



*European Laboratory for  
Modelling of Landscapes and  
Territories over the Long Term*

<http://modelter.zrc-sazu.si>

# IFCN scholarship Activity report 2008

S. Poglajen\*

With the collaboration of L. Nuninger\*\* and E. Fovet\*\*

LEA ModelTER

\* Harpha sea d.o.o. (Koper, Slovenia)

\*\* Chrono-Environnement UMR 6249, MSH C.N. Ledoux USR 3124 (Besançon, France)



MSH C. N. Ledoux - Chrono-Environnement, UMR 6249 - ThéMA, UMR 6049 (Besançon-Dijon, France)  
Institut d'Etudes Anthropologiques et Spatiales - Institut d'Archéologie (Ljubljana, Slovénie)

Besançon, 14<sup>th</sup> May 2008

## **Introduction**

In my PhD thesis titled: *Geographical Information Systems in Roman Countryside Studies: Example of North-western Istria*, I examined several phenomena of Roman countryside settlement of the study area. One of the main topics was hierarchal settlement classification with statistical methods according to established criteria. This enabled distinction of results of further analysis according to different classes of settlements. This work is closely related to work group 2 of the ArchaeDyn project and it falls within the activities of the Franco-Slovenian European Associated Laboratory (LEA) ModeLTER (European Laboratory for Modelling of Landscapes and Territories over Long Term). One of the activities regarding the research of contexts is to establish tools to compare heterogeneous settlement data sets with those collected systematically. The criteria for classifying good quality data are already well defined through the work in Archaeomedes and ArchaeDyn projects. The problem remains the integration of poor quality data as is the data of north-western Istria.

The aim of my short research study in Besançon was to test if there is possibility to compare a good quality settlement distribution data with the pore quality data. The case studies were Languedoc<sup>1</sup> in France with good data and north-western Istria (Croatia, Italy and Slovenia) with bad quality data.

## **Data preparation and testing**

From large ArchaeDyn database the focus was on Languedoc area. Because in Istrian database there are just sites from 1<sup>st</sup> and 2<sup>nd</sup> century we extracted 225 Languedoc sites from the same period.

To get the feel of the data first the identical cluster analysis, as was done for the Istrian sites, was performed. In this test the following variables were used:

1. Size - Actual size of the site (Sup)
2. Mat. - Building material (Mat)
3. Occ. - Length of occupation (Occ)
4. Pre. - Length of previous or former occupation (Ant)
5. Fun - Function (Fon.)

Variables 2-5 are categorical except first variable (Size) which is in original units (hectares). Before statistical operations the variables were standardized so the variables are compatible. With Ward's hierarchical clustering data was checked to see how the data is structured and linked. The result is a tree diagram where we can observe aggregation from individual objects (sites) toward a single group or cluster (Fig. 1).

---

<sup>1</sup> Thanks to the Languedoc team C. Raynaud, F. Favory, L. Nuninger, E. Fovet, who let the data base available for this study.

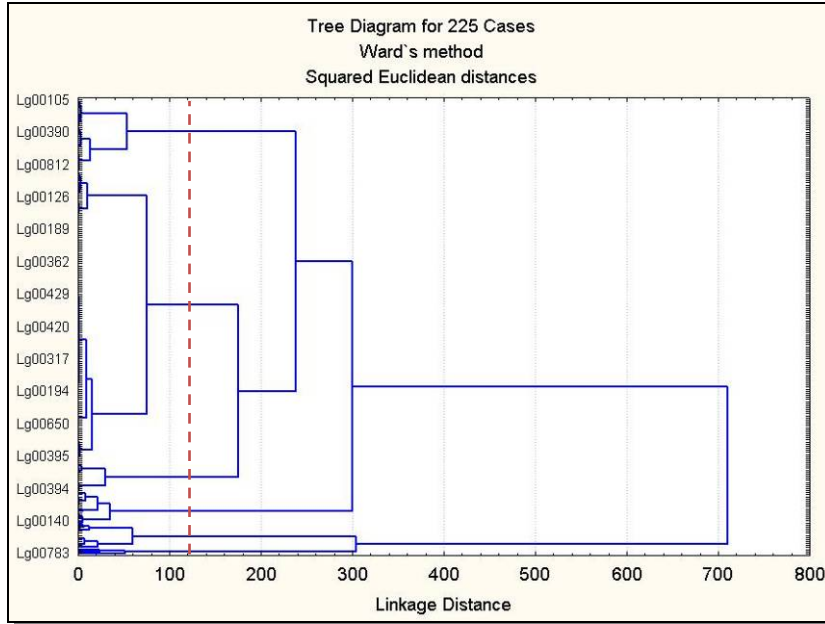


Figure 1: Tree diagram for Languedoc sites from 1<sup>st</sup> and 2<sup>nd</sup> century.

We can see that sites can be effectively divided into six clusters. With this in mind we proceeded to actual clustering with *k*-means method. The initial number of clusters (6) produces six clusters of settlements with more or less different characteristics (Fig. 2). The order of clusters is not representing any meaningful information. Individual groups must be examined to understand their characteristics and can be interpreted (Tab. 1).

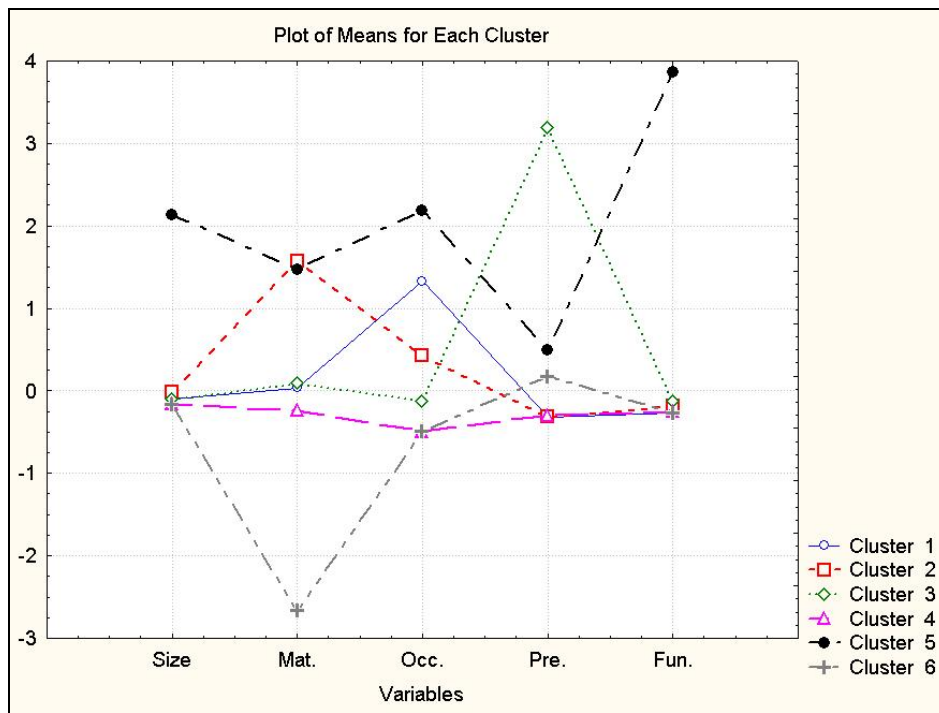


Figure 2: Graph of mean values for each cluster after *k*-means clustering.

Cluster	%	Description	Interpretation?
1	11	Settlements are rather small from less than 0,1 ha to 0,7 ha; majority (72%) of settlements are made of stone and tiles and/or sawed plate; they are very lasting (76% - 300 to 999 years) and without former occupation; all settlements have agricultural or no function.	Colonized settlements (farms?)
2	12	Relatively small settlements (67% - less than 0,3 ha) but with highly luxurious material (78% with decoration elements); relatively lasting (44% from 300 to 999 years) without former occupation; one settlement with craft activity others agricultural or no function.	Villas?
3	7	Small cluster with relatively small settlements (60% - less than 0,3 ha) in material similar to cluster 1; in general short lived (87% up to 2 cent.) but with high degree of former occupation (73% - 1 to 5 cent.); one settlement with craft activity others agricultural or no function.	Old settlements (farms?)
4	59	The biggest cluster of settlements that are of small size (62% under 0,1 ha) and are built with stone and tile and/or sawed plate (82%); not very lasting (76% less than 1 cent.) and with short former occupation (0 - 99 years).	Diverse countryside settlements (farms, auxiliary buildings etc.)
5	6	Large settlements (69% 1 to 5 ha) with highly luxurious architecture and very lasting (77% from 5 to 15 cent.); almost no former occupation (69%); all but one settlement with high degree of function (political and/or religious and/or symbolic).	big villa or vicus or town / or symbolic centre
6	5	Smallest cluster of settlements of small size (76% < 0,1 ha) and all with absence of building material; very short lived (75% < 100 years); almost no former occupation (75%) and no specific function.	No settlement or temporary agricultural building ( <i>tugurium</i> ) or...

Table 1: Description and possible interpretation of each cluster of settlements.

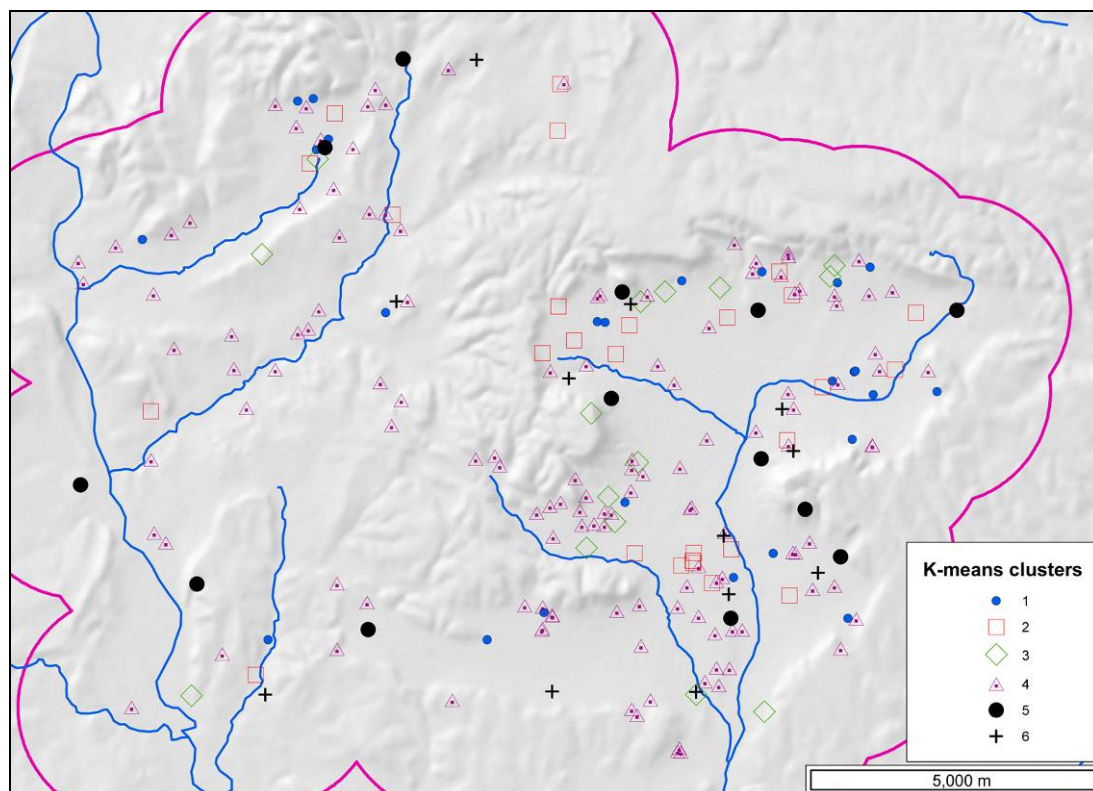


Figure 3: Spatial distribution of k-means clusters in Languedoc.

## **Generalization**

After first test of the data we proceeded with shaping the variables in such a way that we can best describe settlements from both study area (Languedoc and Istria). This meant that we had to generalize or downgrade the Languedoc data. The task was oriented toward distinguishing Roman settlements and also with data base from Istria in mind. Because we narrowed the time span of the data to two centuries and because the data from Istria is not well dated we at first rejected the two variables dealing with time (Occ. and Pre.). Other variables were reclassified or modified.

### **Surface size of the settlements (SizB)**

- 1 – Smallest sites < 0.1 ha (Sup01)
- 2 – Sites between 0.1 and 0.5 ha (Sup03, Sup05)
- 3 – Sites above 0.5 ha (Sup1, Sup2, Supm5, Supp5)

The first value is straightforward; it's intended for the smallest size settlements. The line between value 2 and 3 we set above 0.5 ha because from this size on we can expect to find already a small villa type settlements.

### **Building material (MatB)**

Here we distinguished just two values so that we can make same distinction in Istrian data base. First value represents sites without known building material and/or with basic building material (1) and the second sites with comfortable or luxurious building material and decoration (2).

- 1 – Absence or perishable material and/or basic roman building material (Mat12, Mat3, Mat4, Mat5, Mat6)
- 2 – Advanced or luxurious Roman building material (Mat7, Mat8)

### **Function**

The function variable in Archaedyn project was a combination of symbolic or ideological characteristics and economic or production ones (Gandini et al. 2007). Therefore we split this variable in two variables (symbolic and production) as was done in project Archaeomedes II (Durand-Dastès et al. 1998; van der Leeuw et al. 2003).

### **Symbolic (Symb = Sta.)**

We distinguished five levels of symbolic status characteristics based on proximity to graveyards, epigraphic monuments, sanctuary and existence of defensive structures.

- 1 – No data (Sta1)
- 2 – Graveyard or epigraphic site in surrounding area (Sta2)
- 3 – Previous + graveyard or epigraphic monument on site (Sta3)
- 4 – Previous + sanctuary in surrounding area (Sta4)
- 5 – Defensive structures (Sta5)

### **Production (Prod.)**

This variable was used to establish characteristics about production elements of the settlements. The settlement can be without production elements; they can have elements of agricultural or special craft activities; or they can combine both types of production activities. But because data from Languedoc had little variability I regrouped information into two classes. The first with no data or agricultural production activity and the second with special craft activity.

- 1 – No data or agricultural production elements
- 2 – Specialized craft activity (workshop for: metal, pottery, glass, stone, etc...) or port near by

### ***K-means cluster analyse***

First we applied the same cluster analysis procedure as described above with all the settlements, which resulted in 6 clusters of settlements from Languedoc and Istria. From the plot of standardized mean values we can observe that production variable (Prod) split the clusters exactly according to value (either 1 or 2). Five clusters with value 1 and one cluster of settlements with value 2. Similar is with material variable (MatB). Here there are three clusters of settlements with value 1, two clusters of settlements with value 2 and one cluster (6) where there are settlements with both values for material. The variables size (SizB) and symbolic (SymB) are spread across all six clusters (Fig. 4).

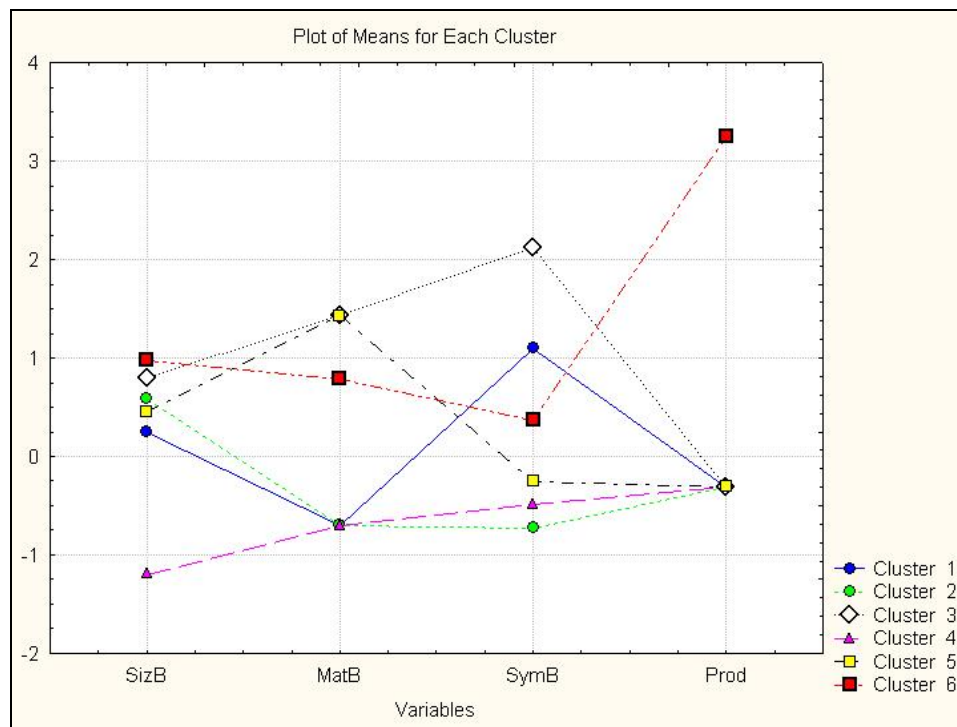


Figure 4: Graph of standardized mean values for each cluster after k-means clustering of all the settlements.

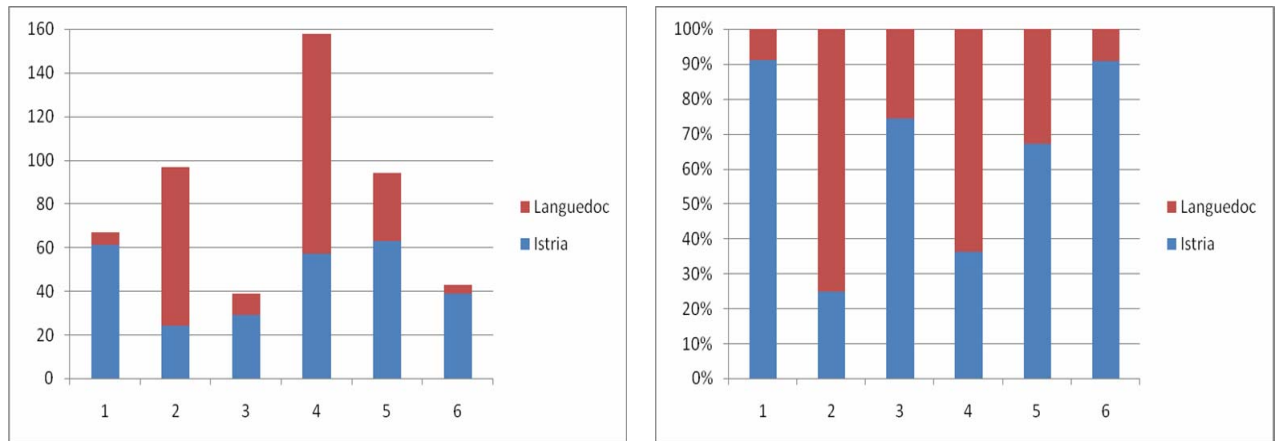


Figure 5: Size of each k-m cluster and distribution of settlements from Istria and Languedoc. On the left the graph shows number of settlements per cluster and on the right proportion of settlements from two regions in cluster.

k-m Cluster	#	Description	Interpretation?	Hier.
1	67	Small to medium size settlements with basic building material (all MatB1); quite high symbolic value (54% Symb3) and all have production 1	Small agricultural settlements (farms) with graveyards	3
2	97	Majority of settlements are of medium size (SizeB2) with basic building material (all MatB1); low symbolic value (all Symb1) and all have production 1	Diverse countryside settlements (farms etc.)	2
3	39	Medium to big size settlements with advanced building mat. (100% MatB2); high symbolic value (3 or 4) and all have production 1	Urban settlements or close to urban centres or close to sacral elements	6
4	158	Small settlements (all SizeB1) with basic building material (all MatB1); low symbolic value (79% with no data) and all have production 1	Small countryside settlements or agricultural buildings or...	1
5	94	Medium size settlements (80% SizeB2) with advanced building material (all MatB2) but low or medium symbolic value (57% Symb1) and all have production 1	Big farms or small villas?	4
6	43	Medium to big size settlements (54% SizeB2) with rich building material (70% MatB2); diversified symbolic value (58% Symb2) and all with production 2	big villa or vicus or town / or symbolic centre	5

Table 2: Description, possible interpretation and hierarchical level of each cluster of settlements.

In figure 5 we can observe that distribution of settlements over clusters is greater in Istria, where in Languedoc settlements are in majority part of three clusters (91% in 2, 4, and 5).

### Languedoc

Clusters 1 and 6 together have just ten settlements in Languedoc. Cluster 3 represents most of the previous cluster 5 (towns??). Settlements from previous cluster 4 are again in majority inside the new cluster 4. They are most numerous and they represent typical countryside

settlements which can be small farms or other type of settlements/buildings. The same goes for cluster 2. Cluster 5 contains most of the settlements from previous cluster 2 (farm/villa) (Fig. 6).

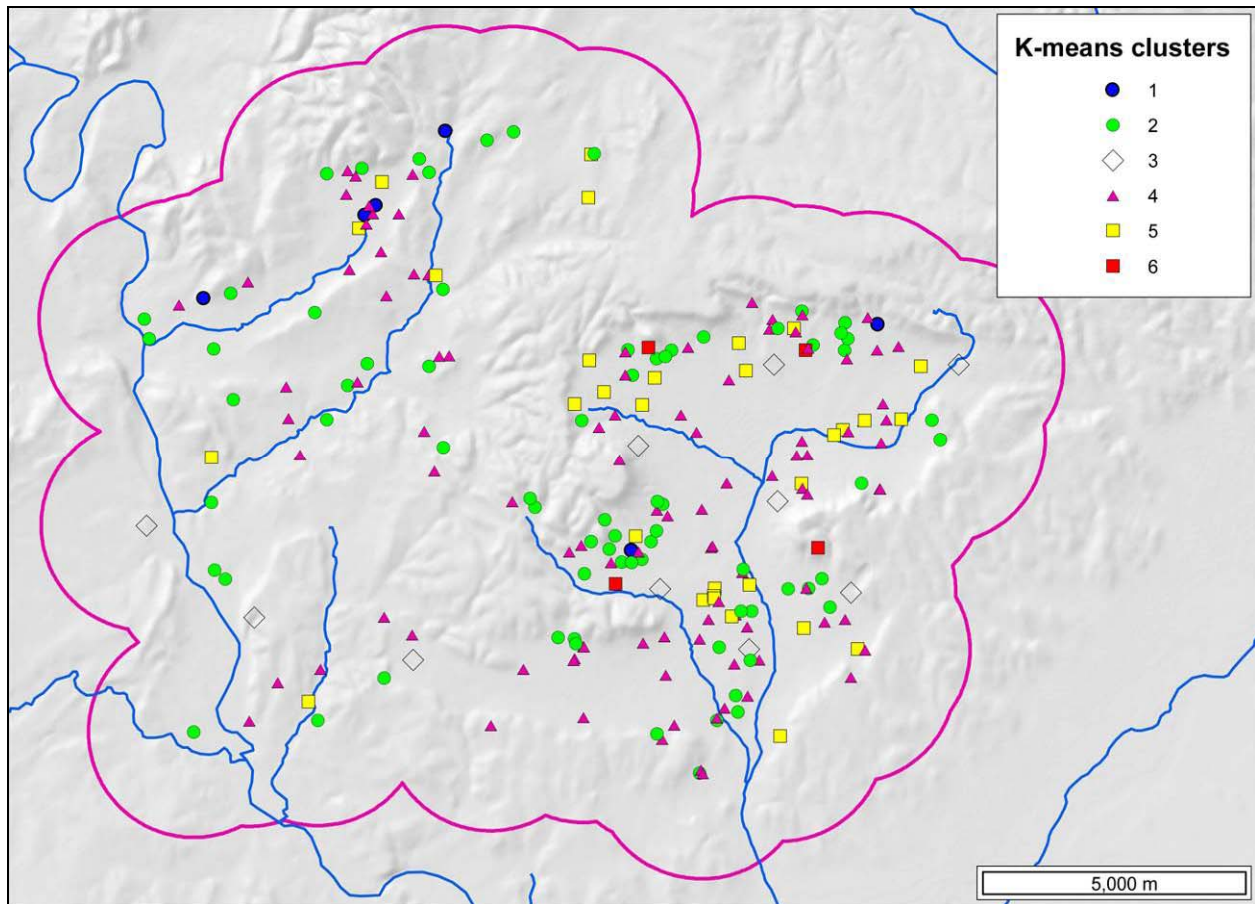


Figure 6: Spatial distribution of k-means clusters in Languedoc.

## Istria

In a general way in Istria there is better spread of the settlements over all six clusters. If we compare these clusters with original clusters from settlement analysis of north-western Istria (Poglajen 2007), we can observe a similar distribution. The first cluster that represents small settlements with high symbolic status is very similar to original Istrian cluster 6. The second cluster we can link with cluster 4 from original analysis and it is comparable in interpretation too. Similar is cluster 4. Cluster 3 is in most cases linked with original cluster 3, which represented suburban settlements with high material and symbolic status; and now also with cluster 6. Next cluster (5) is comparable with original cluster 2 which is interpreted as possible villa type settlements (Fig. 7); and finally cluster 6 is almost identical to original cluster 1 which with high probability represents true roman villas.

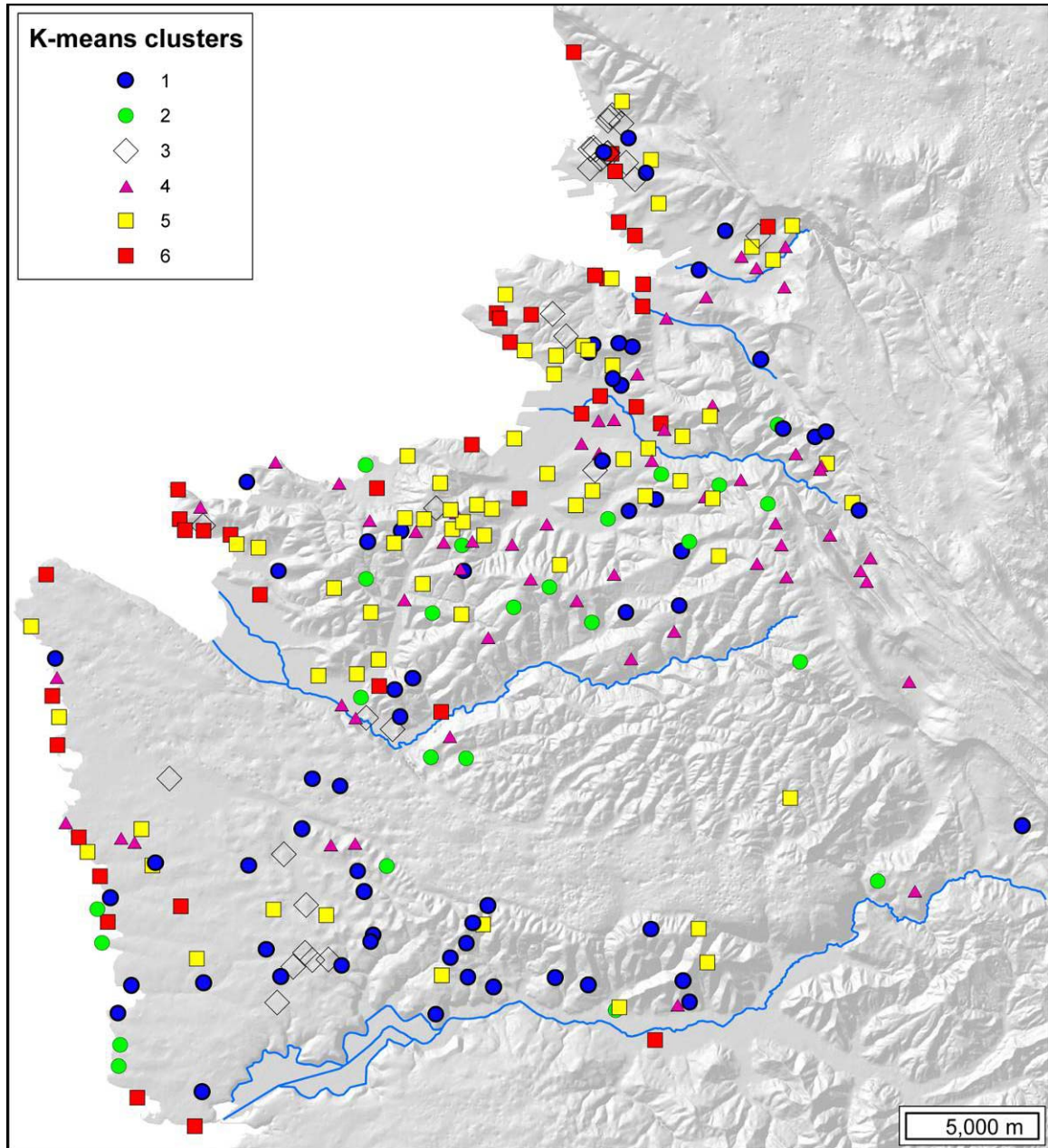


Figure 7: Spatial distribution of *k*-means clusters in Istria.

## AFC analysis

After k-means clustering we proceeded with factor analysis (AFC). Because of small representation of values 3, 4 and 5 in variable SymB, we regrouped this variable. In first analysis (AFC1) the values 4 and 5 were joined, in second analysis (AFC2) values from 3 to 5 were joined and in last analysis (AFC3) the values were in original form. After revising of the results we concluded that the better of the three is the second analysis (AFC2) that resulted in eight clusters of settlements.

Similar as with k-means clustering the spread of settlements over clusters is for Istria relatively even, only in cluster 8 there is just four settlements (1%). In Languedoc, on the contrary the settlements again group in three clusters (94% in 1, 3, 8) (Fig. 8). So it looks that in Languedoc the K-M cluster 4 was replaced with AFC2 cluster 1 and KMC 2 with AFC2 cluster 3.

AFC 2	Nb	KMC	Description	Interpretation?	Hier.
1	125	62% KMC4	Small settlements (64% SizB1) with basic building material (88% MatB1); without symbolic value (94% SymB1) and all have production 1.	Small agricultural settlements (farms) with graveyards	2
2	33	97% KMC4	Majority of settlements are small in size (97% SizeB1) with basic building material (all MatB1); modest symbolic value (97% SymB2) and all have production 1.	Small countryside settlements or agricultural buildings or...	1
3	140	37% KMC2 26% KMC5 25% KMC4	Small to medium size settlements (57% SizB2) with basic building mat. (69% MatB1); without symbolic value (89% SymB1) and almost all with agricultural production.	Farms or small villas	3
4	58	48% KMC1 48% KMC3	Medium size settlements (98% SizeB2) with all types of building material (52% MatB1, 48% MatB2); relatively high symbolic value (72% SymB3) and all have production 1.	Different settlements (suburban or countryside) with high symbolic value	4
5	29	86% KMC6	Medium size settlements (83% SizeB2) with basic to advanced building material (59% MatB2) with medium symbolic value (62% SymB2) and with craft production (86% Prod2).	Big villa or vicus	5
6	64	58% KMC5 39% KMC1	All settlements are of medium size with basic to advanced building material (58% MatB2); medium symbolic value (95% SymB2) and all with production 1.	Farms or small villas	6
7	18	78% KMC6	Big settlements (83% SizB3) with advanced building material (72% MatB2); medium symbolic value (50% SymB2) and with craft production (78% Prod2).	big villa or urban/symbolic centre	7
8	31	32% KMC2 32% KMC4	Settlements of diversified size with basic building material (71% MatB1); without symbolic value (84% SymB1) and all with production 1.	Diverse settlements	??

Table 3: Description, possible interpretation and hierarchical level of each cluster of settlements.

In hierarchical interpretation the first and second clusters are inverted, followed by next clusters (3, 4, 5, 6, and 7). Cluster 8 is hard to interpret; it's a kind of an odd cluster with different type of settlements mainly from Languedoc.

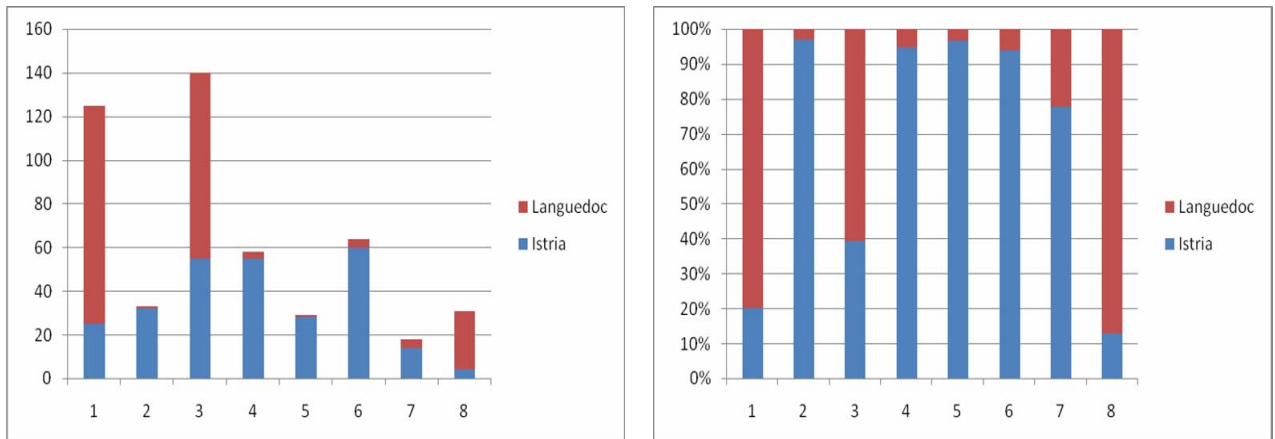


Figure 8: Size of each AFC2 cluster and distribution of settlements from Istria and Languedoc. On the left the graph shows number of settlements per cluster and on the right proportion of settlements from two regions in cluster.

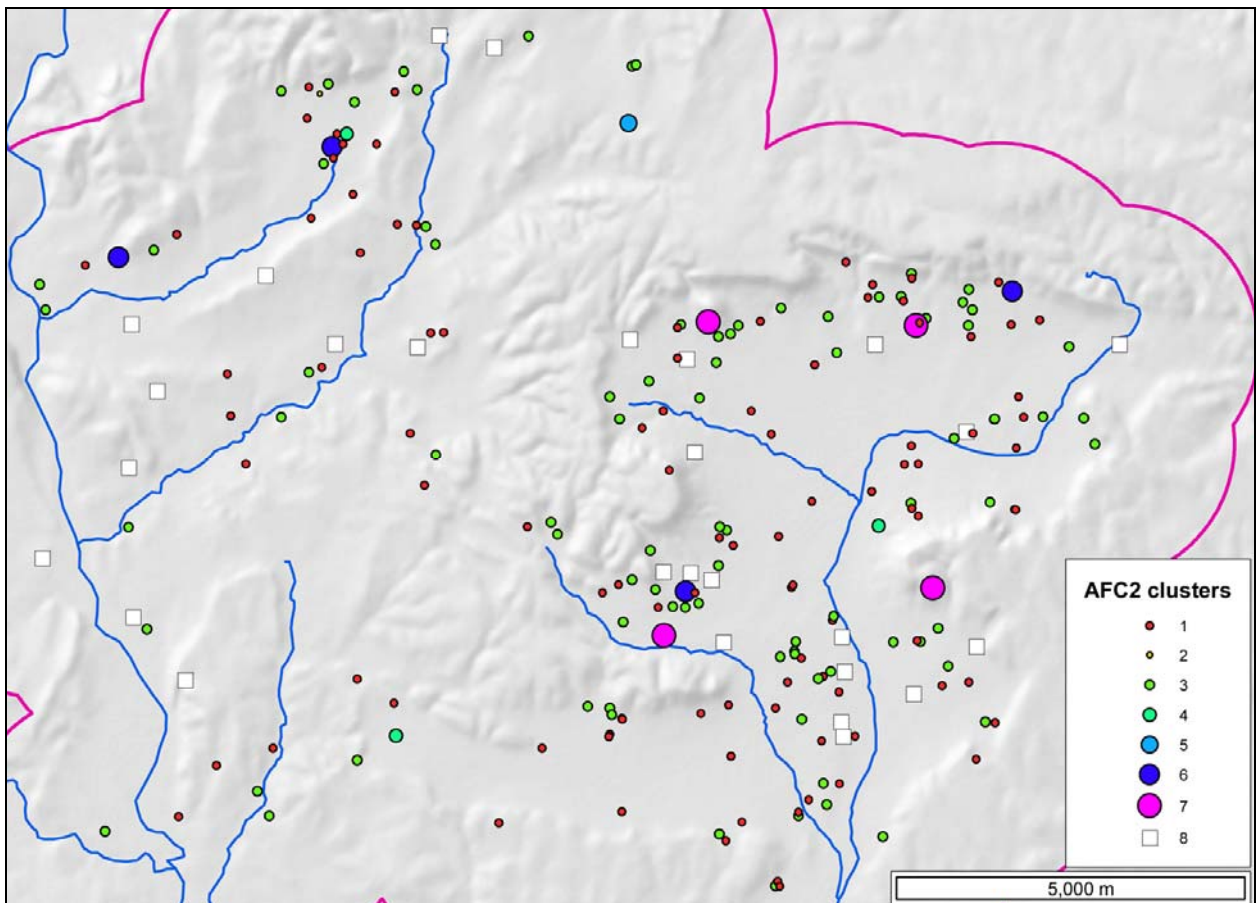


Figure 9: Spatial distribution of AFC2 clusters in Languedoc. The size of the circles corresponds to the interpretation of hierarchy.

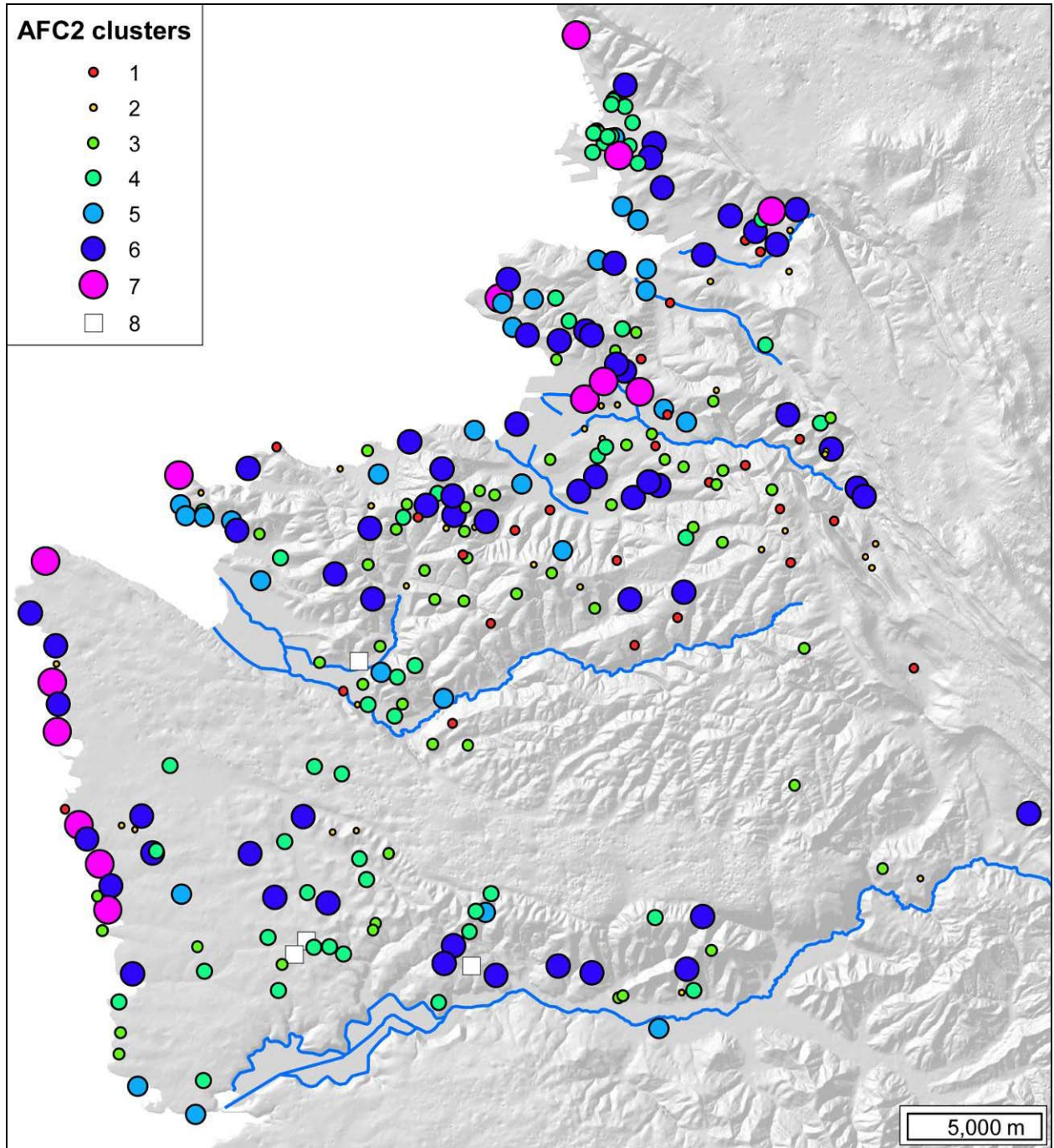


Figure 10: Spatial distribution of AFC2 clusters in Istria. The size of the circles corresponds to the interpretation of hierarchy.

## Conclusion

The comparison of settlement patterns over longer period and between different ecological backgrounds is in itself a challenging task. For proper evaluation the structure of the basic data, the level of quality and detail should be as equal as possible. In this test we made an attempt to compare the settlement data with very different quality and structural range. The first case study area was Languedoc, where data was gathered with systematic survey; the second case study was Istria where data was gathered mainly through bibliographical references and is thus very heterogeneous in quality and detail of information.

The approach was to define variables with general and qualitative classes that can be used to describe sites from both study areas. We ignored original Languedoc variables dealing with time (Occ., Ant) and we replaced function variable (Fon) with symbolic and production variable (SymB, Prod). Clustering analyses were made with two different approaches. First clustering was done with k-means (KMC) method and the second with factor analysis (AFC). The results of both analyses are more or less comparable with original clustering for each area. There is clear distinction of distribution of sites within clusters between Languedoc and Istria. In Languedoc sites in majority fall within 3 clusters but in Istria there are more evenly spread. The problem with AFC2 analysis is that there is maybe too many clusters where some are similar to each other and that makes it difficult to interpret them. This is especially true for odd cluster 8 which encompasses very diverse types of settlements.

We can say that Roman settlement pattern in Languedoc is distinct from Istria. In Languedoc there are a lot of small settlements on one side and real urban centres on the other side of hierarchical spectrum. But in north-western Istria it looks that there is better representation of Roman countryside settlement diversity.

Because poor quality data is usually just approximately dated it's probably not possible to look at settlement dynamics over long period. I think it's better to focus on single periods defined with the area and data in question. (E.g. prehistory, early Roman or late Roman Empire).

Finally it showed again that it's important to check and recheck the final results with original data to ensure that they correspond.

## References

- DURAND-DESTÈS, F., FAVORY, F., FICHES, J. L., MATHIAN, H., PUMAIN, D., RAYNAUD, C., SANDERS, L., van der LEEUW, S. (ur.) 1998, *Des oppida aux métropoles : Archéologues et géographes en vallée du Rhône*. Paris, Anthropos.
- GANDINI, C., BERTONCELLO, F., GAUTHIER, E., NUNINGER, L., TRÉMENT, F. 2007, "Hierarchical typology and settlement patterns modeling at inter-regional scale" – CAA conference, Berlin 2007, forthcoming.
- POGLAJEN, S. 2007, *Geographical Information Systems in Roman Countryside Studies: Example of North-western Istria*, University of Ljubljana, unpublished PhD thesis.
- Van der LEEUW, S., FAVORY, F., FICHES, J. L. (ed.) 2003, *Archéologie et systèmes socio-environnementaux: Études multiscalaires sur la vallée du Rhône dans le programme ARCHAOMEDES. CRA Monographies 27*, Paris, CNRS Editions.