33 A Cappadocian Tablet Problem

A.G. Constantine, J.C. Gower & B. Zielman

Summary
We have three objectives. Firstly to present, in an accessible form, previously published data on 4000–year old cuneiform–inscribed tablets excavated in an Assyrian merchant colony in Kanesh, Cappadocia. Many of the tablets refer to several sites, and it is assumed that this is evidence of their proximity and hence might be used in the construction of a map of the area. Using several simple models for (a) constructing proximity information similar to that found on the tablets and (b) deriving distances from assumed proximities, our second aim is to examine, by means of simulation, the success of map reconstruction of a known region — modern France. Using the information so gained, our third objective is to attempt to construct a map of the region around ancient Kanesh, and to comment on its validity.

33.1 INTRODUCTION

This study is concerned with information contained on cuneiform tablets excavated in Kanesh in Cappadocia. These clay tablets date from circa 1940–1740 B.C., pre–Hittite times in Bronze Age Anatolia. The economic background to this material is given by Larsen (1976). It seems that a network of caravan sites had long been established throughout the Middle East. The material discussed here concerns the part of this network, based on Assur, and, specifically, on that small part of Assur's trade that was with Kanesh. The Old Assyrians had this trade astonishingly well–organised and in its support had established many merchant colonies; Kanesh, whose trade was mainly in tin and textiles in exchange for silver, was one of these. The colonies were in constant correspondence with Assur, and what we have is part of the archive of the establishment in Kanesh. This correspondence has thrown enormous light on many aspects of the organisation of the trade; here we are concerned with the possibility of using the material to draw a map of the area in the vicinity of Kanesh. This is a possibility because the tablets list the names of local sites with which it traded, the exact locations of most of which are unknown. The area is in the mountainous region beyond the upper Euphrates, and it is relevant that this trade made little, if any, use of rivers, because river–distances give a very distorted view of geographical distance. It is known that the caravan trade used donkeys — it is even known how much each donkey carried and how it was loaded.

Tobler and Wineburg (1971) constructed a hypothetical map of the sites using the main corpus of data, published by Bilgic (1951). Bilgic gives for each of 118/9 site–names, which of 612 tablets referred to it. Even in 1977, it seems that about 17,000 tablets had been excavated, and the number has certainly increased since then. Of course, many of these tablets will not refer to any site. Unfortunately much of this material remains to be published. Table 33.1 lists the names of the sites as given by Bilgic and also gives, for each site, the number of tablets on which its name occurs with no other name and the number with at least one other name. Sites 4 and 6 are synonyms, so we immediately reduce the total number to 118 different sites. Bilgic was primarily interested in etymological matters and hence in the site–names themselves. His form of the data is not convenient for identifying whether or not two site–names ever occur together on any tablet, and a table has to be constructed, giving for each tablet the names of the sites referred to; this amounts to reconstructing information given on each of the original tablets. Although Tobler and Wineburg must have made such a table, they did not publish it. Because of the importance of these data, the reconstructed table is given in Table 33.2, where name–numbers refer to the site names given in Table 33.1. There is, however, a

1.	*Abarn(a)*	11	14
2.	Abum	1	4
3.	*Akkad*	14	9
4.	Akku(u)a	7	1
5.	Alsana	-	1
6.	Amku(u)a - See Akku(u)a		
7.	Amalia	1	-
8.	*Amurum*	4	2
9.	Apitipan	1	2
10.	Aprum	-	4
11.	Asataruua	1	-
12.	Asihum	-	1
13.	*Assur*	6	-
14.	Atubazum	-	1
15.	Azamrum	2	-
16.	Eluhut	-	3
17.	Hahhum	30	14
18.	Haka	2	2
19.	Hamiz/s/sanum	2	-
20.	Hanaknak	4	-
21.	Hapura	-	1
22.	Hapurata	2	1
23.	Harana	-	2
24.	Harsiuna	-	1
25.	Harsiua	1	-
26.	Harsamna	1	-
27.	Hartu	1	1
28.	Hattus	9	4
29.	Hurama	9	8
30.	Huturut	-	1
31.	Ibla	-	1
32.	Kaluzanum	-	1
33.	Kammalia	-	1
34.	Kanis	19	14
35.	Kapitra	-	1
36.	Kapra	-	1
37.	Karparta	2	-
38.	Garatum	1	-
39.	Quat(a)ra	3	6
40.	*Katila*	1	1
41.	*Kasur*	2	-
42.	*Kilar*	1	-
43.	Kunanami(at)	4	3
44.	*Kupan*	-	1
45.	Kupurnat	1	2
46.	Kussura	3	2
47.	Lishu	1	-
48.	Lipurna	1	-
49.	Lualkua	-	1
50.	Luara	-	1
51.	Luha	1	2
52.	Luhuzzatia	14	2
53.	Mal'a	2	-
54.	Mallita	-	2
55.	Ma'ama	10	7
56.	Mata	-	1
57.	Mardama	1	-
58.	Nahur	-	1
59.	Nanuua	-	1
60.	Nesa	1	-
61.	Nihria	4	3
62.	Ninassa	1	4
63.	Pahatima	-	1
64.	Palissa	1	-
65.	Badna	4	2
66.	Pinarama	1	-
67.	Puhitar	2	-
68.	Purallum	1	1
69.	*Pura*	4	1
70.	Purushattum	71	29
71.	Purusna	1	-
72.	Puruttum	-	3
73.	Razama	-	4
74.	Sakaria	1	-
75.	Sasu	-	1
76.	Sauit	-	1
77.	*Sitluna*	3	2
78.	Subar	2	-
79.	Supana	1	1
80.	Salanshu(u)a	1	4
81.	Salatu/i(u)ar	15	10
82.	Samuha	-	2
83.	Sana	2	-
84.	Simala	1	2
85.	Sinahutum	4	-
86.	Siris	1	-
87.	Sirmuin	1	1
88.	*Sulupka*	3	9
89.	Suppilulia	1	-
90.	*Surpu*	1	1
91.	*Tadmur*	3	-
92.	*Takkusta*	-	2
93.	Talhat	12	1
94.	Tarakum	1	3
95.	Taskuria	1	-
96.	Tatardu	-	1
97.	Tatania	2	3
98.	Tauinia	2	1
99.	Tegarama	3	2
100.	Tikurna	-	1
101.	Tilimra	-	2
102.	Timelkia	17	5
103.	Tismurna	4	5
104.	Tuhpia	2	2
105.	Tukupta	1	-
106.	Turhumit	26	15
107.	Udum	2	-
108.	Ulama	9	5
109.	Unipkum	-	1
110.	Ursu	2	4
111.	Usuhinum	-	1
112.	Usa	3	2
113.	Wahsusana	50	38
114.	Was/ushania	9	9
115.	*Walhikina*	-	1
116.	*Wanisana*	1	-
117.	*Wilusna*	-	1
118.	Zalpa	22	9
119.	[...]hama	1	-

Table 33.1: For each site, the third column gives the number of tablets that refer to that site alone. The fourth column gives the number of tablets that associate the named site with at least one other site. Names given in italics are possible pseudo–sites that refer to a product named after its place of origin rather than to a nearby geographical location.

complication, because recent research suggests that some names that were originally identified as sites are more likely to name textile products, in the same way that Jersey may refer to the island or to the material, and where the place of manufacture is distant from Kanesh. We refer to these ambiguous sites as pseudo–sites, indicated by italic letters in Table 33.1; excluding pseudo–sites, 101 regular sites remain. Consequently when analysing these data we have compared the results obtained when using all 118 sites and when using only the 101 regular sites.

Only those sites of Table 33.1 that record joint occurrences give information on linkage and possible proximity. Consequently, 33 sites occurring on 55 tablets which refer only to a single site were eliminated from the analysis. This left 85 sites, including 12 pseudo–sites, with joint occurrences on at least one of 457 tablets. For these 85 sites, Table 33.3 gives those that occur together on at least one tablet. This is the so–called table of incidences and it gives the primary information on the linkage of sites. Just because pairs of sites are linked does not imply that there is necessarily a chain that links all 85 sites. Indeed, five of the sites (numbers 21,59,115 and 33,75) form two small disconnected chains that provide little information, leaving only 80 fully linked sites to be mapped. If, further, the 12 pseudo–sites are eliminated, then some of the remaining 73 sites form small independent chains, comprising a set of fragments of size nine (numbers 21,24,31,33,35, 49,56,59,75) which we ignore, leaving a corpus of 64 sites for mapping. The reduction from 80 to 64 reflects losses not only from the direct elimination of pseudo–sites but also both from the chain–fragments and from linkages of main sites through pseudo–sites. Thus our maps, reported in section 4, concern two groups of sites, one of size 80 and a subset of size 64 which excludes pseudo–sites and their side–effects.

Tobler and Wineburg (1971) assumed that the "interaction" between two sites is proportional to their populations, and inversely proportional to the squares of their distances apart. This is the so–called "gravity" model of interaction, expressed in the formula

$$I_{ij} = k\frac{P_iP_j}{d_{ij}^2} \qquad [1]$$

k being a constant of proportionality, I_{ij} being some measure of the interaction between sites i and j, P_i and P_j their populations, and d_{ij} their distance apart. This model, and variants of it, are

Reference		Name-number
CCT	26b	94,111
	27a	10,12
	29	23,63,108,117,118
	37b	1,55
	38a	1,69,118
	42a	14,32,39,73
	42b	44,94
	44	70,100
CCT II	22	8,17,61
	23	43,76
	31	113,114 (TC III 10)
	35	113,114 (TC III 10)
	49a	8,17,61 (CCT II 22)
CCT III	1	103,106
	9	1,110
	31	3,28
	34b	17,118
	36a	72,84
	44b	72,99
	45b	1,88
CCT IV	3a	70,96,113
	4a	27,70,113
	5a	1,88 (CCT III 45b)
	16c	113,114 (TC III 10)
	18a	17,102 (BIN VI 114)
	22b	70,113 (TC II 36)
	27a	103,106 (CCT III 1)
	28b	17,55 (TC 18)
	29a	10,70
	44b	55,98
	48a	70,106 (TC II 36)
Mat	11b	21,59,115
Mat II	23a	106,108
	30a	39,70
VAT	6209	82,114
Jena	281	29,80
	440	81,113 (BIN IV 70)
Liv	3	106,113 (TC II 36)
PSBA VI		2,10,58
Hr	1	28,29,46,52
	31b	4,28
Ark.Derg. IV	4	113,114 (TC III 10)
KTB	15	17,103
	16	88,92 (TC 47)
	S24	22,103

Reference		Name-number
AHDO	1	33,75
Unv.Ank	17	43,106
KTS	36c	1,17
	35a	3.24
	7a	34,62
	19b	34,70 (KTP 4)
	20	34,70 (KTP 4)
	55a	34,81,88
	33a	34,113 (KTP 4)
	19a	43,113
	12	61,80
TTC	14	1,70
	9	5,84
EL	247	3,34
	325a	3,56
	162	16,34
	332	17,70
	243	17,29
	252	29,51
	197	34,113 (KTP 4)
	238	34,113 (KTP 4)
	253	34,113 (KTP 4)
	168	70,112
	336	70,113 (KTP 4)
KTP	4	34,70,104,113
	10	70,81,113
OIP XXVII	62	45,93
	54	54,113,114
	5	80,118
KTHahn	1	55,70,113
	14	23,70,113
	18	39,88
BIN IV	4	1,3
	6	52,81,113
	7	17,87
	23	88,113
	31	108,113 (TC III 271)
	35	70,81
	36	70,106,113 (TC II 36)
	43	70,113 (TC II 36)
	45	30,112,113
	48	29,70 (TC 9)
	64	70,106 (TC II 36)
	65	1,3 (BIN IV 4)
	70	81,106,113

Reference		Name-number
	124	2,73 (TC III 163)
	148	81,88,110
	193	9,39,73 (TC III 163)
	201	1,55 (CCT 37b)
	219	50,55,109,110
	233	81,104
BIN VI	9	40,114
	38	17,133 (BIN VI 114)
	114	17,102,113
	167	70,118
	180	97,118 (TC III 166)
	193	31,35,77
	265	65,118
	269	97,113
TC	9	29,70
	10	29,46,82,101
	18	17,34,55,110
	32	81,113 (KTP 10)
	47	88,92
	53	54,113 (OIP XXVII 54)
	72	1,3,62,113
	81	16,17,39,72,80,102,106
TC II	7	3,90
	23	70,106 (TC II 36)
	26	81,106 (BIN IV 70)
	27	28,29,45,102
	36	70,106,113
	56	1,103
	57	2,10,51
TC III	3	70,113 (KTP 4)
	10	62,113,1114
	60	99,114
	61	3,77
	112	18,36
	143	106,113 (TC II 36)
	149	70,106 (TC II 36)
	158	1,101
	163	2,9,39,73,94
	164	16,18,68,118
	165	62,70,108,114
	166	34,97,118
	169	49,88
	211	1,65,102
	270	79,113
	271	34,108,113

Table 33.2: Tablets with more than one placename. For a key to the code to published texts see Part 2 of the list of references. In the table, the first column gives the reference text and volume number; the second column gives the tablet identification numbers within the publications. Name–numbers refer to the numbers and corresponding names given in Table 33.1. Information given in parenthesis cites a tablet which contains the same linkages in a longer list of names, making the smaller tablet redundant.

popular in demographic studies of population movements (see e.g. Gatrell 1981). If one estimates I_{ij} by n_{ij}, the number of tablets on which sites i and j occur jointly, and if P_i is assumed to be proportional to the total number of occurrences, N_i, of the ith site, inverting the above equation gives estimates of "distances"

[2] $$d_{ij}^2 = k\frac{N_i N_j}{n_{ij}}$$

1.	*Abarn(a)*	3,17,55,62,65,*69*,70,*88*,101,102,103,110,113,118
2.	Abum	9,10,39,51,58,73,94
3.	*Akkad*	24,28,34,56,62,*77*,*90*,113
4.	Akku(u)a	28
5.	Alsana	84
8.	*Amurum*	17,61
9.	Apitipan	39,73,94
10.	Aprum	12,51,58,70
14.	Atubazum	32,39,73
16.	Eluhut	17,18,34,39,68,72,80,87,102,106,110,118
17.	Hahhum	29,34,39,55,61,70,72,80,87,102,103,106,110,113,118
18.	Haka	36,68,118
21.	Hapura	59,*115*
22.	Hapurata	103
23.	Harana	63,70,108,113,*117*,118
27.	Hartu	70,113
28.	Hattus	29,45,46,52,102
29.	Hurama	45,46,51,52,70,80,82,101,102
30.	Huturut	112,113
31.	Ibla	35,*77*
32.	Kaluzanum	39,73
33.	Kammalia	75
34.	Kanis	55,62,70,81,*88*,97,104,108,110,113,118
35.	Kapitra	*77*
39.	Quatara	70,72,73,80,*88*,94,102,106
40.	*Katila*	114
43.	Kunanami(at)	76,106,113
44.	*Kupan*	94
45.	Kupurnat	93,102
46.	Kussura	52,82,101
49.	Lualkua	*88*
50.	Luara	55,109,110
52.	Luhuzzatia	81,113
54.	Mallita	113,114
55.	Ma'ama	*70*,98,109,110,113
59.	Nanuua	*115*
61.	Nihria	80
62.	Ninassa	70,108,113,114
63.	Pahatima	108,*117*,118
65.	Badna	102,118
68.	Purallum	118
69.	*Pura*	118
70.	Purushattum	81,96,100,104,106,108,112,113,114,118
72.	Puruttum	80,84,99,102,106
73.	Razama	94
79.	Supana	113
80.	Salanshu(u)a	102,106,118
81.	Salatu/i(u)ar	*88*,104,106,110,113
82.	Samuha	101,114
88.	*Sulupka*	*92*,110,113
94.	Tarakum	111
96.	Tatardu	113
97.	Tatania	113,118
99.	Tegarama	114
102.	Timelkia	106,113
103.	Tismurna	106
104.	Tuhpia	113
106.	Turhumit	108,113
108.	Ulama	113,114,*117*,118
109.	Unipkum	110
112.	Usa	113
113.	Wahsusana	114
117.	*Wilusna*	118

Table 33.3: Incidence matrix. This table gives those of 85 sites which occur at least once in association with the named site or pseudo–site. To avoid duplication only the first occurrence of each pair is recorded. Thus the first pair given is (1,3), representing at least one joint occurrence of Abarna with Akkad; the pair (3,1) is omitted because it contains the same information.

Tablets mentioning only one site–name can contribute to N_i but not to n_{ij}. Note that when all the populations are equal, the gravity model effectively reduces to considering only information contained in the inverted frequencies.

To convert the distances, into co–ordinates, it is natural to use one of the so–called non–metric multidimensional scaling techniques (see e.g. Kruskal & Wish 1978) which attempt to preserve only the *rank orders* of the distances rather than their absolute values. Reconstructed maps are therefore invariant to monotonic transformations or distortions of the distances, and the geometric constraints (the triangle inequality, for example) override minor "errors" in the data due to random sampling effects. Thus the basic assumption is not that (2) gives absolute estimates of distances but that, at least approximately, it gives estimates that preserve the correct order of the distances. It is a remarkable fact that having ordinal information of this kind is very nearly as good as having the actual numerical distances. Tobler and Wineburg applied one such scaling program, due to Lingoes, to distances between the Cappadocian sites estimated by (2), and constructed a two–dimensional configuration of the sites. Such a configuration is determined up to an arbitrary scale factor, an arbitrary rotation in the plane and perhaps a reflection about an arbitrary axis. Therefore, to determine scale and orientation, the true geographical positions of at least three sites should be known. For data subject to statistical uncertainty many more known sites would be preferable, so that some sort of least squares fit of the computed configuration to the known configuration could be determined. In the Cappadocian situation, only two locations were known with reasonable certainty, together with speculations as to the positions of additional sites based on other evidence such as itineraries.

One method of establishing what confidence should be attributed to maps constructed similarly to that of Tobler and Wineburg, would be to simulate the Cappadocian set-up, but using a known configuration of towns to compare with the computed configurations. Such a study would allow for the empirical investigation of the effects of the assumptions underlying the gravity model (1), the probabilistic effects of random sampling, and also the effect of estimating distances from the data, using models other than the gravity model. In section 3 and following, we report on a simulation study using the departmental capitals (and, where appropriate, department populations) of France to examine the effects of the above–mentioned factors on maps produced by multidimensional scaling methods.

33.2 THE CONSTRUCTION OF MAPS FROM CONTINGUITY DATA

As pointed out in the introduction, the key assumption of Tobler and Wineburg was that the joint occurrence of two names on a tablet reflected geographical proximity of the two sites, and the more joint occurrences, the closer the sites were likely to be. Further, they assumed that a non–metric multidimensional scaling program can be used to construct a map based on the notion that the rank orders of the distances between estimated sites match, as closely as possible, the rank orders of the computed measures of nearness derived from formula (2).

Scaling techniques can be applied to measures of nearness or similarity, for purposes other than the construction of geographical maps. Indeed, the non–metric (or rank order) methods originated in psychology, where it is often assumed that there is an underlying concept of "psychological distance" between, for example, sensory stimuli, and it may be required to determine and map these "distances" from measures of similarity of response to two stimuli. One of the first examples of such an application was Shepard's (1962) scaling of the Rothkopf morse–code data which gave a map, in which nearness was a measure of the confusion between the 36 (26 letters and 10 digits) morse code signals, when each pair was presented in sequence and subjects were asked whether the signals differed. An early application, in archaeology, was Kendall's (1971a) seriation of the Münsingen graves, in which it was desired to produce a one–dimensional configuration (or curve) such that order along the curve reflected the temporal sequence of the graves; here nearness was a measure of the similarity of the ornamentation of pairs of grave objects — especially bronze pins. Both of these applications were highly successful, Shepard's in giving a meaningful interpretation of the reasons for confusion between morse–code signals, and Kendall's in giving a seriation which agreed largely with the archaeological evidence as interpreted by other methods (Hodson 1968).

Three applications by D.G. Kendall of the construction of maps from continguity and similarity data illustrate the success of the methods and introduce some concepts and techniques needed in later sections of the present paper. Kendall (1970, 1971b), obtained an excellent map of Romania by using recommended road–route distances as measures of proximity. Roads between towns rarely follow a straight line, but the ordinal information was sufficiently maintained, so that a good map resulted.

A second example, Kendall (1971b), was concerned with eight parishes in the Otmoor region of Oxfordshire. Here, the data consisted of the numbers of marriages between and within the parishes for the period circa 1600–1850 AD and the number of marriages between members of the parishes and the outside world. At this time, before the coming of the railways, this was an isolated region, whose main external contact was at the Michaelmas and Spring hiring fairs in the local market towns of Abingdon and Bicester. It was assumed that inter–parish marriages would be more common for adjacent parishes than for parishes more widely separated. Kendall defined an appropriate standardized inter–marriage rate between the parishes, and reconstructed a satisfactory map of the Otmoor region using a non–metric scaling program. Complete details may be found in the reference cited.

In a third example, Kendall (1971c), a map of France was reconstructed from an "incidence matrix" which merely listed whether or not departments had a boundary in common. More specifically, the data matrix was an 88×88 matrix, the rows and columns corresponding to the 88 departments of France (omitting Paris and Corsica), and the (i,j)th element was 1 if departments i and j had a common boundary, 0 otherwise. From this, a "distance" matrix was defined with elements 1+ the smallest number of intermediate departments required to make a connected chain between departments i and j. Non–metric multidimensional scaling of this distance matrix gave an excellent map. Indeed, Kendall not only positioned the departmental capitals but also constructed departmental boundaries; this is a refinement that we shall not need.

The second and third examples are particularly relevant to the present investigation. If a marriage between two partners from different parishes is regarded as a "tablet" containing two site–names, and the total number of marriages involving people from a parish is regarded as being proportional to the parish population, then we have a direct analogy with the Cappadocian reconstruction. Although Kendall did not use the gravity model as such, there are some similarities between it and his model for inter–marriage rates. The third example is interesting in that it suggests that contiguities rather than the gravity model may be used for estimating distances. The distance defined as the smallest number of boundaries to be crossed when travelling between two departments is termed a *K*–Metric and is easily computed, as described at the end of section 3, from the table of incidences (i.e. Table 33.3 for the Cappadocian tablets). The *K*–metric is easily seen to satisfy the triangle inequality. This means that it is a true *metric*, and this is obviously a very desirable property when reproducing configurations in Euclidean space. The step from a *K*–metric, defined in terms of crossing boundaries, to a *K*–metric, defined in terms of the smallest number of tablets needed to link two site–names, is a small one.

The incidence matrix of Table 33.3 is easily derived from Table 33.2; it is equally easy to derive the 118×118 matrix giving the frequencies n_{ij} of joint occurrences of sites on the tablets. To save space we do not reproduce the table of frequencies but, as we have seen, they are used in the gravity model. Assuming that joint occurrence of two names on a tablet reflects geographical proximity, the elements of the frequency table might be taken to give a direct measure of "nearness" of the sites, the larger the value, the nearer are the two sites. In both tables of incidence and frequency, zero values have to be handled carefully. This is because a zero value corresponds to an indefinitely large distance. There is also a problem with sites that are connected only by a single link, so that site i is joined only to site j, although j itself may be joined to several other sites. Then, the position of site i is indeterminate and cannot be positioned in a spatial context without using additional information. Frequencies, incidences and the *K*–metric all implicitly handle this indeterminism by placing j remote from the other sites linked to i. This treatment is not unreasonable, because when site i has a single link to site j which itself has many links to other sites, the likelihood is that site i is even further from the other sites than is j; otherwise, one would expect to find other links with i. However, this need not be the correct treatment, as a poorly linked site may merely be an unimportant one that is rarely mentioned but nevertheless is close to tightly linked important sites.

33.3 MODELS AND METHODS

In attempting computer simulations of the Cappadocian tablet situation, one might consider the different sorts of model that underlie the chances of joint occurrences of site–names on "tablets". For example, if a given tablet is assumed to be a record of, say, merchant transactions between the sites mentioned, is the site of origin of the tablet included in the list of names or not? Again, if transactions are recorded between a series of sites, what is the effect on the computer maps when some intermediate sites are by–passed, so that their names do not occur despite their proximity?

Further, it is of interest to examine the effects of different methods of estimating the distances. The effect of population sizes on the distances as calculated by Tobler and Wineburg from the gravity model could be crucial. The *K*–metric used by Kendall in reconstructing the map of France does not require population sizes and therefore is of quite a different nature to the gravity model measure; its use is worth examining in a tablet simulation study. Thus, although non–metric methods are insensitive to small changes that have little effect on the *ordinal* knowledge of the distances, it has to be recognised that some models might have major effects on the estimated ordering, and this needs investigation.

These considerations have influenced the planning of our simulation studies, although other commitments have made us unable to examine everything that we would like to have examined. Firstly we simulated the formation of tablets, using two main models. Then, using the simulated tablet data, we studied four methods for determining distance or association between the towns, from which maps of France were constructed by multidimensional scaling. Only one method of non–metric multidimensional scaling was used, as all the evidence suggests that reconstructed maps are unlikely to be much influenced by this aspect. The remainder of this section gives detailed information on the methods used for tablet–construction and for distance–estimation.

33.3.1 Simulation of tablet formation

In our simulation study, we took the 87 departments of France (omitting the whole of the Paris region, because of its dominant population, as well as Corsica, because of its isolation), together with their populations as given in the 1983 edition of Baedecker's *Almanac*. Each department was represented by its capital town and "tablets" were constructed by the following two methods.

(i) Circles of varying radii (120 km, 160 km, 200 km, 240 km, 300 km, 380 km) were drawn on a map of France with the towns (always departmental capitals) as centres, and all department capitals falling within the circles were noted. For example, a circle centred on Bordeaux might include, depending on its radius, the towns of Angouleme, La Rochelle and Landes. These towns, together with Bordeaux itself, would then be regarded as those comprising a list on a single tablet. The different radii model the region of influence of each capital. Centres were sampled randomly with replacement according to the size of the departmental population, so that, on average, Lille would be sampled more often than most other towns, since it has a large population. Samples of sizes 250 and 500 were taken in each case and the tablets were constructed both including and excluding the centre towns. Circles of too small radii (less than about 80 km) gave disconnected maps and these were rejected. This model allows chance to select a subset of all possible capitals and then permits the effect of different regions of influence to be investigated, but includes all towns within each specified radius. Note that if the radius is made too big, all towns will be included on all tablets, again giving no information on linkage; for France this occurs at around 750 km. This suggests that there is an optimal radius that must depend in some way on the density of settlement.

(ii) Centre towns were sampled with a frequency proportional to their departmental population size, and for each centre town, other towns were sampled with probabilities according to the gravity model (proportional to population, inversely proportional to the square of the distance), tablet size being determined as a binomially distributed variate with $p = 0.04$. In other words, a central town would be sampled (according to population) and then a binomial distribution with $p = 0.04$ sampled. Since there are 87 towns this gives tablets averaging about 3.5 (including repetitions) towns, comparable to the Cappadocian tablets. Once the tablet size, k say, has been determined, the tablet was completed by sampling a further k–1 towns as follows. For a given central town, say i, let $p_{i1},\ldots,p_{i,87}$ denote the probabilities that the other town will be sampled on the tablet according to the gravity model, i.e.

$$p_{ij} = K\frac{P_i P_j}{d_{ij}^2},$$

P_j being the known population of town j, and d_{ij} the known geographical distance between i and j. Assume the probabilities p_{ij} are ordered in increasing order of magnitude, $p_{i1} < p_{i2} < \ldots <_{pi87}$, where K is chosen to ensure that

$$\sum_{j=1}^{87} p_{ij} = 1$$

A random variate, p, uniformly distributed on (0,1) was then sampled and if

$$\sum_{j=0}^{t-1} p_{ij} < p < \sum_{j=1}^{t} p_{ij}$$

town t was added to the tablet. This was done k–1 times for each central town, and repetitions, if any, deleted. This model allows the bigger departments to be be sampled more often than the smaller and uses the gravity model to determine which towns are included on the tablets; tablet size is left to chance but with a predetermined average size.

33.3.2 Determination of distance

Once the "tablets" had been constructed, "distance" or association matrices were constructed and a version of the SMACOF program (De Leeuw & Heiser 1980) used to compute two–dimensional configurations or hypothetical maps. Four different measures of distance, C, D, D^* and K, were used and defined as follows. Denote by $N = (n_{ij})$ the *frequency matrix*, giving the number of tablets, n_{ij}, on which sites i and j occur together. Denote by $\mathbf{A} = (a_{ij})$ the *incidence matrix*, where $a_{ij} = 1$ if sites i and j ever occur together on any tablet and = 0 otherwise; the diagonal values of **A** were taken to be unity. Then

1) $C = (d_{ij})$ was defined directly in terms of the incidences, as $d_{ij} = 1 - a_{ij}$;
2) $D = (d_{ij})$ was defined according to the Gravity Model by

$$d_{ij} = \sqrt{\frac{n_i n_j}{n_{ij}}}, \text{ when } n_{ij} \neq 0,$$

where the n_i are the row totals of N (zero values of n_{ij} are replaced by $\frac{1}{2}$);

3) $D^* = (d_{ij}^*)$ was defined by the Inverse Frequency

$$d_{ij}^* = \frac{1}{n_{ij}}, \text{ when } n_{ij} \neq 0,$$

which is a distance merely using the "nearness" measure n_{ij} and ignoring the estimated populations which were assumed proportional to the n_i in (b) above (zero values of n_{ij} are replaced by $\frac{1}{2}$);

4) the "K–metric", $K = (k_{ij})$ where k_{ij} is equal to the smallest number of tablets required to make a connected chain between sites i and j. The k_{ij} are computed as follows. First replace the zeros by units in the diagonal of the incidence matrix $\mathbf{A}$ defined above, and put $k_{ij} = 1$ if $a_{ij} = 1$. Then compute $\mathbf{A}^2$ and denote by $\mathbf{A}^{(2)}$ the matrix with $a_{ij}^{(2)} = 1$ if the (i,j)th element of $\mathbf{A}^2$ is ≥ 1, $a_{ij}^{(2)} = 0$ if the (i,j)th of $\mathbf{A}^2$ is 0. Then put $k_{ij} = 2$ unless k_{ij} were 1 at the first step when its value remains 1. Continuing in this way compute $\mathbf{AA}^{(2)}$ and define $\mathbf{A}^{(3)}$ to be that matrix with unities in those positions where the elements of $\mathbf{AA}^{(2)}$ are non–zero, and zero otherwise. Then $k_{ij} = 3$ unless k_{ij} was defined to be either 1 or 2 previously. This process is continued until either all elements of $\mathbf{A}^{(p)}$ are non–zero, or $\mathbf{A}^{(p)} = \mathbf{A}^{(p+1)}$. Put simply, the method is equivalent to successively powering–up the incidence matrix $\mathbf{A}$ and noting which elements become non–zero for the first time at each power $\mathbf{A}^m$. When $k_{ij} = m$ such towns cannot be connected by fewer than m tablets. This distance is therefore equivalent to that defined by Kendall in his reconstruction of the map of France, discussed above in section 2. The difference is that whereas Kendall knew for a certainty whether or not two departments had a common boundary, the joint occurrence of two sites on a tablet is no guarantee of proximity and is subject to sampling fluctuations at the time the tablet was made and to the chance of its survival and recovery in modern times.

33.4 RESULTS

How well the distances of a fitted configuration correspond to the given distances is expressed in a badness of fit measure. The most commonly used badness of fit measure is the Stress, which varies between zero and unity. All Stress values in our study were in the range 0.03 — 0.05, indicating good fits. However, low Stress indicates only that a good fit has been found in two dimensions but says nothing about the resemblance of our map to the real map of France. Recovery of the map of France can be assessed by visual inspection or by a formal Procrustes statistic. Procrustes analysis is a least–squares procedure for rotating a given configuration to a target configuration in such a way that, on average, their corresponding points match as closely as possible. The Procrustes statistic is the sum–of–squares of the distances between all pairs of matching points. The minimum of a Procrustes statistic is zero, which indicates the complete agreement of the two maps. In our study, the maximum of a Procrustes statistic is scaled to one, which indicates the poorest possible match between the two maps. Gower (1984) includes an introduction to Procrustes analysis, giving further information and references.

	Radius(Km)	Gravity	Inverted	Incidence	*K-metric*
	380	0.009	0.022	0.058	0.058
	300	0.016	0.017	0.039	0.036
N=250	240	0.035	0.060	0.020	0.028
Tablets	200	0.156	0.080	0.025	0.034
	160	0.460	0.410	0.032	0.032
	120	0.871	0.750	0.231	0.030
	380	0.011	0.012	0.049	0.051
	300	0.011	0.016	0.032	0.036
N=500	240	0.025	0.028	0.025	0.034
Tablets	200	0.136	0.110	0.027	0.030
	160	0.493	0.380	0.041	0.028
	120	0.840	0.720	0.235	0.021

Table 33.4: Procrustes statistics obtained by comparing simulated maps of France with the true map for 250 and 500 "tablets", six radii and four methods of estimating distance.

The results for reconstructing maps of France are shown in Table 33.4. These results refer to the model discussed under (i) of section 3. It made little difference whether or not the centre town was included, so we report only the results excluding the centre town. It is remarkable that a non–metric multidimensional scaling of the crude incidence, C, of joint occurrence usually gives acceptable maps. For the gravity model, D, and the inverted frequencies, D^*, the Procrustes fit is unacceptably poor (high values) with discs of a small (120 km) radius but improves with the larger radii. There is little difference in performance between D and D^*; they both have similar levels and trends in their Procrustes statistics. The results of nonmetric scaling are not much affected by our choice of sample size; the results for

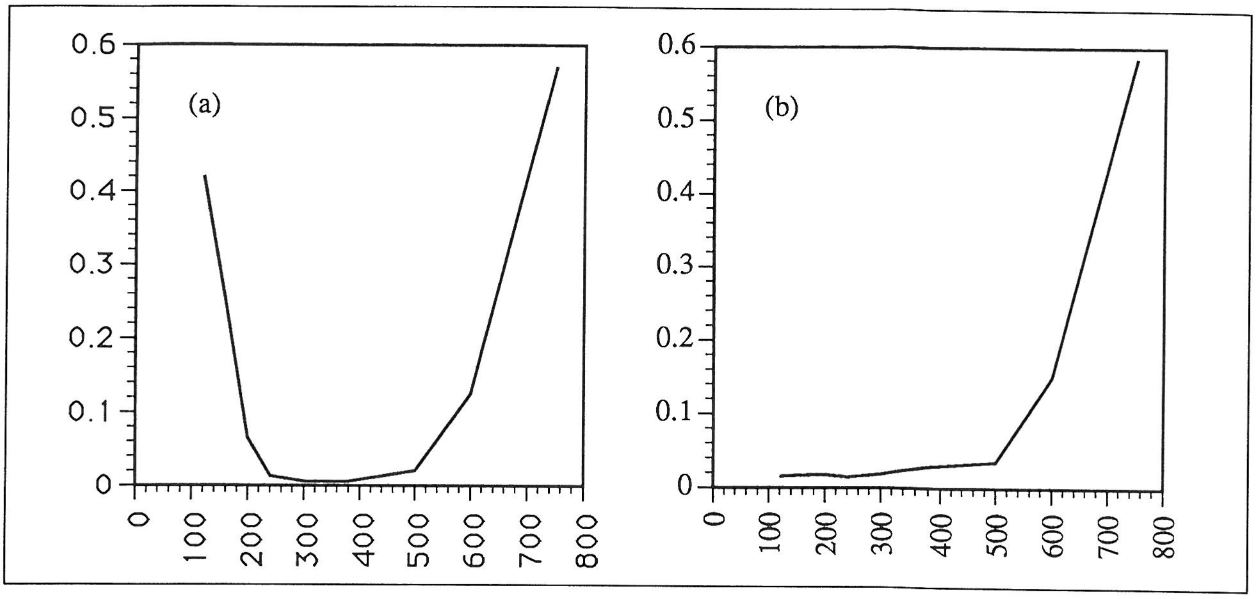
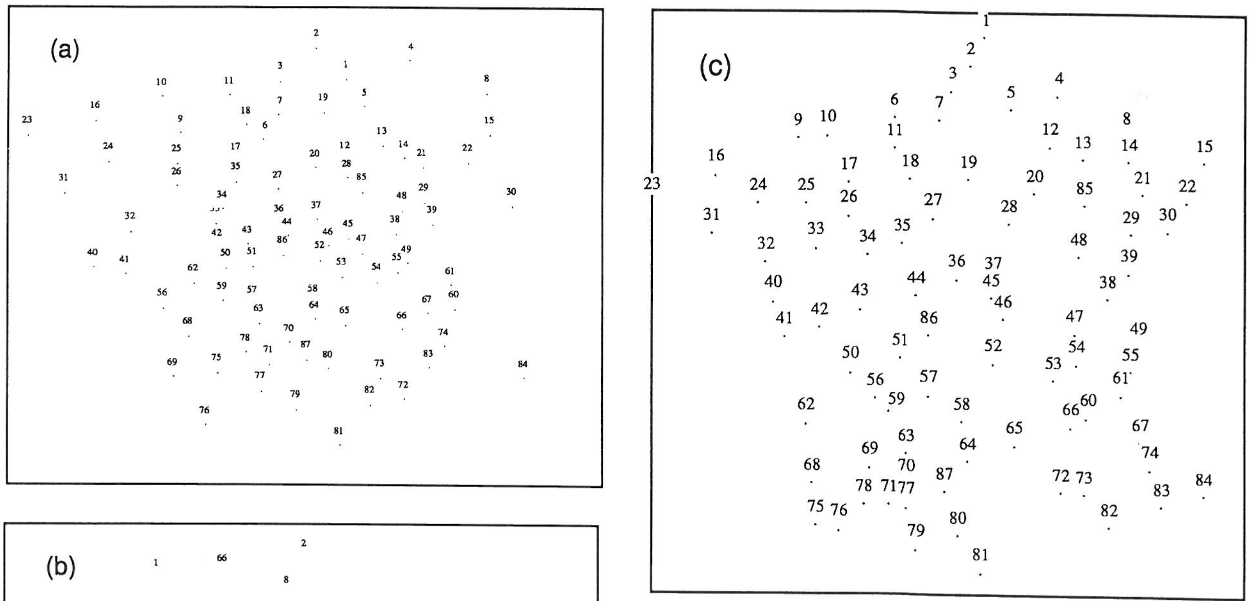


Figure 33.1: Change with increasing radius of disc (Kms) of Procrustes fit of estimated positions of French departmental capitals with their true positions: (a) Gravity model, (b) K–metric



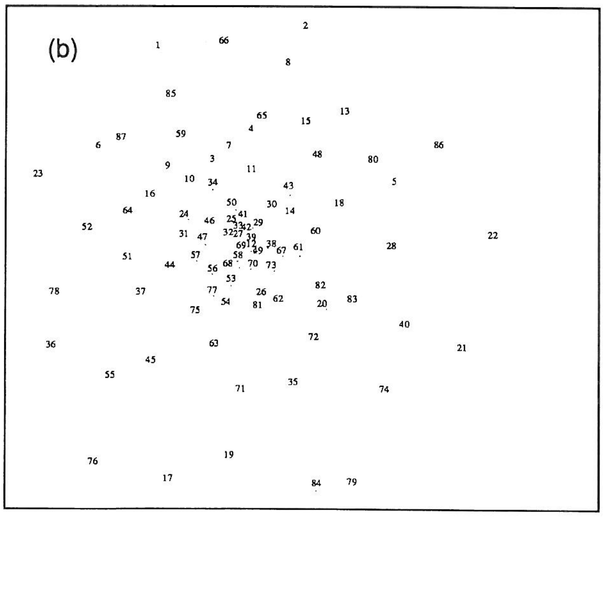


Figure 33.2: (a) Map of France derived from the K–metric distance measure, disc radius 120 km. (b)Map of France derived from the Gravity model distance measure, disc radius 120 km (c) Original map of France .

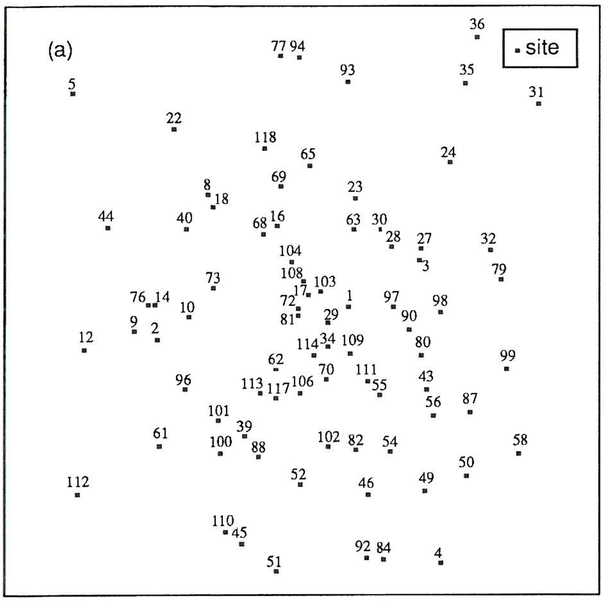


Figure 33.3: (a) Map of Cappadocia 80 sites.

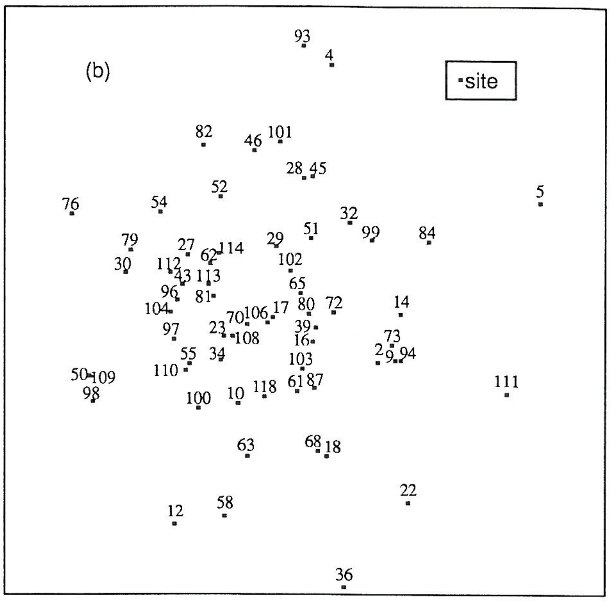


Figure 33.3: (b) Map of Cappadocia 64 sites. Excludes likely textile names and those genuine site–names that are linked, or are mainly linked, via textile names. This map is rotated to fit the positions of the same 64 sites in map (a)

samples of size 500 are only slightly better than those for 250. The *K*–metric yields good results even for small radii. Radii smaller than 95 km gave disconnected maps, and these were rejected. When the radius increases the degree of overlap between "tablets" increases, eventually giving complete connectivity between towns on many tablets, so that there is no information for constructing maps; this effect depends on how distance is estimated. We have incomplete evidence on this aspect but Figure 33.1 shows the U–shaped curves for the *K*–metric and for the Gravity model. This figure gives information additional to the values given in Table 33.4 and shows that complete connectivity occurs, for France, when the radius reaches about 750 Km (gravity model) and 450 Km (*K*–metric). Thus, it seems that the *K*–metric is best for modelling local trade, when the overlap between "tablets" is small. The gravity model behaves poorly for small–scale interaction but is more robust to higher degrees of overlap.

Figure 33.2 illustrates extremes in quality of the maps produced for the same radius (120 Km). Figure 33.2(a) is derived from the gravity model and Figure 33.2(b) from the *K*–metric. Without giving the details of the key relating department names to the number code, no detailed comments can be made, but it is clear that Figure 33.2(a) shows gross distortions with the north and north–west of France tending to become detached. What remains, approximates the correct topology but relative distances are wrong. In general, the longer distances are better reproduced than the shorter ones. In Figure 33.2(b), there is less distortion but there is a tendency for some peripheral departments to migrate inwards; compare 1, 9, 15 and 69. Other peripheral departments have moved outwards; compare 30, 56 and 84. This exemplifies a weakness of the *K*–metric in locating sites that are only weakly linked to the main body. This phenomenon is discussed more fully in the context of the Cappadocian data.

The simulations for the probabilistic model described under (ii) above, gave results remarkably similar to those already described and are not reported. We find this agreement surprising; perhaps probabilities other than p=0.04 would have shown more divergent results, but we did not have the opportunity to explore this aspect in more detail.

33.4.1 Conclusions of the simulation study

All methods performed quite well in the reconstruction of the map of France. For small radii, the *K*–metric turned out to be superior to the other measures. What this really means is that although there must be some, there should not be too much overlap in the sites occurring on the set of "tablets". The Cappadocian tablets exhibit moderate overlap, so it seems appropriate to use the K–metric in an attempt to reconstruct the map

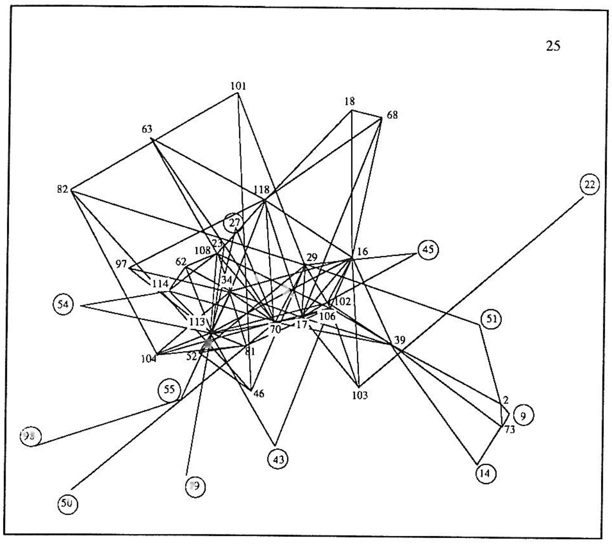


Figure 33.4: A map of 39 sites, showing their linkages. A line joining two sites indicates that they occur together on at least one tablet. Sites that are circled have weak linkages and their positions are especially open to question.

of the area around Kanesh. In the simulations, the underlying structure of the data contains no misleading information. Although towns may be missing from the region covered by a simulated tablet, no provision was made to include extraneous towns from outside the region, or even names which are misidentified as place names (the textile effect). Neither did we examine the effects of excluding a "central" town thoroughly, which would have been desirable to study. Perhaps we should have examined such details, but there are limits to the amount of simulation that can be done economically. Data from regions like ancient Cappadocia are unlikely to be as well–behaved as in our simulations. The main conclusion is that with reliable data it is quite feasible to construct maps from tablet–like information. However we do not know what might be the effect of extraneous features and therefore must be careful about extrapolating.

33.4.2 Back to Cappadocia

As it seems reasonable to assume that the Cappadocian tablets mainly reflect interactions between neighbouring sites, for the reasons explained above, we did not proceed with the gravity model. We checked that our reconstructed map based on the gravity model agreed with that produced by Tobler & Wineburg (1971). Therefore the following results are based entirely on the use of the *K*–metric. The stress values of .017 (all 80 sites) and .046 (64 non pseudo–sites) were very satisfactory and comparable to the values found for the simulations for mapping France. Comparing the 64 sites common to the two Cappadocian maps gave the uncomfortably large Procrustes fit of .60, which suggests that the pseudo–sites may be having a serious effect on the quality of the map. The two maps are shown in Figure 33.3. Figure 33.3(a) is the map of the 80 sites that include the pseudo–sites, while Figure 33.3(b) shows only the 64 non–pseudo–sites. There are major differences between the placements of the sites common to the two maps. It is evident that the pseudo–sites are a major disruptive influence. It is impossible to say whether or not any credence can be attached to the map of the 64 sites. Much depends on our belief in the quality of the data for these sites. Can we be sure that they do not contain unidentified pseudo–sites, or, another potential source of contamination, such as confusion between site names and personal names. All that can be said with confidence is, that if the data are clean and our assumption about proximity and joint occurrence is correct, then the simulations relating to France would support the validity of the map of Cappadocia.

One way of proceeding is to note that the twenty–five sites (4,5,10,12,28,30,32,36,58,61,65, 72,76,80,84,87,93,94,96,99,100,109,110,111,112) which had major displacements between Figures 33.3a and 33.3b might be regarded as unreliable. Put in another way this means that the pseudo–sites seem to have had an undue influence in positioning these 25 sites. Consequently, we produced another map based only on the remaining 39 (=64–25) sites. This is shown in Figure 33.4. We have compared this with a map constructed by Larsen (1976) which is itself informally based on information derived from the tablets, including short itineraries. Larsen's map gives ten names but, for the most part, does not give point–positions but allocates names to regions. Figure 33.4 agrees quite well with Larsen's map for the eight of his sites that are included in our thirty nine (17,29,34,55,104,113,114,118) and which occur in the central region of our map; Kanesh, number 34, is placed at about the centre, as one would expect. Many of the remaining 31 sites tend to occur around the perimeter of our map and on investigation some of these turn out to be linked to only one or two other sites. This again draws attention to a weakness of the K–metric, discussed briefly in the context of Figure 33.2(b) when reconstructing the map of France. A site at the end of a chain is not necessarily further from the centre although it will increase the number of links. To get a firm location one needs a site to be linked to at least three others. Thus the positions of these twelve peripheral Cappadocian sites should be treated with special caution and are specially identified in Figure 33.4. This leaves 27 (= 39 – 12) sites whose positions should be relatively well determined. These include seven of the eight (number 55 is weakly linked) mentioned by Larsen as listed above. The remaining twenty sites may be loosely grouped into two classes: (a) a central group (16,23,52,62,71,81,102,106,108) and (b) a somewhat more tenuous group (2,18,39,46,63,68, 73,82,97,101,103). One would have most confidence in Group (a) augmented by the seven central sites on Larsen's map. There is less confidence in Group (b), while the remaining sites are only very tenuously placed.

Acknowledgments
We thank K. R. Veenhof of Leiden University for his comments, especially for pointing out that some site–names were probably textile names. We also thank Jenny Howlett of CSIRO, Adelaide, South Australia, who fifteen years ago did some preliminary simulations which were forerunners of the work reported here.

References

Part 1: literature cited

Bilgic, E.
1945–51 Die Ortsnamen der "Kappadokischhen" Urkunden im Rahmen der alten Sprachen Anatoliens. *Archiv für Orientforschung* 15:1–37.
De Leeuw, J. & W. J. Heiser
1980 Multidimensional scaling with restrictions on the configuration. In P. R. Krishnaiah (ed.) *Multivariate Analysis V*. North Holland, Amsterdam.
Gatrell, A. C.
1981 Multidimensional scaling. In R. J. Bennett & N. Wrigley (eds.) *Quantitative Geography: A British View*. Routledge and Kegan, London. pp.151–163.
Gower, J. C.
1984 Multivariate Analysis: Ordination, multidimensional scaling and allied topics. In E. H. Lloyd (ed.) *Handbook of Applicable Mathematics: Vol. VI, Statistics*. J. Wiley and Sons, Chichester, pp.727–781.
Hodson, R.
1968 *The La Tene Cemetery at Munsingen–Rain*. Stampfli, Berne.
Kendall, D. G.
1970 A mathematical approach to seriation. *Philosophical Transactions of the Royal Society of London (A)* 269:125–134.
1971a Seriation from Abundance Matrices. In F. R. Hodson, D. G. Kendall & P. Tautu (eds.) *Mathematics in the Archaeological and Historical Sciences*. Edinburgh University Press, Edinburgh pp.215–252.
1971b Maps from marriages; an application of non–metric multidimensional scaling to parish register data. In F. R. Hodson, D. G. Kendall & P. Tautu (eds.) *Mathematics in the Archaeological and Historical Sciences*. Edinburgh University Press, Edinburgh. pp.303–318.
1971c Construction of maps from "odd bits of information". *Nature*, 231:158–159.
Kruskal, J. B. & M. Wish
1978 *Multidimensional Scaling*. Sage University Papers on Quantitative Applications in the Social Sciences, no. 07–011. Sage publications, Beverly Hills.
Larsen, M. T.
1976 The Old Assyrian City–State and its Colonies. Akademisk Forlag, Copenhagen.
Nashef, K.
1991 *Die Orts– und Gewässernamen der altassyrichen Zeit*. Dr. Ludwig Reichert Verlag, Wiesbaden.
Orlin, L.
1970 *Assyrian Colonies in Cappadocia*. Mouton, The Hague.
Shepard, R. N.
1962 The analysis of proximities: Multidimensional scaling with an unknown distance function. I,II. *Psychometrika*, 27:125–140; 219–246.
Tobler, W. & S.Wineburg
1971 A Cappadocian speculation. *Nature* 231:39–41.

Part 2: publications of tablets cited by Bilgic (1951) — amplified with references to Nashef (1991)

CCT Smith, S. *Cuneiform Texts from Cappadocian Tablets in the British Museum*. Part I, London 1921; Part II, London 1924; Part III, London 1925; Part IV, London 1927.
Mat Matous, L. *Tablettes de Kultepe* (unpublished). See also, (i) Matous, L. *Inscriptions cuneiformes de Kultepe*, 2. Prague 1962 and (ii) Donbaz, V. *Keilschrifttexte in den Antiken–Museen zu Stambul: Freiburger Altorientalische Studien, Beihefte: Altassyrische Texte und Untersuchungen Bd. 2*, Stuttgart, 1989.
VAT Museum signature, Berlin (Vorderasitische Abteilung T(h)ontafeln) cited with permission of Liane Jakob–Rost.
Jena Lewy, J. *Die Keilschrifttexte aus Kleinasien, Texte und Materialen der Frau Prof. Hilprecht Collection of Babylonian Antiquities im Eigentum der Universitat Jena, Volume 1*. Leipzig, 1932.
Liv Pinches, Th. G. The Cappadocian Tablets belonging to the Liverpool Institute of Archeology. *AAA* 1, 49 – 80, 1908.
PSBA VI Sayce, A. H. PSBA, VI, 17–25, 1883.
Hr Hrozny, B. *Tablettes de Kultepe* (unpublished) See also, Hrozny, B. *Inscriptions Cuneiformes du Kultepe*, 1. Monografie Archivu Orientalniho 14, Prague 1952.

Ark.Derg.IV Turk Tarih, *Arkeologya ve Etnografia Dergisi.*

KTB Lewy, J. *Die Kultepetexte der Sammlung Rudolf Blanckertz Berlin.* Berlin 1929.

AHDO Lewy, J. Old Assyrian documents from Asia Minor (about 2000 B.C.) *Archives d'histoire du droit oriental.*, 1, 91 – 108, 1937.

Unv.Ank Text in Ankara Museum from Bogaskoy and Kayseri Museum.

KTS Lewy, J. *Keilschrifttexte in den Antiken–Museen zu Stambul: Die altassyrischen Texte vom Kultepe bei Kaisarije.* Constantinople 1926.

TTC Contenau, G. *Trente Tablettes Cappadociennes,* Paris 1919.

EL Eisser, G. and Lewy, J. Die Altassyrischen Rechtsurkunden vom Kultepe, *Mitteilungen der Vorderasiatisch–Aegyptischen Gesellschaft 33 und 35.* Leipsig 1930–35.

KTP Stephens, F. J. The Cappadocian Tablets in the University of Pennsylvania Museum. *JSOR* 11:101 – 136, 1927.

OIP XXVII Gelb, Inscriptions from Alisar and Vicinity. Oriental Institute Publications XXVII.

KTHahn Lewy, J. *Die Kultepe texte aus der Sammlung Frida Hahn.* Berlin, Leipzig, 1930.

BIN IV Clay, A. T. *Letters and Transactions from Cappadocia, Babylonian Inscriptions in the Collection of J. B. Nies 4.* New Haven, 1927.

BIN VI Stephens, F. J. *Old Assyrian Letters and Business Documents, Babylonian Inscriptions in the Collection of J. B. Nies 4.* New Haven, 1944.

TC I Contenau, G. *Tablettes Cappadociennes, Musee de Louvre, Textes Cuneiformes 4.* Paris, 1920.

TC II Thureau–Dangin, Fr. Tablettes Cappadociennes, Deuxieme Serie, Musee de Louvre, Textes Cuneiformes 14, Paris, 1928.

TC III Lewy, J. *Tablettes Cappadociennes, Troisieme Serie, Musee de Louvre, Textes Cuneiformes 19, 20 et 21,* Paris, 1935–37.

Authors' addresses

A.G. Constantine
CSIRO
Division of Mathematics and Statistics
Glen Osmond
Adelaide
South Australia

J.C. Gower
Department of Data Theory
University of Leiden
Pieter de la Court gebouw
Wassenaarseweg 52
Postbus 9555
NL–2300 RA Leiden

B. Zielman
Department of Data Theory
University of Leiden
Pieter de la Court gebouw
Wassenaarseweg 52
Postbus 9555
NL–2300 RA Leiden
E–mail: