

Methods for Regrouping Economic Activities into Meaningful Clusters

Sophie De Mulder, Inge Penninx, Geert Van Haute, Jan Zaman

(Sophie De Mulder, Environment Department Flanders, Koning AlbertII-laan 20, 1000 Brussels, sophie.demulder@vlaanderen.be)

(Inge Penninx, Environment Department of Flanders, Koning AlbertII-laan, 1000 Brussels, inge.penninx@vlaanderen.be)

(Geert Van Haute, Environment Department Flanders, Koning AlbertII-laan 20, 1000 Brussels, jan.zaman@vlaanderen.be)

(Jan Zaman, Environment Department Flanders, Koning AlbertII-laan 20, 1000 Brussels, geert.vanhaute@vlaanderen.be)

1 ABSTRACT

“The Flemish territory is characterized by a large urban sprawl [...]. Even last years, an additional 6 hectares of undeveloped space is being built on daily. As a consequence open space is highly fragmented in Flanders“ (Penninx, De Maeyer, Leroy, & De Mulder, 2021). As a strategic objective, the Flemish spatial government aims at a transition towards a net zero landtake daily by 2040. In this context, our spatial economy research group takes the choices and behaviour of individual companies and their use of space as a starting point. The main goal of the research is informing policy and supporting decision making by discerning spatial patterns, related to economic locations, and more precisely by focusing on the spatial environment of these locations.

Over the years, we developed a the business-oriented approach for local spatial-economic policy and location advice for companies (Giarretta, Zaman, Penninx, & De Mulder, 2019; Zaman, Penninx, & De Mulder, 2020). For this, we need the exact location of the activity and the exact activity of every economic site. However, this information is difficult to gather from the only area-wide economic administrative database for the whole territory of Flanders (VKBO) (Gruijthuijsen et al., 2018). This area-covering database is used for major spatial-economic analyses, but it falls short in precision at the detail level needed for our work. We have carried out quite a lot of research in recent years to get to know the terrain situation by creating a field inventory. A key element of the research is the search for the right spatial synthesis of the data collected at the level of the parcel: through economic ecotopes and market segments we sought to combine the (economic) parcels into meaningful groups with similar characteristics. We described this step in previous papers (Giarretta et al., 2019; Zaman et al., 2020).

Although the past research is interesting for the local policy makers of the mapped area's, we still need to find a way to also make meaningful statements on spatial economic patterns for other areas in Flanders that have not been mapped. Producing this area-covering map for Flanders is rather important, as it will enable us to translate the analyses and the knowlegde we have gathered to (regional) policy. Although being thorough and rather precise, the visual inventory method has some drawbacks: it is time consuming and at this point, it cannot be easily applied to the entire area of Flanders. We therefore opt to first assess if we can extract useful statements regarding economic patterns from administrative databases.

The main research question is whether the synthesis of the mapping data into the economic ecosystems or economic segments can be reproduced with the administrative database. Obviously, the results from the administrative database and the inventory will not be 100% alike. However, we believe it is possible come to spatial economic meaningful groups, even using the administrative database. The purpose of this grouping remains the same as with the inventory work and economic ecotopes and segments: being able to inform policy choices related to economic locations.

In a first step, we examined whether and how the area synthesis (starting from the inventory and resulting into economic ecotopes and segments), that was carried out with manual work, field knowledge and expert opinion can be reproduced through automated methods, specifically through (1) statistical approach and/or machine learning and (2) a spatial predefined spatial clustering. The automated grouping results are reviewed and spatially analysed by spatial planners with territory knowledge. Only in a second step, when the grouping results on basis of the inventory are satisfying, we will rerun the method with the administrative data of the VKBO. In this paper we will discuss the first few steps of the grouping methods, in particular the distance and the activities clustering. We will outline the next steps, using the VKBO-data, assessing if we can come to meaningful economic clusters.

Keywords: machine learning, clustering, Spatial economic patterns, Business Perspective, Flanders

2 RESEARCH ISSUE/INTRODUCTION

In a more general way, we aim at informing policy choices regarding economic locations and the spatial environment of these locations. We therefore chose to investigate the microscale first, in order to then arrive at the mesoscale. Next, we are looking for ways to extend the mesoscale to similar areas in Flanders. To this end, we have two main sources of information: a visual field inventory, covering 3% of the entire Flemish area, and an economic administrative database that covers the entire Flemish area.

The first steps of micro- and mesoscale has been to focus on the organisation of economic activities and spatial patterns in Flanders. One part is the categorisation of economic units and parcels, that were observed in the field, into economic ecotopes and market segments of economic spaces. This categorisation is meant to advise companies and policies of the right location for economic activities: is the type of economic ecotope the one that is best suited to the companies' needs, located in that economic ecotope? Which territorial policies can close the gap between the companies' needs and current economic ecotope? (Zaman et al., 2020)

As the current administrative database for the registration of economic activities (VKBO), was mainly designed for legal purposes (and not spacial identification purposes), the exact location of the activity and the exact activity (many businesses register as many activities as possible) is difficult to extract from the database. A visual check between the exact registered location in VKBO and the reality on the field shows us a discrepancy, that varies according the type of environment (Gruijthuijsen et al., 2018). Therefore, we have been visually recording economic activities on the field since 2017, trying to understand the spatial-economic patterns (for the method see (Giaretta, Penninx, De Mulder, & Zaman, 2018; Giaretta & Zaman, 2017; Giaretta et al., 2019; Zaman et al., 2020). In november 2019, we had mapped an area of almost 38.000 ha in Flanders and Brussels including almost 45.000 economic activities on 34.000 parcels. This led us to the creation of 16 types of economic ecotopes (Zaman et al., 2020). These are economic environments, taking into account the mix of economic activity and the distance between economic activities.

Although the past research is interesting for the local policy makers of the mapped area's, we still need to find a way to also make meaningful statements for other areas in Flanders that have not been mapped. A considerable area was inventoried with a big difference in types of spaces. However, we want to be able to extrapolate the findings to the whole Flemish territory. Producing this area-covering map for Flanders is rather important, as it will enable us to translate the analyses and the knowlegde we have gathered to (regional) policy. Last, but not least, as researchers in the Flemish government we want to generate research that is consistent for the whole territory. For the time being, there are two possibilities to make this switch to the Flemish scale level: an area-wide inventory of the economic use of the parcels or resorting to the administrative data sources. Although being thorough and rather precise, the visual inventory method has some drawbacks: it is time consuming and should ideally be done once a year to capture changes, as companies come and go, move to another location, expand (or decrease) their activities and so on. The first field work dates from 2016 and it is very likely that it is outdated for some economic locations. At this point, a field inventory of the entire Flemish area cannot be easily applied: according to the distribution of competences of the different administrations, the source data of economic information for the entire Flemish territory fall under economic administrations, not under those of spatial planning and environment. Before resorting to a field inventory of the entire Flemish area, we want to asses if we can get any useful results from the administrative database: what type of statements regarding the spatial economic patterns can we extract?

The aim is to create spacial meaningful groups of economic activities, that can be reproduced later on, with each new release of the administrative databases. Of course, the results of this automated grouping will not be exactly equal to the 16 types of economic ecotopes, as the visual and automatic methods are too different. However, both start from the visual inventory of economic activities.

3 THE SEMI-AUTOMATED APPROACH TO OVERCOME AN EXPERT-BIAS

In the visual construction of the area types (starting from the inventory of economic parcels into 16 types of economic area's/ecotopes), the researcher used a semi-automatic approach (Giaretta & Zaman, 2017). The previous attempts, made by experts to group the economic parcels, were based on a purely visual interpretation and revealed a considerable expert bias (Gruijthuijsen et al., 2018). A first pitfall was splitting

the mapped territory into traditional spatial entities such as “industrial park”, “city centre”, “main road”,..., whereas the mapping of economic activities clearly did not stop at the pre-defined borders. The inventory showed for instance quite an amount of productive activities in the mixed fabric of certain areas, which went against the expectations of experts, who tend to see mixed environments only suitable for retail and offices, and industrial parks as the exclusive location for productive and wholesale activities. As a consequence, productive and wholesale activities were not taken into account or acknowledged in mixed environments, according to the expert-based economic area types. In order to overcome these biases, we objectified the occurrence of economic activities into groups, and later interpreted them again.

This method consists of three main steps. First of all, the parcels with economic activities are divided into classes based on distance from each other (0m, 20m, 50m, 90 m) (ArcGIS near tool). Each parcel with an economic use gets an additional attribute field containing the distance to the nearest other parcel with an economic use. This produces a map of parcels with an economic use, grouped by distance. The second step of the operation was to use this map as an indication for drawing polygons based on the key patterns, e.g. distance, density, local field characteristics, grouping the different polygons into three new types. A first group, concerned the merging the neighbouring polygons with a continuous occurrence of economic activities. The second new group consisted of the nearby polygons with less continuous economic activities and even including small clusters of continuous economic activities. The third new area category regrouped the rest of the mapped area, in which economic activities are spread out and fragmented. It is important to note that in this fase of manual drawing, the researcher —who was a spatial expert— applied his knowledge of physical spatial boundaries such as recognising main streets and entrances to back streets, streets with different width, zones for loading and unloading, etc... or spatial appearance such as building typology, presence of greenery or pavements, etc. In the third and last step, in order to define homogeneous areas of similar economic composition, these areas were again subdivided using the presence of different economic activities. A continuous economic area (e.g. along a main road) where one section is mainly retail, and another section has also productive activities would be split in two, to create two distinct, more homogenous types. In this step, the parcel map was initially interpreted visually. In our 2020 paper we explored the types more in depth (Zaman et al., 2020). Again, terrain characteristics were taken into account by the researcher-expert.

The method resulted into a classification of 16 types of economic ecotopes) (Zaman et al., 2020). These are economic environments, taking into account the mix of economic activities and the distance between them.

4 AUTOMATED CLUSTERING (USING THE INVENTORY OF PARCELS WITH ECONOMIC USE)

Although much of the additional field characteristics and implicit knowlegde used by the planner in the semi-automated method could in theory be available digitally, these data are difficult to identify, interpret and to collect for the research areas and the entire Flemish territory. The automation must therefore replace the expert knowledge with something else that also results in a usable spatial economic classification.

Taking aside the extra information the planner used in order to define the final economic types, the semi-automated method consisted of two ‘objectifiable’ steps: firstly, clustering the parcels with economic activities into categories based on distance between the parcels with economic use and secondly, when this distance clustering gives a stable and accountable result, analysing what types of activities are collected in the clusters. Finally, this leads to a new homogenous groups, that are relevant from a spatial economic point of view.

The goal of the economic classification is a map, showing spatial economic relevant results, that covers the entire Flemish territory. Eventually it will be based on data available in administrative (economic) datasets (VKBO). We therefore first tested several methods on distance and activities clustering within the mapped area that contains the data of the terrain inventory. Next, we tried to reproduce the methods with the administrative datasets. This will be discussed in point 5. Every automated step was reviewed by spatial planners with terrain knowledge.

4.1 Distance clustering

We used several methods in order to reproduce these two steps. For the distance clustering we used four different methods: (1) DBSCAN on distance between centroids of parcels, (2) ST_ClusterWithin (3) DBSCAN on GPS routing distance between access points of parcels and (4) Predefined building blocks.

The DBSCAN algorithm is an intuitively quite easy to understand clustering technique. Starting from a random data point, we find out which other data points are within a defined distance (Method 1: 20m, 50, 100, 500, 5000 m and Method 3: ranging from 25 to 250m walking distance between the access points in steps of 25m). For each new data point we then check which other data points are not yet assigned to the cluster and are also within that distance. We repeat this process until there are no more data points in that cluster. Although the method is interesting because it clusters only data points, regardless of the context (cf. biased view of experts in section 3), and because it clusters on both sides of a same road, the randomness of the method has its drawbacks. The algorithm uses both the distance and a minimum number of points to define the core of a cluster. Points that are only reachable from one core point (and not meeting the minimum number of points critirium), are border points. In our data, there are many border points that can be part of different cores, but in reality belong together in one cluster. (Schubert et al., 2017) The DBSCAN clustering seems to jump sometimes to other points, and to forget the continuity of the cluster in other directions. This 'border' condition results in some 'missed' points, that end up in other clusters (See Figure 1). Another problem with DBSCAN is the size of certain parcels. E.g. when we set DBSCAN variable to 25m, as we want to cluster the parcels where the walking distance between access points is 25m or less, neighbouring parcels end up in a different cluster just because they are larger than 25m. However, using a larger walking distance between the access points, such as 100m, tends to smoothen out all differences: everything is put together in one big cluster. Interesting differentiation is then lost (Figure 1, left image). Variants of DBSCAN such as HDBSCAN* work hierarchical and deterministic (McInnes et al., 2017), and will overcome the randomness of the border points, but it remains uncertain whether it will result in accountable spatial clusters.

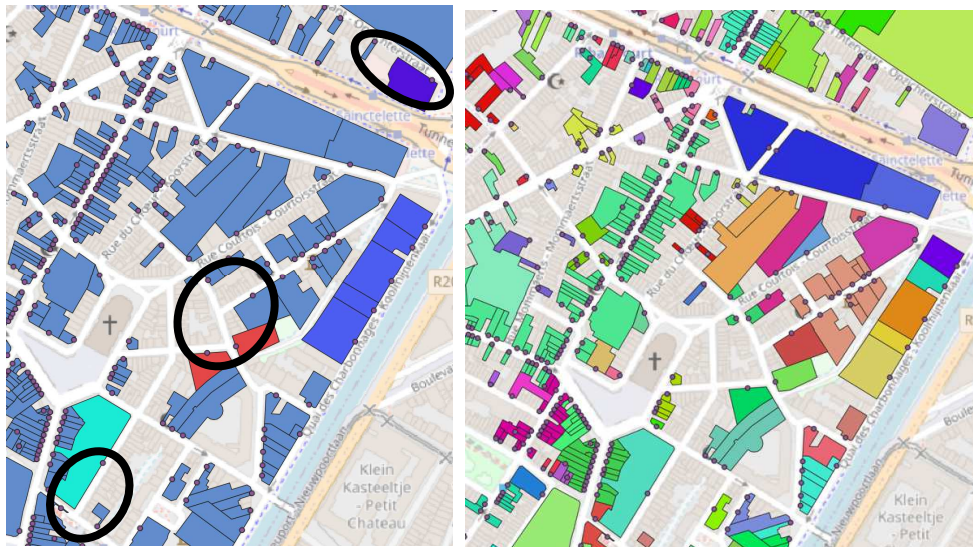


Figure 1 : Distance clusters with DSCAN at 100m (access points) in Brussels centre. Left: some accesses are closer to each other than 100m and yet, are not taken in the same cluster. Right: Distance clusters with DSCAN at 25m (access points): because the parcels are too large, they end up in different clusters, although they are neighbouring. (Background layer: Open Street Map)

The difficulty of dealing with large parcels is mainly due to the reduction of a polygon to a point (centroid or access point). In PostGIS the ST_ClusterWithin function is an aggregate function that returns an array of geometry collections. Each geometry collection represents a set of geometries that are separated by no more than the specified distance (in our case 0,1m, 20m, 90m, 500m). Although this method actually clusters parcels according to a certain distance without randomness, the results show that there are a lot of small clusters, that form a spatial economic point of view should belong together. Next to this, back and front streets are not differentiated.

The fourth method consists of predefining building blocks according to their distance to certain types of roads in OSM¹ (Pedestrian = 25m, Tertiary = 50m in Flanders and 25m in Brussels, Secondary = 70m in Flanders and 50m for Brussels area, Primary = 70m). Only parcels that are within the settlement area are taken into account (Poelmans, Janssen, & Hambsch, 2021), and are categorised according to their distance to pedestrian, tertiary, secondary and primary road.



Figure 2: Distance clusters with ST_Cluster_Within at 0,1m (full colour polygons), 20m (striped polygons) and 100m (dotted polygons). Left: Inner city with primary and secondary shopping street in Roeselare, Flanders: At 0,1m distance, it is logical that roads slice groups that should be together. However, from an spatial economic perspective this does not make sense. Right: A main shopping street in Brussels outer city centre is sliced into several groups. (Background layer: Open Street Map)

After this, parcels are dissolved into building blocks or parts of building blocks. Figure 3 (left) shows the Flemish city of Hasselt that is characterised by two circumferential roads. The bigger circumferential one is the main ring road, a primary road, functioning on a regional level (turquoise colour). The smaller one is a secondary road, connecting the regional and local function (blue colour). It surrounds the medieval city centre, that is now mainly a pedestrian area (pink colour). The two ring roads are linked by several secondary roads. The other building blocks that are not within a certain distance from pedestrian, tertiary, secondary and primary roads are the „rest“ category (green colour). The roads around these building blocks have mainly a local function, e.g. connecting residential area's or some parts of business/industrial parks.

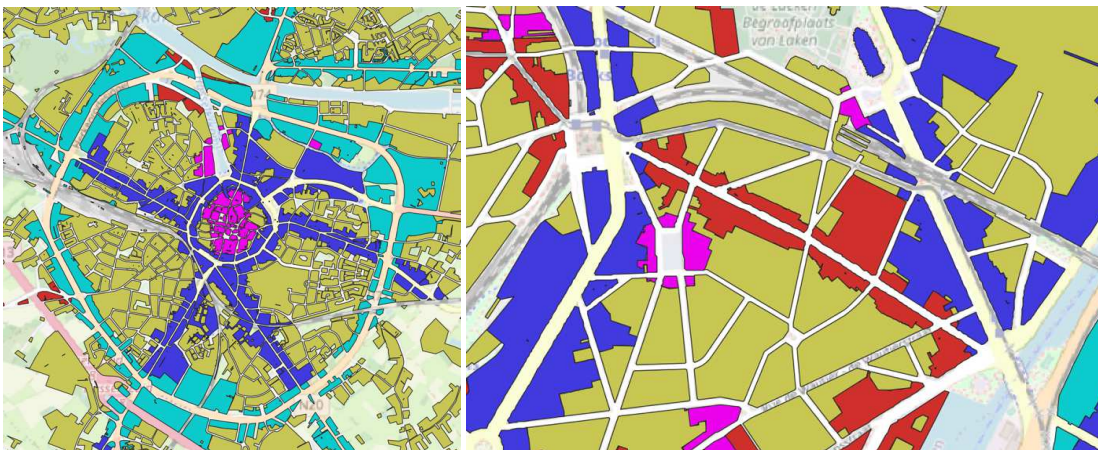


Figure 3: Distance clusters according to predefined building blocks linked to OSM types of roads. Left: the Flemish city of Hasselt and surroundings show that a categorised road network is clearly visible. Right: in the building blocks in the main shopping street in Brussels outer city centre, we distinguish now clearly the main shopping street, as well as the slicing of building blocks, according to back and front streets. (Background layer: Open Street Map)

On the right part of Figure 3, the main shopping street outside the city centre of Brussels, is now clearly defined around a tertiary road (red colour). Because of the predefined distances, the building blocks are as well cut into two parts: a front part, adjacent to the relevant road type, and a back part, that does not have the

¹ Open Street Map in Belgium categorises roads into Pedestrian, Motorway, Trunk, Primary, Secondary, Tertiary, Residential, Unclassified, Track, Path, Footway. For our project we assessed that the types Pedestrian, Primary, Secondary and Tertiary were the best suited, as they relate economic locations to accessibility.

same access and visibility as the front part. The advantage is that we have a spatial clustering that is independent of the presence of economic activities, that has additional attributes linking it to a street name or a street type, and is small. The small size of each cluster combined with the street attributes can allow to merge clusters with the same characteristics in a later stage. It can also prove to be very useful when we need to add other data (on housing, green, inhabitants,...) to get a better understanding of the differences between types for sit location choice.

4.2 Forming (new) groups through clustering activities

After having a distance clustering, the second general step is to regroup the distance clusters into new groups based on activity mix. The activity combinations in these groups should be internally more or less homogenous, but comparing the groups between each other, they should show enough difference. We used again three methods: (1) Factor analysis (FA), (2) Principal Component Analysis (PCA) and (3) hierarchical manually determined thresholds. A FA is an exploratory data analysis method used to find influential underlying factors or latent variables based on a set of observed variables. It assists in data interpretation by reducing the number of variables. It extracts the maximum common variance from all variables and places them in a common score. The FA was done in promax rotation and 16 factors could be determined. Although quite similar to FA, the PCA is a more robust method and is for instance used in Numerical Ecology (Legendre & Legendre, 2013). PCA is common in exploratory data analysis and predictive modelling. It usually helps reducing dimensionality by projecting each data point only on the first few principal components to obtain low-dimensional data, while preserving as much of the variation of the data as possible. The PCA revealed that certain activities such as Health Care; Education; Vehicle cars/trucks; Arts, Culture, Leisure and Sports; and Vacant are clearly distinguishable from each other. The PCA resulted in 9 groups, where for example Vehicle related activities are mainly in factor 9 (F9), Arts, Culture, Leisure and Sports in F7, and so on.

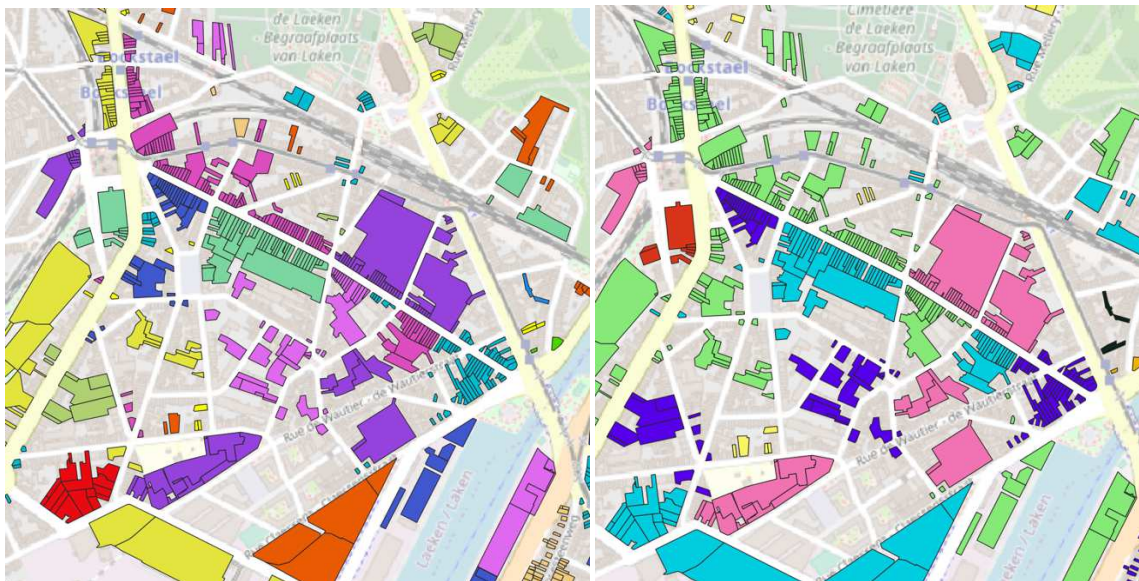


Figure 4: A main shopping street in Brussels' outer city centre. Right: Activities clustering with FA – 16 factors. Each colour is another factor. Left: Activities clustering with PCA in 9 factors. The results of both methods are, when compared to terrain knowledge, quite difficult to interpret. (Background layer: Open Street Map)

However, the new formed groups provided by the two activities clustering methods, FA and PCA proved when reviewing and comparing them with in depth terrain knowledge, to be quite difficult to interpret and explain. On one hand, it is not always clear why a type of activity ends up in a certain factor/group, meaning that the differentiation between factors is not clear. On the other hand, there is not always enough differentiation. For example in the PCA, factor 1 (F1) contains more than 50% of all the economic parcels, and is the most mixed factor. It is not clear why a group of parcels belongs to F1, instead of a more „specialised“ factor. It contains for instance Vehicle related activities, whereas these activities could also be in F9. In the PCA, activities related to Arts, Culture, Leisure and Sports (ASC) should be grouped in the more „specialised“ F7. However, Figure 5 shows this activity is as well in F6. It would make more sense to

have these two groups of F6 in either F7 because of the ASC, or in the same factor 1 as the rest of the street, which is the factor with most mixed activities.

The third method of activities clustering has a less statistical background, and is based on defining certain thresholds within the building blocks, in order to come to a spatial economic relevant result. We used the economic typology maps as touchstone and guide.

The categorisation of the building blocks is based on three aspects: (1) using a count of the total parcels in the building block, the proportion of parcels with economic use versus the parcels without economic use, (2) the average distance between the parcels with economic use and (3) using a count of the parcels with economic use, the proportion of the type of economic activities (e.g. services, industry & production, restaurants & bars,...). The first two aspects are quite similar to the semi-automated approach where the parcels with economic use were divided into continuous, close, discontinuous and solitary categories. As we start from the fourth method for distance clustering, there are no economic data connected to the clusters. Each distance cluster is joined to the data of the individual parcels with economic uses (type of activity, nearest parcel with economic activity, distance). These are used to calculate density and intensity of economic activities in each cluster. Thresholds are manually set to resemble the results from the semi-automated approach. From the count of economic activities and the proportion of each category of economic activities, the type is further developed with conditional statements. When a certain condition is met (e.g. more than 50% retail) the type is defined regardless of the other values, if not the next condition will be checked. The resulting map of types of economic ecotopes resembles the manually produced map, but in this case based on a SQL statement.

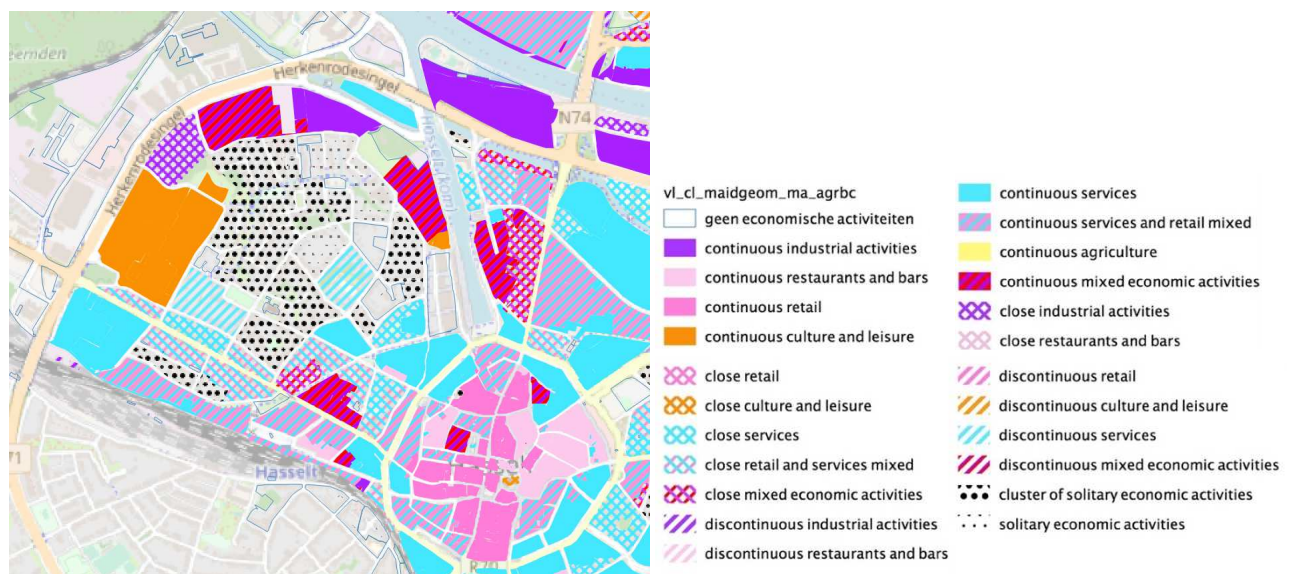


Figure 5: Activities clustering within the predefined building blocks, in the Flemish city of Hasselt. (Background layer: Open Street Map)

Figure 5 shows the results of the clustering in the Flemish city of Hasselt. The economic dense inner medieval centre contains a lot of continuous retail, restaurants and bars, with closer to the small inner circumferential road more continuous services. There are more mixed economy and industrial activities in the vicinity of the bigger circumferential road, which seems logical as there are a couple of regional industrial parks. The secondary road connecting the inner to the outer ring road structures some continuous and close economic activities. The more residential area's accommodate more isolated economic activities. Overall, this description corresponds well to the observations we made in the field. However, the method has as well some disadvantages: (1) it is not certain the results can make sense from a statistical point of view; (2) the thresholds are defined in a way to show relevant differences, based on an expert opinion, and can be biased; (3) it might be that the results cannot be reproduced with machine learning or other forms of artificial intelligence, making it difficult to obtain an extrapolated map for the whole territory of Flanders.

Methods		Distance clustering	Activities clustering	Results +	Results -	How to improve?
		First clustering	Second, new clustering			
1	DBSCAN + Factor analysis	DBSCAN: parcels with economic use clustered as points (20m, 50, 100, 500, 5000 m)	Factor analysis: 2 groups	Distance clustering: unbiased look in clustering method	Distance clustering: Because of the random function in DBSCAN, some points are overlooked (Fehler! Verweisquelle konnte nicht gefunden werden.) Activities clustering: Only 2 factor gives not much differentiation.	Trying another algorithm for distance clustering
2a	Cluster-Within + Factor analysis and PCA	Cluster Within: Parcels with economic use grouped as polygons (0,1m, 20m, 90m, 500m)	Factor analysis: 16 groups	Distance clustering: ClusterWithin groups economic parcels in more are less meaningful groups, based on distance.	Distance clustering: Some groups should be together. Width of (broad) roads slices clusters into smaller groups (Fehler! Verweisquelle konnte nicht gefunden werden.) Activities clustering: -Not all categories seem to be spatially relevant	Trying another distance clustering in order to capture different groups in one group, so that both sides on the same street are in one group
2b			PCA: 9 groups	Activities clustering: PCA method is more robust, used in Numerical Ecology (Legendre & Legendre, 2013)	Activities clustering: -Not all categories seem to be spatially relevant -Clusters are either too big or too small. The size of the parcels matters (Fehler! Verweisquelle konnte nicht gefunden werden.)	
3	DBSCAN	DBSCAN: parcels with economic use clustered as main access points /front entrances (distances ranges from 25 to 250m)	N/A	Distance clustering: Overbridging the distance of small roads, so that the two sides of the roads are in one group is successful	Distance clustering: Because of the random function in DBScan, distance points are overlooked (Fehler! Verweisquelle konnte nicht gefunden werden.) -Defining main entrances difficult at corner parcels and asymmetrical polygons -Clusters are either too big or too small. The size of the parcels matter (Fehler! Verweisquelle konnte nicht gefunden werden.)	Trying another algorithm for distance clustering
4	Predefined building blocks	Based on predefined distance between building blocks (includes all parcels of the settlement area, not just economic use parcels) and the roads. (25m, 50m, 50m, 70m)	Activities added to building blocks. Used of predefined thresholds to categorise the blocks into 23 categories	Distance clustering: -Differentiation is successful, regardless if the parcels are big or small (Fehler! Verweisquelle konnte nicht gefunden werden.) -Type of road is a defining factor for economic activities (e.g. one end of building block is at a main road; other end at a back road) Activities clustering: -Monitoring changes in building blocks is easier, as they stay always the same -All parcels in settlement area are included, which makes it easier to add more characteristics that are non-economic	-It is not sure the results make statistically sense -hierarchical system, we might be biased (e.g. defining thresholds)	Do some statistical analysis

Table 1: Summary of the used methods to make distance clusters and then regroup them into activities clusters

5 FIRST RESULTS OF USING THE SEVERAL METHODS WITH ADMINISTRATIVE DATABASES

As mentioned before, the administrative database (VKBO) that contains information about the location and the type of economic activities shows discrepancies with the data collected on the field (Grujthuijsen et al., 2018). Next to this, each record in the VKBO contains a lot of NACE-codes. Choosing one activity, is not an

easy task. Comparing the administrative database with the field observations, there are a lot more records in the VKBO. On average, the number of economic locations visible on the field accounts for only 48% of the number of businesses registered in the VKBO data. In certain residential areas there could be three times as much activities in the administrative database. There is definitely some extra research to do in order to determine for which categories of economic activity the discrepancy is large and for which the information from VKBO does give a good picture of the actual activity.

Nonetheless, some first tests with the VKBO, using the previous distance and activities methods, have been done, trying to replicate: (1) the PCA results via machine learning (ML) algorithms and (2) the predefined building blocks. The test with ML consisted of two steps. First of all, the results of the PCA within the mapped area of the inventory of the parcels with economic use are scored on accuracy according to ML algorithms K-Neighbours Classifier and Random Forest Classifier. This resulted into high accuracies varying between 0,93 and 0,95. The next step is to label the geographically obtained distance clusters with the activities (NACE-codes) on the parcels as available in the VKBO. Afterwards, a model is again trained through ML algorithms to find, this time on the basis of the NACE activities, the obtained activities clusters. This approach gave results around 0,59 and 0,57 accuracy.

The last test consisted of using all available NACE-codes and linking them to the predefined buildings blocks of the distance clustering. New thresholds had to be developed, as the non-visible businesses are also included in the data. Considering the dominance of service activities, most of the parameters were set rather low. Figure 6 illustrates the first results for the Flemish city of Hasselt.

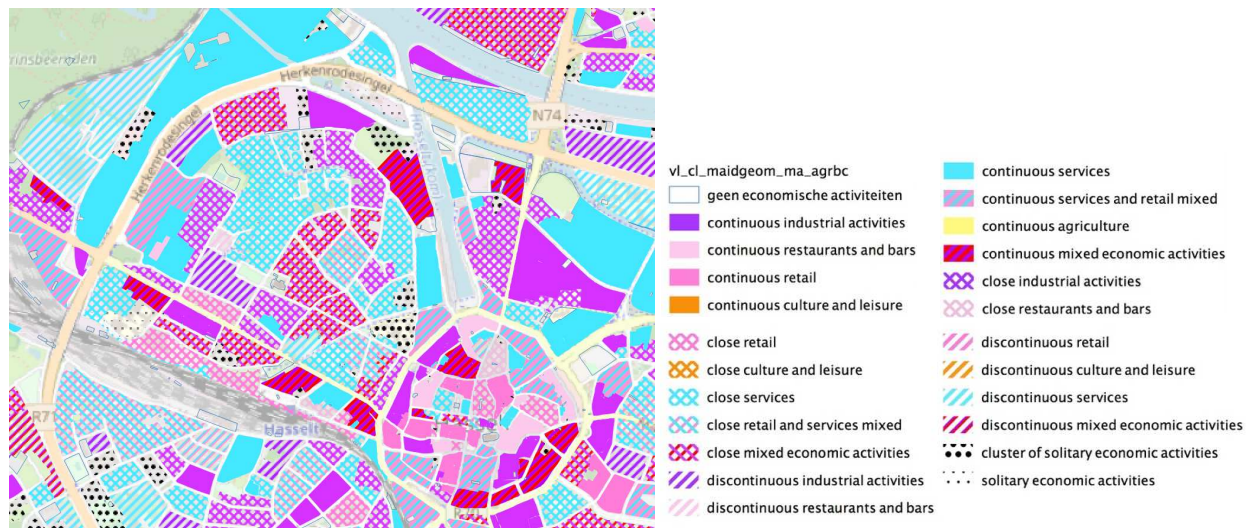


Figure 6: Activities clustering within the predefined building blocks, based on all available NACE-codes, in the Flemish city of Hasselt. (Background layer: Open Street Map)

6 CONCLUSION

Our recent work focussed on comparing potential methods for creating a map of economic ecotopes for Flanders, based on the experience from mapping economic a small part of our territory. We produced different methods to get to an automated or scripted result, based on an inventory of economic activities. This shows that (1) clusters of economic activities exists and (2) that types are statistically different, as the PCA and FA show. However, spatially interpreting these clusters proved to be more difficult. So far, adding some spatial information to the characteristics seems to be necessary. Therefore, we also created (3) a hierarchical, conditional method (based on the building blocks) to obtain similar results to drawing the types manually, and (4) proved that this can also be done using the registry of companies database. It is not yet known if this clustering method with the building blocks makes sense from a statistical point of view or if it can be reproduced for the entire area of Flanders. In conclusion,, for each of these methods, many difficulties need to be solved from the point of view of spatial interpretation, statistics and choice of thresholds, and more exploring of different ways of doing this will be needed. We consider it a big advantage to be able to compare this variety of approaches as we assume it will help us make an educated and informed choice on how to represent the map of types of economic ecotopes in Flanders, and to use it to provide policy advice.

The next steps will include the development of both a statistical and an hierarchical method to produce this map. We will also try to understand what makes the key differences between the field inventory of visible economic activities and the location of companies in the registry (VKBO). Regarding the DBSCAN and ST_ClusterWithin, we think that a more in depth knowledge of how the algorithm (and its derivatives e.g. HDBSCAN, GDBSCAN) works with our type of data will be needed.

7 BIBLIOGRAPHY

- Giaretta, F., Penninx, I., De Mulder, S., & Zaman, J. (2018). Defining economic typologies based on an economic activities database. Paper presented at the Real Corp 2018, Wien.
- Giaretta, F., & Zaman, J. (2017). Can an economic activities inventory fill the knowledge gap about the economic sector in a policy making process? . Paper presented at the Real Corp 2017, Wien.
- Giaretta, F., Zaman, J., Penninx, I., & De Mulder, S. (2019). Enhanced economic typology for spatial economic policy. Paper presented at the Real Corp 2019, Karlsruhe.
- Gruijthuisen, W., Vanneste, D., Steenberghen, T., Penninx, I., De Mulder, S., Zaman, J., . . . Horemans, E. (2018). Assessing Expanding Space Use versus Infill for Economic Activities. Paper presented at the Real Corp 2018, Wien.
- Legendre, P., & Legendre, L. (2013). Numerical Ecology (3rd ed.). Amsterdam: Elsevier.
- Penninx, I., De Maeyer, J., Leroy, S., & De Mulder, S. (2021). Behavioural studies in spatial planning. Paper presented at the Real Corp 2021 – Cities 20.50 – Creating habitats for the 3rd millennium – smart – sustainable – Climate neutral, Vienna.
- Poelmans, L., Janssen, L., & Hamsch, L. (2021). Landgebruik en ruimtebeslag in Vlaanderen, toestand 2019, uitgevoerd in opdracht van het Vlaams Planbureau voor Omgeving. Retrieved from
- Zaman, J., Penninx, I., & De Mulder, S. (2020). How to attract the right economic activities in a certain spatial environment? Paper presented at the Real Corp 2020, Aachen.