VII Conferencia Española de Biometria
Palma de Mallorca 10-12 de marzo de 1999

Measuring the spatial pattern of *Orobanche crenata* weeds by SADIE red-blue analysis.

Joe N. Perry[1], Francisca López Granados[2]

[1]Department of Ent. & Nem., IACR - Rothamsted, Harpenden, Herts. AL5 2JQ, UK
[2]Instituto de Agricultura Sostenible (C.S.I.C.), Aptdo. 4084, 14080 Cordoba, Spain

RESUMEN

A new index and graphical displays, termed 'red-blue' plots, are presented to measure clustering in spatially-referenced count data. The index can detect clusters in the form of patches of several nearby large counts, and of gaps of several nearby small counts. The methods are illustrated using field counts of *Orobanche*, a Spanish parasitic weed.

Palabras y frases clave: weeds, spatial pattern, clusters, patches, gaps, SADIE

1. Introducción

This paper concerns spatial pattern in spatially-referenced ecological count data, where the position of each sample unit is recorded in two dimensions. The SADIE system (Perry *et al.* 1996; Perry 1998a) provides a description of the spatial features of such a set of counts, independently of their numeric properties. New methods are introduced here to detect and measure clusters. The term cluster means a region of either relatively large counts close to one another in two-dimensional space (i.e. a patch), or of relatively small counts (i.e. a gap). The new methods, termed red-blue techniques, utilize coloured graphical displays; an initial description was given by Perry *et al.* (1999).

SADIE uses the Transportation algorithm from the linear programming literature (Kennington & Helgason 1980) that provides notional 'flows' of individuals and fractions of individuals, from 'donor' sample units, with greater than average abundance, to 'receiver' units, with less than average abundance. These flows may be depicted graphically in an 'initial-and-final' plot (Perry, 1998b). The spatial pattern is quantified by permuting the observed set of counts amongst the sample units, to provide data to test the null-hypothesis that the counts are arranged randomly with respect to one another.

To supplement the current SADIE measure of overall spatial pattern an index of clustering is here derived for each sample unit. This index distinguishes whether a large count at that unit is isolated, or a member of a patch, or whether a relatively small count at that unit is isolated or forms part of a gap. The techniques employ previously unused information from the initial-and-final plot, to ascribe, to each sample unit, a measure of the degree to which it contributes to clustering.

The data used for the example (Fig. 1) are counts of individuals of the parasitic weed Crenate broomrape (*Orobanche crenata* Forsk.), collected by Francisca López-Granados, over a 20x15 grid measuring 20m x 40.2m, within a field of Faba bean (*Vicia faba* L.

'Alameda'), in southern Spain (López-Granados and Garcia-Torres, 1991). They range from 0 to 8, with sample mean and variance of 1.16 and 3.06, respectively. The initial-and-final plot for these data (Fig. 2) shows the outflows originating from donor units and terminating at receiver units denoted by asterisks. To maintain visual clarity, the strength of each flow is not marked on the plot.

2. Methods

The new techniques use the average distance of flow associated with each unit. For donor unit *i*, at position (x_i, y_i), the outflow to the *j*th of n_i receiver units, $j = 1, \ldots, n_i$, at position (x_j, y_j), is denoted as v_{ij}. The distance of this flow, $[(x_i - x_j)^2 + (y_i - y_j)^2]^{1/2}$, is denoted as d_{ij}. The average distance of outflow from unit *i*, weighted by the individual flows, is denoted as Y_i, where $Y_i = \Sigma_j d_{ij} v_{ij} / \Sigma_j v_{ij}$. In exactly similar fashion, there is an average distance of inflow associated with each receiver unit *j*; by convention the values of d_{ji} and hence of Y_j are taken to be negative. For units within a cluster, average distances of flow tend to be greater than for units in regions where similar counts are randomly distributed. Hence, clustering is indicated by unexpectedly large values of Y_i or of $|Y_j|$. If ${}_iY$ denotes the average absolute distance of flow for that point under the randomizations described in the introduction, ${}_cY$ denotes the average distance of outflow for the count observed at position (x_i, y_i) over its randomized positions, and ${}_oY$ denotes the overall mean average absolute distance of flow for all points and counts in the randomizations, then a standardized and dimensionless index of clustering, ν_i, is given by: $\nu_i = Y_i \; {}_oY \; / \; {}_iY \; {}_cY$. For inflows, ν_j is defined similarly, again with the convention that it is negative in sign. Values of ν_i larger than about 1.5 indicate patchiness; values of ν_j negatively larger than –1.5 indicate membership of a gap; values close to unity indicate a random placement of that unit in relation to others nearby.

When plotted on a map of the sample units and contoured, the values of the indices ν_i and ν_j indicate visually the location and extent of clusters in the data (Fig. 3). Note that ν_i and ν_j encapsulate spatial and not numeric information; each value is standardized to allow for both the magnitude of the observed count at that particular unit, and for the position of that unit relative to the others. An 'empirical distribution function' plot (Diggle 1983; Perry 1995) may be used as an informal test to detect clustering; more formal tests are provided by, for example, comparing the average values of ν_i for outflows and of ν_j for inflows with corresponding values from randomizations. The example in this paper relates to a rectangular grid of sample units, but this is not a requirement of the method; sample units may be located anywhere in two dimensions. Software to analyse data using the techniques in this paper is freely available from the first-named author.

Results

The overall SADIE index, I_a was 1.29 (P=0.081), indicating moderately strong, but not quite significant aggregation. There were two large gaps and several small patches and gaps (Fig. 3). The clustering index for patches was 1.27 (P=0.084) and for gaps was 1.19 (P=0.126). EDF plots for both outflows (Fig. 4, above) and inflows (Fig. 4, below) approach the 95th centile envelope, confirming moderately strong clustering into both patches and gaps.

References

Diggle, P.J. (1983). *Statistical Analysis of Spatial Point Patterns.* Academic Press, London.

Kennington, J.L., & R.V. Helgason, R.V. (1980). *Algorithms for Network Programming.* Wiley, New York.

López-Granados, F. & Garcia-Torres, L. (1993). Population dynamics of Crenate Broomrape (*Orobanche crenata*) in Faba bean (*Vicia faba*). *Weed Science*, **41**, 563-567.

Perry, J.N. (1995). Spatial analysis by distance indices. *J. Anim. Ecol.*, 64, 303-314.

Perry, J.N. (1998a). Measures of spatial pattern and spatial association for counts of insects. In: *Population and Community Ecology for Insect Management and Conservation.* ed. J. Baumgartner, P. Brandmayr & B.F.J. Manly: Balkema, Rotterdam, pp. 21-33

Perry, J.N. (1998b). Measures of spatial pattern for counts. *Ecology*, 79, 1008-1017.

Perry, J.N., Bell, E.D., Smith, R.H. & Woiwod, I.P. (1996). SADIE: software to measure and model spatial pattern. *Aspects of Applied Biology*, 46, 95 - 102.

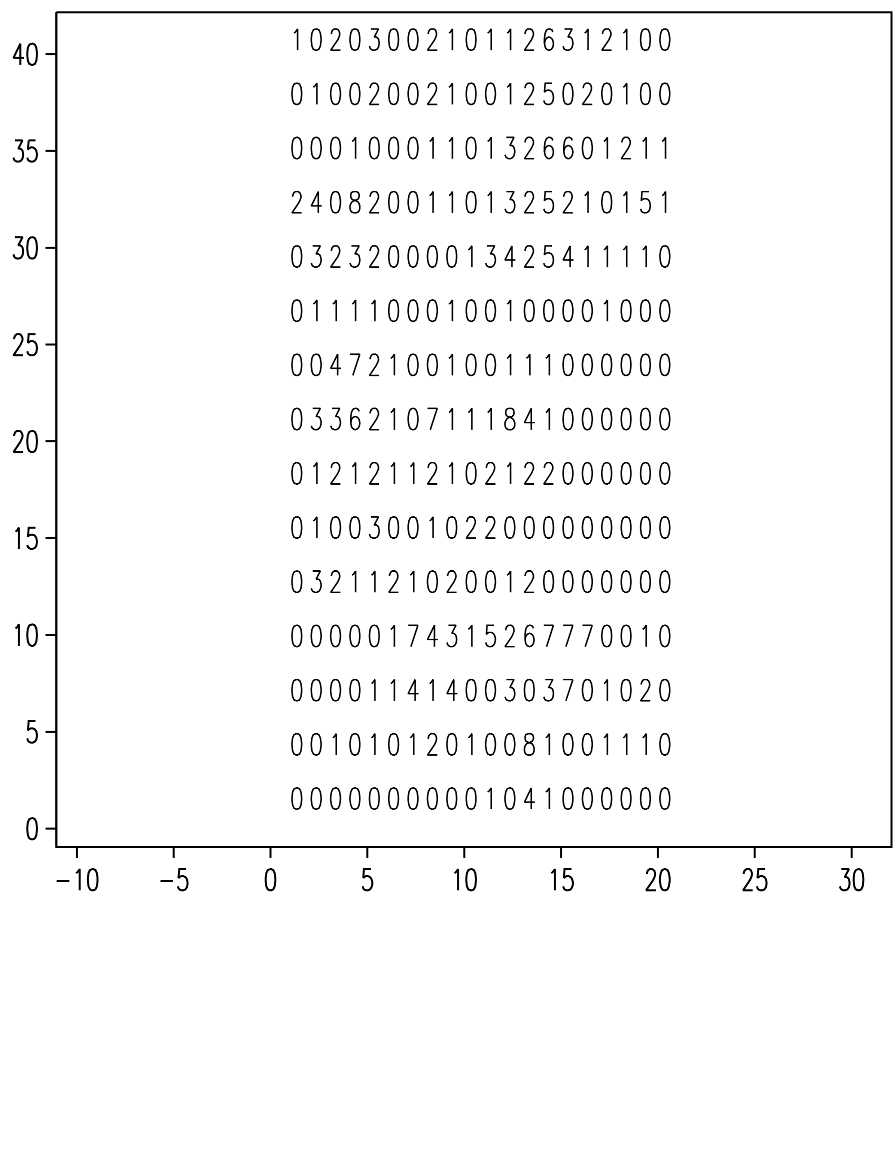
Perry, J.N., Winder, L. & Holland, J.M. (1999). Red-blue plots for detecting clusters in count data. *Ecology Letters* (submitted).

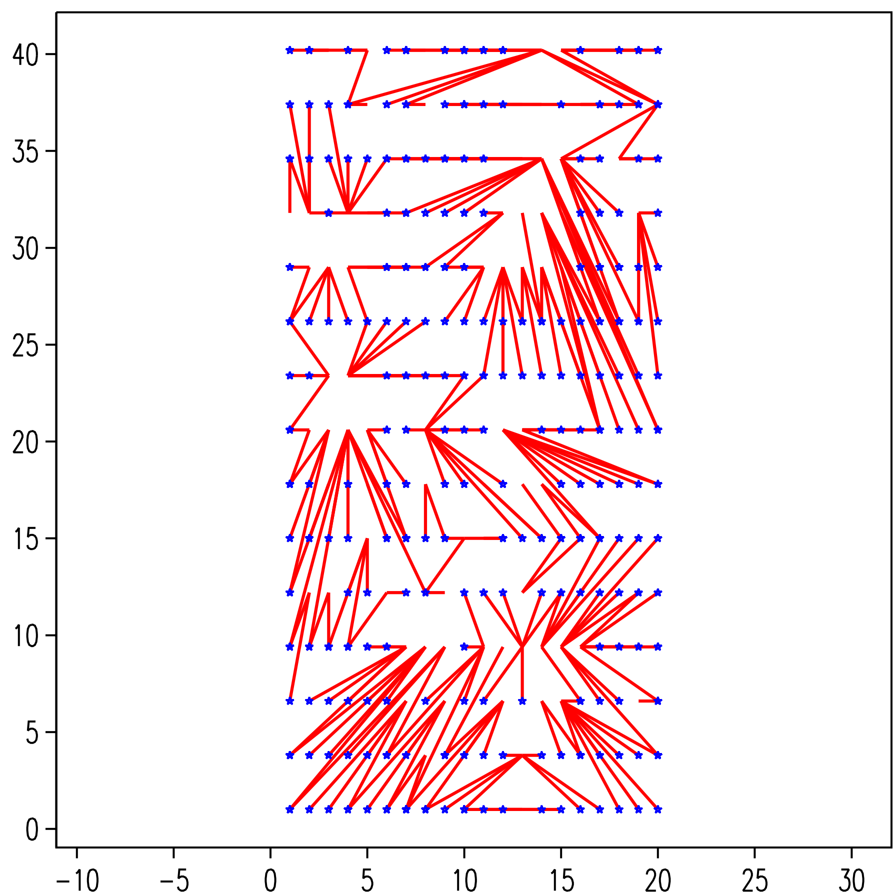
Fig. 1 Counts of *Orobanche crenata* sampled over 300 sample units in a 15 x 20, row x column rectangular grid. Axes show distances in metres.

Fig. 2 Initial-and-final plot for the data in Fig . 1 showing notional flows originating from donor units with above-average counts, and terminating at receiver units, with below-average counts, denoted by asterisks. Note how large average flow distances are associated with the patch of relatively large counts centred on (14.0, 34.6).

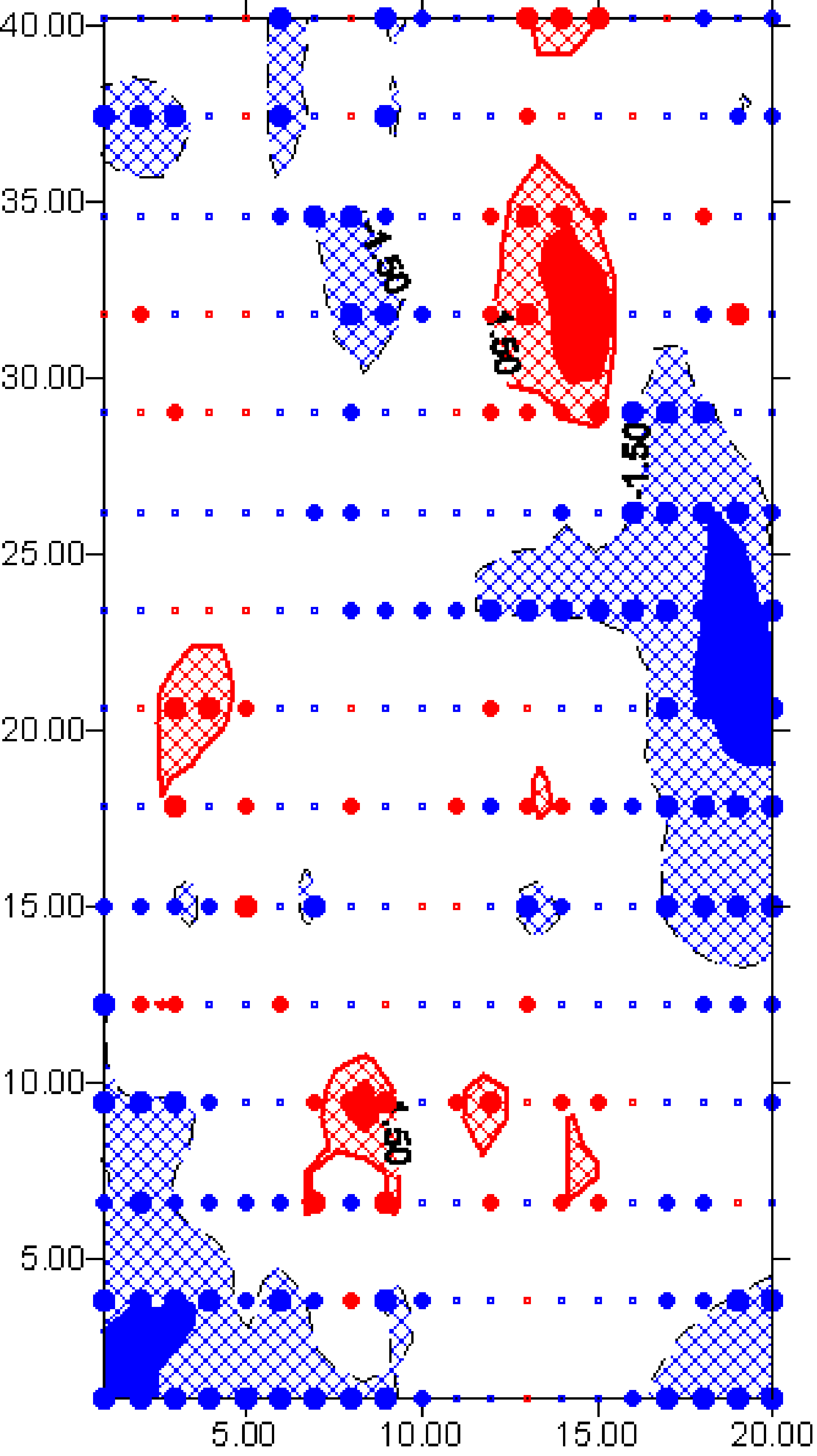
Fig. 3 Map of clustering in the data of Fig. 1. Values of clustering indices $\nu_i<1$, $1<\nu_i<1.5$ and $1.5<\nu_i$, shown as small, medium and large filled circles, respectively; values of $|\nu_j|$ are categorized similarly as open circles. Bold lines: contours enclosing patches (darker shaded areas) of $1.5<\nu_i$, or gaps (lighter shaded areas) of $\nu_j<-1.5$. Dashed contour represents zero clustering, i.e. the boundary between patches and gaps, along which estimated density is the mean, 1.16, for these data.

Fig. 4 Example of EDF plot of ranked Y_i plotted against rank, for each outflow sample unit. Observed values: **x**; mean value of Y_i from 6000 randomizations: red **Δ**; envelope of upper 95th centiles: red **∇**. Observed values easily exceed means from randomizations; they approach closely but never exceed envelope. The EDF for inflow units is shown next, with red symbols replaced by blue, but now the means are blue red **∇** and the lower 5th centiles are blue **Δ**; the envelope is exceeded three times.

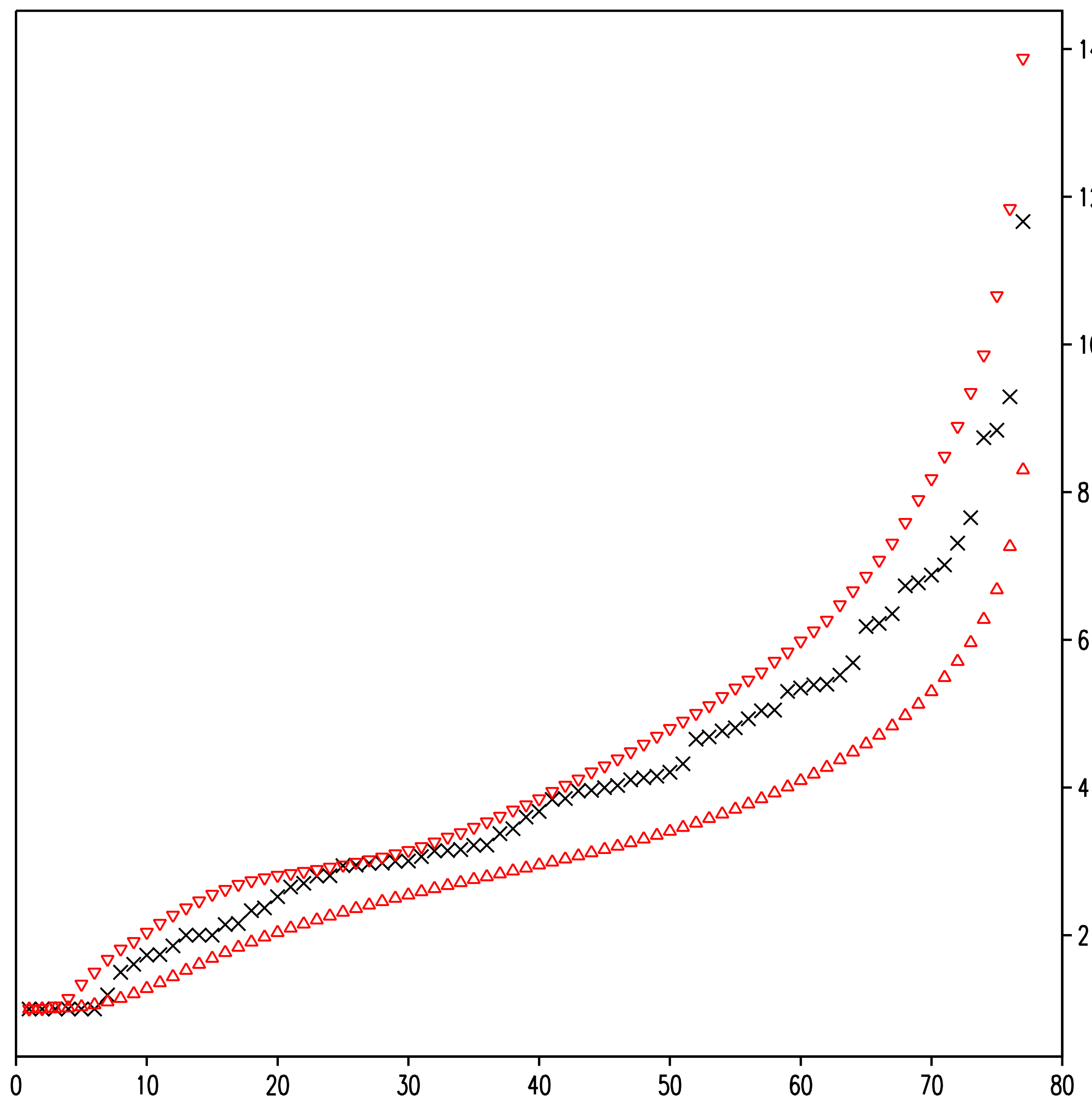

1 0 2 0 3 0 0 2 1 0 1 1 2 6 3 1 2 1 0 0
0 1 0 0 2 0 0 2 1 0 0 1 2 5 0 2 0 1 0 0
0 0 0 1 0 0 0 1 1 0 1 3 2 6 6 0 1 2 1 1
2 4 0 8 2 0 0 1 1 0 1 3 2 5 2 1 0 1 5 1
0 3 2 3 2 0 0 0 0 1 3 4 2 5 4 1 1 1 1 0
0 1 1 1 1 0 0 0 1 0 0 1 0 0 0 0 1 0 0 0
0 0 4 7 2 1 0 0 1 0 0 1 1 1 0 0 0 0 0 0
0 3 3 6 2 1 0 7 1 1 1 8 4 1 0 0 0 0 0 0
0 1 2 1 2 1 1 2 1 0 2 1 2 2 0 0 0 0 0 0
0 1 0 0 3 0 0 1 0 2 2 0 0 0 0 0 0 0 0 0
0 3 2 1 1 2 1 0 2 0 0 1 2 0 0 0 0 0 0 0
0 0 0 0 0 1 7 4 3 1 5 2 6 7 7 7 0 0 1 0
0 0 0 0 1 1 4 1 4 0 0 3 0 3 7 0 1 0 2 0
0 0 1 0 1 0 1 2 0 1 0 0 8 1 0 0 1 1 1 0
0 0 0 0 0 0 0 0 0 0 1 0 4 1 0 0 0 0 0 0
40
35
30
25
20
15
10
5
0
−10
−5
0
5
10
15
20
25
30

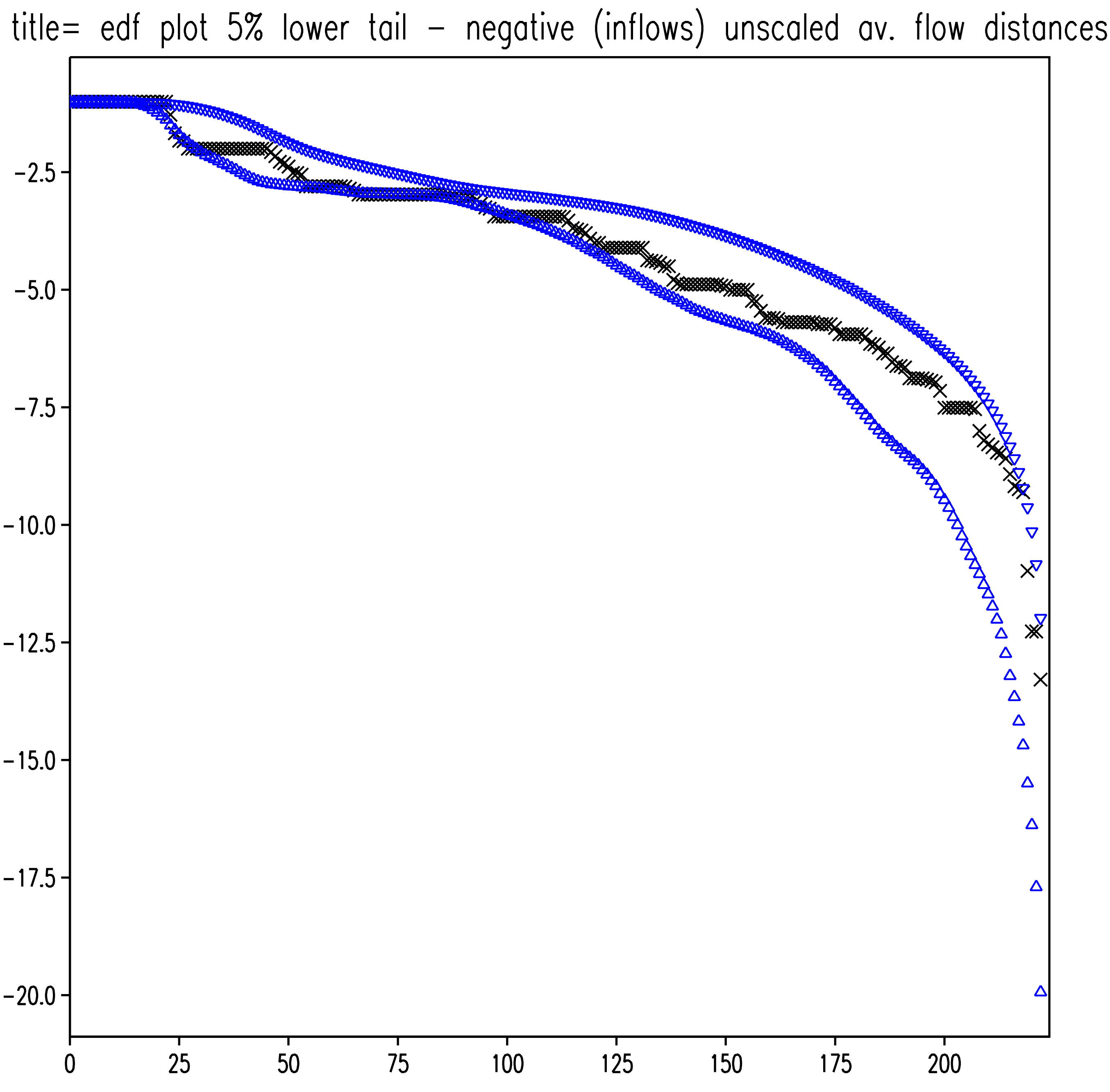


40
35
30
25
20
15
10
5
0
−10
−5
0
5
10
15
20
25
30



40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
5.00
10.00
15.00
20.00
1.50
1.50
-1.50
1.50


0
10
20
30
40
50
60
70
80
2
4
6
8
10
12
14


title= edf plot 5% lower tail - negative (inflows) unscaled av. flow distances
-2.5
-5.0
-7.5
-10.0
-12.5
-15.0
-17.5
-20.0
0
25
50
75
100
125
150
175
200