatial covariance is an interaction among discrete spatial objects, i.e. polygons or cells. This requires the specification of a spatial stochastic process with a neighbour structure, i.e. the weights matrix, in order to correct for spatial autocorrelation and heteroskedasticity.

Pringle et al. (2003) highlighted the importance of deciding whether a crop displays enough variation - both in terms of magnitude and spatial distribution – to warrant change from uniform management. They proposed an opportunity index that accounts for both aspects of yield variation, and discussed some indices including coefficient of variation, fractal dimension, semi-variogram, and Fairfield Smith b' index. The work of Pringle et al. (2003) exemplifies the key motivation for research in this area, i.e. making site-specific management an economically viable alternative to farmers.

Site-specific management involves cutting-edge technologies but also a static view of crop yield that pre-dates the current paradigm of plant growth based on resource capture (Charles-Edwards, 1982; Monteith, 1977; Passioura, 1977). There is potential, therefore to increase the benefits of spatial and information technologies applied to agriculture by a better linkage with appropriate theories allowing for a dynamic view of yield determination. Interactions between neighbouring plants can reduce survival or reproductive outcome of suppressed individuals (Mohler et al., 1978; Yoda et al., 1963). Importantly, differences in early growth are compounded with time, leading to highly dynamic temporal patterns in the structure of populations (Harper, 1977). Statistical descriptors of size inequality in plant populations include moments around the mean, i.e. standard deviation, skewness, and kurtosis. Weiner and Solbrig (1984), however, emphasized the inadequacy of these moments, and proposed the use of methods developed by economists to evaluate inequality in wealth, including the Lorenz curve and Gini coefficient. These have become common tools in population biology, but they have been scarcely used to deal with agronomic issues (Pan et al., 2003; Sharma et al., 1998; Vega and Sadras, 2003).

This paper explores the applicability of Lorenz curves and Gini coefficients to analyse the magnitude of grain yield variation at the paddock scale. The study is not motivated by the most immediate questions in site-specific management, i.e. how to make it work for

farmers, but tries to explore opportunities for developing methodological and theoretical interfaces with plant population ecology.

2. Method

2.1 Experiment

The experiment, climate and soils of the region are described elsewhere (Bongiovanni, 2002; Bongiovanni and Lowenberg-DeBoer, 2001; Bongiovanni and Lowenberg-DeBoer, 2003). Briefly, the experiment was carried out in the area of Rio Cuarto, in the Sandy Pampas of Córdoba Province (33° S, 64° W), a semi-arid environment of central Argentina. The topography is undulating, with flatter sectors above 180-320 m above sea level. The origin of the soils is Aeolian, and these are predominantly of sandy loam texture, and with a high percentage of fine sand. Maize crops were grown in three farms belonging to the Argentine Association of Regional Consortia of Agricultural Experimentation (AACREA), a private organization where professional consultants advise small grower groups. Urea was applied to establish the nitrogen treatments outlined in Table 1. Rainfall was 963 mm in the 1998-99 cropping season and 624 mm in the 2000-01 cropping season, in comparison with a 15-year average of 839 mm. In the 2000-01 cropping season, experimental units at *El Piquete* were split into irrigated and non-irrigated treatments. Yield was measured in 9.8 x 9.8 m units using an AgLeader™ yield monitor. Since the raw data included points that were closer within the same row than between rows, these data yield points were averaged for a within-row distance equivalent to the between-rows distance. This was done with the Geographic Information System (GIS) software SToolbox™, creating 9.8 x 9.8 meter grids over the observations, and rotating them to match the rows. The grids were digitised as polygons after averaging the data within each grid. In this study, yield units comprise therefore 96 m², i.e. 9.8 x 9.8 m.

2.2. Lorenz curves and Gini coefficients

To construct Lorenz curves, yield units within each treatment were ranked from lowest to largest, and the cumulative fraction of total yield (y) was plotted against the cumulative fraction of the population (x). The upper limit of this curve is the y = x line indicating perfect equality. The Gini coefficient G quantifies the area between the Lorenz curve and the line of perfect equality expressed as a fraction of the area under the y = x line. It ranges from 0, when all units are equal, to a theoretical maximum of 1 in an infinite population in which all units but one yield 0 (Weiner and Solbrig, 1984). Polynomials were fitted to describe Lorenz curves (L; R² ~ 0.99), and G was calculated as (Harch et al., 1997; Weiner and Solbrig, 1984)

$$G = 1 - 2 \int_0^1 L \, dx \quad \text{eq. (1)}$$

Coefficient G was corrected by $n(n-1)^{-1}$ to account for the comparatively small number of yield units in the irrigated plots (Pan et al., 2003; Weiner and Solbrig, 1984).

3. Results

The combination of soils, seasons, availability of water and nitrogen generated a large variation in yield among and within plots (Table 1). Figure 1 presents examples of Lorenz curves and frequency distributions of yield for contrasting crops. For the high-yielding crop in Figure 1ab, yield units were concentrated in a range between 5 and 10 t ha⁻¹, with more than half of the population in a very narrow range between 5 and 7 t ha⁻¹; this is reflected in a Lorenz curve close to the $y = x$ line, and a low Gini coefficient (Figure 1 ab). In contrast, the low yielding crop in Figure 1cd had a marked, L-shaped frequency distribution, where 50% of the yield units only contributed ~ 20% of the total yield. The high degree of inequality in this crop is reflected in a large G. Comparison of Figure 1a vs. c thus suggests an inverse relationship between yield and G that was confirmed for the pooled data ($P < 0.0001$, Figure 2). The inverse relationship between yield and G was also evident in the analysis of particular farms and treatments. At *La Morocha* in 1998-99, increasing nitrogen fertiliser from 0 to 129 kg ha⁻¹ decreased G at a rate of 2.5×10^{-4} per kg N ($r^2 = 0.94$, $P = 0.001$), and increased grain yield at an average rate of 13 kg per kg N ($r^2 = 0.86$, $P = 0.008$). At *El Piquete* in 2000-01, irrigation halved G ($P = 0.003$) and doubled yield ($P = 0.0001$) (Figure 2).

The main deviation from the general relationship between yield and G was observed in the crops at *Las Rosas* in 2000-01 (Figure 2). With G around 0.09, yield at *Las Rosas* was almost twice the yield at *La Morocha* (open circles vs. open triangles in Figure 2). As pointed out by Weiner and Solbrig (1984), the Gini coefficient is the most important summary statistic for the Lorenz curve but, like any summary statistic, it does not contain all the information in the data. Different Lorenz curves can have similar G's, as shown in the comparison of Figure 1 e vs. g. It is worth noting how the Lorenz curve captures the bimodality of the population in Figure 1ef, with an inflection point at $x \sim 0.5$ (Figure 1e).

Discussion

Lorenz curves and Gini coefficients are widely used to characterise inequalities in economics and ecology (Beaugrand and Edwards, 2001; Bendel et al., 1989; Figini, 1998; Harch et al., 1997; Rousseau et al., 1999; van Hecke et al., 1995; Weiner and Solbrig, 1984). There are fewer applications of these tools in agronomic studies. Sharma et al. (1998) used Gini coefficients to quantify the effect of stubble management on the functional diversity of bacterial communities of European agricultural soils. Pan et al. (2003) used Lorenz curves and Gini coefficients to characterise variation in wheat crops along a gradient of water availability in China. They found an inverse association between harvest index and G, which is consistent with the inverse relationship between grain yield and G in Figure 2. The range of G for their wheat crops, from 0.067 to 0.190, compares with a range from 0.027 to 0.191 in our maize crops (Figure 2). Vega and Sadras (2003) used Lorenz curves to investigate the links between population size structuring and size-dependent plant growth in field crops of soybean, maize and sunflower. Population inequality was much larger in the irrigated, well fertilised, small experimental plots of Vega and Sadras (2003) (their Figure 6) than in the larger, commercial crops in this study (Figure 1). A number of reasons can account for this

difference, of which the definition of “yield unit” deserves some attention. In the analysis of Vega and Sadras (2003), yield units were individual plants, whereas the present study used 96 m² units. A hypothetical, bimodal population with two plant classes with yield ratio 1:2 will have $G = 0.083$, whereas the analysis of the same crop on an area basis would return $G \sim 0$. While it is an unrealistic target for site-specific crop management, plant-to-plant variation can have dramatic consequences for crop structure and functioning (Crawley, 1983; Hide et al., 1995; Hide et al., 1997; Hirst, 1973; Sadras, 1996). Scaling-up variation from individual plants to yield units of agronomic meaningful size is not straightforward, but it would open opportunities to incorporate plant population theories into the analysis of spatial variation in crops.

Lorenz consistency is a requirement for sound inequality indices. An index is Lorenz consistent when it satisfies the axioms of scale irrelevance, population replication, rank preserving transfer and anonymity (Figini, 1998). These properties make the Lorenz consistent Gini coefficient particularly apt for analysis of spatial variation in crops. Scale irrelevance means that if each income (yield) is multiplied by the same constant, inequality does not change. Systematic, multiplicative errors in yield monitoring would therefore be irrelevant in the calculation of G as a measure of paddock yield variation; the index can be reliably estimated with *relative* yield measures. It is valid therefore to use Lorenz curves and Gini coefficients to compare inequalities in populations with different means (Weiner and Solbrig, 1984), as illustrated in the comparison of plant- (Figure 6 in Vega and Sadras, 2003) and area-based quantification of inequalities (Figure 1 in this paper). It is important to emphasise the difference in Lorenz consistent indices to multiplicative and additive variation.

The additive axiom states that adding a constant to each income (yield) decreases inequality. If management or environmental factors increase yield in all spatial units by the same amount, the overall inequality in the paddock, as assessed with G , will decrease. The decrease in G with conditions that favour crop growth and yield are therefore related to this principle (Figure 2).

The principle of population replication states that inequality remains the same when the size of the population changes, provided the change does not affect the income (yield) shares of the corresponding percentile groups. The consequence of this principle is that yield units could be aggregated or disaggregated with no change in G if the condition of percentile maintenance is satisfied. This property allows for G to be used, within certain limits, for comparisons of fields of different sizes.

The rank preserving equalisation or Pigou-Dalton condition states that if a transfer of income (yield) $t < h/2$ from a person j (yield unit j) with income (yield) $x_j = x_i + h$ to a person i (unit i) with income (yield) x_i takes place, with h a positive real number, *ceteris paribus*, inequality diminishes. The “transfer” of yield between units could be mediated by management practices aiming, for instance, to allocate greater doses of fertiliser to more responsive units in detriment to less responsive ones.

The principle of anonymity states that if individuals (yield units) interchange their incomes (yields), inequality remains the same. The consequence of this property is that G , as other measures of inequality, would remain constant irrespective of where in the field the different units are located; like the coefficient of variation (Pringle et al., 2003); G is non-spatial. The use of G in site-specific management would therefore require its application in combined indices, such as the opportunity index of Pringle et al. (2003).

Another constraint for the use of G as an index of inequality is that the same coefficient could describe contrasting populations (Figure 1ef vs. gh). Despite being the most important summary statistics for the Lorenz curve, G (like any summary statistic) does not contain all the information in the data (Weiner and Solbrig, 1984). This limitation of G accounts for the deviation of crops at *Las Rosas* 2000-01 from the general trend; Lorenz curves are thus essential for a thorough agronomic interpretation of Gini coefficients.

In summary, Lorenz curves and Gini coefficients seem sound tools to characterise the magnitude of yield variation within paddocks, and could be complements to yield maps and other analytical tools. While the Gini coefficient provides information that could be similar to that provided by other statistics such as the coefficient of variation (Pan et al., 2003), there are two main benefits from the use of G and Lorenz curves.

First, looking at spatial variation in terms of inequality is possibly an effective approach for extension purposes. People are generally familiar with the concept of wealth inequality, as we often hear figures describing a small percentage of the population having a disproportionate share of the aggregated society income. Lorenz curves are thus particularly apt to present crop heterogeneity in terms of inequality, and to highlight the relative contribution of low- and high-yielding sections of the field to total paddock yield.

Second, sharing of analytical tools might provide new linking points between site specific- management and other disciplines, chiefly economics and ecology, which have a much more developed framework for dealing with inequalities.

Acknowledgements

We thank Cristiano Casini (INTA Manfredi, Argentina) and Mario Buteler (University Siglo 21, Argentina) for comments on the manuscript, and the Australian Grains Research and Development Corporation for financial support.

References

- Beaugrand, G. and Edwards, M., 2001. Differences in performance among four indices used to evaluate diversity in planktonic ecosystems. *Oceanologica Acta*, 24(5): 467-477.
- Bendel, R.B., Higgings, S.S., Teberg, J.E. and Pyke, D.A., 1989. Comparison of skewness coefficient, coefficient of variation and Gini coefficient as inequality measures within populations. *Oecologia*, 78: 394-400.
- Bongiovanni, R., 2002. A spatial econometric approach to the economics of site-specific nitrogen management in corn production. Ph. D. Thesis, Purdue University, 306 pp.
- Bongiovanni, R. and Lowenberg-DeBoer, J., 2001. Precision agriculture: economics of nitrogen management in corn using site-specific crop response estimates from a spatial regression model, American Agricultural Economists Association Annual Meeting, Chicago.
- Bongiovanni, R. and Lowenberg-DeBoer, J., 2003. Precision agriculture and sustainability. *J. Precision Agriculture*, in press.
- Charles-Edwards, D.A., 1982. Physiological determinants of crop growth. Academic Press, North Ryde, 161 pp.
- Crawley, M., 1983. Herbivory. The dynamics of animal-plant interactions. *Studies in Ecology Vol 10*, 10. Blackwell Scientific Publications, London, 437 pp.
- Figini, P., 1998. On the properties of Field's index of inequality. *J. Income Distribution*, 8(1): 131-141.
- Harch, B.D., Correll, R.L., Meech, W., Kirkby, C.A. and Pankhurst, C.E., 1997. Using the Gini coefficient with BIOLOG substrate utilisation data to provide an alternative quantitative measure for comparing bacterial soil communities. *Journal of Microbiological Methods*, 30(1): 91-101.
- Harper, J.L., 1977. The population biology of plants. Academic Press, London.
- Hide, G.A., Welham, S.J., Read, P.J. and Ainsley, A.E., 1995. Influence of planting seed tubers with gangrene (*Phoma foveata*) and of neighbouring healthy, diseased and missing plants on the yield and size of potatoes. *J. Agric. Sci.*, 125(1): 51-60.
- Hide, G.A., Welham, S.J., Read, P.J. and Ainsley, A.E., 1997. Effects of planting mixtures of different sizes of potato seed tubers on the yield and size of tubers. *J. Agric. Sci.*, 128(2): 173-180.
- Hirst, J., Hide GA, Stedman OJ, Griffith RL, 1973. Yield compensation in gappy potato crops and methods to measure effects of fungi pathigenic on seed tubers. *Ann appl Biol*, 73: 143-150.
- Lowenberg-DeBoer, J. and Swinton, S.M., 1997. Economics of site-specific management in agronomic crops. In: F.J. Pierce and E.J. Sadler (Editors), *The state of site-specific management for agriculture*. ASA, CSSA, SSSA, Madison, pp. 369-396.
- Mohler, C.L., Marks, P.J. and Sprugen, D.G., 1978. Stand structure and allometry of trees during self-thinning of pure stands. *Journal of Ecology*, 66: 599-614.
- Monteith, J.L., 1977. Climate and efficiency of crop production in Britain. *Phil Trans Royal Soc London, Series B*, 281: 277-294.
- Oliver, M.A., 1999. Exploring soil spatial variation geostatistically. In: J.V. Stafford (Editor), *Precision Agriculture '99*. Sheffield Academic Press Ltd, Odense, Denmark, pp. 3-17.

- Pan, X.-Y., Wang, G.-X., Yang, H.-M. and Wei, X.-P., 2003. Effect of water deficits on within-plot variability in growth and grain yield of spring wheat in northwest China. *Field Crops Research*, 80(3): 195-205.
- Passioura, J., 1977. Grain yield, harvest index and water use of wheat. *J. Austr. Insti. Agric. Sci.*, 43: 117-120.
- Pringle, M.J., McBratney, A.B., Whelan, B.M. and Taylor, J.A., 2003. A preliminary approach to assessing the opportunity for site specific crop management in a field, using yield monitor data. *Agricultural Systems*, 76: 273-292.
- Rousseau, R., Van Hecke, P., Nijsen, D. and Bogaert, J., 1999. The relationship between diversity profiles, evenness and species richness based on partial ordering. *Environmental and Ecological Statistics*, 6(2): 211-223.
- Sadras, V.O., 1996. Population-level compensation after loss of vegetative buds: Interactions among damaged and undamaged cotton neighbours. *Oecologia*, 106: 432-439.
- Sharma, S., Rangger, A., von Lützwow, M. and Insam, H., 1998. Functional diversity of soil bacterial communities increases after maize litter amendment. *European Journal of Soil Biology*, 34(2): 53-60.
- van Hecke, P., Moermans, R., Mau, F. and Guittet, J., 1995. Border effects and size inequality in experimental even-aged stands of poplar clones (*Populus*). *Annales des Sciences Forestieres*, 52(3): 193-200.
- Vega, C.R.C. and Sadras, V.O., 2003. Size-dependent growth and the development of inequality in maize, sunflower and soybean. *Annals of Botany*, 91: 1-11.
- Weiner, J. and Solbrig, O.T., 1984. The meaning and measurement of size hierarchies in plant populations. *Oecologia*, 61: 334-336.
- Yoda, K., Kira, T., Ogawa, H. and Hozumi, K., 1963. Self-thinning in overcrowded pure stands under cultivated and natural conditions. *Journal of Biology Osaka City University*, 14: 107-129.

Table 1. Rate of nitrogen fertiliser, area monitored, and range of yield measured in 9.8 x 9.8 m units of maize crops grown in Córdoba, Argentina.

Farm	Season	N rate (kg N ha ⁻¹)	Area monitored (ha)	Yield (t ha ⁻¹)	
				Min.	Max.
<i>Las Rosas</i>	1999	0	2.8	3.8	7.8
		26	2.8	3.1	9.0
		53	2.7	4.4	8.1
		66	2.8	4.6	8.3
		106	2.8	5.0	8.8
		132	2.8	4.7	9.0
	2001	0	2.7	1.3	10.9
		39	2.7	2.7	11.1
		51	2.7	3.2	11.3
		75	2.7	1.9	11.0
		100	2.7	3.2	11.8
		125	2.7	3.2	11.7
<i>La Morocha</i>	1999	0	2.5	3.0	9.0
		29	1.7	4.3	8.8
		52	2.5	4.8	8.8
		77	2.5	5.3	9.3
		102	2.5	4.3	9.0
		129	2.4	5.4	9.0
	2001	0	2.1	1.4	7.0
		27	2.2	2.4	7.8
		52	2.0	2.4	8.1
		76	3.0	0.6	7.5
		102	2.1	0.8	7.4
		131	2.2	1.2	7.9
<i>El Piquete</i>	1999	0	2.1	0.0	9.0
		24	2.1	2.2	9.5
		59	2.1	3.5	8.5
		78	2.1	3.5	9.6
		105	2.1	3.4	9.0
	2001	0	0.9	0.7	11.4
		53	0.8	0.6	9.9
		70	0.7	0.6	4.1
		106	0.9	0.2	6.0
		125	1.0	0.2	3.8
	2001 Irrig*	0	1.6	2.8	6.3
		53	1.6	3.2	9.2
		70	1.8	2.6	6.2
		106	1.6	1.4	5.7
		125	1.6	0.9	6.3

* Irrigated as described in Bongiovanni (2002).

Captions to figures

1. Lorenz curves and frequency distribution of yield of maize crops in Córdoba, Argentina. In the left panels, the straight line represents absolute equality in the population. Data from (a, b) *La Morocha* 1999, 77 kg N ha⁻¹, yield: 7.1 t ha⁻¹, area monitored: 2.5 ha; (c, d) *El Piquete* 2001, 106 kg N ha⁻¹, yield: 1.7 t ha⁻¹, area monitored: 0.9 ha; (e, f) *Las Rosas* 2001, 39 kg N ha⁻¹, yield: 7.5 t ha⁻¹, area monitored: 2.7 ha; (g, h) *La Morocha* 2001, 102 kg N ha⁻¹, yield: 3.5 t ha⁻¹, area monitored: 2.1 ha.
2. Relationship between yield and Gini coefficient of maize crops grown in farmers fields in Córdoba, Argentina.

Figure 1



