19

A method for the analysis of incomplete data and its application to monastic settlements in Italy (4th-6th century)

Béatrice Caseau[1] and Yves Caseau[2]

[1] History Department, Princeton University, USA. [2] Software Research, Bellcore, USA.

19.1 Introduction

This article describes the method that we have used in a study of Italian monastic sites that date from the 4th to the 6th century AD[1]. Most of our information concerning these sites comes from literary sources. Among these ancient monasteries, some have left only a faint memory while others had a sufficient success to survive through the centuries. In some cases, there is no formal identification of the site's location, and only guesses are possible. In other cases, such as Monte Cassino, archaeological excavations have given us an accurate picture concerning the size of the first monastic community. This unequal knowledge about monastic settlements of the Late Antique period has often led to privileged studies of those monasteries of which we have more information. To make sure that we have a precise image of the monastic topography, we propose to use all the evidence we have, even when incomplete. Our first goal was to draw a map of these sites, but that became impossible without a sure knowledge concerning the locations. We turned to correspondence analysis (CA) as the right tool to encompass all the data we had gathered on monasteries. The result is a map depicted in Figure 19.1 that shows the type of site chosen for each monastic establishment in relation to the gender of its inhabitants, the century of its foundation, and the type of monastic life led by its founders.

Correspondence analysis allows us to visualise each site for its relation to other sites of the same kind, since they appear grouped in clusters on the graphic. Our goal is to have a readable graphic showing the multiplicity of sites, a fine typology of monastic settlements and its evolution through the centuries studied. We had, however, to adapt the usual techniques used with correspondence analysis, to make it fit for the type of incomplete or uncertain data, that our sources reveal on Italian monastic settlements.

In order for us to construct an accurate picture of monasticism in this period, we have decided to enter all the information we have, although some of it is uncertain and most of it is incomplete. Had we entered this information as a 'yes/no' table, we would have had to force an answer when we do not have definite evidence. This would have led to a biased graphic. Our method produces results that are meaningful from a statistical point of view as well as from a historical perspective. This method can be applied to other types of data whenever an element of uncertainty has to be taken into consideration or whenever the incompleteness of the information would introduce a bias in the analysis.

First, we shall explain what kind of information we wanted to see analysed and represented. Then, we shall present the graphics and the conclusions we can draw from them. Finally, we shall study the method itself and compare it with what is currently done.

19.2 Analysis of Monastic Settlements

We have selected a sample of 65 Italian monastic communities for which we had some information on the topography and on the type of monastic life. We have given a number to each of the sites (1 to 65). Then, we classified our information into nine sets of criteria that we designated with a letter from A to I. On the graphic, each criterion appears with a letter (the set) followed by a number. The criteria are selected according to what appears in our sources and to the type of questions we wish to raise. In this study we have decided to concentrate on the relations between the type of site and the style of monastic life chosen and on gender issues. The main benefit of this approach is that it is custom-made and flexible. Once we have gathered the information on the sites, we can prioritise the questions we want answered and grant each criteria a specific weight according to the prominent theme we wish to follow.

19.2.1 Description of the Main Criteria Used in the Analysis

19.2.1.1 Regional distribution

If we do not always know the precise location of our monastic sites, we usually can locate them in a region. This is why our first criteria, 'A', classifies the sites into five regions (A1-A5). We have divided Italy in regions of unequal sizes, since we have more information on the regions close to Rome than on the other Italian regions. This criterion is mostly indicative.

A1 : Northern Italy
A2 : Tuscia and Umbria
A3 : Rome
A4 : Valeria
A5 : Southern Italy (Campania, Calabria, Sicilia...)

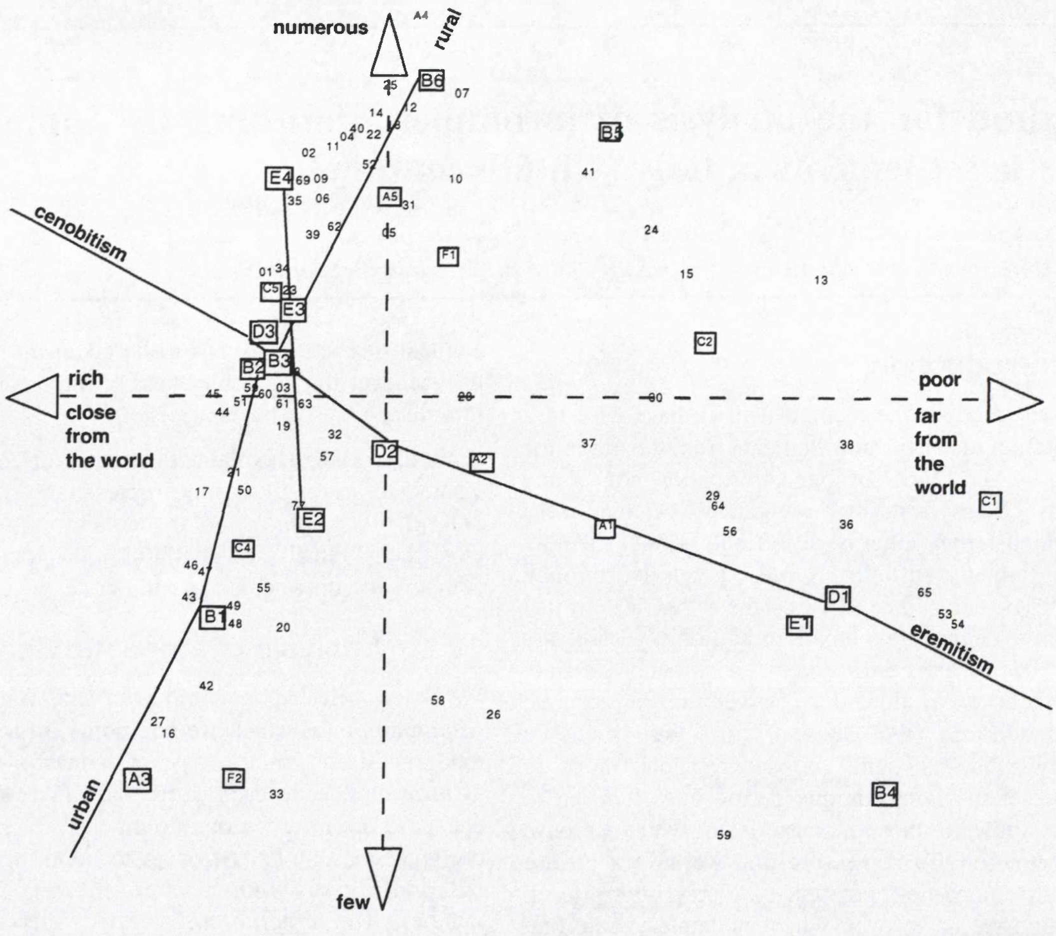

Figure 19.1: Correspondence analysis of monastic settlements

19.2.1.2 Type of site

Our second criterion, 'B', classifies the types of sites recorded. We have established six types (B1-B6). In this classification, the gradient goes from the most to the least accessible. This is the major criteria for our study. We want to be able to see on the graphic what type of site was preferably chosen along these centuries and how this type of geographic location can be related to the other information that we have on monastic establishments.

B1: urban site
B2: peri-urban site (outside the walls, but close to the city)
B3: road or cross-road
B4: rural, in a plain or valley
B5: forest and/or mountain
B6: swamp and/or island

19.2.1.3 Type of habitat

Our third criterion, 'C', concerns the type of habitat. Our monastic settlements fall into five categories, from the use of a natural cave to the building of a specific monastery. Between these two extremes, monks and nuns also chose to live in quickly built, precarious, dwellings, or to show the victory of Christianity over paganism by settling in the ruins of pagan temples, or, simply, to adapt their own house to their new lifestyle. A monk such as Benedict of Nursia began his monastic career in a cave, but he also established monasteries in pagan ruins, and supervised the erection of a monastery.

This criterion is important to bring into the spotlight the variety of monastic establishments in this early period of monasticism. We are particularly looking forward to see the correlation between this criterion and the type of monastic life chosen.

C1: natural habitat (cave...)
C2: precarious habitat (cabin...)
C3: Roman ruin (temples...)
C4: personal house
C5: monastery (building created as a monastery)

19.2.1.4 Type of community

Our fourth criterion, 'D', has a gradient based on solitude. It classifies monastic settlements into three categories. Monks and nuns could either live as a solitary hermit, or belong to a group of hermits living separately, but close by, and under a common spiritual guidance, or, form a community by living together under the same roof. The same monk or nun could change from hermit to cenobite, or the other way around. Since our study deals with settlements, we have entered the dwellings of hermits, even when we know that they eventually left their caves to live in monasteries.

D1: hermit
D2: group of hermits
D3: cenobites

19.2.1.5 Number of monks/nuns

The following criterion, 'E', concerns the number of monks or nuns per site. Monastic establishments were often relatively small during this period. We have defined four categories, 1 person, 2 to 3, 4 to 10, more than 10 persons. This is the criterion for which our uncertainty is maximum. We have no statistics on these early monastic establishments. For each site, we have noted the number of monks or nuns cited in our sources. This is only an indication, since numbers must have fluctuated in the lifetime of each establishment. This criterion has been granted only a small weight in the analysis, since it tells us more about the loquacity of the sources on each site than about the real numbers of monks and nuns. Nevertheless, it is significant because it gives us an indication concerning the relative size of the unit we are dealing with.

E1: 1
E2: 2 to 3
E3: 4 to 10
E4: more than 10

19.2.1.6 Gender

The following criterion, 'F' does not require any explanation. It concerns gender. In most cases, we know the gender of the inhabitants of a monastic settlements. In rare cases, we have presumptions. Gender obviously influenced the choice of a site, as we shall see later in the commentary of the drawing.

F1: monks
F2: nuns

19.2.2 Analysis of the Main Criteria

The result of the correspondence analysis is the map (cloud of points) depicted in Figure 19.1. Our cloud has the shape of a reversed 'V'. We have drawn lines to connect the major criterion and we have added a short comment about the general meaning of those lines. These lines show the structure of the graphic. A short vertical line links the numerical criteria (E). The monasteries with more than ten monks or nuns (E4) are located on top of the drawing, while those with fewer monks or nuns (E2) are under the horizontal axis. Naturally E1, one monk or nun coincides with eremitism (D1) and with hard to access locations such as swamps or islands, and, also, with a preference for natural habitat. Another line is drawn between the types of communities. Eremitic settlements (D1) are located at the bottom and at the right of the drawing. Cenobitic communities (D3) are, at the opposite corner: at the top and left portion of the drawing. A third axis structures the cloud of points and reveals how urban (B1), suburban (B2-B3) or rural (B6) each settlement was.

We can distinguish different clusters:

- eremitic monasticism at the bottom and right of the picture. E1, one monk or nun, D1, hermit, C1, cave and B4, island or swamps, form a group of sites. Martin, the fourth century monk and future bishop of Tours, for example, spent some time as an hermit on the island of Gallinaria. Some of these sites were used as refuge.
- B5, mountain and forest, is located at the top right of the drawing, as well as C2, precarious habitat. This group of settlement is still often eremitic. However, a few semi eremitic and cenobitic monasteries were founded in mountain sites, which explains the attraction of these criteria to the top of the drawing.
- F1, men, is also located in that top right part of the drawing, although it is more central. Eremitism tended to attract more men than women, while F2, women, is located in the bottom left part of the drawing very close to urban sites and particularly to Rome (A3). We have for Rome more evidence than for many other Italian cities. And the number of women monastics in that city was even higher than what is indicated in that drawing, if we believe the number given by pope Gregory the Great of three thousand nuns refugees in Rome in 597. (Ep. VII, 23). These women were fleeing other cities or their country estates. Most of them had adopted a monastic lifestyle either in their own houses, or close to a basilica. We find them in the lower left part of our drawing, with the other urban sites (B1) and with personal houses (C4).
- The most closely knit cluster occupies the central and top left portion of the drawing. There we find B2, suburban site, B3, crossroad site, buildings created as monasteries, C5, and cenobites, D3. This forms a group whose success shall grow in later centuries. Settled close to the lands that feed the growing number of novices, but close enough to urban centres, these settlements managed to keep in touch with the urban clergy, while building their influence on the country side

19.2.3 Adding New Criteria

Our second experiment includes all the previous criteria and introduces three additional ones. One dates the monastic settlements. and the other the chronology. We note for each site the earliest time at which some monastic presence is recorded. This does not presume of the fate of the monastic settlement. Here again, the treatment of uncertain data adopted in our method reveals to be very useful, since we only know a date, for the foundation, in very few cases. Most of the time, however, we can make plausible guesses. This criterion allows us to see the general evolution of monastic settlements throughout the period of late antiquity.

19.2.3.1 G : Date of foundation

G1: IVth century
G2: Vth century
G3: VIth century

Our second purpose is to be able to compare well-known sites to the less successful settlements that disappeared either during this period or later in the Middle Ages. We also wish to place our topographical data in correspondence with known types of monasticism. One criterion measures the influence of Saint Benedict of Nursia; the other analyses the different options adopted by the monks and nuns of this period to live a monastic life.

This criterion (H1) has received little weight in our analysis. Still, it is very interesting. In the graphic, this criterion, 'H1,' shows where Benedict's monasticism stands in comparison with other monastic settlements. This well-known abbot has often been considered as the main Italian monastic figure of this period, thanks to the success of his Rule in later centuries. In his lifetime Benedict had some influence over a number of sites, either as an abbot or as a prominent holy man. This is the occasion for us to compare the sites of Benedict's foundations with contemporary or older foundations.

The foundations made by Saint Benedict of Nursia in the sixth century and the communities that we can consider under his influence are located at the top of the drawing. Benedictine monasticism is found at the joining points of two tendencies: eremitism and peri-urban monasticism. Its central position shows that it realised the synthesis of the prevalent historical tendencies of monasticism at that period: an evolution from eremitism towards cenobitism, from a radical poverty to a relative wealth, but also from urban settlements to rural ones. By its origins, it is close to eremitism. However, it is also relatively open to the world. Midway between the crossroads site (B3) and mountain site (B5), it is both quite remote and accessible.

The third criterion, 'I', proposes an analysis of the monastic communities based on their separateness from the world. An aristocratic lady who chose to wear simple clothes and transforms her house into a 'monastery' did not have the same type of life than the same woman leaving everything to settle in a cave, in some remote mountain. In the first case, she might have visits from her old friends, in the other case, she has decided to concentrate, in an absolute manner, on solitary contemplation. In the first case, she often convinced her closest servants of joining her in the experience, transforming somehow their previous relationship; in the second case, she lived alone or sometimes with one disciple. Most of all - whatever her intimate wish to change her life in a radical manner - in the first case, she still lived in her own house, while in the second case, she has fled human contact by choosing a remote, hard to access location. In between those two extremes, two other possibilities were common. One could settle in another building than the one owned by one's family and serve the near-by basilica, in different ways: chant, hospitality... One could also leave the busy cities and settle in a quiet, yet accessible place, from which contacts with fellow human beings were possible. The rupture with the world was obvious, by the choice of location, but contacts were still easy to establish. This criterion is very interesting if related to the type of sites and to gender.

19.2.3.2 I : Type of monasticism

I1: domestic monasticism (in one's home)
I2: basilical monasticism (close to a basilica)
I3: semi-desert (not too far from human settlements, villages, towns, major roads, practice of hospitality and evangelisation ...)
I4: desert (very hard to access spots)

The result presented in Figure 19.2 has the same shape than the previous one. The introduction of the criterion concerned with Benedict's influence has slightly changed the shape of the cloud at the top of the drawing. 'H1' logically attracts a number of monastic settlements which were closely connected with Benedict.

Eremitism attracted monks from an early period. This is why G1, fourth century, is located on the lower part of the drawing. It is closer to urban monasticism, especially in Rome, since the earliest attempts at monastic life seemed to have been happening in urban centres under the influence of groups of ascetics aware of the development of this lifestyle in the Eastern provinces of the Empire. The position of G2, fifth century, and G3, sixth century, a little bit more on the right shows that eremitic settlements, and, more generally, difficult to reach locations had a growing success. They are represented by a small but quite constant number of examples through the centuries we study. This should not hide the fact that the tendency of these centuries is for monasteries to be build in rural areas, sometimes not far from urban centres. The scene of sixth century monasticism is no longer only dominated by urban foundations and a few eremitical settlements, but by possibly larger units established thanks to the donations by wealthy landowners.

If this graphic allows us to pinpoint the general evolution and to note tendencies in Italian monasticism, it also enables us to see the variety and richness of monasticism during this study period. We have clusters of monastic settlements of a similar nature. We also have a number of isolated points in the cloud. This is a testimony about the numerous subtleties in the choice of lifestyle and of sites, which is a characteristic of that formative period of monasticism. Thanks to a method that does not force us to reduce reality to a simplified model we have been able to enter, for each case, those precise nuances. Let us now turn to the technical aspects of that method.

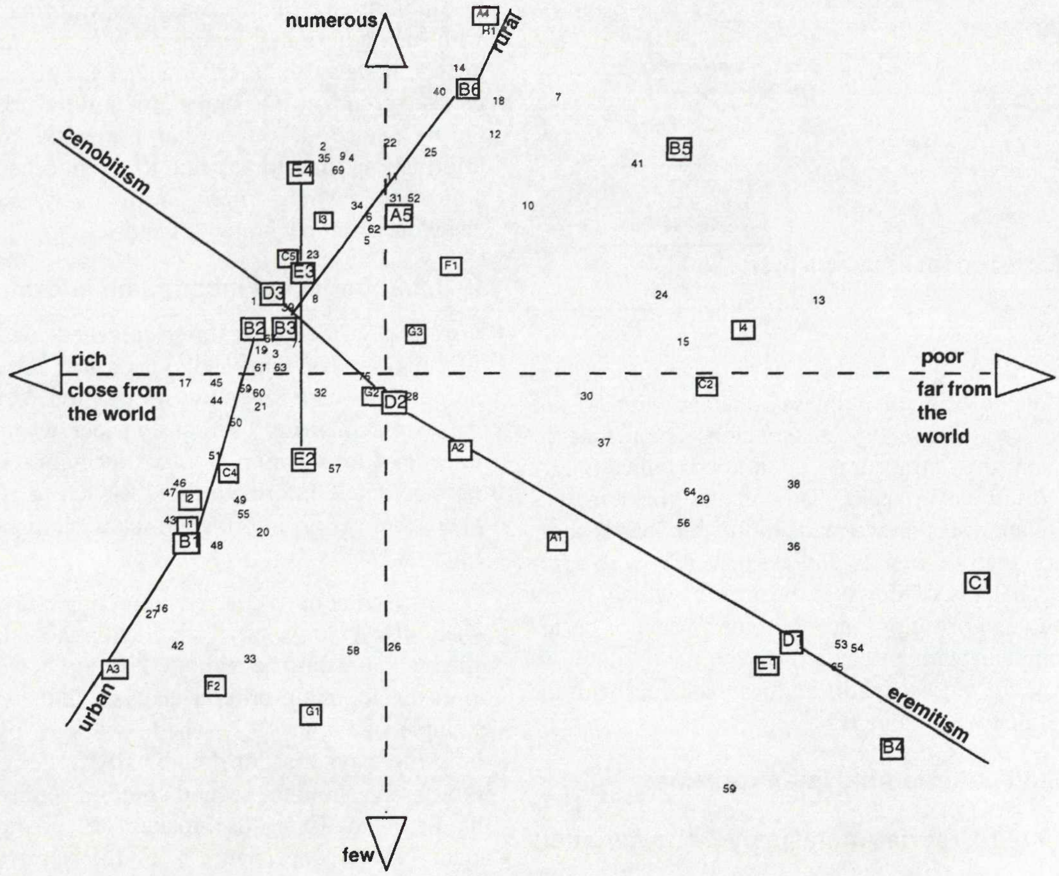

Figure 19.2: Correspondence analysis including historical information

19.3 A Method for Multi-Criteria Analysis of Incomplete Data

19.3.1 Correspondence Analysis for Historical Data – Problem Description

We want to analyse a sample of n individuals (monastic settlements in our study, with $n = 65$). For each individual i, we can make a sequence of observations $o_1, o_2, \dots o_m$, which correspond to a given question. For instance, the observation o_{11} in our study corresponds to the question, 'Is the monastic settlement in a cave'. We define a value $o_j(i)$ which is the probability that the answer to the question is yes, to the best of our knowledge. This value is a real number that has a range from 0 (we are sure that the answer is no) to 1 (we are sure that the answer is yes), where $o_j(i) = 0.5$ indicates that we know nothing about the question o_j for the individual i. The data that we want to analyse can, therefore, be seen as an $n \times m$ matrix of real numbers.

Observations are not independent since many of them are exclusive. We define criteria $C_1, \dots C_k$ as sets of observations that focus on the same aspects and are, therefore, clearly dependent. For instance, we define the criteria $C_1 = \{o_1, o_2, \dots o_5\}$ to be the geographical distribution of settlements. In this case, $o_1, o_2, \dots o_5$ are mutually exclusive, but it is not necessary. The structuration of observations into criteria is important because multi-criteria analysis is more complex and biases can be introduced if we are not careful, as we shall see in Section 19.3.2.

19.3.1.1 Correspondence Analysis

Numerous statistical tools are available to analyse a data matrix similar to the one we have shown, which can be seen as a set of points in an m dimension space. The most common 'multi-dimension' method is principal components analysis (PCA), which projects the set of points on the two (or more) more meaningful dimensions (to maximise the dispersion of the cloud of points, that is, to minimise the loss of information). PCA can be defined as a diagonalisation of the covariance matrix, followed by a projection.

Correspondence Analysis is a variation that consists of normalising the input matrix before applying the PCA (cf. Figure 19.3). CA is well suited to data categorisation, such as the result of an archaeological excavation. If the input matrix for each site contains the number of artefacts of each type the normalisation step of CA allows us to compare meaningfully sites of different size (only the relative frequency of each artefact group counts).

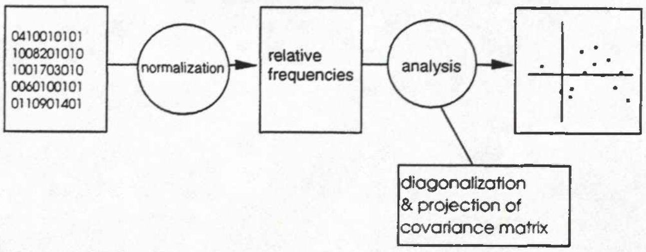

Figure 19.3: Correspondance analysis

The result of the analysis is a two-dimensional projection of the original point set (the individuals) and the original vector base (the observations) such that the 'closeness on this projection' is a good indication of similarity (high covariance). This is the mathematical basis for the analysis that we made in the previous section. The question that we want to address here is how to apply CA to the analysis of our historical data, which is not simple data categorisation and, therefore, cannot be fed into the computer without caution. In the next section we will illustrate the biases and difficulties that occur with the kind of data that we obtained.

19.3.2 Multi-Criteria Analysis Problems

19.3.2.1 Multi-Criteria Analysis by Concatenation

Let us first consider the case of data obtained by concatenation of multiple categorisation. We suppose that for each individual we have obtained a vector (cf. Figure 19.4), which is the concatenation of the multiple categorisation vector. For instance, we may want to analyse excavation data with different (non-comparable) kinds of output (such as bones, tools and plant fossils). For each type of data, we can perform a classification into many groups. For instance, we may have 4 types of bones, 10 types of tools and 4 types of plant fossils. We see each of the three categories as a different 'criterion' for the site excavation and we obtain our 'Multi-criteria' site data by concatenating the categorisation count vectors.

If we run a correspondence analysis on this data without caution, the result may be disappointing. The structure that we have established is not taken into account by the analysis and multiple problems can occur. For instance, if we consider the two sites of Figure 19.5, they have exactly the same distribution patterns for the two first criteria. However, the normalisation step of the correspondence analysis will make these two sites very distinct. What the analysis will show is the correlation between the finding of bones, tools and plants. Although interesting, it is not what we want to accomplish with a Multi-criteria analysis.

Applying CA to this 'raw' data actually produces a classification on meta-information. It will incorporate the fact that we have more information about one criterion than about another for a given site. This can distort considerably the result of the analysis, as we shall see in the next section.

Similarly, the relative weight of information associated with each criterion is a moderately significant factor. With historical data, some criteria can be divided into many precise groups while some other only correspond to a binary choice. If we do not normalise the amount of information that we assign to each criterion, we may arbitrarily favour criteria with many categories and disfavour criteria with few categories.

19.3.2.2 Impact of Incomplete Information

Historical evidence is often plagued with incomplete information. For some individuals and for some criteria, we either do not have any information or we have only approximations. Let us first consider what to do in the absence of information. A usual technique is to represent the incomplete information by 0s such as in Figure 19.6, where there is no information available about the second criterion.

This technique is derived from single-criteria analysis, where it would be perfectly legitimate, since an entry filled with 0 would be ignored. It is not, however, a correct approach for multi-criteria analysis. Let us consider the example of Figure 19.7, where two sites present exactly the same categorisation for the first criterion, but where information about the second criterion is only available for the first site. Using 0s violates our previous condition about the independence of the criteria and the normalisation process of correspondence analysis will make the two sites more distinct that they should be since the similarity of the first criteria will be missed.

Once again, we need to guard ourselves from producing a meta-information analysis, where sites would be classified according to the presence or absence of information.

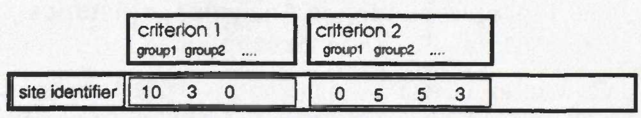

Figure 19.4: Multi-criteria analysis

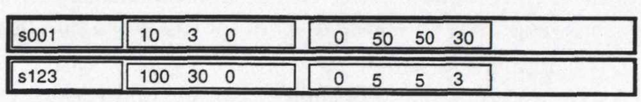

Figure 19.5: Two sites with similar structures

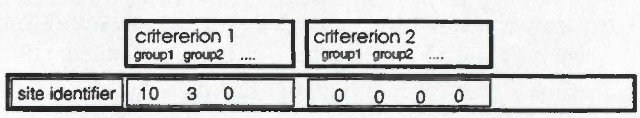

Figure 19.6: Site with no information about criterion #2

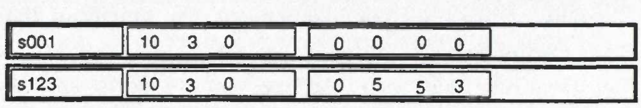

Figure 19.7: Sites with similar structure for criterion #1

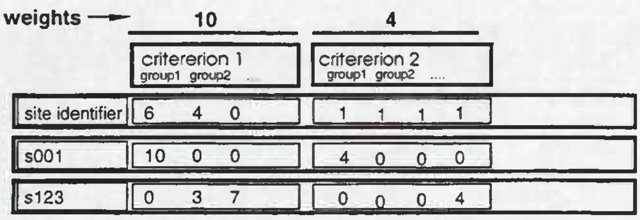

Figure 19.8: Sites representation with our proposed method

19.3.2.3 Uncertain Information

The last concern that we need to address is the uncertainty of information that we have explicitly represented in our modelization, through probabilities. This is especially important with historical data such as textual evidence that is by nature almost always incomplete or uncertain. We need a way to carry this uncertainty into the matrix representation that we give to the correspondence analysis software.

Since we start with a vector of probability of answers for each criterion and since we want to apply correspondence analysis to multi-criteria categorisation data, it is natural to produce a categorisation integer vector that approximate the probability distribution. Let us consider the example of the geographical criteria $C_1 = \{o_1, o_2, \ldots o_5\}$. If we know that the settlement was located in Rome (observation o_3), we have the probability vector (0,0,1,0,0) that we translate into [0,0,10,0,0]. If we believe that it is in Rome, but there is also a possibility that it was actually in Umbria, we use [0,2,8,0,0]. The idea is to approximate a probability distribution with an integer sequence of constant sum.

This allows taking uncertainty into account with a lot of flexibility that ranges from exact knowledge to a total lack of information. In the previous example, the vector [2,2,2,2,2] is the precise representation of the absence geographical data.

19.3.3 A Method Based on Data Pre-Processing

We can now easily describe the method that we have built since it is a consequence of all the previous observations. The first step is to identify the criteria that we want to analyse and to define them as a set of observations. We then collect the values of each observation for each individual in our sample and we record the degree of confidence with the answer.

For each criterion, we allocate a number of token that is the weight that we want to attribute to the criteria in the analysis. A good practice is to start with an even distribution (same for each criterion) and then modify it gradually to emphasise certain aspects. As we shall see later, this has only a limited influence on the output produced by correspondence analysis.

We then distribute the tokens for each individual (i.e. settlement) and each criterion, according to the observations and their degrees of confidence. When we know the exact answer, we put all the tokens in the corresponding information (for instance, the first criterion and the individual *s001* in Figure 19.8). On the opposite, we distribute them evenly when no information is available. This allows to break from the constraints imposed by a strict set of observations. For instance a precarious house may be seen as somewhere between a house (observation o_{15}) and a precarious habitat (observation o_{13}), which we will represent by putting half the tokens in each of these two observations.

This method enforces all the constraints that we have established previously (independence of criteria's weights and fair representation of incomplete information). The result, that we will prove in the next section, is that the analysis is free of many of the biases observed with simpler approaches. An interesting corollary of this method is that answers of 'yes/no' questions should not be represented by one observation (1/0) but rather by two ([2,0] will represent yes, [0,2] will represent no, and [1,1] will represent unknown).

19.3.4 Validation of the Method on Monastic Settlements Data

19.3.4.1 Comparison with other approaches

We compared three different approaches for the same set of data. The first method is the one that we present in this paper. The second method is a hybrid method, where we pick one answer (observation) only for each criterion. That is to say, we choose the most likely answer for each criterion and we represent it with a single 1 entry. The result is a 0–1 matrix, where the independence between criteria is verified, but where some of the decisions are made arbitrarily (when no information is available).

The last approach is the simpler one, where the input matrix is obtained by using a 1 when the answer for the observation is yes and 0 otherwise. We have found this method to be commonly used, even in cases where the biases presented in the previous sections are obviously present.

To evaluate the quality of the analysis, we focused on two indicators. The first indicator is a quantitative coverage number, which it is the sum of the two first eigenvalues produced during the diagonalisation of the covariance matrix. This tells us 'how much' information is represented on the two-dimensional projection of the point cloud. A higher number indicates a better quality analysis.

The second indicator is qualitative, and deals with the existence of a structure in the way the observations are represented in the output of the analysis. When the different exclusive observations of a same criterion are aligned, it gives a basis to interpret the two-dimensional representation. A good example is the population criteria,

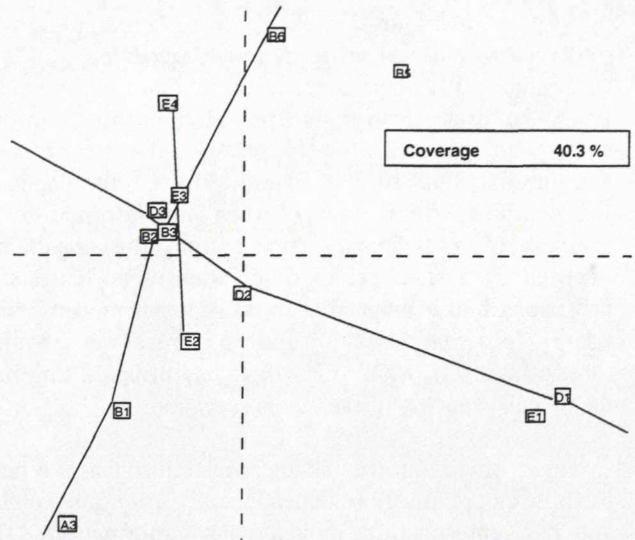

Figure 19.9: Analysis output with our proposed method

where the various observations (1, 2-3, 4-10, 10+ members for the monastic settlement) naturally lead to a 'numerous vs. individual' axis. In the previous analysis (Section 19.2), we have identified three axes, corresponding to the population criterion, the cenobitism/eremitism criterion and the type of site. The remarkable alignment that we have observed makes the interpretation of the result easier and more convincing. Figure 19.9 represents the structure obtained in our analysis, together with the coverage indicator.

Figure 19.10 shows the output for the second hybrid method. We can easily notice the degradation from both the quantitative and the qualitative indicators.

The results of the simpler method are shown in Figure 19.11. These results are poor, since there is no visible structure (thus interpreting them is difficult) and the coverage indicator is very low, which gives little significance to the findings that we can extract from this two-dimensional representation.

19.3.4.2 Stability

We have tried different samples to evaluate the stability of the analysis and the sensitivity of the size of the input. We used samples of sizes ranging from 15 to 77. Our experience suggests that one needs at least 40 individuals in the sample to get stable results. Samples with fewer individuals have strong biases and are not reliable. On the other hand, the results obtained with samples of size 50, 65 and 77 were strikingly similar.

Our experimentation with statistical method is far from being completed. More work is needed to measure stability more accurately, using larger samples. We also plan to use more subtle strategies to explore the use of CA as a tool to evaluate causal dependencies, especially between the environment and the type of monasticism.

19.3.4.3 Impact of Criteria's Weights

The distribution of the criteria's global weight is, as we have said, a global parameter of the analysis. However, it is important to note that it has little effect in the quality of the analysis, as measured previously. In Figure 19.12, we show the results obtained with an even distribution of the criteria's weights.

19.4 Conclusion

We have presented a method for using correspondence analysis with multi-criteria incomplete and uncertain data. This method consists of pre-processing the data so as to eliminate the possible biases caused by the uncertainty and the relationships between criteria. Because of this pre-processing, we have obtained results that are stable and meaningful from a statistical point of view.

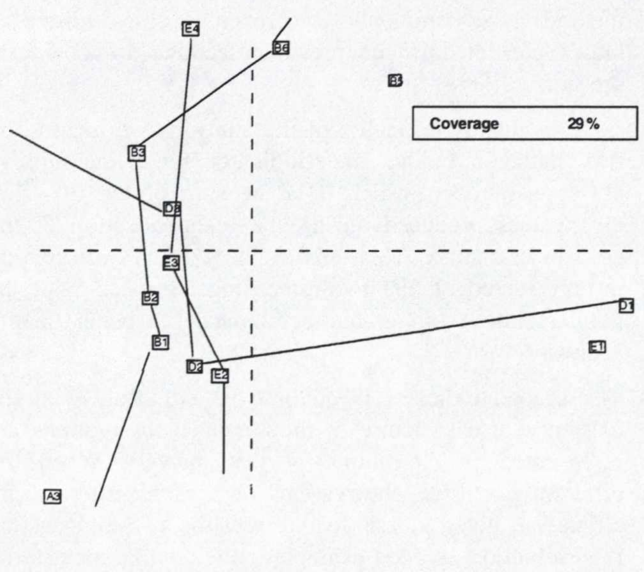

Figure 19.10: Analysis output with the second method

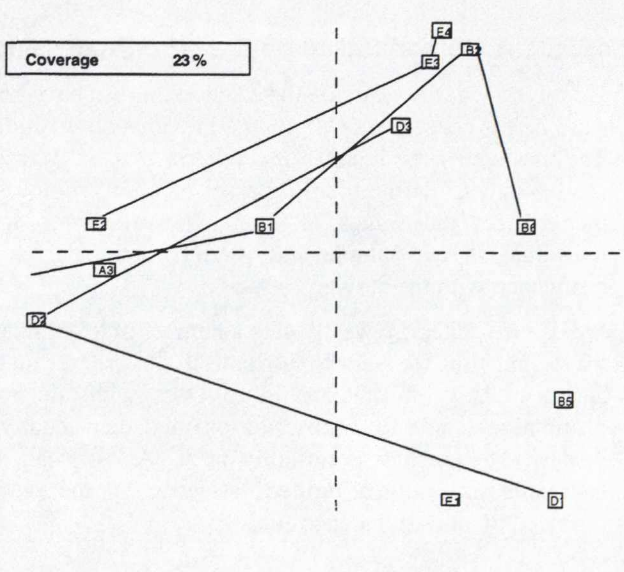

Figure 19.11: Analysis output with the simpler method

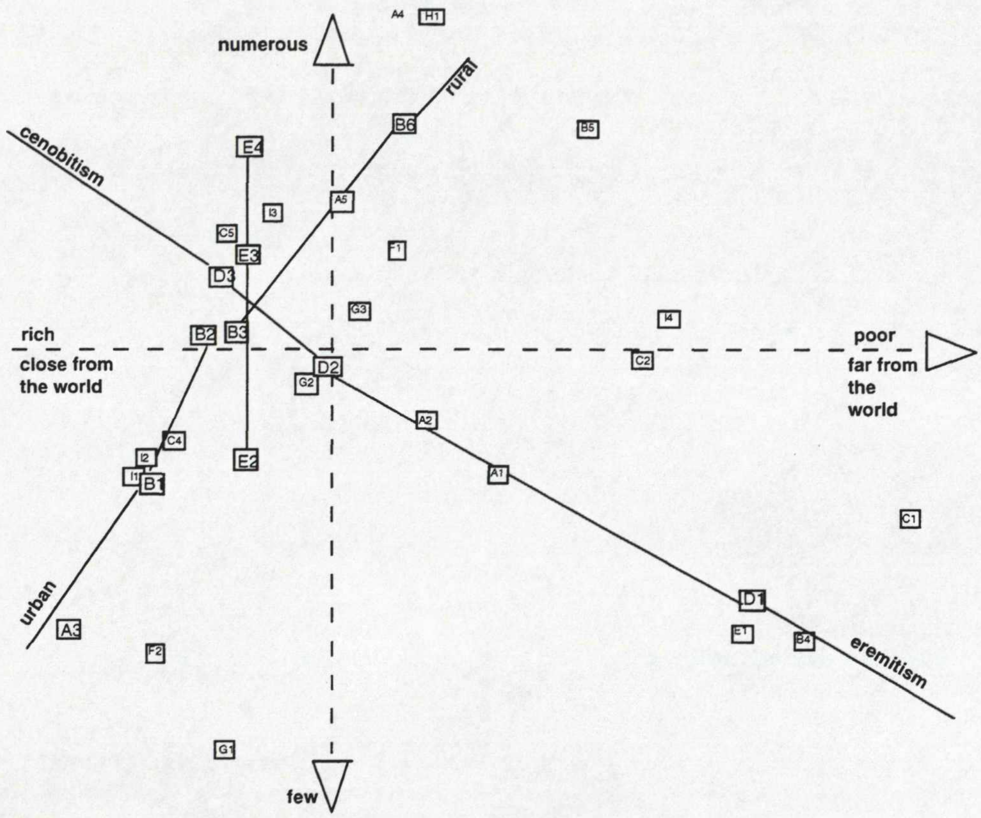

Figure 19.12: Results with even distribution of weights for criteria

At the same time, and this is not a coincidence, the results are also very interesting from an historical perspective. From such graphics, we can draw a few conclusions concerning some characteristics of ancient monasticism in Italy. For example, more women than men were living in communities of an urban type, which is not totally surprising. If we follow the chronology (G1 to G3), we can also note that the tendency is to create monastic centres outside cities, possibly in areas where it was easier to feed a growing community. None of these remarks are revolutionary. A careful analysis of the texts could provide similar ideas. The purpose of such a graphic, however, is to make visible and to confirm such ideas. It should work as an explicit representation of the monastic settlements and therefore, correspond to the results of a careful historical analysis. It can, also, help confirm hypothesis when the number of sites, or of objects studied, is very important. In our study of monasticism, it is particularly useful to identify tendencies and to compare well-known sites to the less successful settlements that disappeared either during this period or later in the Middle Ages. This type of method provides an opportunity to do so.

We believe that this method could be applied to a large number of problems in history and archaeology that deal with multi-criteria data and uncertain information.

Notes

[1] This study was first conducted in Paris, at the Ecole Normale Supérieure, under the supervision of François Djindjian in the seminar 'Informatique et Mathématiques appliquées en archéologie'. We wish to thank him for introducing us to that field of research. We also wish to thank the CIT at Princeton University for their help with SAT.

References

SAS INSTITUTE INC. 1989 *SAS/SAT User's guide.* Version 6. Fourth Edition, vol.1, Cary, NC : SAS Institute Inc., 615–675.

BENZECRI, J. P. 1973 *L'analyse des données. t.2: L'analyse des correspondances*, Paris.

GREENACRE, M. J. 1984 *Theory and Applications of Correspondence Analysis*, London.

HAYASHI, C. 1950 'On the Quantification of Qualitative Data from the Mathematico-Statistical Point of View', *Annals of the Institute of Statistical Mathematics*, 2 (1), p.35–47.

HORST, P. 1935 'Measuring Complex Attitudes', *Journal of Social Psychology*, 6, 369–374.

KOBAYASHI, R. 1981 *An Introduction to Quantification Theory*, Tokyo.

KOMAZAWA, T. 1982 *Quantification Theory and Data Processing*, Tokyo.

LEBART, L., MORINEAU, A. AND WARWICK, K. M. 1984 *Multivariate Descriptive StatisticalAnalysis: Correspondence Analysis and Related Techniques for Large Matrices*, New-York.

TEENHAUS, M. AND YOUNG, F. W. 1985 'An Analysis and Synthesis of Multiple Correspondence Analysis, Optimal Scaling, Dual Scaling, Homogeneity Analysis and Other Methods of Quantifying Categorical Multivariate Data', Psychometrika, 50, 1985, 91–99.