



**Statistics Netherlands**

Division of Social and Spatial Statistics

Statistical analysis department

# **Delineating population clusters by polygons. And research to a grid approach**

*Summary.*

*Statistics of population clusters in the Netherlands is introduced again by the delineating of clusters on behalf of de EU Census 2001 and are published in 2006. The introduction of new datasets and methods made it possible to delineate clusters of more recent date. The preliminary delineation of population clusters 2006 will be based on polygons. Statistics based on this delineation will be published at the end of 2007. A tentative method for deriving population clusters based on grids is researched. Results of this method are compared with the blob method for the Netherlands to get insight in the accuracy of this grid model approach.*

## **1. Introduction**

First attempts on delineating urban and rural areas in the Netherlands were performed in the Census of 1920. A division of inhabited places (localities) and places with scattered houses was introduced in the results of this Census. The practical delineation of inhabited places was made the mayor's of each community. Although this is in principle a spatial delineation the result is administrative, because the figures are published for the part of an agglomeration within the community only. From 1947 on the delineation was made up by clustering neighbourhoods with a built-up area within a community. Also ribbon development of buildings along canals and rivers is addresses as a locality. (Vliegen, 2006).

Delineating population clusters more defined started with the EU-census of 2001. The availability of digital maps and a common model gave rise to this approach. The delineation of the 2001 population clusters could no earlier be performed than 2005 and was published in 2006. A definitive delineation of clusters will be based on the latest published land use map. The land use map of 2006 will be published in 2009. The availability of the land use map of 2003 and register data of addresses and dwellings based on address points of more recent date gave rise to research the development of a preliminary delineation of the 2006 population clusters. Both the delineation based on polygons and grids were subject of research.

With the availability of GIS-methods and register datasets, combined with address points it is now possible to publish statistics of 2006 population clusters at the end of 2007. These statistics will be derived from a preliminary delineation. The delineation of clusters designates addresses of dwellings and inhabitants to lie inside or outside a certain population cluster and therefore acts as a clustering polygon.

It turned out that the preliminary delineation based on non regular polygons was fairly good to perform within acceptable time. Since this delineation is more according to the definitive delineation, based on non-regular polygons of the land use map, this polygon method will be used to publish statistics.

Besides the polygon method a grid method, based on the vicinity as a measure of concentration was studied. The method of vicinity and results from the study will be presented.



**Statistics Netherlands**

Division of Social and Spatial Statistics

Statistical analysis department

## 2. Population cluster delineation

The delineation of the population cluster in the Netherlands is based on the concept of a 'locality', introduced by the UN in the former century. A 'locality' is defined as

*'a distinct population cluster, that is, the population living in neighbouring buildings which either:*

- a) form a continuous built-up area with a clearly recognisable street formation; or*
- b) though not part of such a built-up area, from a group to which a locally recognised place name is uniquely attached;*
- c) though not coming within either of the above two requirements, constitute a group, none of which is separated from its nearest neighbour by more than 200 metres'.*

Practical contribution to the delineating of a locality are: *'..certain land-use categories should not be regarded as breaking the continuity of a built-up area. These categories are: industrial and commercial buildings and facilities, public parks, playgrounds and gardens, football fields and other sport facilities, bridged rivers, railway lines, canals, parking lots and other transport infrastructures, churchyards and cemeteries etc.'* (UNCS, 1998).

This definition of locality is related to the definition of a 'continuous built-up area'.. The 'continuous built-up area' is defined as

*'area occupied by buildings and other structures. Areas covered by the kinds of structures described below are regarded as built-up areas:*

- a) Buildings: residential buildings and related structures; industrial, commercial or utilities buildings and installations, such as factories, depots, stores, administrative buildings and churches.*
- b) Areas around and used by these structures: car parks; parks, lawns and gardens around a building; mines, spoil heaps and other tips connected with the extraction industries; disused industrial and mining sites.*
- c) Transport infrastructure: linear installations (railways, roads and motorways, water courses and canals crossed by a bridge or ferry); non-linear installations (airports, marshalling yards, motorway interchanges).*
- d) Public urban amenities: parks, public gardens and games and sports grounds, including installations, cemeteries, etc.'* (Eurostat 1992).

The definition of the continuous built-up area does not exclude clusters made up solely by industrial areas.

A frequently built-up structure in the western and northern part of the Netherlands is formed by buildings mainly on one side of a street along roads, canals and rivers (ribbon development). These structures can stretch for kilometres and place names could be attached to these structures. Ribbon development and similar structures of buildings are not considered to be part of a distinct concentration of population (Hall, 1980). The delineation of the population cluster in the Netherlands is therefore derived from the UN definition as a:

1. *'...distinct population cluster, that is, the population living in neighbouring buildings which ...form a continuous built-up area with a clearly recognisable street formation'.*
2. Certain land-use categories do not split up the population cluster.



**Statistics Netherlands**

Division of Social and Spatial Statistics

Statistical analysis department

3. Different population clusters of which the residential areas are within 200 meters of each other are considered to form one population cluster. An exception is made when residential area's are separated more than 200 meters by a canal or river, but are connected directly by a bridge or a tunnel (f.i. Rotterdam) they are also considered to form one population cluster.

### **3. Method of delineation by polygons**

The delineation is performed in a GIS using different digital maps of which the Land Use map of The Netherlands is the most important one. Other maps are the National Road Network, the addresses of dwellings and inhabitants and the geometric location of the addresses. Aerial photographs flown for the interpretation of the land use map of 2006 are available. The dwelling register contains 4 types of dwellings and their address. The recreational dwellings do not belong to a residential area and were excluded from the dwelling register. For this investigation the total number of inhabitants per address from the demographic register also is available.

#### ***3.1. Procedure***

The latest land use map of 2003 is published in 2006. The land use map of 2006 will be published in the second quarter of 2009. The preliminary borders of the population cluster of 2006 can be formed by using the 2003 land use map and extend these borders using the location of both addresses of dwellings and inhabitants. In the following text dwellings and inhabitants will be addressed by naming dwellings only.

Different subjects in the definition of clusters will be delineated as follows.

#### ***' , the population living in neighbouring buildings which ...form a continuous built-up area***

The Land Use map distinguishes 24 different types of land use and several types of different use of water bodies (CBS, 2003). For distinguishing most of the types of land use a threshold area of 1 hectare is maintained. Residential areas consist of mostly dwellings and related structures as car-parks and streets. Clustering procedure distinguishes two stages.

First stage:

- The residential area's are selected and clustered within 200 meter. Residential areas of the land use map contain at least 5 dwellings, each separated not more then 50 meters from another. The threshold area of this type is 1 hectare.
- Adjacent areas of retail business, social-cultural facilities, public facilities, industry and building lots which are neighbouring the residential areas are also selected. Building lots for roads and waterworks are excluded.

Second stage:

- Thus forming a clustered area, other land use areas are added when they are falling completely within or are adjoined to the nucleus formed by the area's in the first stage. Areas in falling within are cemeteries, parks, recreational areas of one-day outings and allotments. Areas of sport, paved roads and built-up area of airports may be adjoined. Specific area's such as golf courses, purification plants and area's of airports not containing buildings are not included.

Main roads, railways and water bodies falling within the thus formed proto population clusters are considered to be part of the population cluster.



**Statistics Netherlands**

Division of Social and Spatial Statistics

Statistical analysis department

*“with a clearly recognisable street formation”*

The pattern of paved roads and streets is provided by the map of the National Road Network produced by the Ministry of Traffic and Waterworks. A recognisable street formation within a potential population cluster is formed by at least two parallel roads connected to each other. Thus excluding potential clusters formed by ribbon development. Residential area along single roads which stretch away from the population clusters are excluded from the initial population cluster starting from the beginning of this road with the crossing with the street formation of the cluster.

Extension to preliminary borders for 2006

Thus formed population cluster borders are interpreted on the published areas of land use from the land use map of 2003. To get a preliminary border of the population clusters of 2006 additional registers are used. These are the register of address coordinates, the register of dwellings and the registers of addresses of inhabitants. The register of dwelling provides the residential dwellings. The location of inhabitants is added because not all inhabitants live in dwellings. Some live in buildings or caravans which are not subject to the dwelling register.

The dwellings and inhabitants are clustered using a buffer method taken into account the maximum distance between buildings according to the definition of the land use map. Interpretation of the changing of the borders is performed on screen using these buffered polygons, the preliminary clustered land use polygons of the 2003 land use map, the National Road Map and aerial photographs. Extensions of the population clusters are drawn on screen in a GIS (figure 1).

*...distinct population cluster*

At last, threshold values for the size of a population cluster are applied. After the process of delineating the population clusters addresses in the registers of dwellings and inhabitants are assigned to be within or outside population clusters.

A minimum number of 25 dwellings is applied as a threshold to form a distinct cluster. This is in correspondence with the Dutch directive on the forming of distinct neighbourhoods by the communities where each neighbourhood should at least contain 25 houses (CBS, 1989). In the system of funding of the communities by the government residential areas are delineated based on 500x500m grids. In this system a residential area is formed by at least 1 grid containing 25 or more addresses.

Not all inhabitants live in dwellings and there are be some flaws in the naming of addresses in the dwelling register as recent research showed (van Leeuwen, 2007a and 2007b). In 19 clusters, with each less than 25 dwellings, more than 50 inhabitants resided. These clusters are also added to the population cluster. Thus a distinct population cluster contains at least 25 dwellings or 50 inhabitants.

Finally each new population cluster is given a name as shown on the topographical map, scale 1:25.000.

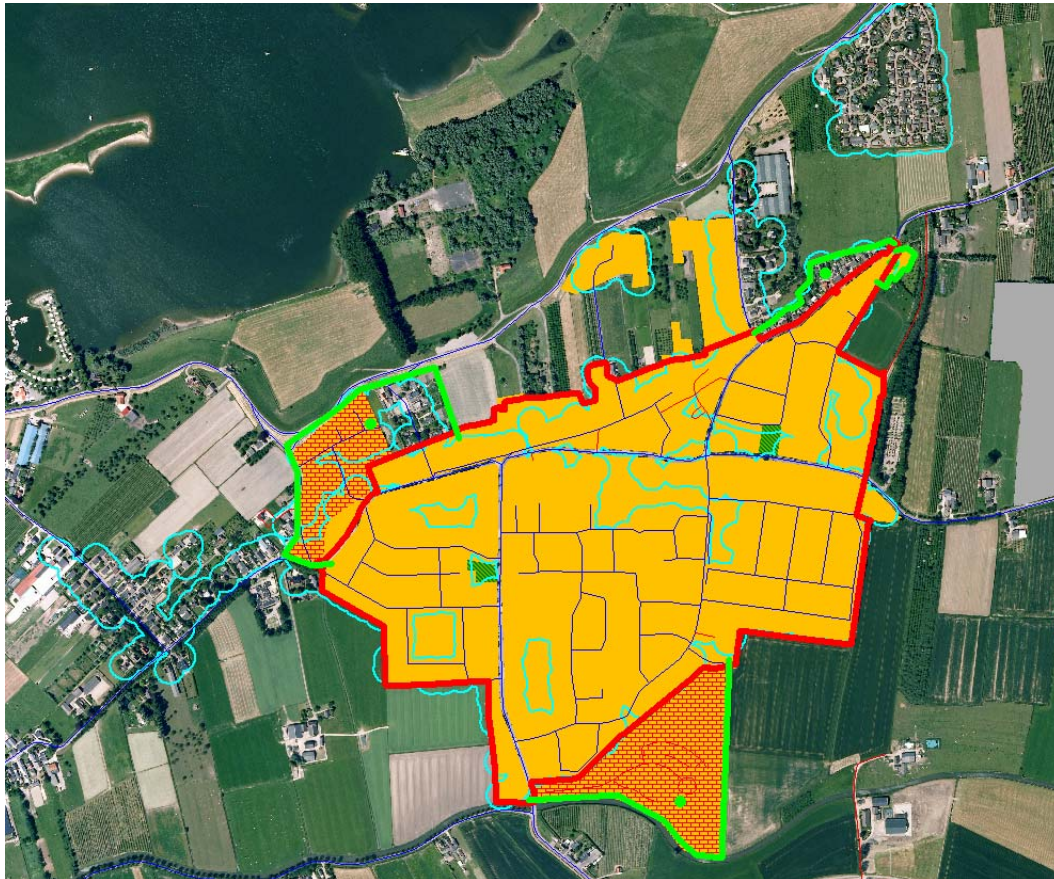


**Statistics Netherlands**

Division of Social and Spatial Statistics

Statistical analysis department

**Figure 1: Delineating population cluster borders 2006 on screen using aerial photographs 2006, land use map 2003 (solid colours), dwellings adjacent within 50 meter (light blue). National Road Map (blue and red lines), delineation of 2001 (thick red line) and extension 2006 (thick light green line).**



### ***3.2. Statistics***

The borders of the population clusters are preliminary. A definitive border will be derived from the published land use map in 2009. These preliminary borders form a clustering polygon which is used to assigned addresses of the different registers to the different population cluster numbers. Statistics on demographical and socio-economical subjects will be derived from the registers and published at the end of 2007. Some first results are published in table 1.



**Table 1: Number of population clusters and inhabitants for 2001 and 2006 clusters in order of size of the number of inhabitants.**

	2001 (definitive)			2006 (preliminary)		
	<i>cluster</i>	<i>inhabitants</i>		<i>cluster</i>	<i>inhabitants</i>	
	<i>x1</i>	<i>x1000</i>	<i>%</i>	<i>x1</i>	<i>x1000</i>	<i>%</i>
Number of inhabitants						
less than 500	593	158.1	1.0	694	175.0	1.1
500 – 2.000	701	744.2	4.7	699	739.9	4.6
2.000 – 5.000	324	1,062.8	6.6	334	1,090.4	6.7
5.000 – 10.000	164	1,167.2	7.3	165	1,163.5	7.1
10.000 – 20.000	115	1,606.4	10.0	116	1,619.9	9.9
20.000 – 50.000	80	2,349.4	14.7	82	2,470.4	15.1
50.000 – 100.000	27	1,790.9	11.2	27	1,873.1	11.5
100.000 and more	22	5,553.0	34.7	22	5,734.3	35.1
Outside population cluster		1,554.8	9.7		1,497.9	9.2
Netherlands	2,026	15,987.1	100	2,139	16,334.2	100

The number of inhabitants living in cities with 20.000 and more inhabitants is relatively increased. Subject of study will be whether this is mainly due to the extension of cities or the increase of concentration of dwellings within the cities.

The population cluster set for 2006 contain 2,139 different population clusters (table 1). The number of small clusters is dramatically increased. This may be due to better methods in spotting small areas by the method of buffering of the location of dwellings. Also dwellings sometimes mix with concentration of recreational houses. A confrontation of 2001 register data within these clusters will show whether new population cluster in 2006 would have been population clusters in the 2001 set already.

## 4. Study on delineation by grids

A method taking into account the vicinity of a grid is introduced. Subject to study is the better goodness of fit of this approach in relation to a method based on the density of a cell itself.

### 4.1. Density of the surrounding area

The method of density of the surrounding area takes into account the surrounding of cell being processed. The number of dwellings within a cell is subject to the location of the cell itself. A cell with a small number of dwellings may be the result of a more or less coincidental location of the cell and will not distinguish enough the quality of a cell whether it is part of a larger population cluster or not.

The method of taking into account the surrounding of a cell processed is in use by Statistics Netherlands since the early 90's of the former century (van der Stadt, 1992). This method is applied to distinguish different types of urbanisation within the Netherlands. The concentration of living, working and shopping is represented by the concentration of addresses of all this kind of activities.



**Statistics Netherlands**

Division of Social and Spatial Statistics  
Statistical analysis department

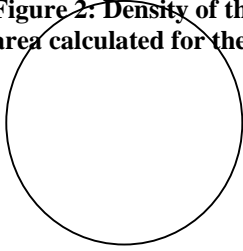


**Statistics Netherlands**

Division of Social and Spatial Statistics

Statistical analysis department

**Figure 2: Density of the surrounding area calculated for the dark grid cell.**



Calculation of the density of the surrounding area (DSA) is illustrated in figure 2. The surrounding area is defined as a circle around the centre of the cell.

The DSA for the cell concerned (dark grey) is the sum of all addresses in this cell and the cells from which the centre lies within the search radius divided by the total surface of the cells. In this example the sum of addresses present in the grey cells divided by the total area of all 13 cells. The unit of the DSA is expressed in addresses/km<sup>2</sup> and is assigned to the centre cell (dark grey).

When, for instance, the number of addresses change within one cell, the DSA of all thirteen cells which has this cell in its neighbourhood, is altered, expressing the change of concentration in the surrounding of this cell.

To demonstrate the impact of the spreading of addresses on the value of the DSA, figure 3 is compiled. Only the dark and light grey cells contain 10 addresses. The density of each cell separately is 10 addresses/ha and do not change. Other cells do not contain addresses. All cells are 100m x 100m in size.

The average DSA for four cells are calculated in different constellations. The radius is 200m which means 13 cells of 100m x 100m are taken into account to calculate the DSA for each cell.

The first situation is a rural area with no other addresses within the search radius than the addresses in the cell itself. The DSA is very low because there are no addresses in the surrounding of each cell.

The second example is an example of ribbon development. Addresses are concentrated, but only in one direction. The DSA of the outer cells is only influenced by two of the other cells.

The example of clustered buildings shows a concentration in more than one direction. All four cells are within the search radius of each other.

In the next example these clustered cells border a larger city. All cells are influenced by one or more cells from this larger city, but still there is area without addresses in the neighbourhood.

In the last example the four cells are completely surrounded by addresses and the DSA is similar to the density of the cell itself.





**Figure 3: Change in density of the surrounding area for four cells.**  
**Each 100mx100m cell, light or dark grey, contain 10 addresses.**  
**Dark grey cells are numbered from left > right, top > down.**

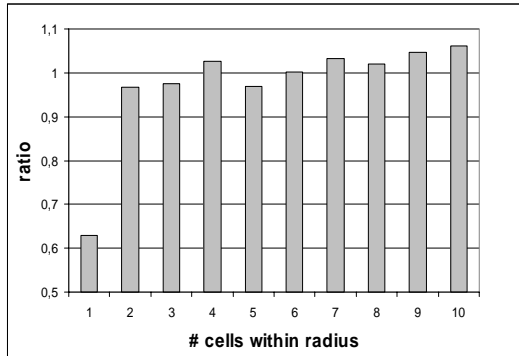
		Addresses within search radius per cell (dark grey)				Average density of the surrounding area for four cells addresses/ha
		Cell 1	Cell 2	Cell 3	Cell 4	
scattered buildings		10	10	10	10	0.8
ribbon development		30	40	40	30	2.7
clustered buildings		40	40	40	40	3.1
edge of a city		50	50	80	80	5.0
centre of a city		130	130	130	130	10.0

#### ***4.2. Determining the size of the grid***

Residential areas, separated more than 200m are considered to form different populations clusters, when not directly connected by areas of industrial or sport activities. Addresses of these clusters do not influence each other. Addresses however within the radius of 200m are supposed to belong to one cluster. Therefore the search radius is chosen to be 200m.



**Figure 4: Ratio of surface within search radius and the surface of the sum of cells in respect to the number of cells within the radius**



The search radius determines the size of the grid cell. The size of the cell may not be too large in respect of the search radius. The total area covered by large cells may well exceed or underestimate the surface of the area covered by the search radius. A very small cell size in respect to the search radius may cause computational problems.

A grid size as large as the search radius, for instance, will contain only 5 centres of the grid within the search radius. The total area covered by the grid cells will be 0.6 times the surface covered within the search radius. A grid size half the size of the radius will cover very well area considered (figure 4).

Given a search radius of 200m, a grid size of 100m is chosen, thus defining 12 cells of 100mx100m within the search radius of a selected cell.

### ***4.3. Comparing different results of the grid approach with polygon population clusters.***

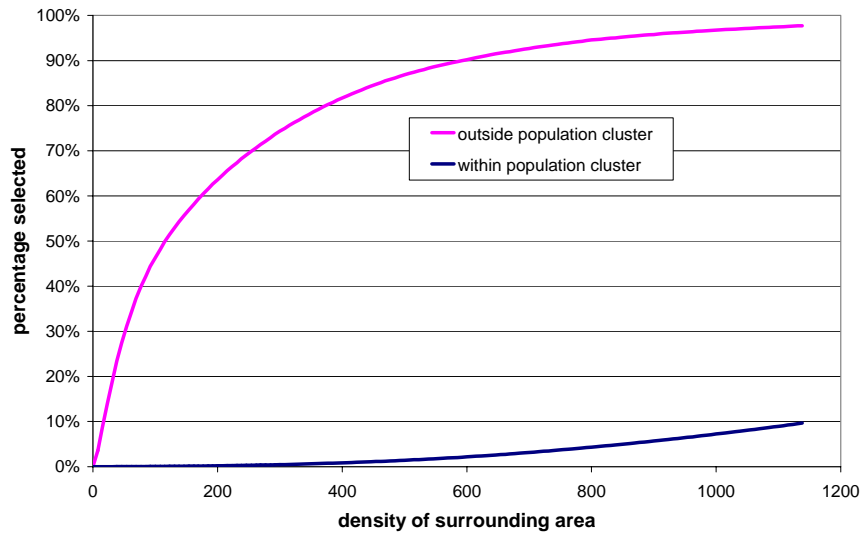
The polygons of the population clusters given, the dwellings lying within and outside the population cluster can be counted. For each grid cell the number of dwellings lying within or outside a population cluster is totalised, as well as the number of addresses in a grid cell. For each cell the density of the surrounding area based on the total number of addresses is calculated.

Figure 5a shows the cumulative number of dwellings selected by increasing DSA of the grid cell. The number of dwellings selected lying outside the polygon population cluster increases sharply. A value of the density of the surrounding area of approx. 600 addresses/km<sup>2</sup> selects grids containing 90% of the dwellings outside the population cluster and 2,1% of dwellings lying within the population cluster.

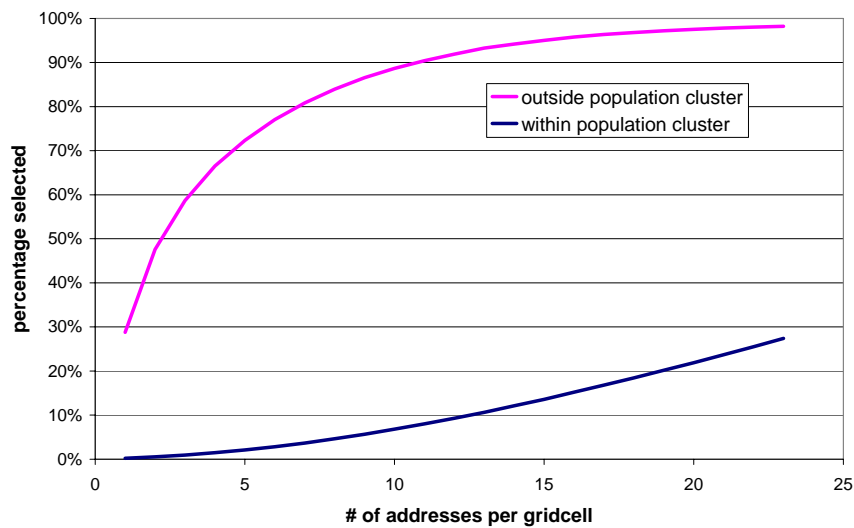
Figure 5b shows the cumulative number of dwellings by increasing number of addresses per grid cell. When the cumulative number of dwellings lying outside the clusters increase the selected number of dwelling lying inside the population clusters also increases. At a value of 11 addresses/km<sup>2</sup>, 90% of the dwellings outside the population cluster and 8 % of the dwellings outside the population cluster are selected.



**Figure 5a: Cumulative percentage of selected dwellings lying outside and lying inside population clusters by values of densities of the surrounding area. Radius 200m, grid size 100m.**



**Figure 5b: Cumulative percentage of dwellings lying outside and lying inside population clusters by values of the total number of addresses per grid cell. Grid size 100m.**



## Conclusion

Definitive delineation of population clusters will be performed on the polygons of the land use map. The result of the delineating of population clusters by preliminary polygons using new datasets fitted very well. The visual interpretation of the street network, exclusion of ribbon development and interpretation of the minimum area of 1 hectare could be performed within reasonable time. The preliminary delineation of population clusters by polygons is the type of delineation that will be used.



## Statistics Netherlands

Division of Social and Spatial Statistics

Statistical analysis department

Different methods of prediction the delineation by grids turned out to be in favour of the method using the DSA. Although it turned out to be a good prediction method in distinguishing clusters of dwellings within a stretch of ribbon development, it overestimates the dwellings belonging to population clusters very near the clusters.

## Literature

- CBS, 1989. De landelijke wijk- en buurtindeling 1988. Voorburg/Heerlen.
- CBS, 2003. Classificatie Bodemgebruik  
<http://www.cbs.nl/nl-NL/menu/themas/dossiers/nederland-regionaal/publicaties/artikelen/archief/2003/2000-classificatie-bodemgebruik.htm>
- Den Dulk, C.J., H. Van de Stadt and J.M. Vliegen, 1992. Een nieuwe maatstaf voor stedelijkheid: de omgevingsadressendichtheid (*A new measure of urbanisation: the address density of the surrounding area*). *Maandstatistiek van de bevolking, juli 1992*. 's-Gravenhage, SDU/Uitgeverij, pp. 14-27.
- Eurostat 1992. Definition of urban agglomeration in the European Community. Working party of 26/28 of February 1992 on Demographic Statistics. Luxembourg
- Hall, M. en C. Morgan, 1980. Definition and measurement of urban areas. A report on The Netherlands to the Commission of the European Communities. Reading. Department of Geography.
- United Nations Statistical Commission, United Nations Economic Commission for Europe, Conference of European Statisticians and Committee on Human Settlements. 1998. Recommendations for the 2000 Censuses of Population and Housing in the ECE Region. Statistical Standards and Studies No. 49. New York.
- Van der Stadt, H. and Vliegen, M., 1993. A new measure of urbanisation: the address density of the surrounding area. In: Beets, G.C.N. et al (eds.), *Population and families in the Low Countries 1993: late fertility and other current issues*. Lisse, Zwets en Zeitlinger, pp.151-165.
- Van Leeuwen, 2007a. Controle en verbetering van het Woningregister. Internal note. CBS, Voorburg.
- Van Leeuwen, 2007b. Opzet van een Adresrug. Internal note. CBS, Voorburg.
- Vliegen, M., 2006 Standardised Interfaces for Geographical Analyses: their significance or function for statistical output. Experiences in the Netherlands.