

Sahebali, M. W. W.; Sadowski, Bert M.; Nomaler, O.; Brennenraedts, R.

**Conference Paper**

## Infrastructure Rollout and Fibre Provision: The case of NGN in the Netherlands

30th European Conference of the International Telecommunications Society (ITS): "Towards a Connected and Automated Society", Helsinki, Finland, 16th-19th June, 2019

**Provided in Cooperation with:**

International Telecommunications Society (ITS)

*Suggested Citation:* Sahebali, M. W. W.; Sadowski, Bert M.; Nomaler, O.; Brennenraedts, R. (2019) : Infrastructure Rollout and Fibre Provision: The case of NGN in the Netherlands, 30th European Conference of the International Telecommunications Society (ITS): "Towards a Connected and Automated Society", Helsinki, Finland, 16th-19th June, 2019, International Telecommunications Society (ITS), Calgary

This Version is available at:

<https://hdl.handle.net/10419/205210>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Infrastructure Rollout and Fibre Provision: The case of NGN in the Netherlands

*M.W.W. Sahebali\*, Bert M. Sadowski\*\*, O. Nomaler\*\*\*, R. Brennenraedts\**

*\* Dialogic, Utrecht, The Netherlands*

*\*\* School of Innovation Sciences, Eindhoven University of Technology, The Netherlands & University of Trento, Italy.*

*\*\*\* UNU-MERIT, Maastricht, The Netherlands*

## **Abstract**

Within the growing literature on next generation network (NGN), much research has focused on infrastructure competition and spatial effects driving investment incentives in NGN provision. However, less attention has been paid to the dynamic factors explaining NGN rollout. The purpose of this paper is to examine the geographical effects of NGN rollout by utilizing basic data mining techniques in conjunction with exploratory spatial data analysis. In explaining NGN rollout, the paper derives a dynamic geographical model on NGN provision and examines it empirically by focusing on the spatial and temporal effects driving NGN development in the Netherlands. The paper confirms previous research on market uncertainty and the techno-economics of NGN development, but shows, in addition, that more specific factors related to local effects and demand uncertainty are vital in explaining NGN rollout.

## 1. Introduction

As next generation access networks (NGN) are expected to generate greater economic benefits to society (Röller & Waverman, 2001; Van Der Wee, Verbrugge, Sadowski, Driesse, & Pickavet, 2015), less is known about the factors driving next generation networks (NGN) across countries (Briglauer & Gugler, 2013; Cambini & Jiang, 2009). As there is some agreement that infrastructure competition has been vital for basic broadband development (Cambini & Jiang, 2009), high capacity broadband provision seems to follow different dynamics (Bourreau, Doğan, & Manant, 2010; Briglauer, Cambini, & Grajek, 2018; Fourie & de Bijl, 2018). Apart from infrastructure competition from basic broadband (Briglauer & Cambini, 2018), there are a number of other country specific factors driving high capacity broadband provision. In addition, it seems that municipal areas earlier attract NGN provision compared to more rural areas (Briglauer & Cambini, 2018). For the Netherlands characterized as a country with a small rural sector (Feijóo, Ramos, Armuña, Arenal, & Gómez-Barroso, 2018), NGN development has followed a path in which some suburban and semirural areas received NGN much earlier compared to greater municipalities (Sadowski, Nucciarelli, & de Rooij, 2009).

As new innovative technology, NGN differ from basic broadband (and older copper) technologies as they provide for higher speed (above 100 Mbit/s) and allow a greater variety of broadband services to be provided (Sadowski, 2017a). Considering the economic advantages this technology can provide the rollout of this new innovation technology has surprisingly been slow in Europe (European Commission, 2017) and United States (FCC, 2016). This has partly been due to the techno-economics but also high uncertainties surrounding the implementation of this technology (Nucciarelli & Sadowski, 2018). Previous research has demonstrated that the different techno-economics of NGN implementation and the dynamics in broadband markets are important factors in explaining NGN provision (Briglauer & Cambini, 2018). However, empirical studies have to a lesser extent focused on more detailed analyses of geographical factors and demand uncertainty in driving NGN provision.

From the literature on NGN development (Bourreau & Cambini, 2012; Bourreau, Cambini, & Hoernig, 2015), we derive a dynamic model which takes the geographical development of NGN rollout in terms of spatial and temporal effects into account. The paper utilizes basic data mining techniques in conjunction with exploratory spatial data analysis to examine NGN provision for about 7.8 million household addresses in the Netherlands. In contrast to previous studies (Grubestic, 2008), we focus in our model on NGN rollout and the spatial and temporal effects driving this development. We are able to confirm previous studies (Briglauer & Cambini, 2018) that NGN development has been driven by other factors compared to basic broadband.

In the article, we first examine the theoretical discussion on NGN development in the literature (Section 2). Second, we present our model on NGN migration and the spatial and temporal effects driving NGN rollout (Section 3). Afterwards, we present our analysis of NGN rollout in the Netherlands (Section 4). Finally, we summarize the argument and present our conclusions (see section 5).

## 2. The spatial effects driving NGN development

### 2.1. Broadband provision and investment incentives

A growing literature on broadband provision and investment incentives (Bourreau et al., 2010; Briglauer, Frubing, & Vogelsang, 2014; Cambini & Jiang, 2009) has provided new insights into the ways access regulation, the degree of network competition and alternative public policies can affect broadband

development across countries. Within the first strand of literature, research has addressed the investment incentives of incumbent firms (Faulhaber & Hogendorn, 2000) and/or of alternative operators (Bourreau et al., 2015; Brito & Tselekounis, 2016; Matsushima & Mizuno, 2018) depending on the conditions of access. These studies have just recently addressed differences in access rules as affecting NGN rollout across geographical locations (Bourreau et al., 2015) while leaving other spatial determinants to further research. A different stream of literature has been focused on discussing the extent to which wholesale competition between vertically-integrated firms (Bourreau, Hombert, Pouyet, & Schutz, 2011; Ordovery & Shaffer, 2007) might lead to a more competitive landscape in retail markets. It demonstrated that vertically-integrated firms can discriminate between markets even in a situation characterized by a high level of wholesale competition. In addition to this literature, this study focuses on spatial determinants affecting investment incentives across regions. Within a third area, the literature has focused on analysing in greater detail the spatial factors across different regions in their effects on investment incentives of companies (Grubestic, 2006; Mack & Grubestic, 2014) without linking this discussion to NGN provision yet. In this context, the study in this paper explores the NGN rollout depending on variety of socio-demographic and demographic factors by using local spatial statistical techniques to develop a spatial taxonomy for NGN provision in the Netherlands.

## 2.2. NGN rollout: a spatial model

Similar to Bourreau, Cambini et al. (2015), we assume a geographical coverage competition between two incumbent operators (firm 1 and firm 2), investing in the coverage of new NGN infrastructure, and an entrant (firm E) not investing anymore. There are different locations  $z \in [0, \bar{z}]$ , each with identical demand. In addition, the fixed costs  $c(z)$  of covering a location  $z$  is differentiable and increasing depending on the density of population. Locations with a low index  $z$  are more densely populated and therefore are related to lower fixed costs of coverage. In contrast, a high index  $z$  indicates a less densely populated area with higher fixed costs over coverage. Furthermore, both incumbents have similar investment costs  $c(0) = 0$  (Bourreau et al., 2015).

In order to provide NGN infrastructure across different locations, both firms  $i = 1, 2$  have to decide in which area  $[0, z_i]$ , they want to roll out their network. In case, they want to cover the area the total cost of investment is  $C(z_i) = \int_0^{z_i} c(x)dx$ . The cost of investment has the following properties  $C'(z_i) = c(z_i)$  and  $C''(z_i) = c'(z_i) > 0$  (Bourreau et al., 2015). The access price  $a$  is exogenous set by the regulator (Bourreau et al., 2015).

In geographical competition, each incumbent firm  $i = 1, 2$  chooses a certain coverage  $[0, z_i]$  to maximize profits depending on the rival's coverage  $[0, z_j]$ , defined as  $j = 1, 2$  with  $j \neq i$ . In case firm  $i$  chooses  $z_i > z_j$  the firm will be the only access provider in the area. In case, there are two providers in the area  $[0, z_j]$ , either firm  $i$  or firm  $j$  can be the access provider, and they jointly can generate expected profits in the location  $\pi^d(a)$ . The expected profit of firm  $i$  is

$$\Pi_i(z_i, z_j) = \begin{cases} z_i \pi^d(a) + (z_j - z_i) \bar{\pi}^{(j)}(\bar{a}) + C(z_i) & \text{if } z_i \leq z_j \\ z_j \pi^d(a) + (z_i - z_j) \bar{\pi}^{(i)}(\bar{a}) + C(z_i) & \text{if } z_i > z_j. \end{cases} \quad (\text{Bourreau et al., 2015})$$

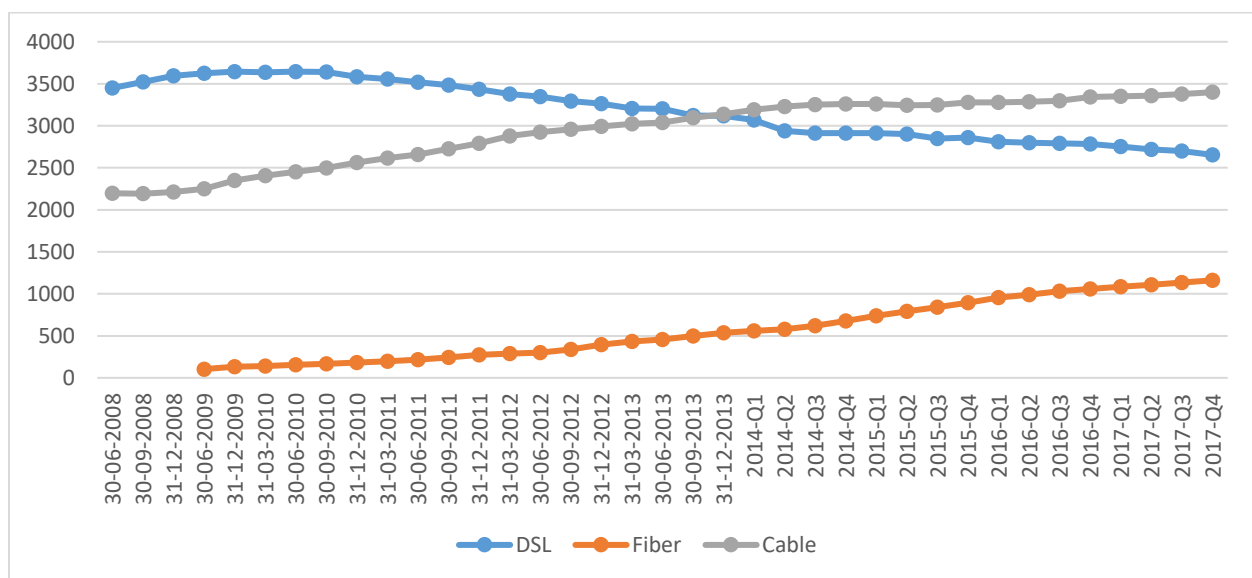
In addition to this model, we assume that both firms evaluate the progress of rolling out NGN according to rollout in neighbouring local areas. In order to calculate the propensity of firms to provide NGN, we use local Moran's I.

### 3. The evolution of NGN development in the Netherlands

#### 3.1. Historical evolution of NGN and broadband providers

Within the Netherlands, the broadband infrastructure consists primarily of xDSL connections, coax-modem-based broadband and fibre connections. Figure 1 shows the changes in the availability of these different broadband connections, with a decrease in xDSL technologies since 2009 and a steady increase in fibre and cable-modem connections over time. The penetration of fibre technologies in terms of households connected increased between 2010 and 2014 from 0.5 million to 2 million. Currently fibre penetration is at about 29% with over 2.5 million households connected (ACM, 2018).

**Figure 1: Growth of Broadband Technologies over time in the Netherlands since 2008**

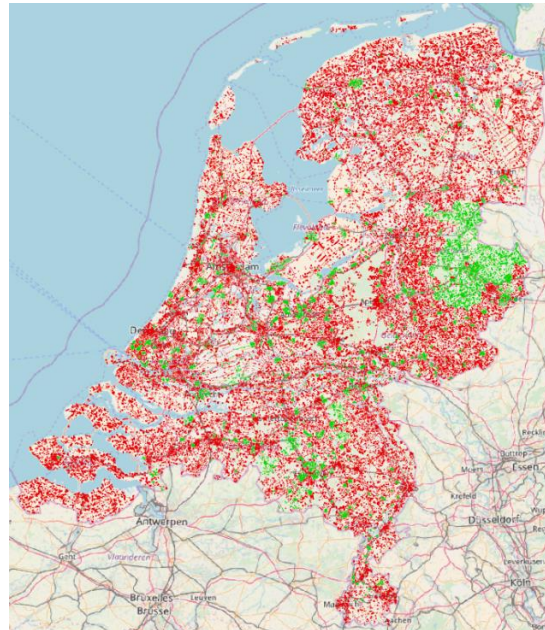


Source: (ACM, 2018)

As the backbone of the broadband network in the Netherlands is mainly based on fibre technologies, a key issue has been how the differences in terms of internet speed can be explained, for example in rural areas. A key factor explaining differences across areas has been the absence of cable networks and long distances to the backbones and street cabinets (Van der Haar, Driesse, Brennenraedts, Pegtel, & Cazemier, 2017). Interestingly

Figure 2 shows that a great deal of fibre connections can actually be found in rural areas, and in the west part of the country. Furthermore, these technologies are apparently rather absent in the centres of large cities and most prevalent in medium-size towns in the Netherlands.

**Figure 2: Fibre connectivity in the Netherlands (Dialogic Mappingtool, 2017)**



(Explanation: Green area depict the existence of fibre connections, red areas their absence).  
(OpenStreetMap contributors, 2018)

Within the Netherlands, fibre providers are the following firms: KPN, EQT, Ziggo, Tele2, Eurofiber, and regional FttX providers (Besteman et al., 2017). In the Dutch general broadband market, KPN and Ziggo have about equal market shares of 40% up to 45% each (ACM, 2018). This leaves a remaining 10% to 20% to other parties such as Tele2, local network owners and network owners with their own specific types of networks, such as fibre. With respect to fibre rollout, three options can be distinguished: Firstly, there are fibre deployment projects led by incumbent firms. Secondly, a variety of partnerships between cable companies, regional distribution network operators and investment funds. Thirdly, citizens opt for building up a local bottom up initiative (Sadowski, 2017b). Currently, the fibre market is mainly served by incumbents and investment funds.

There are different players in the Dutch business market: Ziggo deploying mainly an HFC network and targeting this market with FttO (Bremmer, 2017). Another player in the FttO market is Eurofiber deploying infrastructure after which any provider is allowed to offer broadband services over the network. There are also initiatives from other companies at business parks that cooperate to have their business park connected to fibre. In areas where there is no business case for fibre or wired broadband infrastructures seems infeasible, alternative technology like wireless broadband are deployed.

In analysing the difference between rural and urban *areas*, we defined different localities (*bevolningskernen* CBS (2011)) based on areas with at least 50 residents and at least 25 households in an uninterrupted recognizable pattern as urban. Applying this distinction to the dataset used in this paper, provided us with 875,591 household addresses that are located in rural areas. For the rural households, 14.7% had a fibre connection, and for urban household this share was 30.1%.

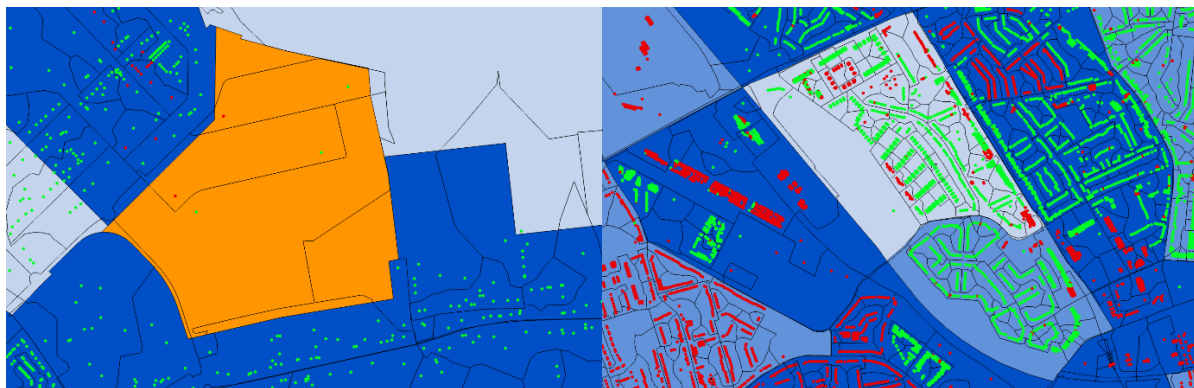
## 4. The empirical model

### 4.1. Spatial association

In order to test if broadband availability deviates from the null hypothesis of spatial randomness, one approach is to determine statistically significant patterns of spatial association in terms of spatial autocorrelation (Grubestic, 2006). The objective behind this approach is to define locations that share similar characteristics with respect to broadband availability and competition. In the following, the statistics, global Moran's  $I$  and local Moran's  $I$  have been estimated to examine additional spatial factors affecting locations.

We specified the global Moran's  $I$  in the following way: Pairs of households were assigned a weight ( $w_{ij}$ ) of 1 when they are in the same region and otherwise 0. Thus these weights have been determined by contiguity instead of distance. In that way a Pearson correlation could be obtained for fibre connections in the same region. In general,  $I$  could yield a value between -1 and 1. A positive  $I$  indicated that within a specified region, households with a fibre connection are clustered with other households which also have a fibre connection, and non-fibre households are clustered with non-fibre households. This would imply a pattern of clustering which follows the boundaries of the pre-determined region. An  $I$  of 0 means that the spatial distribution of fibre and non-fibre households is completely random. If  $I$  is negative this means that there is an alternating pattern between fibre houses, so for every non-fibre house there exists a fibre house in the same area and vice versa. Figure 3 illustrates this pattern by calculating the Moran's  $I$  for all neighbourhoods while taking the streets as regions within which household are considered as neighbours. Green points indicate houses with a fibre connection while red dots indicate households without fibre. The neighbourhoods are coloured according to their global Moran's  $I$ , going from orange ( $I = -1$ ) to white ( $I = 0$ ) to dark blue ( $I = 1$ ). In the orange neighbourhood, the streets have a similar number of fibre households as the non-fibre households within that street. In the light blue neighbourhood (Figure 3, right image) there is a more random pattern between fibre and non-fibre households within the streets. The dark blue neighbourhoods have streets which have clear clusters of either fibre or no fibre. The global Moran's  $I$  can thus contribute to the analysis by assessing the extent of clustering and the boundaries of clusters in regions.

**Figure 3: Global Moran's  $I$  for streets within neighbourhoods**



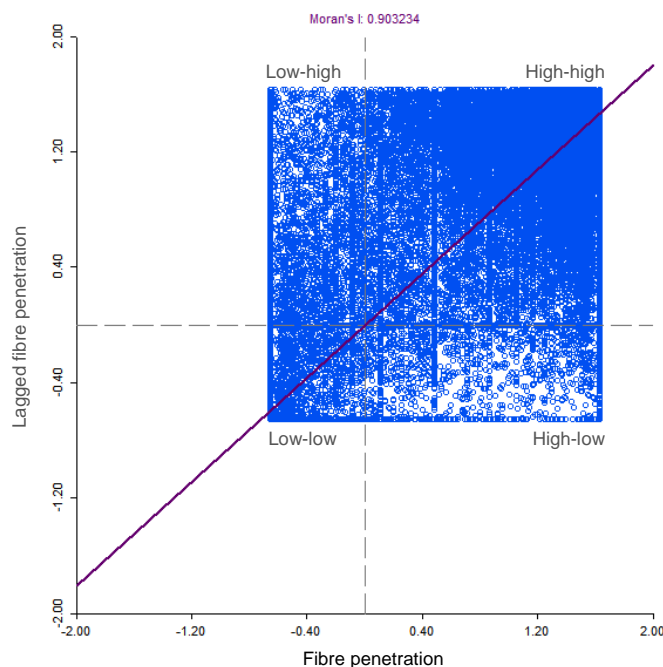
Whereas the global Moran's  $I$  describes the overall regions, the local Moran's  $I$  splits up this statistic for each specific observation.  $I$  is calculated for every observation  $i \in S$ , with  $S$  being the set of observations,  $z$  as the deviation from the mean,  $j \in S$  and  $j \neq i$  (Anselin, 1995). In that way each observation can be



evaluated according to their neighbouring areas. In contrast to the analysis of the global Moran's I, Queen contingency has been applied to regard regions as neighbours for the local Moran's I. This means that regions will be regarded as neighbours when they share a common border. These neighbours can be direct neighbours in a first-order analysis, or the neighbours of neighbours in a second-order analysis (including lower orders). In Grubestic (2006) and Mack (2014) the results of the local Moran's I have been categorized according to the following, so-called, LISA (local indicator of spatial association) categorization (Anselin, 1995) in which the deviation from the mean in a neighbourhood is compared to the level of fibre houses in neighbouring regions:

- High-high: the core
- Low-low: the periphery
- Low-high: islands of inequity
- High-low: islands of availability

**Figure 4: Moran scatterplot**



The names have been derived from the quadrants in a Moran scatterplot (Figure 4). The points plotted on the plane represent individual observations, and the slope of the line fitted through these points equals the global Moran's I. The first low/high LISA categorization stands for the value of that specific observation itself (Figure 4, horizontal axis). The second category refers to the average value in neighbouring regions, therein functioning as a lagged variable (Figure 4, vertical axis). The high-high and low-low categories indicate that the observation is part of a larger cluster, as neighbouring observations have a similar level of fibre houses as the observation itself. Most interesting are thus the low-high observations where the observation itself has a low share of fibre houses, but the neighbouring observations have a high share of fibre houses. The same holds for high-low, which represent hubs of high fibre connectivity surrounded by observations with a low share of fibre households. For the first two categories the spatial autocorrelation is high, whereas the local Moran statistic would be low for the latter two categories, which can be

considered as spatial outliers. The contribution of the local Moran's I in the analysis here lies in its ability to identify specific types of rollout regions.

## 4.2. Variables, Data and Results

In order to conduct an empirical analysis and estimate the propensity that a firm will provide NGN at a specific location we used a variety of data sources to generate our dataset. In classifying the different variable, we applied the concepts of technological uncertainty, market uncertainty and demand uncertainty involving the development of broadband infrastructure (Nucciarelli & Sadowski, 2018).

The main variables to be used in the regression are summarized in Appendix 1. First the dependent variable will be described after which all independent variables will now be discussed per category.

### 4.2.1. Dependent variable: Spatial cluster association

The dependent variable of the empirical model is based on what type of LISA cluster the neighbourhood belongs to. This variable is thus categorical, with its classification being based on the penetration of fibre connections within both the area itself and its neighbouring areas. Data on the fibre penetration indicates whether households have a fibre connection or not. This only describes whether a connection is present. Therefore, it does not imply that the household has subscribed for the fibre connection, which would be adoption. The fibre availability variable is coded binary as 0 for no fibre and 1 for fibre. This is aggregated into a fibre penetration variable by taking the mean for every specified region. This variable only contains information which has been made available by fibre deployers for this dataset. All households for which no information on fibre connectivity is available are considered as non-fibre households. Therefore, the dataset might suffer from underreporting for some regions. The main fibre deployers included in the dataset are KPN and EQT. Next to that the dataset contains data on business parks which have a fibre connection and some regional providers. For the province of Zeeland there appears to be a considerable lack of data, this region will therefore not be well represented in the analysis. Also, most of the fibre-rollout initiatives are underrepresented in the dataset. As a consequence, the LISA categories in this analysis will be based on a variable which actually represents the availability of fibre connections rolled out by major fibre deployers.

### 4.2.2. Techno-economic uncertainty

For the techno-economics of fibre, different proxies have been used which determine the costs of the physical deployment of fibre. Firstly, the distance required to connect a house to the fibre network is used. The sub-loop length is herein determined by the distance to the Eurofiber backbone. The Eurofiber backbone is not the only backbone in the Netherlands. Yet, since it was the only backbone information available for this analysis and the Eurofiber network is open to use for other providers, it will be regarded as a proxy for backbone infrastructure in general.

Next to that the density of houses and distance to the next house have been used. These could give the impression of measuring the same phenomenon. But the distance to the next house has been included as an additional variable to control for remoteness of houses, as it is able to capture the proximity of houses at different scale than the density measure for neighbourhood. This measurement will, for example, be capable of highlighting spatial outliers.

Physical obstacles, such as the proximity to trees and water bodies have also been included in the regression. These are often expected to hamper the rollout process, as fibre cable ducts need to be diverted around the obstacles. The water bodies have been calculated as the percentage of the

neighbourhood's surface which is covered by water. The proxy used for trees is determined by the average tree height in a radius of 20 m from every address. The dataset on tree height was obtained from the Rijksinstituut voor Volksgezondheid en Milieu (RIVM). The RIVM has derived this dataset from a height scan of the Netherlands. The dataset contains trees with height ranging from 2.5 to 50 m, measured from the ground. This dataset has been provided in a raster where every pixel represents an area of 10 by 10 m (Ministerie van Binnenlandse Zaken, 2017).

Another physical aspect which might influence the cost is the ground quality. This could force the infrastructure deployer to put the cables deeper in the ground but is also used by municipalities to constitute the additional fees deployers have to pay for the recovering of the ground and tileworks. The ground quality variable used here determines the percentage of bad quality soil (meaning clay and peat) in a municipality. The descriptive statistics of these variables in the rural and urban dataset can be found in the appendix (see Appendix 2 and 3).

#### 4.2.3. Market uncertainty

The presence of alternative infrastructures is, according to Bourreau et al. (2017), another factor which influences the deployment of broadband. Data on the presence of possible substitutes have been used to estimate a Herfindahl Hirschman Index (HHI) as is done in Fourie & De Bijl (2017), to determine the market uncertainty in the Dutch fibre market. As there is no specific data available about the local market shares of specific providers, the analysis here will look at the market share of broadband technology types. In the regression here the HHI has been calculated per neighbourhood as:  $HHI = p_{fibre}^2 + p_{cable}^2 + p_{copper}^2 + p_{wireless}^2$ , with  $p$  being the share of houses in that neighbourhood with that specific technology. Herein the fastest available technology is considered, so that all shares together sum up to 100. This means that the technologies supersede each other in the order (from high to low) of: fibre, cable, copper over 30 Mbps and wireless broadband. So, for every neighbourhood it is calculated what the share of houses is with these four specific technologies as the most superior connection to that house. The shares are stored as percentages, the maximum possible HHI is therefore 10,000. On average the dataset yields an HHI of 7,856, which can be considered as high concentration per neighbourhood. This could be seen as an indication that, on a neighbourhood basis, there is a high concentration of the same technology. Moreover, the presence of a coax connection has also been included as an independent variable, given the high performance of coax under the latest DOCSIS standard and the expectation that consumers are insensitive to the connection type, yet sensitive to speed (Fourie & De Bijl, 2017). The variable included is defined by the number of coax connections that every individual address has. This is based on a selection of providers for which the data was available, resulting in 7 million addresses with one coax connection and just 92 with more than one coax connection.

#### 4.2.4. Demand uncertainty

Demand uncertainty actually determines the rollout in terms of adoption. Adoption studies already indicate that income and level of education are determining factors, therefore the analysis will explore whether such characteristics also hold for deployment (Flamm & Chaudhuri, 2007; Quaglione, Agovino, Berardino, & Sarra, 2018). The data available for this part of the analysis does not have a high level of detail, as the data is mostly available at neighbourhood or municipal level only. Hence, the independent variables on demand uncertainty will classify the type of neighbourhood a house is located in, in terms of demographical characteristics. The demographical variables of gender, marital status, age and level of education are used here because they have been found to have effects on broadband rollout in previous

studies (Dwivedi & Lal, 2007; Flamm & Chaudhuri, 2007; Quaglione et al., 2018). The other variables have been included to observe whether other socio-cultural determinants of a neighbourhood also can have an impact on the demand for fibre.

However, including all predictor variables listed in Appendix 1, in their current form, would lead to a logit model with high multicollinearity. Using the rule of thumb of interpreting VIF values above 5 as multicollinearity, most of the social predictors appear to be related. Herein VIF values of over 10 could be found for cultural and religious predictor variables. This has led to the decision of applying an exploratory factor analysis on the demand uncertainty variables, in order to extract uncorrelated factors.

From this it can be observed that the age group variables and marital status are correlated with each other. The percentage of woman is positively correlated with the highest age group, indicating that women are generally older than men. This will be the first set of demographic indicators. A second set of factors will be obtained from a combination of religious, education level and cultural diversity variables. Taking an eigen values threshold of around 1, scree plots (see Appendix 7) suggest that the three factors should be extracted for both sets.

The loadings of the variables onto the factors can be observed in Appendix 6 for set 1 and Appendix 8 for set 2. From the first set of factors, three main factors have been extracted. Factor 1 appears to represent a scale from *students to families*, factor 2 appears to represent *age*, and factor 3 goes from *married to divorced*. The second set of factors is supplemented with variables on cultural diversity. The factors that can be identified here are, *Dutch to multi-cultural* for factor 1, *education level* from high to low for factor 2 and *Reformed to Catholic* for factor 3. Given the geographic split between Reformed and Catholic Christians in the Netherlands, factor 3 can also be loosely interpreted as going from north to south.

There are several incomplete cases in the data, especially in rural areas. Those appear to be caused by the high number of missing values for the variables on the country of origin. The variable on non-western background has fewer missing values and is highly correlated with all these variables. Therefore, a KNN imputation ( $k = 11$ ) has been used to impute these missing values on the country of origin. Essentially this can be regarded as mean imputation, but in this case the mean is taken only from the  $k$  observations which appear to be nearest to the missing observation. This proximity has been taken in terms of geographic coordinates and in terms of the second set of variables used for the factor analysis.

Both sets of factors explain a considerable amount of the variance, as the first set accounts for 70% of the variance and the second set for 64% of the variance. As a result of the factor analysis, all factors have been standardized around a mean of 0. Including these factors, instead of these variables separately, into the regression has solved the multicollinearity problem, since all VIF values are now below 5.

## 5. Results

### 5.1. Global clustering

Starting with the Global Moran's I, the results are shown in Table 1. Herein houses within the same region are taken as neighbours (have a weight of 1) and pairs of houses that are not in the same region are not regarded as neighbours (here the weight for that pair is 0). Global Moran's I values that are close to 0, normally indicate a total random pattern. Negative values that approach -1, indicate complete dispersion, meaning that for every house with fibre in the same signified region, there is also a house without fibre. In that case, half of that region generally has fibre. A positive spatial correlation approaching 1, on the

other hand, would support the claim of perfect clustering according to the specified region. Therein, all the houses within the same region are either connected to fibre or not connected to fibre at all. This spatial correlation coefficient has been calculated for the entire Netherlands, taking the respective region designations, from macro to micro scale, of municipalities, zip code 4 areas, neighbourhoods and zip code 6 areas.

Table 1 shows that zip code 6 regions, which is practically at street level, yield a high correlation. This indicates that in most cases the entire street is being connected to fibre or is not connected to fibre. Cases where only a part of the street is connected by fibre are thus rare. For neighbourhoods the Moran's I is lower, but still considerably high. This would indicate that there is a high correlation between connectivity by fibre for houses within the same neighbourhood. The value on the neighbourhood level is also higher for urban areas than for rural areas, 0.814 ( $p < 10^{-6}$ ) against 0.686 ( $p < 10^{-6}$ ) respectively. Increasing the specified region size, it can be observed that the coefficient drops sharply. Zip code 4 regions are generally larger than neighbourhoods. The lower value for its autocorrelation in comparison to neighbourhoods also indicates that fibre rollout is less clustered according zip code 4 areas. For municipalities, this value is even lower, as not many municipalities have a fibre penetration which is close to 100% or close to 0%. This implies that fibre projects indeed seem to be rolled out at a per-neighbourhood basis. This observation conforms with the project plans of rollout actors, in which they mostly target projects on a per-neighbourhoods or borough basis.

**Table 1: Moran's I per region category**

Region category	Moran's I
Municipality	0.344***
Zip code 4	0.695***
Neighbourhood	0.804***
Zip code 6	0.896***

Significance  $p$  is denoted by \*\*\* for  $< 10^{-6}$ , \*\* for 0.01, \* for 0.05, and no symbol for  $> 0.05$ .

These correlation values can also be split out per province, allowing a view which highlights the variation between provinces. In Table 2, Moran's I values have been listed per province, for different regional units. The differences between provinces quickly become apparent here. The most interesting regions are, however, the provinces with lower correlation. In these provinces the regions are characterized by the fact that only a part has fibre deployed. This could be interpreted in different ways. Either this region is currently being rolled out, or a fibre project has been cancelled and stopped there. Nonetheless, from a practical point of view, the batch non-fibre houses in this region could be least inexpensive to connect next for an upcoming fibre project in the near-future, as they are close to a fibre infrastructure that is already in place. This phenomenon would be most plausible in regions where the correlation is low on a zip code 6 level, but it could also be the case for low spatial autocorrelation on neighbourhood level. This is less likely for regions with a high Moran statistic, because for those areas there is either a small number of non-fibre houses in relation to the houses with fibre or there is a small number of fibre houses. In the former case it would possibly imply that these houses are unconnected due to the preference of the house owner and in the latter case it could imply that the connected houses are either businesses or early adopters. Following this reasoning on a neighbourhood level, the provinces of Flevoland and Utrecht could be expected to receive new fibre rollout projects in a short term. For the province of Zeeland, it can be observed that fibre is mainly rolled out at a street level and that the clustered presence of fibre within

municipalities is close to random. Although, the reason for observing this effect might also be caused by the lack of data on the Zeeland province.

**Table 2: Moran's I for different region categories per province**

Area	Municipality	Zip code 4	Neighbourhood	Zip code 6
Drenthe	0.159 ***	0.582 ***	0.725 ***	0.901 ***
Flevoland	0.043 ***	0.202 ***	0.371 ***	0.843 ***
Friesland	0.556 ***	0.682 ***	0.774 ***	0.859 ***
Gelderland	0.284 ***	0.550 ***	0.701 ***	0.872 ***
Groningen	0.187 ***	0.687 ***	0.802 ***	0.922 ***
Limburg	0.619 ***	0.716 ***	0.856 ***	0.915 ***
Noord-Brabant	0.353 ***	0.689 ***	0.829 ***	0.901 ***
Noord-Holland	0.095 ***	0.647 ***	0.815 ***	0.865 ***
Overijssel	0.445 ***	0.667 ***	0.768 ***	0.891 ***
Utrecht	0.347 ***	0.624 ***	0.648 ***	0.864 ***
Zeeland	0.009 ***	0.102 ***	0.220 ***	0.786 ***
Zuid-Holland	0.346 ***	0.731 ***	0.804 ***	0.918 ***

Significance  $p$  is denoted by \*\*\* for  $< 10^{-6}$ , \*\* for 0.01, \* for 0.05, and no symbol for  $> 0.05$ .

It is expected that fibre deployment follows different dynamics in urban than in rural areas. Therefore, it will cluster differently for these different modes of regions. In Table 3 the global Moran's I is split out for urban and non-urban per province. The provinces of Flevoland, Friesland, Overijssel and Utrecht appear to have a lower clustering here for urban regions than for rural regions on both the neighbourhood and zip code 6 level. This means that rural areas are clustered less randomly according to neighbourhood and zip code than urban areas. Hence, less disparity is expected in urban areas than in rural areas of those provinces, implying that rollout in rural parts of those provinces would require more attention.

**Table 3: Moran's I for neighbourhood level per province split out by rural/urban houses**

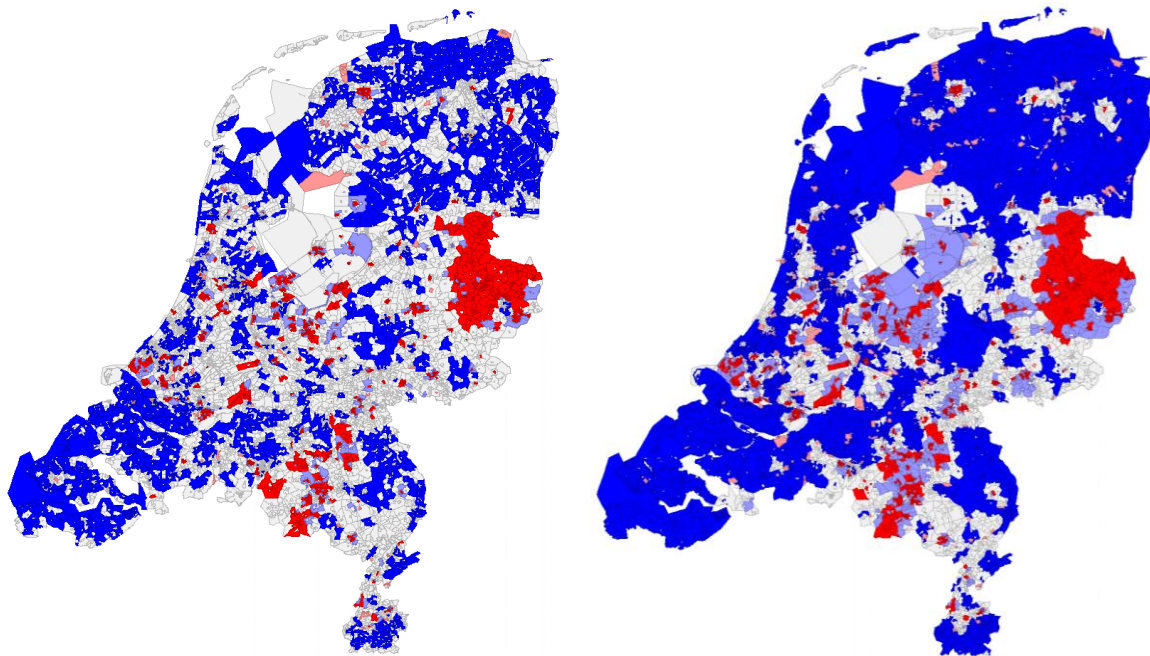
Area	Neighbourhood Rural	Neighbourhood Urban	Zip code 6 Rural	Zip code 6 Urban
Drenthe	0.448 ***	0.744 ***	0.797 ***	0.895 ***
Flevoland	0.620 ***	0.268 ***	0.913 ***	0.679 ***
Friesland	0.710 ***	0.781 ***	0.939 ***	0.847 ***
Gelderland	0.627 ***	0.711 ***	0.830 ***	0.869 ***
Groningen	0.623 ***	0.809 ***	0.865 ***	0.924 ***
Limburg	0.551 ***	0.879 ***	0.823 ***	0.926 ***
Noord-Brabant	0.657 ***	0.836 ***	0.852 ***	0.905 ***
Noord-Holland	0.608 ***	0.823 ***	0.762 ***	0.872 ***
Overijssel	0.868 ***	0.769 ***	0.949 ***	0.884 ***
Utrecht	0.721 ***	0.659 ***	0.866 ***	0.860 ***
Zeeland	0.291 ***	0.348 ***	0.526 ***	0.922 ***
Zuid-Holland	0.566 ***	0.813 ***	0.897 ***	0.920 ***

Significance  $p$  is denoted by \*\*\* for  $< 10^{-6}$ , \*\* for 0.01, \* for 0.05, and no symbol for  $> 0.05$ .

## 5.2. Local clustering

The local Moran's I (see left Figure 5) per neighbourhood provides an even better insight into how fibre rollout spreads between neighbourhoods, as it breaks up the Global Moran's I into localized statistics for every specific neighbourhood. The LISA categorization in that figure is explained in the subscript caption of that figure, with the number of neighbourhoods with that denomination between parentheses. The second order Moran's I (see Figure 5 right) also takes the neighbours of neighbours into account. This could be regarded as a prediction further into the future, and it makes it more clear which areas are likely to receive fibre. In Figure 6 the same results are displayed at a higher resolution, i.e. at street level instead of neighbourhood level, making the clusters more pronounced.

**Figure 5: Local Moran's I for neighbourhoods according to first order (left) and second order (right) neighbours**



White: Not Significant ( $p > 0.05$ ) (N=9133; 4003), Red: High-high (2191; 2329), Blue: Low-low (1554; 6018), Light blue: Low-high (242; 610), Light red: High-low (63; 223)

### 5.2.1. High-high: the core

The high-high regions highlight the locations of fibre hotspots in the Netherlands, these are the cores of fibre rollout in the Netherlands where fibre rollout appears to spread to neighbouring neighbourhoods. This implies that a project has targeted a group of neighbouring neighbourhoods or even that neighbourhoods learn from one another, causing fibre deployment to spread. One of the main hotspots is the area in Overijssel where EQT (formerly CIF) rolled out as a batch of projects. Whereas Overijssel has one main fibre cluster, the other provinces have a lot of unconnected clusters. Among these, there are larger clusters around the centres of medium-sized cities. But there are also more remote smaller clusters in villages scattered around. The reason for this pattern of unconnected cores remains unclear. It could be attributed to the low density of houses in the regions between the cores. Yet, another explanation can be that all these separate cores do not cluster because they have initially been started by different project owners, mostly being different municipalities or industrial parks.



#### 5.2.2. Low-low: the periphery

The periphery areas, where fibre penetration is equally low as the neighbours, appear mainly as large clusters in rural areas. These regions are the least likely to receive fibre deployment from their neighbours. These regions would, therefore, benefit most from new projects that target neighbourhoods in this cluster specifically. An interesting observation, however, is that larger cities in the Netherlands (e.g. Amsterdam, Rotterdam, Den Haag, 's-Hertogenbosch) also appear to belong to this peripheral cluster. As broadband has previously been found to agglomerate in metropolitan areas, in a US case, it could also be expected for the case of fibre in the Netherlands (Grubestic, 2006). Yet, since this is not observed here, it seems that other factors either inhibit the diffusion of broadband rollout to these regions or fibre is not competitive enough to fully penetrate these metropolitan areas.

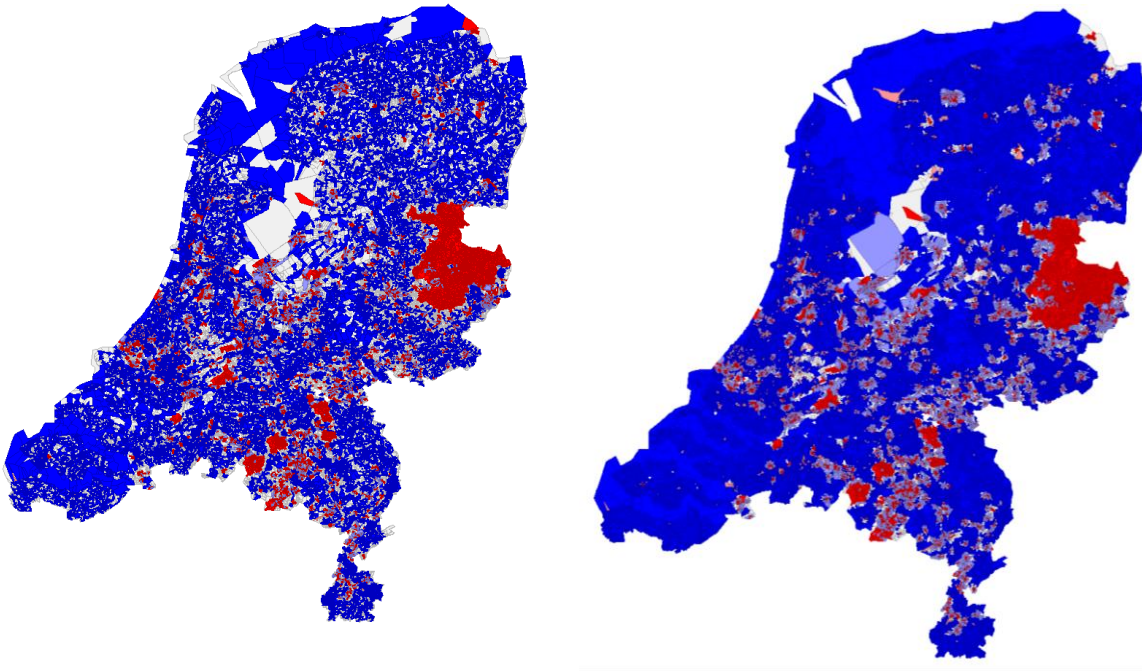
#### 5.2.3. Low-high: islands of inequity

The low-high areas are most interesting here, because in these areas, fibre can be expected next if it is assumed that fibre will spread. However, for these areas, there might also be a specific reason why the share of fibre connected houses there is low in relation to the neighbouring neighbourhoods. For some of the low-high areas visual inspection shows that fibre has already been deployed there. In the case that fibre has been deployed here before 2017, it indicates that the dataset used in this research is incomplete. But, in the cases that fibre has only been deployed recently there, it indicates that the LISA statistic rightfully expected deployment. Several of these regions are mostly the areas of remote houses just outside of the city, recreational areas or industrial areas. Therefore, for some of these areas it is not reasonable to expect fibre deployment, given the low number of residential houses in the area.

#### 5.2.4. High-low: islands of availability

For high-low areas, it might be the case that the neighbouring areas might be next in line for deployment. These are areas which have not shown to significantly cluster with surrounding areas yet, indicating that these are separate projects in the middle of the periphery. These could be recently initiated projects, or the project is too remote or has not been successful enough to spread to neighbouring neighbourhoods. Whereas Figure 5 would be sensitive to the remoteness of neighbourhoods, Figure 6 would control better for this, as it regards a neighbourhood as a cluster of streets, instead of a singular island.

**Figure 6: Local Moran's I for postal code 6 (PC6) regions according to first order (left) and second order (right) neighbours**



White: Not Significant ( $p > 0.05$ ), Red: High-high, Blue: Low-low, Light blue: Low-high, Light red: High-low

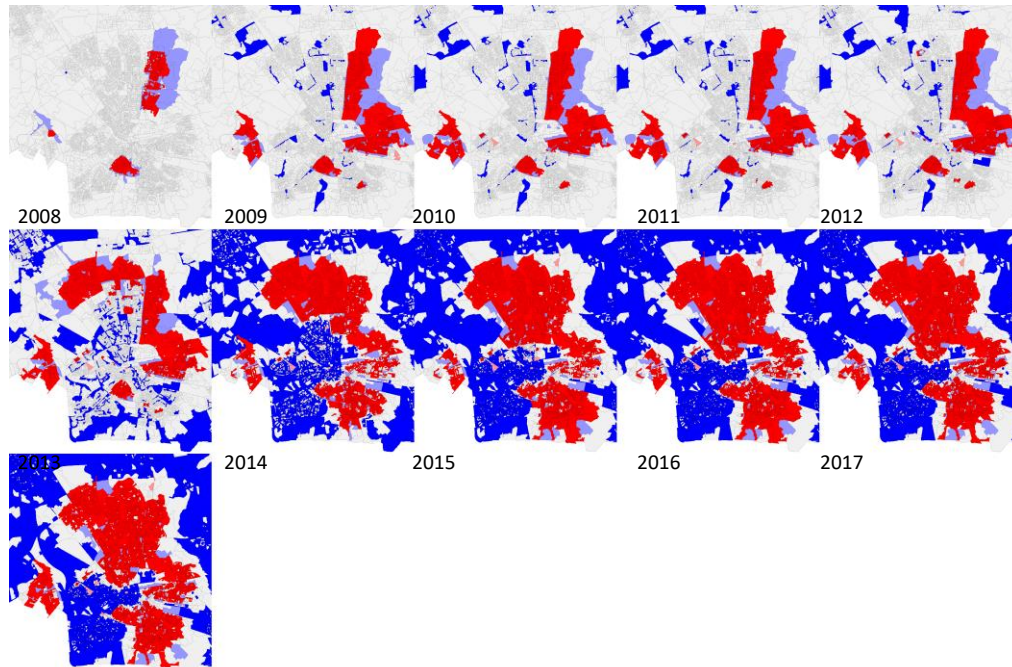
### 5.3. Prediction of the model: Local rollout over time

In order to demonstrate how this clustering propagates over time, a different dataset has been obtained on the rollout of fibre per household, accompanied by the rollout date. This data will demonstrate the rollout on a more detailed level over time for households in Eindhoven. Note that in this case only the household connections in Eindhoven itself have been included. In Eindhoven the rollout of fibre started around 2007, initiated by a public-private partnership (PPP) named OnsNet. OnsNet had previously deployed fibre in Nuenen making it the first Dutch municipality to connect all its households with fibre. The success of these networks have been attributed to PPPs' characteristics of openness, ubiquity and demand aggregation (Sadowski et al., 2009; Van Der Wee et al., 2015). Along with many other fibre networks in Noord-Brabant, Eindhoven's fibre network has eventually been taken over by KPN around 2012 and in 2017 it has been merged with the KPN network (OnsBrabantNet, 2017). While KPN initially prioritized further upgrades of its copper network above the rollout of FttH, the company has recently announced to realize another million FttH-connections by 2021.

Figure Figure 8 shows how the LISA categorization on street level changed over the years in the city. In the 2008 image there appeared to be three main cores. The north-east core is close to the Nuenen municipality, which might explain the location of this fibre core. For the southern core there does not seem to be an apparent reason for why a fibre core started here. This core is a combination of high and low income neighbourhoods and is closer to the centre of the city, but that does not explain its location (CBS, 2018b). The core in the west is in the city quarter of Meerhoven. This area consists mainly of recent housing development, and the houses which were connected there around 2008 have been built in the

same year (Kadaster, 2018). Thus, it can be assumed that these houses received a connection immediately at their completion.

**Figure 8: Yearly LISA for the Eindhoven region**

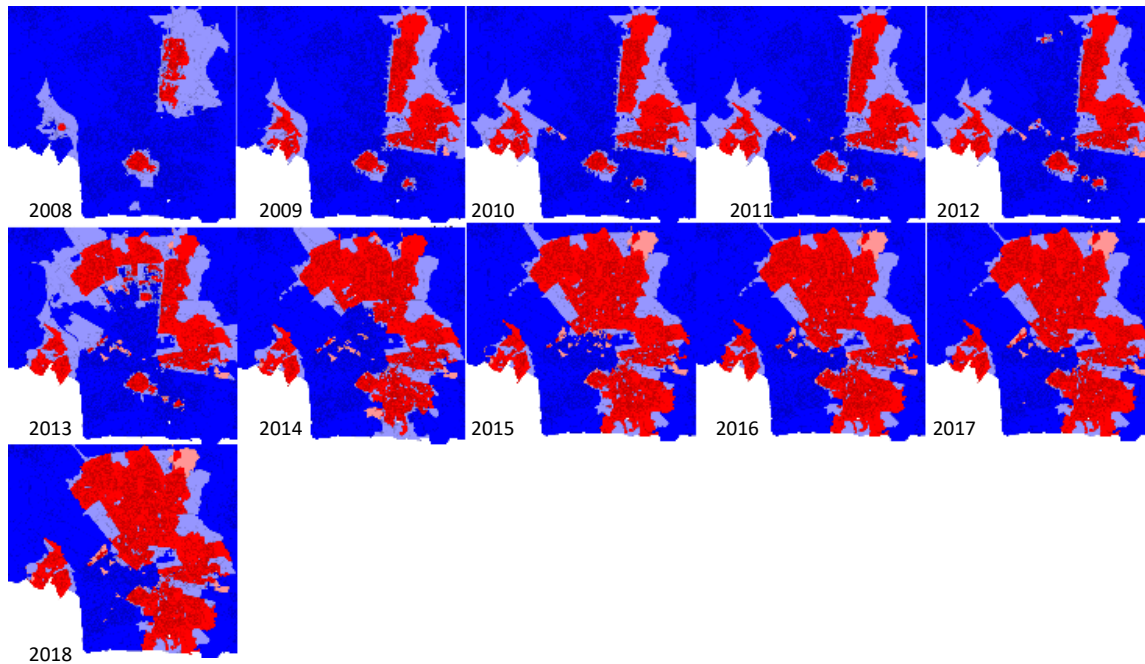


The plots range over years (left to right, up to down) with the upper left image representing the start of 2008 and the lowest image representing the start of 2018. White: Not Significant ( $p > 0.05$ ), Red: High-high, Blue: Low-low, Light blue: Low-high, Light red: High-low

Regarding the diffusion of fibre over the years, there are a few observations which stand out. Firstly, the islands of inequity indeed seem to indicate where fibre will be deployed in the coming years. There are some areas where this is not the case, as they stay light blue over the years, but these areas are mostly uninhabited areas, such as recreational parks and industrial areas. For those sparsely inhabited areas it also appears that fibre rollout to neighbouring regions is not hindered by them and just spreads its way around these areas. Secondly, fibre deployment in Eindhoven appears to have started with a small number of separate projects in different parts of the city which eventually merged together into one main core. Between 2008 and 2009 two additional projects seem to have started, one in the east and one in the south. Thereafter, fibre deployment seems to have spread from these cores to neighbouring streets until all cores have merged together into one cluster. This dynamic data also reveals that 2013 was the year with most deployment activity and that almost all activity appears to have ceased after 2015, leaving the south west and centre of Eindhoven unconnected. This observation is in line with the trend presented in Figure 1. These areas where no fibre gets deployed in the near future also appear to be characterized by a certain feature of the LISA categorization.

Of course, this is a demonstration of solely one city and dynamics for other cities will differ, across both time and space, from these observations on Eindhoven. Nevertheless, it does reveal a certain rollout strategy which is being applied in the Netherlands, as it shows that the LISA category of islands of inequity indeed can be a realistic indicator of the direction of fibre deployment.

**Figure 9: Yearly LISA for the Eindhoven region considering 2<sup>nd</sup> order neighbours and without significance**



The plots range over years (left to right, up to down) with the upper left image representing the start of 2008 and the lowest image representing the start of 2018. Red: High-high, Blue: Low-low, Light blue: Low-high, Light red: High-low

Figure 8 has demonstrated that some areas will never change, because they are industrial or recreative areas. However, since the neighbours of these areas do seem to be affected, the same results have been recalculated, but then 2<sup>nd</sup> order neighbours are also taken into account when calculating the local Moran coefficient. These results are presented in Figure 9. In this figure significance is not included in the visualization, to make the categorization look more pronounced. What can be observed in that figure, is that the islands of inequity are even better precursors for rollout. In addition, the islands of availability appear to play a more prominent role now. Whereas the islands of inequity are now indicating the future expansion of fibre clusters, the islands of availability indicate the commence of new clusters in surrounded by periphery regions. These islands of availability will eventually expand further and form core clusters. This phenomenon can be observed especially in the period from 2012 up to 2016. It first occurs in the northern region of Eindhoven, where an island of availability expands into a larger core region from between 2012 and 2013. After that, the same phenomenon can be observed in the small area just north of the city centre of Eindhoven. But here there appears to be a scattered group of islands of availability, which eventually merge together into one core cluster.

From these spatial results it can be concluded that fibre rollout does diffuse spatially in the Netherlands. Where clusters are formed, they appear to expand and eventually merge together. But time also comes out as an important dimension. The differences between some of the years in Figure 8 indicate the rapid behaviour in which rollout can unfold itself. Simultaneously, this is also an indication that many high detail dynamics are involved in the progress, which might come to the fore at a higher resolution of time.

#### 5.4. The empirical model

The spatial analysis has yielded exploratory results, which can be enriched and explained by qualitative data. Bringing together the spatial results with quantitative data can further enhance the perspective on the rollout situation in the Netherlands. Using the LISA categorization as a dependent variable in a

regression model with proxies for the different investment uncertainties as independent variables, it can be explained what types of areas constitute a certain LISA clustering categorization.

Equation 1 gives a simplified representation of the empirical model, with  $TE$  being the list of techno-economic variables,  $DU$  being the list of demand uncertainty variables and  $MU$  being the list of market uncertainty variables (see Appendix 1). Moreover,  $P$  is the chance for a neighbourhood to belong to a specific LISA category  $l$ , with  $l \in [High\ high, Low\ low, Low\ high, High\ low]$ . The response variables are categorical and linearity is assumed between predictors and logistic odds ratio. The *High high* category is taken as a reference group here.

**Equation 1: Formula logistic regression**

$$\ln\left(\frac{P(Y_i = l)}{P(Y_i = High\ high)}\right) = \sum_{a=1}^{TE} (X_a \beta_a)_l + \sum_{b=1}^{DU} (X_b \beta_b)_l + \sum_{c=1}^{MU} (X_c \beta_c)_l$$

In the following, the empirical model is tested with data on the neighbourhood level wherein all variables, at address level, have been aggregated by taking the mean. We further control for rural and urban areas, as some neighbourhoods are located in both rural and urban areas. In a similar fashion we use a variable which assesses the extend of industrialism for a region by taking the percentage of addresses within a neighbourhood that are located at an industrial park.

This empirical model can be used to distinguish what factors inside the model have constituted the difference between different area types and to confirm the qualitative observations of the LISA results quantitatively. The multinomial logistic models estimated in Table 4 compares the odds of being in a specific LISA category against the odds of belonging to the reference category, i.e. *the core (High high)*. The coefficients might therefore explain what effects have contributed to the difference between these regions.

Relating this model (Equation 1) to the geographical model referred to in section 2.2 the market uncertainty variables can be regarded as an identifier of whether a duplicated infrastructure area (DIA), where both firms are covering the same set of locations with an NGN, or a single infrastructure area (SIA), where only a singular firm has deployed NGN for a set of locations (Bourreau et al., 2015), is more likely for a certain type of cluster. Therein both fibre and coax are seen as NGN technologies, taking fibre to be firm  $i$  and coax as firm  $j$ . The techno-economic variables, on the other hand, can be regarded as a proxy for the costs making up the cost function  $C(z_i) = \int_0^{z_i} c(x)dx$  for every area  $z_i$ . Whereas the geographical model referred to in section 2.2 assumes identical demand, the model in Equation 1 differentiates in demand by assessing the demand uncertainty.

The first column of results (Table 4) indicates what makes an area belong to the periphery instead of the core. These results can also be regarded as the difference in regions with likely large-scale investment in a fibre network as opposed to regions where this is absent. It appears that the peripheral area is more likely to be industrial or rural. An area is also significantly more likely to belong to the periphery than to the core, if the percentage of water bodies is higher, the ground quality is worse, or if there are longer distance to the backbone and to the next house. Based on the spatial results, these would thus be the characteristics of an area which is not expected to have fibre anytime soon. The periphery also differs from the core in that the share of coax connections is higher. Within both the cores and the periphery a SIA is thus more likely.

The islands of inequity and the core mainly differ in terms of the distance variables. This means that these regions do not share an equal level of fibre penetration yet because there is a larger distance between houses, even though they are generally closer to the Eurofiber backbone. In comparison to the core, islands of inequity can thus be described as regions with a lower density of houses, larger distances to the next house and lower tree heights. But these regions are also characterized by being more rural or industrial than the cores. This matches with the visual observations above. These determinants could be the reason why these regions do not have an equal penetration of fibre yet as their neighbouring regions have. Again, this highlights that it is less evident for rural houses and other geographic outliers to get connected to fibre.

The core and islands of availability can be distinguished according to different factors. In comparison to the core these islands of availability are more likely to be industrial and have a lower height of trees. This could signify industrial areas, but these could also be countryside villages, given the higher likelihood of a coax connection, the longer distance to the backbone and the positive coefficient for low quality ground. For the islands of availability a DIA situation is thus more likely. The reason why these high-low areas have not become a core cluster yet could be that these are indeed too remote. One should however acknowledge the large class imbalance which is specifically the case for the islands of availability ( $N = 63$ ), when compared to the core ( $N = 2191$ ).

**Table 4: Odds ratios results for LISA multinomial regression model at neighbourhood level for standardized independent variables**

<b>Predictors</b>	<b>Odds ratios</b>					
	<i>Low low: periphery</i>		<i>Low high: islands of inequity</i>		<i>High low: islands of availability</i>	
Intercept	0.946		0.189	***	0.002	***
Industrial (compared to non-industrial)	3.667	**	8.702	***	36.918	***
Urban (compared to rural)	0.112	***	0.097	***	0.246	*
<b><i>Techno-Economics</i></b>						
Distance to backbone	1.183	**	0.514	**	1.496	*
Density of houses	1.161	*	0.664		0.810	
Distance to the next house	5.156	***	1.948	*	1.717	
Low quality ground	1.599	***	1.272		1.604	**
Multiple family dwellings	1.041		1.228		0.762	
Water bodies	1.647	***	1.075		0.983	
Average tree height	0.887	*	0.891		0.338	***
<b><i>Demand Uncertainty</i></b>						
Average income	1.167		0.872		0.477	*
Multicultural	1.028		0.871		0.603	
Lower education	1.062		0.942		0.411	**
Catholic	1.166	**	1.568	***	1.428	*
Family	0.877		0.790		1.074	
Age	1.344	***	1.116		1.069	
Divorced	1.487	***	1.172		1.003	
<b><i>Market Uncertainty</i></b>						
HHI	0.667	***	0.454	***	0.327	***
Coax connection	8.462	***	1.283		35.034	***
N ( <i>High high</i> =2191)	1554		242		63	
Pseudo R2 McFadden	0.431					

Significance  $p$  is denoted by \*\*\* for  $< 10^{-6}$ , \*\* for 0.01, \* for 0.05, and no symbol for  $> 0.05$ . These odds ratios are in reference to the *High high: core* category.



## 6. Summary and conclusions

The results of the analysis have demonstrated that the rollout of fibre networks in the Netherlands does cluster together and rollout appears to be a neighbourhood-driven process. This means that rollout in the Netherlands is not a random process and is mainly project-based. This clustering is more prevalent in urban areas than in rural areas, indicating that there is more randomness and variability in rural neighbourhoods. The results on the local Moran's I and the LISA categorization also indicate that regions might learn from one another and that projects expand in geographical clusters. This can help identify areas where it is spatially most rational to rollout next. Yet even more important, it shows which areas are least obvious to be rolled out in the future, especially if there is no incentive to innovate in rural areas. These spatial results have demonstrated that spatial diffusion and relations of fibre should not be left unconsidered when it pertains to rollout.

The empirical model has taken these results another step further. Whereas the spatial analysis was able to identify the types of regional clusters for broadband availability, the empirical model has substantiated this, using statistics on investment uncertainties. Through the results of this model it can thus be understood how different areas of coverage are characterized in terms of different types of costs. These results help explain how these uncertainties have shaped certain regions in such a way that they belong to a specific cluster area. The periphery areas are set apart from the core, because they are generally remote and located on industrial parks or in rural areas. The islands of inequity do not have an equal level of penetration as their neighbouring regions because they represent the areas close to the core where addresses are more dispersed. The islands of availability are more characterized in a way that it presents well-developed villages. In the view of the model by Bourreau et al. (2015), cores and the periphery can be related to the situation in an SIA whereas islands of availability can be related to a DIA.

A general conclusion from all results presented here is that a spatial dimension can bring out new dynamics in a rollout problem. The availability and use of general statistics is common, but by acknowledging the possible spatial dimension of these data, new insights can be obtained.



## 7. Appendices

### Appendix 1: Variables used in the regression

	Level of analysis	Units	Year	Source
Weight	Municipality	<i>Ratio</i>	2017	Glaskaart Stratix 2017
Rurality (Rural, Urban, Industrial park)	Household	<i>Category</i>	2011 (Rural/Urban) 2017 (Industrial park)	(CBS, 2011)
<b>Techno-Economics</b>				
Distance to Eurofiber backbone	Household	<i>m</i>	2017	(Eurofiber, 2017)
Distance to next house	Household	<i>m</i>	2017	Dialogic dataset
Average height of trees within a radius of 20m	Household	<i>m</i>	2017	(Ministerie van Binnenlandse Zaken, 2017)
Low quality ground	Municipality	<i>% (per group)</i>	2017	(CBS, 2018a)
Percentage of waterbodies	Neighbourhood	<i>%</i>	2017	Kerncijfers wijken en buurten 2017
Density of houses	Neighbourhood	<i>Amount / (ha)</i>	2016	Kerncijfers wijken en buurten 2016
Distance to next house	Household	<i>m</i>	2017	BAG
Percentage of multiple family buildings (apartments, flats)	Neighbourhood	<i>%</i>	2016	Kerncijfers wijken en buurten 2016
<b>Demand Uncertainty</b>				
Households (N= 8666312)				
Gender	Neighbourhood	<i>% (per group)</i>	2017	Kerncijfers wijken en buurten 2017
Age	Neighbourhood	<i>Amount (per group)</i>	2017	Kerncijfers wijken en buurten 2017
Average Income	Neighbourhood	<i>€</i>	2015	Kerncijfers wijken en buurten 2015
Education level	Postal code 4	<i>% (per group)</i>	2014	(CBS, 2017)
Religion	Municipality	<i>% (per group)</i>	2015	(CBS, 2016)
(Non)-western background	Neighbourhood	<i>% (per group)</i>	2017	Kerncijfers wijken en buurten 2017
<b>Market Uncertainty</b>				
HHI	Neighbourhood (or other area)	<i>[0, 10000]</i>	2017	Dialogic dataset
Availability of coax	Household	<i>Number of coax connections</i>	2017	Dialogic dataset

## Appendix 2: Datasets

The data sources used for the analysis are the following: a dataset with fibre availability for all addresses in the Netherlands, CBS datasets on regional variables, and a longitudinal dataset on fibre rollout in the city of Eindhoven.

The first data source available for this research is based on a broadband dataset provided by the research and consultancy firm Dialogic. This dataset comprises the main part of the analysis, as it is used to run a logistic regression on the successful rollout of fibre. The data contains the BAG-data of 8.9 million households in the Netherlands along with the availability of a fibre connection and other connection types for those households in 2017. In order to improve the accuracy of the dataset provided by Dialogic, the *Glask kaart* published by Stratix (2017) is used to assign weights to the factors of the regression. The Stratix dataset indicates the percentage of households covered by fibre per municipality, so for the cases where the initial dataset underreports in comparison to the Stratix dataset, the deviation between these datasets has been used as a weight in the regression. These weights will ensure that observations in underreported municipalities will have less influence on the estimates of the regression.

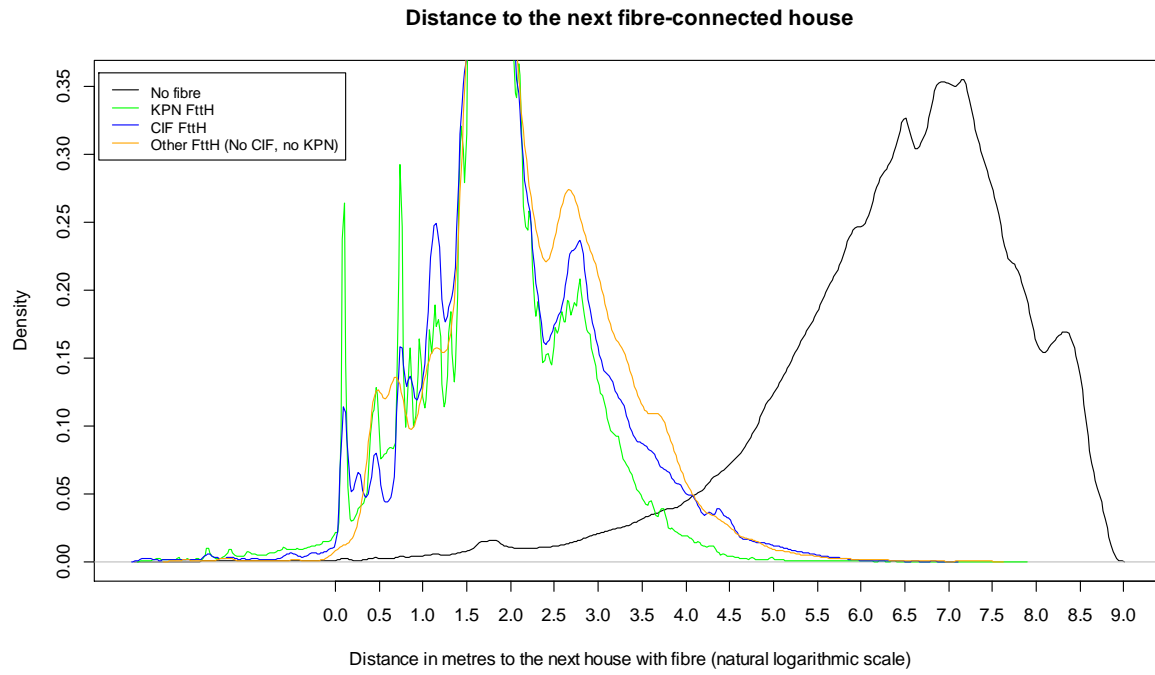
The second set of data has been obtained from public datasets, available through Central Statistics Bureau (CBS) and the Data portal of the Dutch Government. This is used to model the social and additional regional data. Most of the variables have been obtained from the CBS' *Kerncijfers Wijken en Buurten*, which contains regional statistics on a neighbourhood level.

All initial data has been stored into separate PostGIS databases for the sake of flexibility and the ability to perform geographically related queries. The eventual statistical analyses have all been carried out in the statistical software package of R version 3.4.0. Geographic visualizations of the data have been constructed using QGIS 3.0.2 (2018). For the local Moran's I calculations and visualizations thereof, Geoda 1.12.1.131 (Anselin, Syabri, & Kho, 2006) has been used.

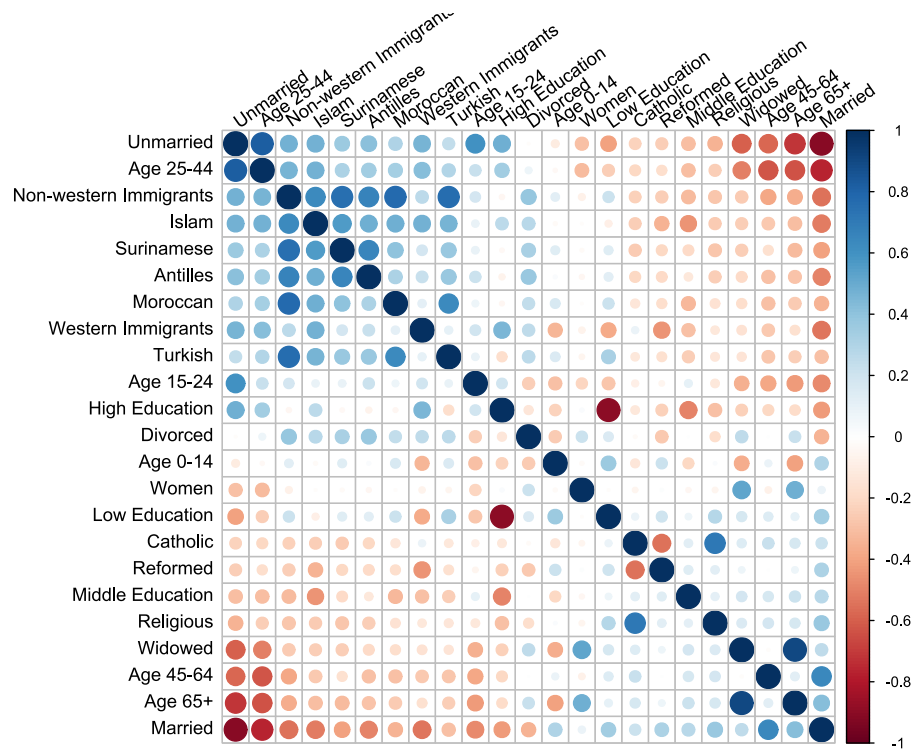
**Appendix 3: Descriptives of techno-economic variables**

<b><i>Variables</i></b>	<b><i>Minimum</i></b>	<b><i>Mean</i></b>	<b><i>Maximum</i></b>	<b><i>Standard deviation</i></b>	<b><i>Missing values</i></b>
Distance to Eurofiber backbone	0.0	1387.0	37088.0	2667.4	0
Average tree height	0.0	5.0	49.0	3.4	0
Low quality ground	0.0	24.6	100.0	31.6	705295
Percentage of waterbodies	0.0	3.7	100.0	7.4	0
Density of houses	0.0	25.1	234.0	26.0	335941
Distance to next house	0.0	9.1	690.0	20.1	983
Percentage of multiple family buildings (apartments, flats)	0.0	35.9	100.0	30.6	436105

#### Appendix 4: Kernel density plot of distance to the next fibre connection



## Appendix 5: Correlogram for the factor variables



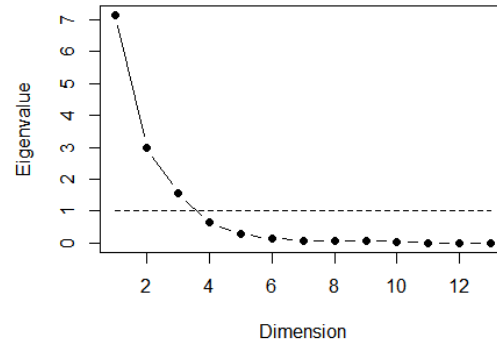
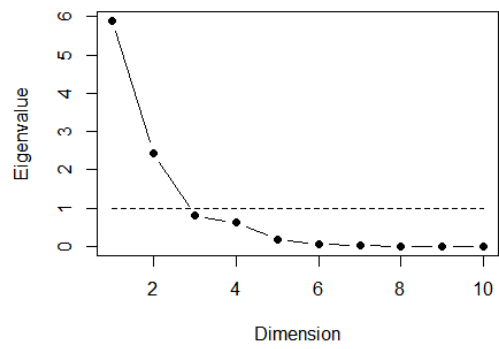
The variables have been automatically sorted by the first principal component order. Both the size and the colour of the cells indicate the Pearson correlation coefficient.

**Appendix 6: Loadings of the factor analysis for the first set**

<b><i>Variables</i></b>	<b><i>Factor 1</i></b>	<b><i>Factor 2</i></b>	<b><i>Factor 3</i></b>
Age 00-14		-0.43	
Age 15-24	-0.55		
Age 25-44	-0.76	-0.30	
Age 45-64	0.70		
Age $\geq 65$	0.45	0.81	
Women		0.51	
Unmarried	-0.93		
Married	0.95		-0.31
Divorced			0.97
Widowed		0.97	
Proportional variance	0.34	0.24	0.12

Loadings are cut off at 0.3.

**Appendix 7: Scree plots of set 1 (left) and set 2 (right)**



**Appendix 8: Loadings of the factor analysis for the second set**

<b>Variables</b>		<b>Factor 1</b>	<b>Factor 2</b>	<b>Factor 3</b>
High education			-1.00	
Middle education			0.54	
Low education			0.88	
Catholic				0.99
Islam		0.58	-0.35	
Reformed			0.32	-0.68
Religious				0.66
Western immigrants		0.30	-0.42	
Non-western immigrants		0.97		
Background from	Morocco	0.72		
	Netherlands Antilles	0.64		
	Surinam	0.69		
	Turkey	0.72		
Proportional variance		0.30	0.20	0.14

Loadings are cut off at 0.3.



## 8. References

- ACM. (2018). *Telecommonitor Q1 - Q2 2017*. Retrieved from <https://www.acm.nl/sites/default/files/documents/2018-01/telecommonitor-rapportage-eerste-halfjaar-2017.pdf>
- Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*, 27(2), 93–115. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>
- Anselin, L., Syabri, I., & Kho, Y. (2006, January). GeoDa: An introduction to spatial data analysis. *Geographical Analysis*. Wiley/Blackwell (10.1111). <https://doi.org/10.1111/j.0016-7363.2005.00671.x>
- Besteman, S., Van Der Haar, A., Mascini, C., Aelmans, M., Brennenraedts, R., & Hermanussen, L. (2017). *Belang van connectiviteit voor digitale transformatie*. Retrieved from [https://www.ispconnect.nl/wp-content/uploads/2017/11/Belang van connectiviteit voor digitale transformatie.pdf](https://www.ispconnect.nl/wp-content/uploads/2017/11/Belang_van_connectiviteit_voor_digitale_transformatie.pdf)
- Bourreau, M., & Cambini, C. (2012). International Journal of Industrial Organization Access pricing , competition , and incentives to migrate from “ old ” to “ new ” technology ☆, 30, 713–723. <https://doi.org/10.1016/j.ijindorg.2012.08.007>
- Bourreau, M., Cambini, C., & Hoernig, S. (2015). Geographic access markets and investments. *Information Economics and Policy*, 31, 13–21. <https://doi.org/10.1016/j.infoecopol.2015.04.003>
- Bourreau, M., Cambini, C., Hoernig, S., Parcu, P. L., Rossi, M. A., & Silvestri, V. (2017). *The future of broadband policy, part 2: Technological neutrality, path dependency and public financing*. <https://doi.org/10.2870/571682>
- Bourreau, M., Doğan, P., & Manant, M. (2010). A critical review of the “ladder of investment” approach. *Telecommunications Policy*, 34(11), 683–696. <https://doi.org/10.1016/j.telpol.2010.09.002>
- Bourreau, M., Hombert, J., Pouyet, J., & Schutz, N. (2011). Upstream competition between vertically integrated firms. *Journal of Industrial Economics*, 59(4), 677–713. <https://doi.org/10.1111/j.1467-6451.2011.00469.x>
- Bremmer, D. (2017, December 6). Waarom de aanleg van glasvezel in Nederland stokt. AD. Retrieved from <https://www.ad.nl/economie/waarom-de-aanleg-van-glasvezel-in-nederland-stokt~acaaab8b/>
- Briglaue, W., & Cambini, C. (2018). Does regulation of basic broadband networks affect the adoption of new fiber-based broadband services? *Industrial and Corporate Change*, 1–22. <https://doi.org/10.1093/icc/dty012>
- Briglaue, W., Cambini, C., & Grajek, M. (2018). Speeding up the internet: Regulation and investment in the European fiber optic infrastructure. *International Journal of Industrial Organization*, 61, 613–652. <https://doi.org/10.1016/j.ijindorg.2018.01.006>
- Briglaue, W., Frubing, S., & Vogelsang, I. (2014). The Impact of Alternative Public Policies on the Deployment of New Communications Infrastructure-A Survey. *Review of Network Economics*, 13(3), 227–270. <https://doi.org/10.1515/rne-2015-0004>
- Briglaue, W., & Gugler, K. (2013). The deployment and penetration of high-speed fiber networks and services: Why are EU member states lagging behind? *Telecommunications Policy*, 37(10), 819–835.

<https://doi.org/10.1016/j.telpol.2013.05.003>

- Brito, D., & Tselekounis, M. (2016). Access regulation and the entrant's mode of entry under multi-product competition in telecoms. *Information Economics and Policy*, 37, 20–33. <https://doi.org/10.1016/j.infoecopol.2016.10.005>
- Cambini, C., & Jiang, Y. (2009). Broadband investment and regulation: A literature review. *Telecommunications Policy*, 33(10–11), 559–574. <https://doi.org/10.1016/j.telpol.2009.08.007>
- CBS. (2011). *Bevolkingskernen in Nederland, 2011*.
- CBS. (2016). Kerkelijke gezindte en kerkbezoek naar gemeenten. Retrieved June 15, 2018, from [https://www.cbs.nl/-/media/\\_excel/2016/51/kerkelijke gezindte en kerkbezoek naar gemeenten.xlsx](https://www.cbs.nl/-/media/_excel/2016/51/kerkelijke_gezindte_en_kerkbezoek_naar_gemeenten.xlsx)
- CBS. (2017). Persoonskenmerken op postcode 4-digit niveau, 2014. Retrieved June 6, 2018, from <https://www.cbs.nl/nl-nl/maatwerk/2017/19/persoonskenmerken-op-postcode-4-digit-niveau-2014>
- CBS. (2018a). CBS StatLine - Maatstaven Financiële-verhoudingswet (Fvw). Retrieved June 6, 2018, from <http://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=60039fvw>
- CBS. (2018b). Kerncijfers wijken en buurten 2015. Retrieved June 20, 2018, from <http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=83220NED>
- Dwivedi, Y. K., & Lal, B. (2007). Socio-economic determinants of broadband adoption. *Industrial Management and Data Systems*, 107(5), 654–671. <https://doi.org/10.1108/02635570710750417>
- Eurofiber. (2017). Glasvezelnetwerk - Zakelijk glasvezel - Eurofiber. Retrieved June 6, 2018, from <https://www.eurofiber.nl/glasvezelnetwerk/>
- European Commission. (2017). *Broadband Coverage in Europe 2016*. <https://doi.org/10.2759/679477>
- Faulhaber, G. R., & Hogendorn, C. (2000). The market structure of broadband telecommunications. *The Journal of Industrial Economics*, 48(3), 305–329. <https://doi.org/10.1111/1467-6451.00125>
- FCC. (2016). *Measuring Fixed Broadband Report 2016*. Washington. Retrieved from <https://www.fcc.gov/reports-research/reports/measuring-broadband-america/measuring-fixed-broadband-report-2016>
- Feijóo, C., Ramos, S., Armuña, C., Arenal, A., & Gómez-Barroso, J. L. (2018). A study on the deployment of high-speed broadband networks in NUTS3 regions within the framework of digital agenda for Europe. *Telecommunications Policy*, 42(9), 682–699. <https://doi.org/10.1016/j.telpol.2017.11.001>
- Flamm, K., & Chaudhuri, A. (2007). An analysis of the determinants of broadband access. *Telecommunications Policy*, 31, 312–326. <https://doi.org/10.1016/j.telpol.2007.05.006>
- Fourie, H., & de Bijl, P. W. J. (2018). Race to the top: Does competition in the DSL market matter for fibre penetration? *Telecommunications Policy*, 42(9), 778–793. <https://doi.org/10.1016/j.telpol.2017.11.003>
- Fourie, H., & De Bijl, P. W. J. (2017). Race to the top: Does competition in the DSL market matter for fibre penetration? *Telecommunications Policy*. <https://doi.org/10.1016/j.telpol.2017.11.003>
- Grubestic, T. H. (2006). A spatial taxonomy of broadband regions in the United States.

<https://doi.org/10.1016/j.infoecopol.2006.05.001>

- Grubestic, T. H. (2008). Spatial data constraints: Implications for measuring broadband. *Telecommunications Policy*, 32(7), 490–502. <https://doi.org/10.1016/j.telpol.2008.05.002>
- Kadaster. (2018). BAG Viewer. Retrieved June 20, 2018, from <https://bagviewer.kadaster.nl/lvbag/bag-viewer/index.html?geometry.x=156999.23125&geometry.y=384128.43875&zoomlevel=4>
- Mack, E. A. (2014). Broadband and knowledge intensive firm clusters: Essential link or auxiliary connection? *Papers in Regional Science*, 93(1), 3–29. <https://doi.org/10.1111/j.1435-5957.2012.00461.x>
- Mack, E. A., & Grubestic, T. H. (2014). US broadband policy and the spatio-temporal evolution of broadband markets. *Regional Science Policy & Practice*, 6(3), 291–308. <https://doi.org/10.1111/rsp3.12042>
- Matsushima, N., & Mizuno, K. (2018). Strategic investments under competition for access provision. *Telecommunications Policy*, 42(2), 127–144. <https://doi.org/10.1016/j.telpol.2017.09.005>
- Ministerie van Binnenlandse Zaken. (2017). Boomhoogte in Nederland - Datasets - Data.overheid.nl. Retrieved May 15, 2018, from <https://data.overheid.nl/data/dataset/3e1b0709-f5e6-4817-8cce-e37ece2f6548>
- Nucciarelli, A., & Sadowski, B. (2018). *Strategic and policy*. FrancoAngeli.
- OnsBrabantNet. (2017). Ook in de toekomst de beste dienstverlening - OnsBrabantNet. Retrieved June 20, 2018, from <https://www.onsbrabantnet.nl/blog/nieuws/2017/Ook-in-de-toekomst-de-beste-dienstverlening/>
- OpenStreetMap contributors. (2018). Copyright and License. Retrieved from <https://www.openstreetmap.org/copyright>
- Ordoover, J., & Shaffer, G. (2007). Wholesale access in multi-firm markets: When is it profitable to supply a competitor? *International Journal of Industrial Organization*, 25(5), 1026–1045. <https://doi.org/10.1016/j.ijindorg.2007.01.001>
- QGIS Development Team. (2018). QGIS Geographic Information System. Open Source Geospatial Foundation Project. Retrieved from <http://qgis.osgeo.org/>
- Quaglione, D., Agovino, M., Berardino, C. Di, & Sarra, A. (2018). Exploring additional determinants of fixed broadband adoption: policy implications for narrowing the broadband demand gap. *Economics of Innovation and New Technology*, 307–327,. <https://doi.org/10.1080/10438599.2017.1350358>
- Röller, L.-H., & Waverman, L. (2001). Telecommunications infrastructure and economic development: A simultaneous approach. *American Economic Review*, 91(4), 909–923. <https://doi.org/10.1257/aer.91.4.909>
- Sadowski, B. M. (2017a). Advanced users and the adoption of high speed broadband: Results of a living lab study in the Netherlands. *Technological Forecasting and Social Change*, 115, 1–14. <https://doi.org/10.1016/j.techfore.2016.09.009>
- Sadowski, B. M. (2017b). Consumer cooperatives as an alternative form of governance: The case of the broadband industry. *Economic Systems*, 41(1), 86–97. <https://doi.org/10.1016/j.ecosys.2016.04.004>
- Sadowski, B. M., Nucciarelli, A., & de Rooij, M. (2009). Providing incentives for private investment in

municipal broadband networks: Evidence from the Netherlands. *Telecommunications Policy*, 33(10–11), 582–595. <https://doi.org/10.1016/j.telpol.2009.08.005>

Stratix. (2017). Glaskaart Nederland 2017Q1. Retrieved February 20, 2018, from <https://www.stratix.nl/glaskaart/>

Van der Haar, A., Driesse, M., Brennenraedts, R., Pegtel, W., & Cazemier, M. (2017). *Nederland op Glasvezel*. Retrieved from <http://www.fibercarriers.nl/wp-content/uploads/2017/10/fca-nederland-op-glasvezel.pdf>

Van Der Wee, M., Verbrugge, S., Sadowski, B. M., Driesse, M., & Pickavet, M. (2015). Identifying and quantifying the indirect benefits of broadband networks for e-government and e-business: A bottom-up approach. *Telecommunications Policy*, 39, 176–191. <https://doi.org/10.1016/j.telpol.2013.12.006>